UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação

Tese

Dengue Incidence Rate Estimation Using Aerial and Street-level Urban Imagery with Deep Learning Models

Virginia Ortiz Andersson

Pelotas, 2019

Virginia Ortiz Andersson

Dengue Incidence Rate Estimation Using Aerial and Street-level Urban Imagery with Deep Learning Models

Tese apresentada ao Programa de Pós-Graduação em Computação do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Ciência da Computação.

Advisor: Prof. Dr. Ricardo Matsumura Araujo
Coadvisor: Prof. Dr. Cristian Cechinel

Pelotas, 2019

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

A111d Andersson, Virginia Ortiz

Dengue incidence rate estimation using aerial and street-level urban imagery with deep learning models / Virginia Ortiz Andersson ; Ricardo Matsumura Araujo, orientador ; Cristian Cechinel, coorientador. — Pelotas, 2019.

154 f. : il.

Tese (Doutorado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2019.

1. Dengue fever estimation. 2. Visual computational sociology. 3. Street-level and aerial images. 4. Deep convolutional neural networks. 5. Machine learning. I. Araujo, Ricardo Matsumura, orient. II. Cechinel, Cristian, coorient. III. Título.

CDD : 006.32

Elaborada por Aline Herbstrith Batista CRB: 10/1737

AGRADECIMENTOS

Agradeço à Universidade Federal de Pelotas (UFPel) e ao Programa de Pós-Graduação em Computação, da mesma insitituição, pela oportunidade de cursar o Doutorado em Computação.

Ao meu orientador, Prof. Dr. Ricardo Matsumura Araujo, pela dedicação na orientação do presente trabalho. Da mesma forma, agradeço ao Prof. Dr. Cristian Cechinel pela coorientação.

Aos colegas do grupo de pesquisa *DataLab*, pela contribuição de conhecimentos e ideias.

Ao meu noivo, Henrique Avila Vianna, pelo apoio incondicional emocional e intelectual, e aos meus familiares, pela presença e compreensão.

ABSTRACT

ANDERSSON, Virginia Ortiz. **Dengue Incidence Rate Estimation Using Aerial and Street-level Urban Imagery with Deep Learning Models**. Advisor: Ricardo Matsumura Araujo. 2019. 154 f. Thesis (Doctorate in Computer Science) – Technological Development Center, Federal University of Pelotas, Pelotas, 2019.

Motivated by sociological theories that present the physical appearance of the urban environment as an influential factor in the behavior of inhabitants, the new *Visual Computational Sociology* research area has investigated computer vision models to infer latent variables such as demographic, socioeconomic, cultural, and health indicators from aerial and street-level urban imagery. Just like crime events can be inferred from the appearance of the urban environment, occurrences of diseases, such as dengue fever, can be explained from visual data as well. This work proposes the use of aerial and street-level images to estimate dengue fever incidence rates, in an automated way, to increase the estimation effectiveness of dengue and its variants in urban regions. Specifically, it was proposed using computer vision techniques capable of extracting attributes from urban images automatically and neural network models for multiple regression to estimate latent variables of dengue incidence using urban environment visual attributes as predictors. For this, experiments were carried out using street-level and aerial images, together with historical dengue fever data obtained from the Brazilian capitals Rio de Janeiro (RJ), São Paulo (SP), and Salvador (BA). Results showed evidence that: (i) street-level image features can be used for estimating dengue incidence rates, although models using aerial image features present better results; (ii) the combination of aerial and street-level features contribute to better results in estimating dengue incidence rates; (iii) models generalize poorly to other cities, slightly improving the results when using transfer-learning techniques and multiple cities in training and (iv) Deep Convolutional Neural Networks (Deep Convnets) are suitable for use in the proposed model, since its features presented better results compared to designed descriptor techniques. At last, it is expected that the proposed models will contribute to an improvement in the state of the art of dengue estimation models, and the obtained results contribute to public health policies in urban centers, through better results or in optimizing their accomplishment.

Keywords: Dengue Fever Estimation. Visual Computational Sociology. Street-level Images. Aerial Images. Deep Convolutional Neural Networks. Machine Learning. Computer Vision.

RESUMO

ANDERSSON, Virginia Ortiz. **Estimativa de Incidência de Dengue Utilizando Imagens Urbanas Aéreas e no Nível da Rua com Modelos de Aprendizado Profundo**. Orientador: Ricardo Matsumura Araujo. 2019. 154 f. Tese (Doutorado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2019.

Motivada por teorias sociológicas que apresentam a aparência física do ambiente urbano como um fator influente no comportamento dos indivíduos habitantes, a recente área de pesquisa *Sociologia Computacional Visual* investiga modelos de visão computacional para inferir variáveis latentes, como indicadores demográficos, socioeconômicos, culturais e de saúde a partir de imagens urbanas aéreas e no nível da rua. Da mesma forma que as ocorrências criminais podem ser inferidas a partir da aperência do ambiente urbano, ocorrências de doenças, como a dengue, também podem ser explicadas a partir de dados visuais presentes nas imagens. Este trabalho propõe o uso de técnicas de visão computacional capazes de extrair atributos de imagens urbanas automaticamente e modelos de rede neural para regressão múltipla de variaáveis latentes da incidência da dengue usando atributos visuais do ambiente urbano como preditores. Foram realizados experimentos com imagens aéreas e de rua, juntamente com dados históricos de dengue nas capitais Rio de Janeiro (RJ), São Paulo (SP) e Salvador (BA). Os resultados mostraram evidências de que (i) *features* de imagens no nível da rua podem ser usadas para estimar as taxas de incidência de dengue, embora os modelos que utilizam *features* de imagens aéreas apresentem melhores resultados, e (ii) a combinação de *features* aéreas e de nível de rua contribuem para melhores resultados na estimativa das taxas de incidência de dengue; (iii) modelos generalizam insuficientemente para outras cidades, melhorando ligeiramente os resultados ao usar técnicas de *transfer learning* e mais cidades no treinamento; e (iv) Redes Neurais Convolucionais Profundas (Deep ConvNet) são adequadas para uso no modelo proposto, uma vez que apresenta melhores resultados em comparação com técnicas de descritores projetados. Finalmente, espera-se que os modelos propostos contribuam para uma melhoria no estado da arte dos modelos de estimativa de dengue, e os resultados obtidos contribuam para as políticas de saúde pública nos centros urbanos, por meio de melhores resultados ou na otimização de sua realização.

Palavras-chave:

Estimativa de Dengue, Sociologia Computacional Visual, Imagens no Nível da Rua, Imagens Aéreas, Redes Neurais Convolucionais, Aprendizado de Máquina, Visão Computacional

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
CNN	Convolutional Neural Network
ConvNet	Convolutional Neural Network
DENV	Dengue Virus Serotype
e-SIC	Sistema Eletrônico de Informações ao Cidadaão
ESRI	Environmental System Research Institute
FC	Fully-connected Neural Network Layer
GIS	Geographic Information System
GIST	“Gist” of a scene, Espatial Envelope
HOG	Histogram of Oriented Gradients
IBGE	Instituto Brasileiro de Geografia e Estatística
LIRAa	Levantamento Rápido de Índices de Infestação pelo Aedes aegypti
LR	Linear Regression
MAE	Mean Absolute Error
MSE	Mean Squared Error
NN	Neural Network
QGIS	“Quantum” GIS
ReLU	Rectified Linear Unity
RMSE	Root Mean Squared Error
SGD	Stochastic Gradient Descent
SINAN	Sistema de Informação de Agravos de Notificação

CONTENTS

1 INTRODUCTION

Dengue is an infectious disease transmitted by mosquitoes, which causes morbidity and significant economic losses (ZELLWEGER et al., 2017). According to the World Health Organization (WHO, 2018), it is estimated that 500,000 people are infected with dengue all year round, including their most severe case, and approximately 2.5% of these people die from the disease. Any of four viruses serotypes DENV-1 to DENV-4, from the genus *Flavivirus*, can cause classic dengue fever and its variants dengue hemorrhagic fever and dengue shock syndrome. The DENV-1 to DENV-4 viruses are spread through the female mosquitoes, more specifically the *Aedes aegypti* and *Aedes albopictus*. The transmission of dengue occurs with the mosquito feeding on an infected person, followed by an incubation phase inside the mosquito that will become infectious to other people (GUZMAN; ISTÚRIZ, 2010).

According to WILDER-SMITH; GUBLER (2008), the observed increase in the incidence of cases of dengue fever is highly correlated with global population growth. A growing population entails societal and demographic changes such as urbanization, deforestation, construction of dams and irrigation systems, and can lead to precarious infrastructures of housing, sewage, and waste management systems. The *Aedes* mosquitoes are closely associated with humans, water, and domestic environment features. Some of these features, especially in the urban environment, create ideal conditions for mosquitoes breeding and feeding in humans (THAMMAPALO et al., 2008).

Estimating regional dengue incidence is an essential part of creating policies to prevent the disease from spreading. Traditionally, dengue fever estimation is done by physically inspecting locations for specific environment characteristics that contribute to the vector habitat, or using historical data and socioeconomic and climatic factors correlations. Such methods usually involve the use of demographic data, such as in the attack ratio (AR) index, which divides the number of dengue cases by the population at risk. According to (COELHO; DE CARVALHO, 2015), to calculate the AR index is necessary to estimate the population exposed to the disease, and the number of the population at risk is virtually impossible to determine without regular virological data collection. Moreover, the Demographic Census, which acquires population data, is

expensive and is collected with a significant period of at least ten years (IBGE, 2016). The difficulty in collecting demographic data, especially in developing countries, pose difficulty in inferring such demographic-based indexes.

While traditionally estimation is done by physically inspecting locations or using historical data, recently remote sensing techniques have been proposed, including the use of aerial photography (AMARASINGHE et al., 2017). Aerial photography and street-level images depict the urban scenarios and are close representations of the human environment available in digital form. Urban imagery is gradually turning ubiquitous since the popularity of services offered by Apple, Bing, and Google through aerial maps and street view, and their use in daily life is gradually increasing, mostly to aid navigation (GEBRU et al., 2017).

A relatively new research field, the *Visual Computational Sociology* is dedicated to investigate the ability of direct inference of latent socioeconomic variables from urban images through the use of computer vision techniques, specially Deep Convolutional Neural Networks to relate a city's physical appearance with social statistics, such as income (GEBRU et al., 2017), demographics (De Nadai et al., 2016), behavior (ARIETTA; EFROS, 2014) and to make decisions based on implicit and explicit visual cues (KHOSLA et al., 2014). Since the urban imagery contains data that can correlate with such latent variables, the information on images can also be used to infer latent health variables, such as dengue fever incidence, without the use of demographic data.

1.1 Motivation

The urban environment contains a large amount of information we can directly perceive, especially by our sight, and additional information we can infer, implicitly connecting the elements present in the environment we observe. We visualize a scene in an urban environment detecting houses, cars in the streets, signs, and commercial buildings. By observing this same scene, it is possible to infer, according to our acquired knowledge, other characteristics of the place, such as safeness, wealthiness, and healthiness of the residents or the purpose of the environment, and if it is a commercial, industrial or residential place (KHOSLA et al., 2014). Those implicit pieces of information are non-visual attributes (ARIETTA; EFROS, 2014) or "latent variables", which are present in society and the urban environment and are not directly visible, such as demographic, socioeconomic, cultural, and health indicators of the region observed.

Street-level images depict typically urban scenes from a person's perspective and are the closest depiction of the human environment available in digital form. Their popularity increased with navigation services, such as Google Street View (ANGUELOV et al., 2010; GOOGLE, 2017), that make available access to street images, contribut-

ing to their ubiquity. Recently, cars with onboard cameras are gradually becoming more popular, contributing to the popularity of street-level image data category (GEBRU et al., 2017).

The urban environment can also be represented digitally by aerial photography, providing a bird's-eye view of Earth, with images taken from a variety of aircraft. Aerial photography has become widely available in the last decade due to the popularity of mapping services such as those offered by Apple, Bing, and Google, with a large fraction of cities covered by such services. Aerial images provide a unique source of geographic information by allowing a large area to be covered in a single image.

Since the emergence of street-level and aerial imagery data, computer vision models have been used in conjunction with street-level and satellite imagery to relate the physical appearance of a city with social statistics such as income, perceived safety, social behavior, and to make decisions based on implicit and explicit visual cues. These models are inspired by environmental theories that show the influence that the environment exerts on social factors and human behavior, such as *Broken Window Theory* (WILSON; KELLING, 1982), and "Routine Activity of Places" (SHERMAN; GARTIN; BURGER, 1989). This relatively new research area received denominations from the authors DOERSCH et al. (2012) as "*Computational Geo-cultural Modeling*", from ARIETTA; EFROS (2014) as "*City Forensics*" and from GEBRU et al. (2017) as "*Visual Census Estimation*" and "*Visual Computational Sociology*". Recently, SHAPIRO (2017) referenced this line of research as "*Street-level datafication*", which links visual environment qualities, geographic information, and social valuations and risks. Other denominations for this area appears as "*vision-based urban perception*", in WANG et al. (2018).

These Visual Computational Sociology models are being used to navigate and search for optimized urban routes, improvements in planning and understanding of urban aspects, to propose alternatives and automate the census and remote sensing applications such as object detection, crop yield estimation, and crowd estimation (DOERSCH et al., 2012; ARIETTA; AGRAWALA; RAMAMOORTHI, 2013; ARIETTA; EFROS, 2014; KHOSLA et al., 2014; De Nadai et al., 2016; GEBRU et al., 2017,?; MNIH; HINTON, 2010; CHENG; HAN, 2016; GEIPEL; LINK; CLAUPEIN, 2014; MEYNBERG; CUI; REINARTZ, 2016).

The same way that socioeconomic factors, such as the occurrence of crimes, can be explained by environmental criminological theories, the presence of diseases, such as dengue fever, can also be attributed to environmental factors and physical urban structures (COHEN et al., 2003; WILDER-SMITH; GUBLER, 2008; KIKUTI et al., 2015).

According to KRYSTOSIK et al. (2017), an investigation at streets and neighborhood level that incorporates environmental attributes such as aspects of housing, street

configuration, presence of standing water, and other environmental risks is of interest to the research area. The use of urban imagery, such as street-level images, has proved to be a useful proxy for demographic data, and the elaboration of models of disease inference.

1.2 Research Goals

Given the need for a fine-scale analysis of neighborhoods and streets related with dengue cases, and the emergence of Visual Computational Sociology research area, which applies computer vision models to infer latent variables from urban imagery images, the following Subsections 1.2.1 and 1.2.2 presents an overview of the general objectives and specific objectives, respectively, of this research Thesis.

1.2.1 General Objective

The general objective of this present Thesis is:

- To use street-level image domain in conjunction with aerial imagery to estimate dengue incidence rates in urban regions.

1.2.2 Specific Objectives

The Specific Objectives of this Thesis, enrolled as Hypotheses in Section 3.2, are:

- To evaluate the use of aerial and street-level image features as dengue incidence rate single estimators.
- To evaluate the use of aerial and street-level image features combined as dengue incidence rate estimators.
- To evaluate the geographic portability of the models, trained and evaluated in one city to be applied in other different city.
- To evaluate the suitability of Deep Convolutional Neural Networks (Deep ConvNet) for use in the proposed models as urban image feature extractors.

The next Section 1.3 enumerates the contributions resulted from the application of methodology and experiments to the achievement of the listed objectives, and Section 1.4 presents the published works resulted from the presented models experiments.

1.3 Contributions

- In the field of *Visual Computational Sociology*, it is proposed a Deep Convolutional Neural Network based model for estimating real positive values of latent

health variables, specifically dengue fever rates, from actual incidence data, as a regression model, without normalization or stratification.

- Also in the field of *Visual Computational Sociology*, it was produced results with models that merge features obtained from different domains, that may be useful in future analysis and research.
- The provision of anonymous datasets of aggregated dengue incidence rates per fine-grained regions, labeling their belonging georeferenced location points, in Brazilian cities.
- Models trained with real dengue fever data, especially for aerial image data domain, can contribute to future works by *transfer learning* techniques, with *Deep Convolutional Neural Networks*.
- In dengue fever prediction, it is proposed a model for estimating dengue incidence rates solely from urban images, a ubiquitous data domain, without the need of social-economic government data e.g., population, income, which can be scarce in developing countries.
- Also, in dengue fever prediction, the proposed models can be used in public health policies targeted at fine-scale street-level analyses.

1.4 Published Works

The following works were published using the resulting methodology and models proposed in this present Thesis, sorted by relevance order:

- ***Combining Street-level and Aerial Images for Dengue Incidence Rate Estimation.*** In: IJCNN - International Joint Conference on Neural Networks, 2019, Budapest, Hungary. Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN-2019), 2019. p. 1-8. - ***Qualis A1***
- ***Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images.*** In: IJCNN - IEEE International Joint Conference on Neural Networks, 2018, Rio de Janeiro. Proceedings of the 2018 IEEE International Joint Conference on Neural Networks (IJCNN-2018), 2018. p. 1-8. - ***Qualis A1***

The following works were published using the same methodology and model basis proposed in this present Thesis, with the use of street-level images for estimation of other urban latent variables, specifically crime incidence rates:

- *Investigating Crime Rate Prediction Using Street-Level Images and Siamese Convolutional Neural Networks.* In: Communications in Computer and Information Science (PRINT), v. 720, p. 81, 2017. After the presentation at the *First Latin American Workshop on Computational Neuroscience* (LAWCN). - *Qualis B5*

- *Towards Crime Rate Prediction through Street-level Images and Siamese Convolutional Neural Networks.* In: ENIAC - National Meeting of Artificial and Computational Intelligence, 2017, Uberlândia. Proceedings from National Meeting of Artificial and Computational Intelligence. Porto Alegre: SBC, 2017. V. 1. p. 1-8. - *Qualis B4*

2 FOUNDATIONS AND RELATED WORKS

2.1 Computer Vision Designed Descriptors

Designed descriptors are a set of computer vision algorithms and techniques developed to obtain features or concise representations from images to solve specific problems such as object detection, image classification and scene recognition (SRINIVAS et al., 2016). Notable examples of descriptors are *Spatial Envelopes* or *GIST* from OLIVA; TORRALBA (2001) and *Histogram of Oriented Gradients* (HOGs) from DALAL; TRIGGS (2005). Those were the standard choice of computer vision techniques before the advent of Deep Learning approaches, where feature vectors are learned by deep convolutional architectures. The next Subsection 2.1.1 briefly describes the descriptor *GIST* and Subsection 2.1.2 describes the *HOGs*.

2.1.1 Spatial Envelopes (GIST)

Proposed by (OLIVA; TORRALBA, 2001), the global descriptor Spatial Envelope or *GIST* descriptor is a low dimensional representation of an image without requiring low-level or mid-level object segmentation i.e. without requiring segmenting and processing individual objects or regions. According to the authors, the "gist" is "*an abstract representation of the scene that spontaneously activates memory representation of scene categories*", such as a beach, a city, a forest.

Following (OLIVA; TORRALBA, 2001) idea, the spatial structure of an image scene can be represented as a set of perceptual dimensions: naturalness, openness, roughness, expansion and ruggedness, and this set of dimensions can be estimated using spectral information. There are different models for GIST descriptors, including the original (OLIVA; TORRALBA, 2001)'s GIST, where an image is divided into 4x4 equally-sized non-overlapping regions, in which the magnitude spectrum of *Windowed Fourier Transform* (WFT) is computed. Next, the feature dimension of the resulting WFT is reduced using principal component analysis (PCA). It is usually applied to scene recognition problems. GIST descriptors were used in the feature extraction methodology of DOUZE et al. (2009), FANG; SANG; XU (2013), KHOSLA et al. (2014), NAIK et al.

(2014) and ORDONEZ; BERG (2014).

2.1.2 Histogram Of Oriented Gradients (HOGs)

The HOGs were proposed by (DALAL; TRIGGS, 2005) to characterize the appearance and shape of an object in an image by the distribution of local intensity gradients or edge directions. In short, the authors of HOG method divide an image window into small cells, compute the gradients of the cells, distribute weights between spatial and orientation features and accumulate a local 1-D histogram of the gradient directions over the pixels of this cell. Also, each cell can belong to a block in which all cells are normalized according to an accumulated "energy". HOG descriptors were used in street-level images feature detection by the authors (DOERSCH et al., 2012), (ARIETTA; AGRAWALA; RAMAMOORTHI, 2013), (ARIETTA; EFROS, 2014), (KHOSLA et al., 2014), (PORZI et al., 2015), (NAIK et al., 2014).

2.2 Deep Convolutional Neural Networks

According to GOODFELLOW; BENGIO; COURVILLE (2016), Deep Convolutional Neural Networks (Deep ConvNets) (LECUN et al., 1998) are neural networks that use convolution operations, in one or more layers, for processing data with grid-like topology. They learn attributes or features hierarchically, i.e., performing feature learning, which best represents the grid-like data for a given task, e.g., object classification in images. The concept of depth or deep learning on ConvNets, according to the authors, refers to the methodology of breaking a learning problem into a series of more simplified mappings, described by different layers.

The traditional methodology for pattern recognition problems in computer vision involves the use of hand-designed feature extraction, to concentrate the relevant input values, followed by a trainable classifier algorithm, with the disadvantage of becoming increasingly complex whenever a new feature extraction method was proposed (LECUN et al., 1998; SRINIVAS et al., 2016). Later, during the competition of the *ImageNet Large Scale Visual Recognition Challenge* in 2012 (RUSSAKOVSKY et al., 2015), KRIZHEVSKY; SUTSKEVER; HINTON (2012) proposed a model based on convolutional neural networks (ConvNets) that surpassed the results of other models based on classic computer vision techniques in the task of classifying an image into 1 of 1000 labels. The proposed model was the *AlexNet* architecture, which made ConvNets the state of the art in computer vision models. The feature learning methodology of Deep ConvNets contrasts with feature design methodology in computer vision classic approaches, such as *GIST* (OLIVA; TORRALBA, 2001), "*Scale-invariant feature transform*" *(*SIFT) (LOWE, 2004) and *HOG* (DALAL; TRIGGS, 2005).

2.2.1 Generalities of Deep ConvNets

Deep ConvNets are composed of convolutional layers, that apply some convolution-family operation in data, together with layers that execute other operations such as pooling, batch normalization, and activations, in a hierarchical arrangement following a concept based on the representation of the cat's primary visual cortex, composed of "simple fields" and "complex fields". The "simple fields" are convolutional layers that are responsible for learning basic shapes such as borders and contours, and "complex fields" are pooling layers responsible for learning a collection of these basic forms, previously found, to identify parts of objects (GOODFELLOW; BENGIO; COURVILLE, 2016; HUBEL; WIESEL, 1962; SRINIVAS et al., 2016).

In convolutional layers, the *Convolution* is a linear operation between two functions $f(x)$ and $g(x)$ to produce a third function $h(x)$ that describes overlapping regions or modifications during several multiplications between $f(x)$ and a delayed $g(x)$ function through time t (BRACEWELL, 1986; O'HAVER, 1997).

The mathematical description of the general convolution, that is observed through time, is described in Equation 1,

$$(f * g)(t) = \int_0^t f(\tau)g(t-\tau)d\tau \tag{1}$$

where the integral of the function $f(\tau)$ multiplied by a spatially reversed function $g(t-\tau)$ observed with τ through time t gives the convolution function $(f*g)(t)$ between $f(t)$ and $g(t)$ (BRACEWELL, 1986; O'HAVER, 1997; HOSSACK, 2016; GOODFELLOW; BENGIO; COURVILLE, 2016).

In machine learning, the function $f(t)$ is referred as the grid-like multidimensional input data, such as an image, and the function $g(t)$ that multiply $f(t)$ by means of shifting through time is know as *convolution kernel*, or simply *kernel*, and the output of $(f*g)(t)$ is the *feature map*. In addition, the convolution operation is executed in more than one dimension, i.e. axis, at a time, and most common implementations present a different convolution-family operation know as *cross-correlation*[1], where some transformations in *kernel* function such as "flipping" are not used (GOODFELLOW; BENGIO; COURVILLE, 2016).

The Equation 2 describes the cross-correlation operation used in most neural network libraries, for input function I and kernel K with i, j displacements as,

$$(I \star K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \tag{2}$$

defining the sum of the product in both input axes, from $I(m,n)$ input function,

[1] According to GOODFELLOW; BENGIO; COURVILLE (2016), there is a convention in machine learning field of calling both convolution and cross-correlation as "convolution".

displaced in i, j units, and $K(m, n)$ kernel function, without reversal. While the learning process is performed, the kernel K - a multidimensional array of parameters - is adapted while loss function minimization criteria, to obtain the feature map from $(I \star K)(i, j)$ that best fit the problem criteria (BRACEWELL, 1986; GOODFELLOW; BENGIO; COURVILLE, 2016).

According to LECUN et al. (1998) and GOODFELLOW; BENGIO; COURVILLE (2016), the motivation for the use of convolutional layers instead of ordinary fully-connected layers are: (i) the possibility of reducing the connections, resulting in sparse weighting, (ii) parameter sharing and (iii) equivariance.

The convolutions operations perform a series of linear activations in the convolutional layers. The pooling layers in Deep ConvNets are responsible for executing a summary statistic in a particular location in output, replacing this output value by the statistic. It also has nonlinear operators, such as the traditional sigmoid, hyperbolic tangent (*tanh*), *rectified linear unit (ReLU)*, and *Leaking ReLU* functions, which are added to the output of convolutional layers, to insert a nonlinearity to the convolution operation, generalizing better to real data. The final architecture is trained using an optimization algorithm, such as *Stochastic Gradient Descent* (SGD) (BOTTOU, 2010), to find the best weights for all blocks that best fit the target task, minimizing a loss function (SRINIVAS et al., 2016).

The general structure of Deep ConvNets, taking *LeNet* and *AlexNet* (LECUN et al., 1998; KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architectures as basic references, is composed of blocks that are organized following a concept based on the representation of the cat's primary visual cortex, composed of "simple fields" and "complex fields" (HUBEL; WIESEL, 1962; SRINIVAS et al., 2016). In simplified terms, in ConvNets' "simple fields" are hidden layers that are responsible for learning basic shapes such as borders and contours, and "complex fields" are layers responsible for learning a collection of these basic forms, previously found, to identify parts of objects (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figure 1 represents the *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture, presented here for the purpose of exemplifying the operation of Deep ConvNets. The "simple fields" respond to oriented edges and grids, similar to convolutional filters, and the "complex fields" are sensitive to edges and grids, but with spatial invariance, i.e. don't vary if they are translated into space (HUBEL; WIESEL, 1962; SRINIVAS et al., 2016). The "simple fields" gave the idea of "convolutional layers" - "C" blocks represented in Figure 1 - responsible for learning a set of convolutional cores or filters. The "complex fields" are implemented as "Max-Pooling layers" - "P" blocks in Figure 1 - which perform the max-pooling operation responsible for applying a maximization filter, replacing each $n \times n$ region with its highest value. This operation selects the highest activation in the region, contributing to spatial invariance (as in complex

fields) and decreases the size of the parameters to be learned by the next layers.

New architectures, in addition to the *AlexNet*, are being developed, driven mainly by the need to obtain better results in applying benchmark datasets and competitions such as the *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) (RUSSAKOVSKY et al., 2015). The result of these competitions, usually involving image classification, and supervised machine learning, contributes to new approaches in building these architectures. Also, the use of Deep ConvNets may differ, depending on the task to which they are being applied.

2.2.2 Transfer Learning

According to SRINIVAS et al. (2016), the use of Deep ConvNets as a computational vision methodology can be done by training a complete architecture "from scratch[2]", or by using a previous developed architecture and pretrained weights, technique called *transfer learning*.

Transfer learning works by exploring the ability of first convolutional layers to generate attributes common to most natural images (YOSINSKI et al., 2014). Thus, pretrained ConvNets can use (i) *fine-tuning*, when using previous architecture and pretrained weights, and updating gradients from all or only some layers, i.e., "*frozen weights*" during training. In addition, these layers can be removed, replaced, or added to the original architecture to suit the new network task; or (ii) using activations of a previously trained ConvNet as an image feature extractor, similar to designed descriptors methods (YOSINSKI et al., 2014; SRINIVAS et al., 2016).

When the proposed task for architecture is very different from the general task of already built and previously trained architectures, it is necessary to train the complete

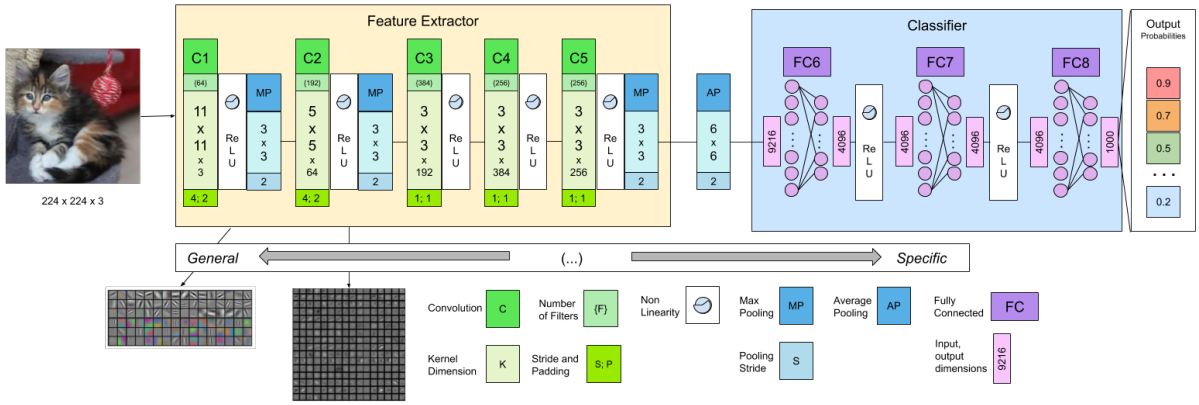


Figure 1 – Deep Convolutional Neural Network model example depicting *AlexNet* architecture. Adapted from KRIZHEVSKY; SUTSKEVER; HINTON (2012); SRINIVAS et al. (2016).

[2]Term to inform that no previous training weights from different datasets were used to initialize the training process.

Table 1 – Main benchmark datasets for training and validating *Deep* ConvNets.

Dataset	Author	Overview	Total Images
MNIST	(LECUN et al., 1998)	Manually written characters.	70.000
CIFAR-10/100	(KRIZHEVSKY, 2009)	Small images ($32 \times 32 \times 3$) labeled with 10 and 100 classes of distinct natural elements.	80 million
ImageNet	(DENG et al., 2009)	Images labeled according to WordNet hierarchies of approximately 1000 classes.	3.2 million
MS-COCO	(LIN et al., 2014)	It has approx. 91 different object types, totaling 2.5 million labeled instances.	328.000
OpenImages	(KRASIN et al., 2017)	It has labels in *bounding boxes* and global labels.	9 million
SUN	(XIAO et al., 2010)	Intended for computer vision problems involving categorization and scene comprehension. It has 899 categories.	130.519
Places	(ZHOU et al., 2014, 2018a)	Intended for categorization and understanding of scenes. It has 434 categories.	10 million

Deep ConvNet text in all its layers, using a specific dataset for the task. This methodology is costly because it requires a large amount of training data to obtain good results (SRINIVAS et al., 2016).

2.2.3 Benchmark Datasets

Benchmark datasets, in the computer vision research field, are databases containing a large number of previously labeled images, specifically designed for the training and validation of Deep ConvNet architectures, applied to computer vision problems. The primary research datasets were grouped in Table 1. Among the datasets listed, the most used imagery set in state of the art, with transfer learning techniques, is the *ImageNet* dataset. *Places* dataset is most used in related works of the Visual Computational Sociology area since its main goal is to categorization and understanding of natural scenes.

2.2.4 Architectures Implementations

Among the developed Deep ConvNet architectures that presented the best performances in computer vision tasks competitions, e.g. ILSVRC, according to KARPATHY (2017), those presented in Table 2 stand out. In addition to these architectures, recent implementations have also been added, which somehow optimize consolidated proposed architectures.

Table 2 – Notable Implementations of Deep ConvNet Architectures.

Architecture	Author	Overview
LeNet	(LECUN et al., 1998)	Introduced the successful use of Deep ConvNets for digit recognition.
AlexNet	(KRIZHEVSKY; SUTSKEVER; HINTON, 2012)	Popularized the use of Deep ConvNets in computer vision problems after the results obtained in the ILSVRC (RUSSAKOVSKY et al., 2015), making it the standard methodology in the field.
VGGNet	(SIMONYAN; ZISSERMAN, 2015)	Containing 16 convolutional layers, the authors have shown that small depth filters in the network are important components for better performance.
Inception-v3	(SZEGEDY et al., 2015)	Optimized computational resources compared to previous architectures.
ResNet	(HE et al., 2016)	Implemented optimizations with normalization of parameters between layers. Has different versions varying number of layers. Is the state of the art between current implementations.
SqueezeNet	(IANDOLA et al., 2016)	Small CNN, which achieves the same accuracy of architecture *AlexNet* with fewer parameters.
Inception-v4	(SZEGEDY et al., 2017)	Proposed combination of *ResNet* and *Inception* architecture.
DenseNet	(HUANG et al., 2017)	With a *feed-forward* approach, where all layers are connected together, seeks to correct various problems other architectures, e.g. *vanish-gradient*, reducing the number of parameters used.

2.3 Visual Computational Sociology

Computer vision and machine learning models have been used extensively for the discovery of environment-related attributes in street-level images. According to DUBEY et al. (2016), the visual analysis of scenes in urban images has different objectives, such as perception responses, understanding of the characteristics of cities through the urban visual scene, the study of the connection between urban appearance and socioeconomic factors, and the comparison between different environments. Although most of works in the Visual Computational Sociology use street-level imagery to infer latent variables, the use of aerial photography and satellite images with computer vision models is proposed for land use classification, poverty incidence identification, geo-localization and house prices. In the next Subsections 2.3.1, 2.3.2, 2.3.3, works using street-level, aerial and satellite imagery, and fusion models with aerial and street-level

images, respectively, are summarized to cover the use of computer vision models to infer latent variables in Visual Computational Sociology models.

2.3.1 Street-level Imagery Models

In "*What makes Paris look like Paris?*", DOERSCH et al. (2012) proposed a methodology to automatically extract visual elements from street-level images, such as windows, balconies, traffic signs, and lamps, that are representative for a specific geographical region. In "*On Relating Visual Elements to City Statistics*", ARIETTA; AGRAWALA; RAMAMOORTHI (2013), applied the methodology of DOERSCH et al. (2012) to estimate probabilities of social and economic statistic occurring in an area, based on street-level images. The authors collected georeferenced statistics with latitude and longitude coordinates of (i) occurrence of theft, (ii) housing prices, (iii) presence of graffiti and (iv) presence of trees, from the U.S. cities San Francisco, Oakland, Seattle, Chicago, Los Angeles, Boston and Philadelphia. The authors applied the patch clustering methodology to street-level imagery, aiming to discover discriminative patches from the statistics. Next, authors used Support Vector Regression (SVR) (DRUCKER et al., 1997), by learning a set of weights over the detection scores for each visual feature. Combining the weighted detection scores, the authors computed the probability of observing the statistic based on visual features detected in the image.

Later, ARIETTA; EFROS (2014) extended ARIETTA; AGRAWALA; RAMAMOORTHI (2013) work and applied the proposed prediction algorithm in US cities, discovering that there was a predictive relationship between visual elements from an environment and non-visual variables like crime and theft rates, housing prices, population density, graffiti density and perception of danger. They compared the use of HOG (DALAL; TRIGGS, 2005), and color descriptors with the Deep Convolutional Neural Networks (ConvNet) activations from the fifth convolutional layer (C5) of the *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture, and concluded that the HOG + color descriptors were more visually consistent but captured less from the city semantics. The authors also presented a scalable distributed processing framework to speed the extraction of visual elements from images and implemented their model to relate visual city attributes to any geolocated statistic presented in that city.

In "*Looking Beyond the Visible Scene*" (KHOSLA et al., 2014), the authors explored the ability to use street-level imagery that contains urban visual scenes to predict the distance of surroundings establishments such as hospitals and fast-food restaurants, and crime rates using the visual scene of a location. The authors applied HOGs and Deep Convolutional Neural Networks (ConvNet) activations from the seventh fully-connected layer (FC7) of *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture as descriptors. They relate the street view images with the distance of the closest establishments and trained a linear Support Vector Regression (SVR)

(DRUCKER et al., 1997) algorithm on the image features obtained by each descriptor. The results achieved in finding hospitals and fast-food restaurants ranged from 0.58% to 0.61% of average accuracy, and 72.0% for crime rate prediction.

In "*Learning High-level Judgments of Urban Perception*", (ORDONEZ; BERG, 2014) proposed the use of an automatic prediction of the sense of places, i.e. the perception about a location, using the *Place Pulse 1.0* dataset (SALESSES; SCHECHTNER; HIDALGO, 2013), to predict the human judgments of safety, uniqueness, and wealth of a location. Their model was divided into a machine learning binary classifier e.g., safe or non-safe and a regression model, that approximates the quantitative rate values from the *Place Pulse* dataset. The authors proposed the use of GIST (OLIVA; TORRALBA, 2001), SIFT (LOWE, 2004) descriptors with Fisher Vectors (PERRONNIN; SANCHEZ; MENSINK, 2010) encoding, and Deep ConvNet activations from *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) FC6 layer. The results showed that the use of the FC6 layer as descriptors surpassed the use of classic computer vision techniques.

The authors NAIK et al. (2014) in "*Streetscore: predicting the perceived safety of one million streetscapes*" proposed the use of street view images to predict the perception of safeness of an environment. The ground truth data was the *Place Pulse 1.0* dataset (SALESSES; SCHECHTNER; HIDALGO, 2013), gathered by user ranking. Authors used different feature extraction methods and chose the best performance features: GIST (DALAL; TRIGGS, 2005), Geometric Texton Histograms, and Geometric Color Histograms. Next, the image features obtained with the methods were used to train a predictor using Support Vector Regression (SVR) (DRUCKER et al., 1997) algorithm. The features were used alone and combined (named by the authors as *StreetScore*). The achieved accuracy of StreetScore binary safe or non-safe classifier was 78.42%.

In "*Predicting and Understanding Urban Perception with Convolutional Neural Networks*", (PORZI et al., 2015) proposed the use of Deep Convolutional Neural Networks (ConvNets) to predict human perception responses of safeness from a place automatically. The authors proposed a Deep ConvNet architecture based on *AlexNet* architecture (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), where the first to second and third layers are maintained and latent detectors layers are attached to the previous feature extraction layers. The safety predictor is a linear function applied to the output of the latent detector. The authors used the *Place Pulse 1.0* dataset without transforming the user rates in ground truth rankings, obtaining as the most relevant results among all experiments, 70.25% of accuracy with the proposed Deep ConvNet against 66.37% of HOG descriptors and SVM classifier.

In "*Deep Learning the City: Quantifying Urban Perception At A Global Scale*", (DUBEY et al., 2016) proposed two Siamese-like Deep ConvNet architectures

(CHOPRA; HADSELL; LECUN, 2005), named *Streetscore-CNN* (SS-CNN) and *Ranking SS-CNN* (RSS-CNN) to predict the human judgment of the six attributes with pairwise comparisons. The SS-CNN is trained for binary classification and the RSS-CNN network learns an additional set of weights to retrieve a ranking distribution instead of a binary classification. They proposed different approaches using Transfer Learning techniques, initializing the feature extraction layers with pretrained weights from *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), *PlacesNet* (ZHOU et al., 2014) and *VGGNet* (SIMONYAN; ZISSERMAN, 2015), obtaining 72.4% with SS-CNN and 73.5% with RSS-CNN of accuracy training with VGGNet architecture.

De Nadai et al. (2016) in "*Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life*" proposed a "liveliness," i.e., level of activity estimator for neighborhoods in Rome and Milan, Italy, using street-level imagery related to population density. The authors used categories of population density obtained from mobile phone activity data to proxy the liveliness in the neighborhoods. The safeness prediction model was implemented by fine-tuning the ConvNet *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) trained with *Places205* (ZHOU et al., 2014) and retraining with *Places Pulse 1.0*. By transferring knowledge, the resulting network could predict the appearance of safety in the new images. The authors found that the appearance of safety is positively correlated with higher densities of people present in neighborhoods per area.

KANG; KANG (2016), in "*Urban Safety Prediction Using Context and Object Information via Double-Column Convolutional Neural Network*", propose the prediction of safety scores of US cities, using the fusion of context and object information via double-column Deep ConvNet, being one column for context information extraction and another column for object information extraction. For context information, authors use the whole street-view image, and object information is extracted from the highest saliency patch. Ground truth safety scores were obtained from *Place Pulse 1.0* (SALESSES; SCHECHTNER; HIDALGO, 2013) dataset. Authors used a Deep ConvNet based on *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture. Object context column receives the whole street-level image, and the object information column receives the most significant image patch. In the proposed method, vectors resulting the last max-pooling layers were concatenated, replacing the original *AlexNet* fully-connected layers to 8192, 4096 and 1 nodes respectively. As a result, the authors achieved root mean square error (RMSE) of 0.74, the best result between all experimented approaches, with 0.90 of Pearson correlation coefficient.

In "*Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US*" GEBRU et al. (2017) and "*Fine-Grained Car Detection for Visual Census Estimation*", (GEBRU et al., 2017) proposed a computer vision pipeline to predict several statistics outcomes e.g. race, education levels, income, voter preferences,

per capita carbon emission and crime rates, through the detection and classification of cars in Google Street View images and regression methods to correlate the quantities and classifications with social city statistics.

Next, GEBRU et al. (2017) takes the visual census estimation step, using the output of the car classifier to cluster different data related to the type of car, e.g., most expensive cars and high or low percentage of foreign cars and to estimate demographic statistics and voter preferences. The authors trained a logistic regression model to estimate education levels and race, and a ridge regression model to estimate income and voter preferences. The authors found strong correlations between the carmaker and race and between car body type and voter preference, suggesting that most Democrats own sedans, and Republicans prefer extended-cab pickup trucks.

LIU et al. (2017), in "*Place-centric Visual Urban Perception with Deep Multi-instance Regression*" presented a regression model based on a deep multi-instance approach method to predict safety scores from street-level images in specific regions. The authors propose the use of $20,000$ crime occurrence location points, with 8 street-level camera views directions in 5 different cities of the USA. Each location point is labeled with a weakly supervised method of assigned the crime labels by first estimating primary instances, and secondly, clustering the instances to label regions with the highly scored values. Next, the authors apply a deep convolutional neural network based on *AlexNet* architecture to predict crime scores. Authors achieved $R^2 = 0.81$ for their built dataset.

In "*StreetNet: Preference Learning with Convolutional Neural Network on Urban Crime Perception*", (FU; CHEN; LU, 2018), authors propose the use of Deep ConvNets to learn a perceptional crime rank, i.e., inference of certain types of crimes most likely to occur in an area, from street-level images, in US cities. The crime ranking labeling was conducted using the georeference, type, and timestamp of event, from official cities crime records. Crime types were ranked with a descending order, with timestamp crime occurring time nearest the street-level image timestamp. However, authors do not explain how they gather timestamps for street-level images. To evaluate their Deep ConvNet model, authors compare specific metrics for label-ranking results, reporting that their approach performs better than *AlexNet* and *VGG* architectures pretrained with *ImageNet* and *Places* datasets when designed for the same task. Furthermore, authors show that perpendicular street-level images perform better than using 4-direction, stating that there is heavy noise in parallel images, and reported the best classification results in a small radius, with sample points near events.

In "*Measuring social, environmental and health inequalities using deep learning and street imagery*", SUEL et al. (2019) addresses the problem of inferring categories of different latent variables from street-level images using the same proposed model. Authors gather Census information from London, U.K., such as mean income, occupancy

rating, education, self reported health, unemployment, and crime deprivation. Each latent variable was categorized in a 10-decile distribution, where decile one corresponding to the worst-off 10%, and decile 10 the best-off 10%, being one categorized decile of each latent variable attributed to a postcode location. For each location, street-level images in different orientations were obtained from Google Street View API (GOOGLE, 2017). Authors use *VGG16* (SIMONYAN; ZISSERMAN, 2015) architecture as a feature extractor, and fed each feature into three fully-connected layers, concatenating the last layer to classify images in a specific decile, for each latent variable. Best decile predictions obtained for mean income with mean average error $MAE = 1.1$ and Pearson correlation coefficient $r = 0.86$, and worst results using the model presented Pearson correlation coefficient $r = 0.57$ for crime deprivation, and mean Pearson correlation between all outcomes equal $r = 0.77$.

2.3.2 Aerial and Satellite Models

In the research article "*Combining Satellite Imagery and Machine Learning to Predict Poverty*", from JEAN et al. (2016), the authors use daylight satellite imagery obtained from Google Static Maps API (GOOGLE, 2018a) to estimate poverty regions in African countries, using as a proxy the ground truth information of nightlights intensities instead of georeferenced economic data. The authors propose the use of Deep Convolutional Neural Networks (ConvNets) pretrained with *ImageNet* (DENG et al., 2009) datasets and fine-tune it on daylight satellite images to predict nightlight intensities. This network is used next as a feature extractor for nightlight intensities from daylight images. Next, features extracted from this network are used as predictors in ridge regression models, with cluster-level data from a parallel survey performed to map poverty in the localities. Based on these extracted features, the models can predict expenditures and assets in the local regions, with r^2 ranging from 0.41 to 0.75.

Later, OSHRI et al. (2018) in "*Infrastructure Quality Assessment in Africa using Satellite Imagery and Deep Learning*", studied the use of Deep ConvNets to infer infrastructure quality in African countries using satellite images from *Landsat 8* and *Sentinel 1* and ground truth data for infrastructure from *Afrobarometer Round 6* (BENYISHAY et al., 2017). The authors' methodology consists of detecting infrastructure in studied regions as a multi-binary classification problem, where the input is a satellite image, and the outputs are binary labels indicating the quality of different types of infrastructure detected in the site. The authors used *ResNet* (HE et al., 2016) architecture pretrained with *ImageNet* (DENG et al., 2009). Modifications were made in the first convolutional layers, extending filters to more than three channels, initializing RGB channels with *ImageNet* weights and the non-RGB channels with *Xavier* initialization (GLOROT; BENGIO, 2010). They obtained the best results on the classification of electricity and sewerage, with AUROC values greater than 0.85, and infrastructure

elements that require zoom levels, e.g., market stalls, health clinics, police stations, from satellite images presented worst results.

ALBERT; KAUR; GONZALEZ (2017) in "*Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale*" propose the use of Deep ConvNets of architectures VGG-16 (SIMONYAN; ZISSERMAN, 2015) and ResNet (HE et al., 2016) to classify types of urban land use in satellite images from Google Maps with large scales, using *Urban Atlas* [3] dataset as ground truth for land use labels. The authors investigate the use of transfer learning in the ConvNets architectures with different domains, e.g. *ImageNet* (DENG et al., 2009), and related domains as *DeepSat* dataset (BASU et al., 2015), with different European cities, and performed the classification on satellite Google Maps image tiles with resolution 224×224 and zoom level of 17, labeled according to a 100×100 grid structure interpolated in *Urban Atlas* original polygons. As results, the authors found that *ResNet-50* architecture pretrained with *DeepSat* dataset and fine-tuned with satellite images collected for the experiments yielded better results in classification, a zoom scale of $250m$ performed better than $50m, 100m, 150m$ zoom levels, and a dataset composed of multiple cities for training, presented gains in reported accuracies, 70% to 80%, when classifying different cities, in contrast with training with a unique city.

In "*Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery*" (BENCY et al., 2017), authors propose the use of features obtained from satellite images from Google Maps in house pricing datasets, performing fine-tune in Deep ConvNets from *Inception v3* (SZEGEDY et al., 2016) architecture pretrained with *ImageNet* (DENG et al., 2009) to infer house prices. The authors compared the use of features obtained from Deep ConvNets with known methods in housing price and tested different zoom scales in satellite images, with several estimators, including linear regression, random forest, and multilayer perceptron (MLP) regressors. The authors found that house level attributes obtained from deep features, and neighborhood-level features from housing price dataset are complementary to infer price variations. Furthermore, the zoom level inspection in satellite imagery from Google Maps showed that BENCY et al. (2017) proposed method benefits from larger neighborhood areas.

NAJJAR; KANEKO; MIYANAGA (2018) in "*Crime Mapping from Satellite Imagery via Deep Learning*", the authors propose the use of Deep ConvNets to generate maps indicating urban crime rates using satellite imagery, obtained from Google Static Maps API (GOOGLE, 2018a) and crime occurrences from US cities Chicago, Denver and San Francisco, collected from their police departments. The authors distributed the crime occurrences inside a grid data structure, disposed of in the studied area, with corresponding regions of $900m^2$ each. Each cell received a safety score according to the sum of crimes committed within its boundaries, a methodology similar to related

[3] `https://www.eea.europa.eu/data-and-maps/data/urban-atlas`

works (ANDERSSON; BIRCK; ARAUJO, 2017). The labels of the crime rate were categorized into three levels, according to k-means clustering. Images were collected from Google API with different zoom levels, from 17 to 20, from cell center coordinates, and used with *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) ConvNet architecture, pretrained with *ImageNet* (DENG et al., 2009) and *Places205* (ZHOU et al., 2018b) datasets, and different models were trained to test the performance of transfer learning from different datasets and different zoom levels of satellite images. The authors report as a result the accuracy of 0.79 using zoom levels of 17 and *Places205* pretrained model.

Recently, "*Poverty Incidence Identification of Cities and Municipalities using Convolutional Neural Network as Applied to Satellite Imagery*" from MESINA; ISANAN; MADERAZO (2019) presents the use of Deep ConvNets to classify the wealth of developing countries through satellite imagery. In their work, authors use *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture to identify poverty levels in municipalities of the Philippines using Google Earth (GOOGLE, 2018b) satellite images. Poverty ground truth data for municipalities was obtained from the Philippines' economic and development authorities, and georeferenced coordinates of their boundaries obtained from GADM database[4]. The authors performed data augmentation in 100 original examples, obtaining a total of $72,000$ images, and wealth index labels were categorized in range of 10, i.e., $0-10$, $11-20$, until 41 and beyond, and changed the fully-connected 1000 classifier layer from *AlexNet* to 5 classes. As a result, the authors obtained an average accuracy of 0.84 in classifying wealth ranges.

In "*Predicting Food Security Outcomes Using ConvNets for Satellite Tasking*", GANGULI; DUNNMON; HAU (2019) proposed an ensemble of models to assist policymakers using food security metrics (FSM) in developing world countries. The first model consists of a Deep ConvNet that predicts FSM from satellite images correspondent to the studied regions, and the second model uses the resulting predicted values as reward grids in a reinforcement search problem for automated satellite direction modification algorithm. To predict FSM values from satellite images, authors trained a *VGG16* (SIMONYAN; ZISSERMAN, 2015) architecture with *DeepSat* (BASU et al., 2015) dataset to classify satellite images with land types and use labels. Next, authors extracted features from *SustLab*[5] images from the pretrained *VGG16* network, and built ridge regression models to predict FSM categories of values. Results obtained for the authors' models varied between 30.8% to 46.4% of test accuracy. Finally, the resultant grid structure with FSM categories was used as a reward grid in a satellite tasking problem. For more information on the subject, which is not addressed in this work, see reference.

[4] `https://gadm.org/`

[5] Proprietary dataset. Reference not provided by authors.

2.3.3 Street-level and Aerial Fusion Models

In the Scene Analysis research field, related to Visual Computational Sociology, models can benefit from the combined multi-view aerial and street-level images, in tasks such as object detection, scene understanding, reconstruction, and classification. According to LEFÈVRE et al. (2017), the varying viewpoints, different scales, illuminations, sensory modality, and possibly time acquisition from satellite and street-level images, present a challenge for models that merge information from such different domains. In this Subsection, related works that propose the use of aerial and street-level images combined are presented.

WORKMAN et al. (2017), in "*A Unified Model for Near and Remote Sensing*" addresses the issue of fusing aerial and street-level images in a Deep ConvNet model to predict latent variables, which authors denominate fusing "overhead" (aerial) and "ground-level" (street-level) to obtain "*geospatial functions*". For the ground-level feature images, authors use *VGG16* (SIMONYAN; ZISSERMAN, 2015) architecture pre-trained with *Places* (ZHOU et al., 2018b), with the resulting feature vector of 820 positions reduced to 50. The same architecture is used for overhead images, where authors use only the convolutional layers and reduce the dimensionality of each output. Next, the features are fused by average pooling the ground-level feature map with kernel size 6×6 and stride 2 and concatenating it with the overhead feature on the 7th *VGG16* convolutional layer. To predict the latent variables (geospatial function) using the ground-level and overhead concatenated feature maps, authors compute the *hypercolumn* (HARIHARAN et al., 2015) features for each $H \times W$ pixel location, which consists of outputs of all convolutional units above that location, stacked in one vector. This hypercolumn feature is then fed to a multilayer perceptron (MLP) to classify the categories of building, land use, and building age. The best results achieved were 77.40% of accuracy for land use classification.

In "*Integrating Aerial and Street View Images for Urban Land Use Classification*", CAO et al. (2018) presented approaches to merge street-level and aerial images using Deep ConvNets models to classify urban land use. The authors construct ground feature maps using semantic features extracted from street-level images and interpolating the features in the spatial domain, fusing street-level features, and aerial images sources from different views. To obtain the ground feature maps, authors used a pre-trained *ResNet* architecture (HE et al., 2016) with *Places365* dataset (ZHOU et al., 2018b) without the original classifier layer to extract feature vectors from the 4 cardinal Google Street View images, concatenating each feature vector. Next, each semantic feature vector in each point is interpolated to adjacent areas, using Nadaraya-Watson regression with Gaussian kernel to construct the ground feature maps, the same methodology presented in WORKMAN et al. (2017). To fuse the ground-level feature maps and aerial images, and further classify urban land use, authors propose

the use of an encoder-decoder based Deep ConvNet architecture, known as *SegNet* (BADRINARAYANAN; KENDALL; CIPOLLA, 2017). Their strategy uses two encoders, one for aerial images and other to the ground-level density maps, and fusion is performed by concatenating each corresponding feature maps produced by each convolutional layer using channel dimension. The authors' strategy was to search for the best convolutional layer to concatenate the ground-level and aerial features. As a result, they found that the central layer immediately before the 3rd. max-pooling layer achieved the best performance, and the overall result of 78.1% of accuracy in land use for Brooklyn, New York.

Moving from land use classification to more sociological perceptions, WANG et al. (2018) in "*Urban Perception of Commercial Activeness from Satellite Images and Streetscapes*" authors proposed the use of classic computer vision methods to predict commercial activeness of urban districts. Commercial locations were obtained from a *Point-of-Interest* dataset. Street-level images were obtained from Baidu API (BAIDU, 2019), while aerial imagery was obtained from Google Earth service (GOOGLE, 2018b). The proxy data for commercial activeness was defined as the sum of the user-generated comments on commercial entities. Authors performed the feature extraction of street-level and aerial images using *HOGs* (DALAL; TRIGGS, 2005), *GISTs* (OLIVA; TORRALBA, 2001), autoencoders and multifractal spectra, and applied *Bag-of-Features* (BoF) technique to obtain a histogram-based descriptor for the studied regions, that group similar visual features in histogram bins. The BoF descriptor, created from all feature extraction applied methods, is then used as predictors in a *Support Vector Regression* algorithm with RBF kernel to infer real values of commercial activeness. Accuracy is measured by *Mean Absolute Percentage Error* (MAPE), and authors achieved 59.83% to 60.17% of accuracy using all features combined.

In LAW; PAIGE; RUSSELL (2018), "*Take a Look Around: Using Street View and Satellite Images to Estimate House Prices*" authors propose the use of satellite image data and street-level images to improve the estimation of house pricing models, called "hedonic[6] price models" in London, U.K.. They propose non-linear and linear fusion methods of known housing attributes, e.g., age, size, and accessibility, as well as visual features from Google Street View (GOOGLE, 2017) and Bing Aerial images (MICROSOFT, 2019), used as proxies for visual desirability of neighborhoods. In their models, house prices to be predicted were set on a logarithmic scale, and housing attributes were normalized. Two main Deep ConvNets models were used to learn feature extractors for street-level and aerial images with a hedonic price model, training the proposed architecture end-to-end with a non-linear and linear fusion approach. In the non-linear approach, features from housing numeric attributes vector are concatenated to the vector-valued output of ConvNets from street-level and satellite images input. Next,

[6]Related to pleasant and unpleasant sensations.

in this same model, the resultant feature vector is input into a fully-connected 2-layer network with 60 and 30 nodes, respectively. The linear approach combines each Deep ConvNet for street-level and satellite images as a single network, with a direct scalar output, without non-linearity functions. The results reported by the authors shows that models using hedonic and urban imagery features yielded the best predictions, and the use of aerial images improves the prediction of models using street-level imagery. The mean square error (MSE) reported by authors with the full model, i.e., with hedonic, aerial and street-level features, was 0.08.

HOFFMANN et al. (2019) in "*Model Fusion for Building Type Classification from Aerial and Street View Images*", authors explore data fusion strategies for models that use aerial and street-level imagery, for building type classification problems. They propose the use of two fusion approaches, one being a geometric-level model fusion, and other a decision-level model fusion. In geometric-level approach, authors explore *VGG* (SIMONYAN; ZISSERMAN, 2015) architecture fine-tuned with *ImageNet* (DENG et al., 2009) with different setups, by stacking two-stream from last convolutional vector-values horizontally, and stacking the last dense layer vector-values vertically. In decision-level model fusion, authors also explore *Inception* (SZEGEDY et al., 2016) architecture fine-tuned with *ImageNet* and *Places365* (ZHOU et al., 2018b), first by blending decision layers from the independent models by taking the mean of their softmax layer vector-value result, and next by stacking decision layers with an additional machine learning algorithm, through concatenating the resulting softmax vector. For each building type annotation from Open Street Maps (OSM) (OpenStreetMap contributors, 2017), authors collected aerial images from Bing Maps API (MICROSOFT, 2019) with different zoom levels and street-level images from Google Street View (GOOGLE, 2017) pointing to each building, covering a variety of US cities. As results, authors found that better building type classification is achieved with fusion feature vectors from dense layers, and decision-level models yielded the best results, when highly detailed zoom level, e.g., 19, models are added to the ensemble, compared with models that contain only street-level imagery, achieving 76% of building type accuracy with decision-level model.

In the recently published work "*Google Street View Image of a House Predicts Car Accident Risk of its Resident*", KITA; KIDZIŃSKI (2019) proposed a methodology to predict car accidents using images of insurance company client houses, gather by Google's Street View (GOOGLE, 2017) and Static Maps API (GOOGLE, 2018a) services. The authors investigated current models that used zip-code as one of the attributes and showed that different house features could be found in the same zip-code. For that, they process more than $20,000$ addresses, manually annotating information about physical aspects of client houses, such as age, type, condition, and wealth of residents, from street-level images, and from aerial imagery they gather neighborhood

type, building density and street quality. Next, they investigate the importance of the created attributes for risk prediction models, comparing the state of the art insurance model with the new street image attributes. Results reported by authors indicate that the urban imagery derived attributes can be predictive of car accident risk, inexistent models, independently of the use of consolidated variables. Authors are also the first to address the issue of the indiscriminate and unauthorized use of street and aerial imagery data for social mining through deep convolutional neural networks and machine learning, that could predict private aspects of the residents and house owners.

2.4 Dengue Fever Estimation Using Environmental Attributes

Satellite sensory data are commonly used to obtain environmental attributes as predictors for estimation models of dengue, using regression analysis and machine learning algorithms. Several works proposed the use of continental and ocean temperature, relative humidity, vegetation indices, and cumulative precipitation, combined with socioeconomic variables (BUCZAK et al., 2012, 2014; TEURLAI et al., 2015; STOLERMAN; MAIA; KUTZ, 2016; ANGGRAENI et al., 2017; ASHBY et al., 2017; DEB et al., 2017; GUO et al., 2017; LAUREANO-ROSARIO et al., 2017; LI et al., 2017; SCAVUZZO et al., 2017; LAUREANO-ROSARIO et al., 2018). According to the mentioned authors, these attributes correlate with the increase or decrease of dengue incidence in a region and can be measured by remote sensing, usually performed by satellite. Working with a wide area, without differentiation of localities, is one of the problems with models that use satellite sensory data because they do not carry out fine-scale evaluations or focus on specific sites where the mosquito reproduces and lives, which contribute to its proliferation. This impairs actions to combat mosquitoes and disease prevention, as it ignores high-granularity aspects present in neighborhoods, streets, and homes.

Previous works on health behavior associated with urban characteristics investigated US Census Bureau building information. They manually collected attributes as predictors of sexually transmitted diseases (STD), premature mortality, body mass index (BMI), depression and mental health, and vector related infectious diseases such as dengue fever. Previous to computer vision models, urban visual characteristics to be used as predictors were obtained through manual inspection and classification on-site, or through video resources, as *ad hoc* research or for census data collection purposes. Notable works with such methodology are "*Broken Windows and the Risk of Gonorrhea*" COHEN et al. (2000), "*Neighborhood Physical Conditions and Health*" (COHEN et al., 2003) for STDs and premature mortality, mental health and depression in "*Stressful Neighborhoods and Depression: A Prospective Study of the Impact of Neighborhood Disorder*" (LATKIN; CURRY, 2003) and "*Relation between neighbor-*

hood environments and obesity in the multi-ethnic study of atherosclerosis" (MUJAHID et al., 2008) for BMI and environment association. Works involving the investigation of physical environmental characteristics related to the incidence of dengue fever are described next.

The need for a more specific view of propitious sites for dengue vectors to breed and live has prompted researchers to investigate environment attributes at the neighborhood, street, and residence levels. In "*Environmental factors and incidence of dengue fever and dengue hemorrhagic fever in an urban area, Southern Thailand*", THAMMAPALO et al. (2008) investigated with exclusively manual techniques of data collection, from on-site visits of health agents to localities, urban factors that could correlate with the incidence of dengue in Songkhla county, southern Thailand. As possible attributes, the authors manually collected data on characteristics of the buildings, such as type, social function, and the presence or absence of appropriate water distribution, sewage, and garbage collection systems. As a result, high correlations were found between indices of dengue and houses built with bricks, stores interconnected with air passage wells, and the absence of an appropriate garbage collection system.

In "*Using geographically weighted regression (GWR) to explore varying spatial relationships of immature mosquitoes and human densities with the incidence of dengue*" (LIN; WEN, 2011), the authors propose a predictive model of dengue incidence using demographic density in Kaohsiung and Fengshan cities, Taiwan and *Aedes* vector larval density indexes by inhabitants of the region Breteau Index which estimates the density of *Aedes* mosquitoes per 100 houses or dwellings in the vicinity. The data was collected manually, from visits of health agents in the neighborhoods and places studied, trained to collect data. The authors used Ordinary Least Squares (OLS) and GWR regression models to analyze the correlation between the chosen attributes and the dengue indices of the regions. As a result, the authors explained that high rates of dengue were associated with high rates of *Aedes* densities by residents in some regions. However, in other areas, high rates of dengue were correlated with low *Aedes* indices, by residents, leaving the study inconclusive.

In "*Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing-based case study*" (KHORMI; KUMAR, 2011), the authors investigated the relationships between the quality of the neighborhoods, population, and risk of dengue incidence in Jeddah, Saudi Arabia, using urban satellite imagery. To classify the quality of the city's neighborhoods as "low" or "high", the authors analyzed the images manually according to empirical criteria of the region, measuring the width of the streets, the density of houses by area, and the roof area per house. They used Geographically Weighted Regression (GWS) to estimate dengue risk, using the total number of dengue cases in each district of Jeddah. The population in each district and the environmental attributes that characterize the neighborhoods

were set as predictors. As a result, the model proposed by the authors revealed that there were strong positive associations between the attributes of population density and quality of the neighborhoods and the probability of contamination with dengue. Districts with low quality, according to their criteria, had a 71% risk of contamination with dengue compared to high-quality ones. In addition, low population densities indicated a low risk of dengue outbreaks in the region.

Subsequently, in "*Near real-time characterization of urban environments: a holistic approach for monitoring dengue fever risk areas*", SARFRAZ; TRIPATHI; KITAMOTO (2014) proposed a method to classify land-use types to investigate the incidence of dengue in Phitsanulok Province, Thailand, using urban satellite images of the Google Earth services (GOOGLE, 2018b). The authors extracted buildings from satellite imagery using pixel-based image sorting techniques. Attributes related to land use were obtained through spatial analysis. As a result, the authors showed that the majority of dengue cases were found near irregular, compact, and dense buildings, located near roads, surrounded by dense vegetation. Buildings in low-density housing, with bare soil and medium vegetation, had low rates of dengue.

In "*Assessment of land use factors associated with dengue cases in Malaysia using boosted regression trees*" (CHEONG; LEITÃO; LAKES, 2014), the authors used satellite imagery to study the relationships between land use characteristics and dengue fever occurrences focusing on large areas. They used land features manually cataloged through on-site field trips, such as agricultural land use, water bodies, and the presence of forests as environmental predictors. Through Boosted Regression Trees (BRT) models, the authors showed that human settlements and water bodies land factors contributed the most for dengue risk prediction.

2.4.1 Dengue Assessment With Computer Vision and Urban Imagery

Recently, authors have used computer vision techniques and urban aerial photographs obtained by unnamed aerial vehicles (UAV) devices or "drones" to assist in the classification of attributes that influence the occurrence of dengue. In "*Supplementing Dengue via the Drone System*" (AMARASINGHE et al., 2017; SUDUWELLA et al., 2017), the authors used aerial photographs obtained by UAVs with Histogram of Oriented Gradients (HOG) descriptors (DALAL; TRIGGS, 2005) and Support Vector Machines (SVM) for learning and classifying areas with water. In this work, the authors propose the detection of possible areas with water retention using images obtained with the UAV flying over specific sites indicated by health agents in Sri Lanka provinces.

Similar methodology is presented in CASE (2017) master thesis, "*Mosquitonet: Investigating the Use of Unmanned Aerial Vehicles and Neural Networks for Integrated Mosquito Management*", where the author proposes the use of Deep Convolutional

Neural Networks (Deep ConvNets) to classify the presence of *Aedes Albopictus* and objects related to their presence, using aerial images obtained from UAV devices, in New York City, USA. Common backyard objects, e.g., containers, umbrellas, and flower pots, were previously manually mapped, together with the occurrence of *A. Albopictus*, and used as ground truth data. Among other object-related experiments, a Deep ConvNet of *VGG16* (SIMONYAN; ZISSERMAN, 2015) architecture was used, pretrained with *ImageNet* (DENG et al., 2009), to the binary classification of presence or not of *A. Albopictus*. The small samples for the positive or negative presence of the mosquitoes foreclose an accuracy report in results, but the author claims that the Deep ConvNet activations showed objects in images similar to the manually collected for classifying a household.

Motivated by the need to obtain fine-scale street-level data in places where access and retrieval is difficult, in "*A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Petit-Goave, Haiti*", CURTIS et al. (2013) proposed a method of obtaining spatial data at the street and residential level using video, named *Spatial Video*, with the objective of collecting information present on this scale, including places with standing water, garbage accumulation, presence of animals, specific characteristics of the local population and other cultural phenomena. Health experts then review these videos, and features of interest to the sites are detected by those experts and manually georeferenced for future analysis of the disease on the spot.

Subsequently, in "*Supporting local health decision making with spatial video: Dengue, Chikungunya and Zika risks in a data-poor, informal community in Nicaragua*" CURTIS et al. (2017) applied the *Spatial Video* technique to identify street-level attributes to map possible *Aedes* mosquito breeding sites where agents execute Spatial Video in different streets to produce risk maps of diseases caused by the same mosquito, using the same *Spatial Video* technique, (KRYSTOSIK et al., 2017) add the "geonarratives", where health agents accompany the vehicle performing the *Spatial Video* narrating the attributes of interest, e.g., accumulation of trash, floods, residences with no proper conditions, which are found in the streets and monitored neighborhoods. These narratives are peer-reviewed and manually georeferenced for future risk mapping of mosquito-borne diseases and lack of sanitary conditions.

Spatial Video techniques and geonarratives manually obtain the attributes of interest in places where aerial photographs and satellite images are not reachable. These attributes are obtained manually from the review and investigation by a specialist. The automation of this process of obtaining attributes has been recently researched through the application of computer vision techniques. In "*Study of CNN Based Classification for Small Specific Datasets*", LE et al. (2018) proposed, to aid in the mapping of dengue, the automatic detection and classification of attributes of interest e.g., accu-

mulated litter, puddles, and pagoda constructions into images using ConvNets. The work of LE et al. (2018) detects and classifies attributes for dengue mapping, and is not a predictive model of dengue.

Until the published works of ANDERSSON; BIRCK; ARAUJO (2018); ANDERSSON; CECHINEL; ARAUJO (2019), no related work has been found that automatically detects attributes in street-level urban images and satellite imagery and also performs dengue rates estimation using computer vision techniques and machine learning.

3 HYPOTHESES

3.1 General Objective

The general objective of this Thesis is to use street-level image domain in conjunction with aerial imagery to estimate dengue incidence rates in urban regions.

3.2 Enrolled Hypotheses

The following specific hypotheses are proposed for the present Thesis:

- **H. 1.** Using street-level urban imagery is as effective as using urban satellite imagery for estimating dengue-associated rates.
- **H. 2.** The combination of street-level and urban aerial imagery contributes to better results in estimating dengue-associated indices compared to the independent use of each image domain.
- **H. 3.** Dengue estimation models trained in a single city are able to generalize estimation to other cities.
- **H. 4.** It is possible to extract, from the models that use street-level and aerial images data, informative attributes for dengue prevention not yet explored in urban areas that correlate with the occurrence of dengue.
- **H. 5.** Deep Convolutional Neural Networks (Deep ConvNet) are suitable for use in the proposed models because they can learn and extract attributes from urban images, and estimate dengue-associated rates using these attributes.

4 METHODOLOGY

4.1 Data Collection, Pre-processing and Description

4.1.1 Dengue Fever Data

The main city chosen for the investigation and implementation of the proposed approaches, in the present Thesis, was Rio de Janeiro, located in the southeast region of Brazil. Two more Brazilian capitals that appear in the *Rapid Survey of Aedes aegypti Infestation Indexes (LIRAa)* performed in 2017 (Ministry of Health, 2017) were also chosen for testing hypotheses and additional experiments, being São Paulo, capital from the southeastern state of São Paulo, and Salvador, capital from the northeastern state of Bahia. The locations of the chosen cities are depicted in Figure 2, in the details (a) and (b) over the Brazilian territory.

In Brazil, dengue fever (DF) notification records are maintained through the SINAN (Notification of Grievance Information System) database system, managed by the Health Surveillance Secretariat, of the Brazilian Ministry of Health. In general, dengue fever records from SINAN datasets contain information about grievance collected by health field agents at the site of infection through a questionnaire conducted with the possibly affected individuals. Common fields are the type of grievance, investigation start date, personal information about the patient such as residence address,

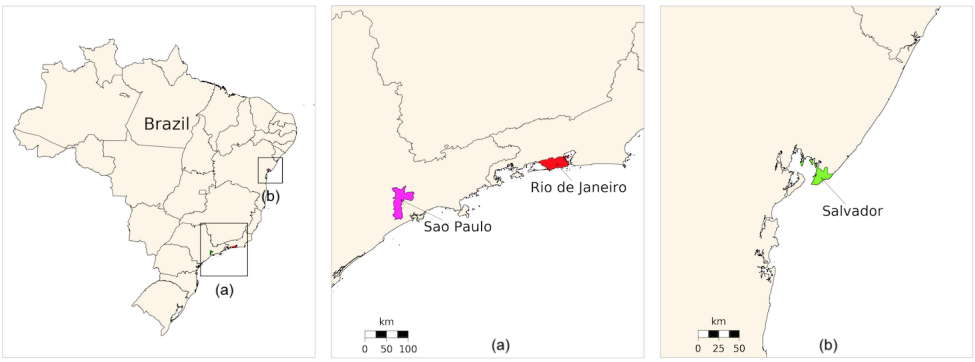


Figure 2 – Depiction of the geographic location of the Brazilian studied regions: in (a) Rio de Janeiro and São Paulo; and (b) Salvador.

Table 3 – Total dengue cases, autochthonous dengue cases and used records, for each studied city in the given year range.

City	Total	Autochthonous	Used	Year Range
Rio de Janeiro	226,478	226,478	226,478	2010 - 2014
São Paulo	58,156	43,048	42,445	2015 - 2017
Salvador	10,856	1,615	1,331	2015 - 2017

the serotype of dengue virus, autochthonous flag, i.e., cases where individuals were infected in the place they inhabit, and final classification (SUS, 2018).

For Rio de Janeiro city, SINAN's autochthonous dengue notification cases between the years 2010 and 2014 were obtained through the city's data portal "Datario". From this data, all cases were pre-processed by the Municipal Secretariat of Health, converting the occurrence addresses as georeferenced data, with latitude and longitude coordinates (DATARIO, 2017; SMS-RIO, 2017).

The SINAN's dengue notification cases for São Paulo and Salvador, between the years 2015 and 2017, were requested through the Electronic System Of Citizen Information Service (e-SIC) to the Brazilian Ministry of Health (CGU, 2017). Original records contain reported occurrence nominal addresses, and were converted to latitude and longitude coordinates using Google Geocoding API (GOOGLE, 2019a) for the proper use in this work. Some addresses could not be geographically resolved, therefore some autochthonous occurrences were not included in the final datasets.

For all cities, only the autochthonous cases were used, allowing the site of infection to be obtained more accurately from the available data. Rio de Janeiro city contains the majority of dengue fever reported cases, followed by São Paulo and Salvador. Table 3 enumerates total cases of dengue in each studied city, the number of autochthonous cases, and cases used between the year range of occurrences. The geographic distribution of the dengue fever autochthonous cases, in the given years' range, for each studied city, is depicted in Figure 3.

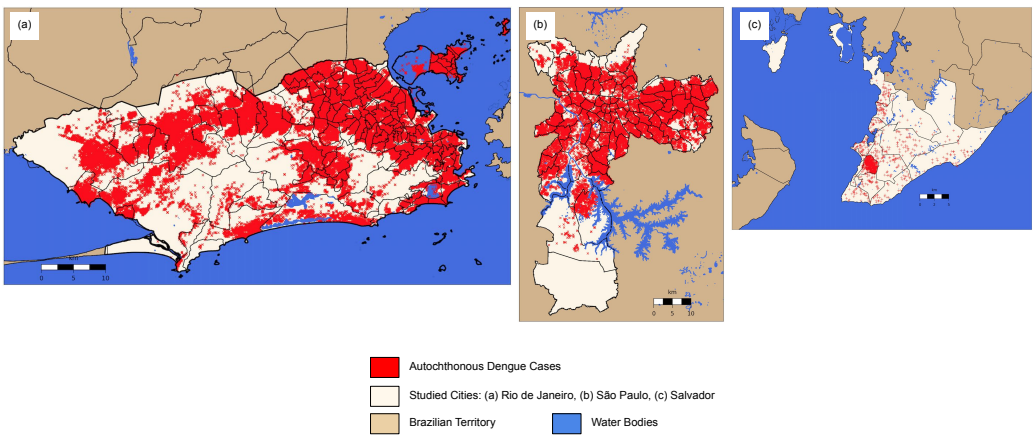


Figure 3 – Geographic distribution of autochthonous dengue cases in studied cities, between 2010 to 2014 for Rio de Janeiro, and 2015 to 2017 for São Paulo and Salvador.

4.1.2 Street-level and Satellite Data Imagery

4.1.2.1 Urban environment coverage

In order to obtain uniform coverage of the studied territories with urban images from Rio de Janeiro, São Paulo, and Salvador, *Environmental System Research Institute (ESRI) shapefiles* containing the shape of the streets were collected using *OpenStreetMaps Hot Export Tool* services (OpenStreetMap, 2018). The *shapefiles* contains geometry and attribute information related to spatial elements stored as a set of vector coordinates in a dataset (ESRI, 1998). They represent the roads and paths in cities with points, lines and polygons that shape the streets and their boundaries, i.e., the city limits, as depicted in Figure 4(a), with two-dimensional (2D) vertices composing the streets in Figure 4(b).

Preprocessing of *shapefiles* was performed using the QGIS tool to clip vectors outside city boundaries and to select only elements relative to paths and roads (QGIS Development Team, 2009). Next, all 2D vertices of street lines and polygons were transformed into latitude and longitude coordinates. From this georeferenced vertices, new points between them were created, using the *haversine* distance Equation 3,

$$a = \sin^2(\varphi_B - \varphi_A/2) + \cos\varphi_A * \cos\varphi_B * \sin^2(\lambda_B - \lambda_A/2)\theta \tag{3a}$$

$$c = 2 * \text{atan2}(\sqrt{(a)}, \sqrt{(1-a)}) \tag{3b}$$

$$d = R * c \tag{3c}$$

if the distance d satisfies the condition $d > 100m$, with $A(\varphi_A, \lambda_A)$ and $B(\varphi_B, \lambda_B)$ the original vertices from which the distance will be calculated, with φ the latitude and λ the longitude of each point respectively, and $R = 6378.137$ the equatorial radius of the earth in kilometers (GEONET, 2017; NASA.GOV, 2019).

Figure 4 – (a) Rio de Janeiro streets' shape, with the detail of 2D vertices in (b) composing the streets.

Table 4 – The number of georeferenced points computed for each studied city, covering all urban roads and paths in the territory.

	Rio de Janeiro	São Paulo	Salvador
Total City Points	407,305	576,556	112,146

The interpolated points created from the difference of vertices are depicted in Figure 5. Thereon, a dataset of georeferenced points, containing latitude and longitude coordinates for all segments of roads and paths of the urban mesh was created for each studied city, with total computed points enumerated in Table 4. All conversions and haversine distance computing were implemented with *Python Matplotlib Basemap* library (HUNTER, 2007).

4.1.2.2 Obtaining Aerial and Street-level imagery

Each city dataset that contains georeferenced points, composing the urban shape, created following the methodology described in Section 4.1.2.1, was used to gather urban imagery correspondent to each location point, to create a representative urban imagery dataset that covers most of the territory of each city. To achieve this, the *Google Street View* and *Static Maps* (GOOGLE, 2017, 2018a) services were used to request RGB images for each location point, for street-level and aerial images respectively.

With Google Street View API, four requests were made per georeferenced point for Rio de Janeiro, São Paulo, and Salvador cities, corresponding to the cardinal directions of north, east, south, and west. The Street View API takes the parameter *heading* set in the queries as $S = \{0, 90, 180, 270\}$ as the cardinal directions, respectively. Other parameters set were $size = 640 \times 640$ for each image, and $pitch = -0.76$ that specifies the amount of "up" or "down" from the angle of the camera. In Google Static Maps API, requests were made per georeferenced point, with parameter *center* set to the point's latitude and longitude values, *zoom*, that sets the level of "enhance" of the map, being detail values of "Buildings" equal to $A = \{19, 20, 21, 22\}$. All aerial images were

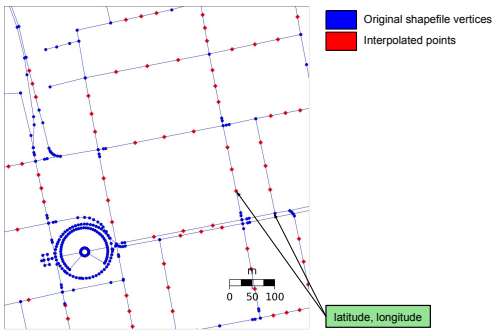


Figure 5 – Original and interpolated latitude and longitude points, originally from *shapefiles*, covering the city and forming the street mesh.

requested to have a resolution of 640×640. Figure 6 depicts the use of Street View and Static Maps APIs to request the four street-level images and four zoom level aerial images for each georeferenced point that compose the street's shape.

Although all the points in datasets are geographically valid, not all of them have valid street-level or aerial images, for all camera and zoom levels or only some of them. The output of invalid requests, i.e., geographic localities that Street View or Static Maps APIs cannot return any image, are gray images displaying an error message at the center, and their resulting files are less $17kb$ each. Image files in which the size was less or equal $17kb$ were filtered and set as unavailable for street-level and aerial domains. The location points of incomplete or unavailable images were set as not usable for the study if one or more camera views were missing, or one of $zoom = 19, 20, 21$ values unavailable for aerial images.

Until the present date, requests for images from specific date and year are not supported by both Street View and Static Maps Google APIs. In order to verify the year of the street-level and aerial downloaded images, a random sample of 10 points was selected for each city. Next, the year of the images from the corresponding geographic points was obtained through manually comparing the images available in Google Street View and Google Earth services with the downloaded images using the *time machine tool*. Figure 7 shows the resulting years for the samples in each city. The street-level and aerial images were gathered using APIs between years of 2016 to 2018, and a small portion of points of Rio the Janeiro had their imagery requested in 2019, due to the unavailability of images at the time of the first request.

A temporary street-level and aerial imagery datasets were built using the previously described methodology for the experiments of this present Thesis, for each city, totaling $256,380$ valid points for Rio de Janeiro, $281,444$ valid points for São Paulo and $86,455$ valid points for Salvador. Table 5 summarizes the total valid points, street-level, and aerial images that resulted after filtering invalid and missing images. Figure 8 depicts some collected image examples from Rio de Janeiro, São Paulo, and Salvador cities of

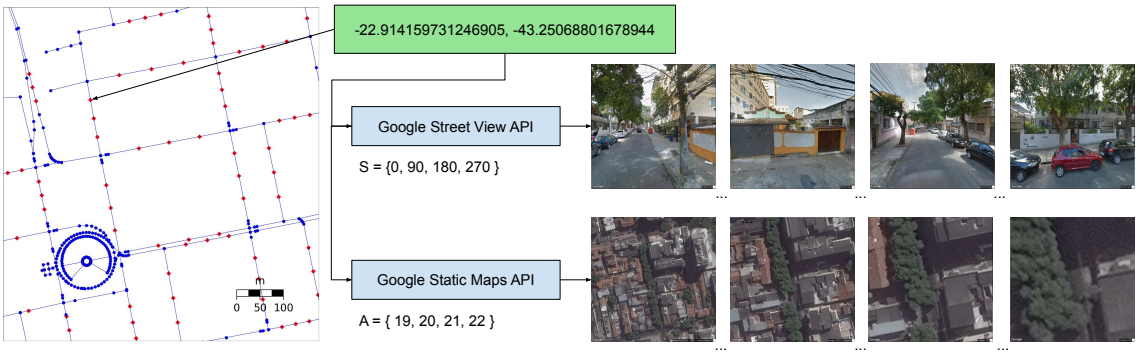


Figure 6 – Example of requests made to obtain street-level and aerial imagery from Google services, using the latitude and longitude location points from shapefiles.

Table 5 – Summary of valid points and total aerial and street-level images available for the study in each city.

City	Valid points	Street-level images	Aerial images
Rio de Janeiro	256,380	1,025,520	1,025,520
São Paulo	281,444	1,125,776	1,125,776
Salvador	86,455	345,820	345,820

street-view and aerial domains, with an example of invalid requests returning no image from a specific zoom level from a geographic point in Figure 8(f).

4.1.3 Dataset labeling methodology

In order to relate the georeferenced points that compose the streets shape with the distribution of dengue fever rates in the studied territories, a grid-like data structure was implemented following a methodology in criminology and crime prediction known as *Quadrat thematic mapping*, where a count of crimes or a calculated density value can be assigned to the area covered by the cell (BOWERS; JOHNSON; PEASE, 2004; ECK et al., 2005; CHAINEY; TOMPSON; UHLIG, 2008; ROSSER et al., 2016).

The developed grid structure divides the territory into equally spaced cells with a given resolution, delimited by arbitrary lower and upper-latitude and longitude corner coordinates, with an independent instance for each city. Their line and column coordinate index access each cell, and each cell can be assigned as being inside the territory polygon or not. This structure provides latent variables with georeferenced points to be distributed between the cells, allowing for a rate of incidence of the variable for each grid cell region to be calculated as the sum of all occurrences found inside the cell. As well as the latent variables, all georeferenced street points can be distributed between each cell. When a georeferenced street point is associated with a cell, i.e., it lies inside that cell, the street point is labeled with the corresponding total latent variable of that cell. Thus, each street point is associated with a variable occurrence rate defined by its surroundings, limited by the defined cell size.

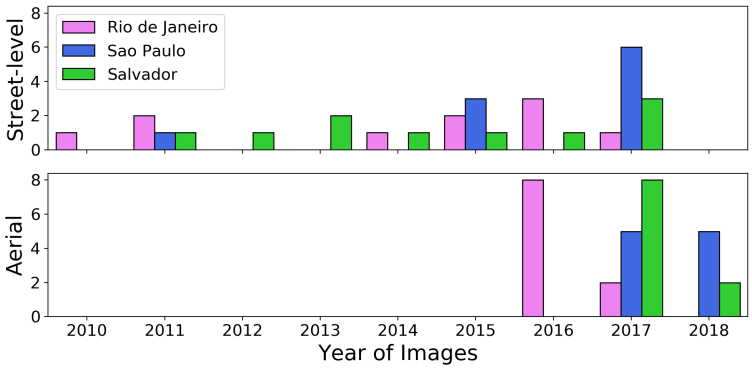


Figure 7 – The year of the original capture of street-level and aerial imagery found in a random sample of 10 georeferenced points for each city.

(a) Street-level from Rio (b) Aerial from Rio de Janeiro

(c) Street-level from São Paulo (d) Aerial from São Paulo

(e) Street-level from Salvador (f) Aerial from Salvador

Figure 8 – Street-level and aerial imagery samples from the studied cities. Columns left to right from street-level are 0 (north), 90 (east), 180 (south), and 270 (west) cardinal directions. For aerial, left to right images depicts zoom levels equals to 19, 20, 21, and 22. In (f), an example of unavailable aerial images for $zoom = 22$.

Figure 9 illustrate an example of grid structure instance over Rio de Janeiro, with a resolution of 80×80. Each cell is indexed by their $(line, column)$ and contains information about their boundaries, such as latitude and longitude corner coordinates. Each latent variable with geographic coordinates is distributed between the cells, and the assigned label of that cell is the total amount of their occurrences inside the cell. Similar to the latent variables, street georeferenced points were also distributed between the cells and label according to the cell total variable value.

Following the methodology described, a grid data structure of size 80×80 was instantiated for each studied city, dividing the territories into equally spaced cells, limited by a provided lower and upper latitude and longitudes that comprehend the administrative limits of the cities selected using the Google Maps service (GOOGLE, 2019b). Each dengue fever georeferenced point in the dengue occurrences datasets was distributed between the cells, with the sum of all dengue occurrences found inside the cell assigned as the label of that cell. Similarly to the dengue fever occurrences, each street location point was distributed among the cells whose position corresponded to and were labeled with the dengue value associated with that cell.

The application of the grid structure provides a blocky *hot-spot* map of dengue fever occurrences, as depicted in Figures 10(a), 11(a) and 12(a), showing the distribution of dengue rates over the administrative limits of Rio de Janeiro, São Paulo and Salvador. Since the methodology goal is to relate dengue rates with urban areas represented by street points, only cells that contained georeferenced street points were used, and cells that covered areas such as lakes and forests without streets or roads were removed

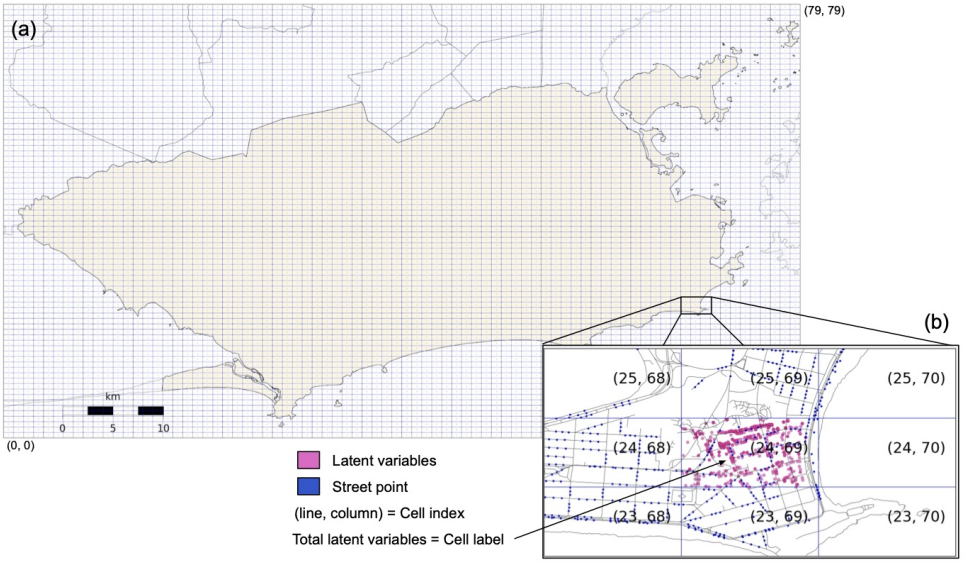
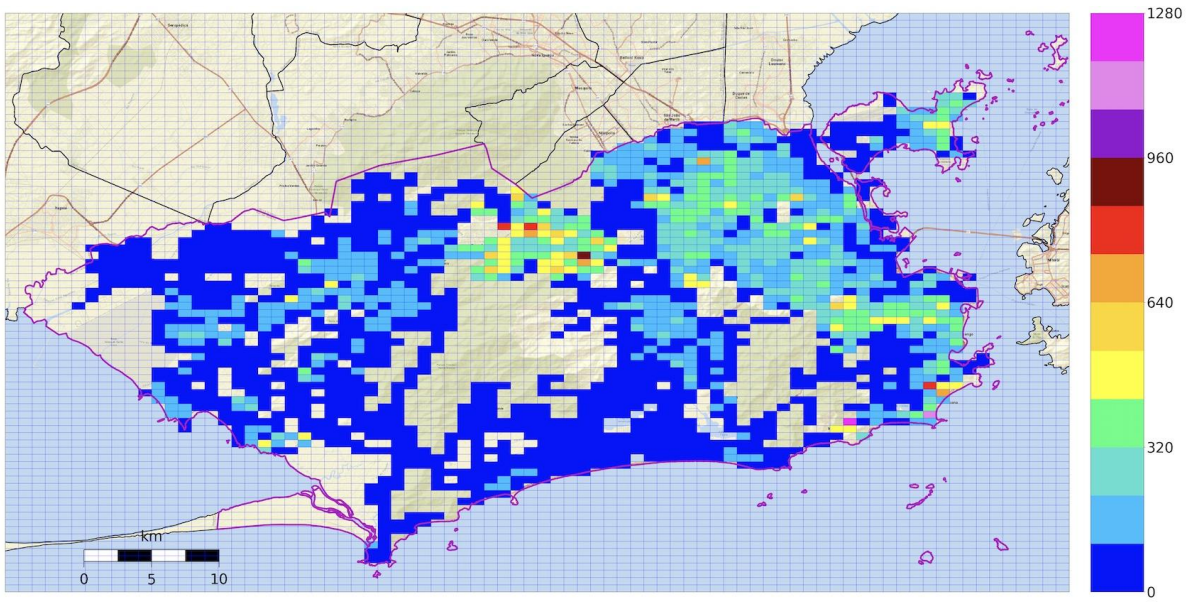


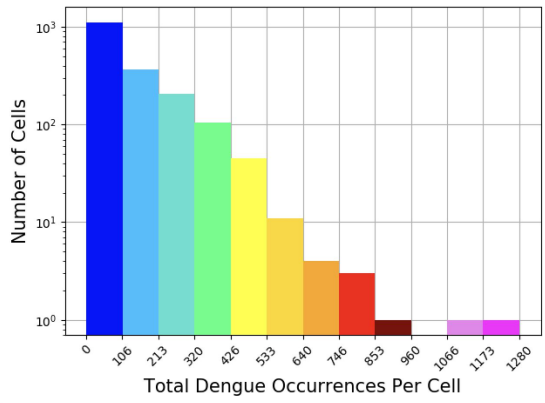
Figure 9 – A grid data structure with size 80×80 over Rio de Janeiro territory. Each cell can be accessed by their $(line, column)$ coordinates and are labeled according to the total latent variable found inside their boundaries. All streets points are labeled according to their cell's labels.

from the map. Each distribution varies from 0, i.e., no dengue occurrences, to a maximum of total dengue cases per cell. Shades of blue are used to denote low values of notifications in the area; shades of green and orange represent average incidence rates and higher rates by shades of red. Very high incidence rates in Rio de Janeiro were highlighted using shades of magenta.

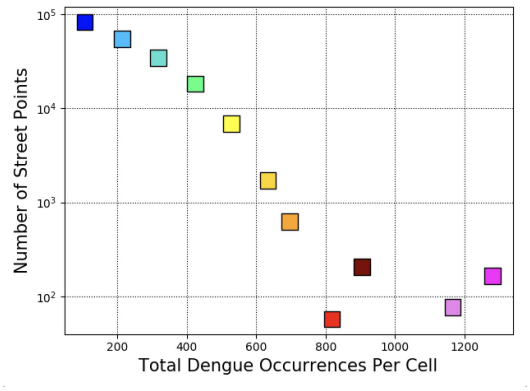
For each *hot-spot* map, a histogram of the data presents the total number of cells per dengue occurrences intervals, depicted in in Figures 10(b), 11(b) and 12(b). It is noticeable that the vast majority of values fall in low and medium rates, with few cells showing high and very high dengue rates. Figures 10(c), 11(c) and 12(c) shows



(a)



(b)



(c)

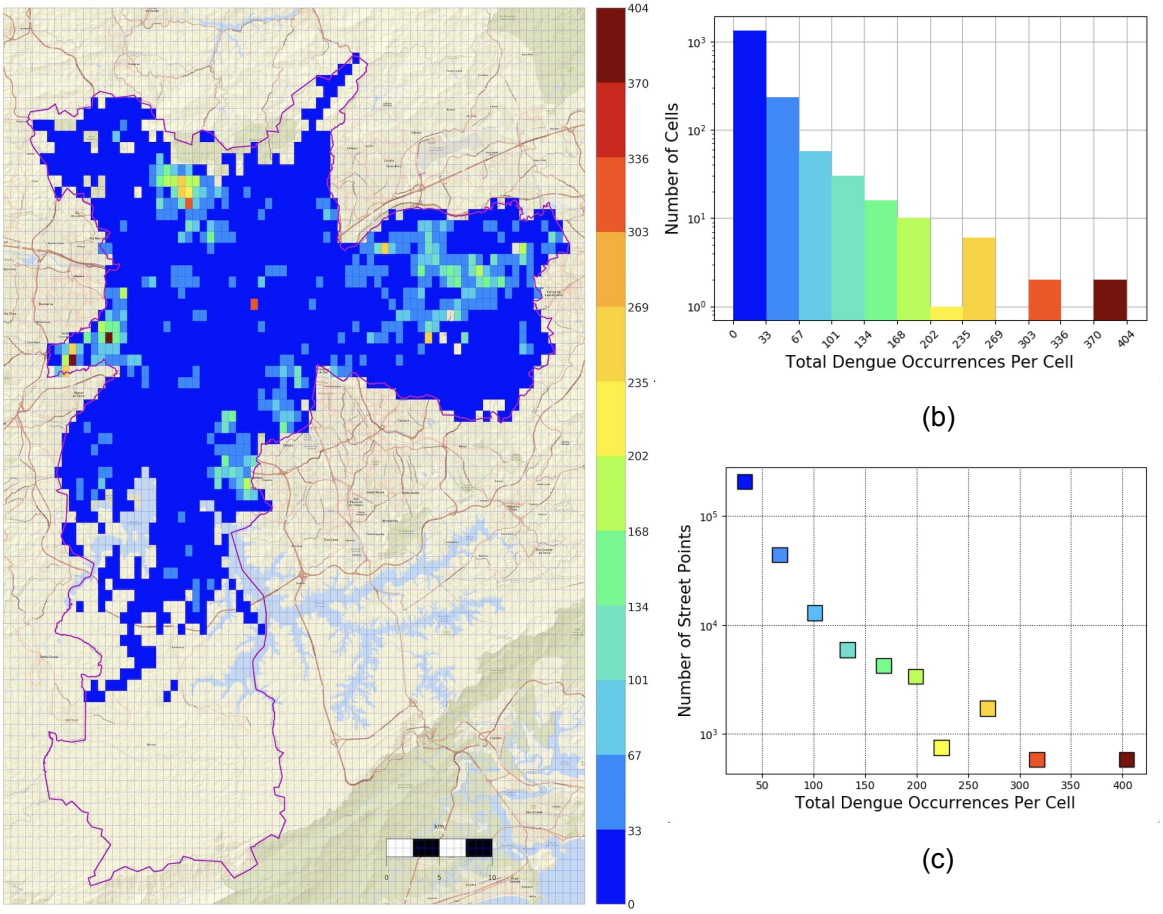
Figure 10 – (a) The 80 $\times$ 80 grid data structure with dengue fever notification cases in Rio de Janeiro. (b) Histogram depicting the distribution of total occurrences by cell, in absolute values. (c) The total number of street location points regarding the total dengue occurrences distribution.

Table 6 – Dengue fever distribution statistics for each city grid structure.

Statistics × City	Rio de Janeiro	São Paulo	Salvador
Maximum Dengue Value	1,280	404	80
Mean Dengue Distribution	83.63	16.83	1.23
Std. Deviation	128.9	32.3	5.92
10% Maximum Dengue Value	128.0	40.4	8.0

the distribution of streets regarding the total dengue occurrences per cell distribution, showing a similar relation between the number of street points and dengue occurrences labels.

Parameters used for creating the 80×80 grid structures for Rio de Janeiro, São Paulo and Salvador are summarized in Appendix A. Finally, Table 6 summarizes the dengue occurrences in each structure and overall statistics for each dengue distribution per cell.



(a) (b) (c)

Figure 11 – (a) The 80 × 80 grid data structure with dengue fever notification cases in São Paulo. (b) Histogram depicting the distribution of total occurrences by cell, in absolute values. (c) The total number of street location points regarding the total dengue occurrences distribution.

4.2 Problem Formulation

In the present Thesis, the problem of estimating dengue fever rates from urban imagery is proposed as a regression problem, where a model receives as input images from specific geographic locations and predicts a dengue fever rate corresponding to a real positive number or zero. The approach chosen was regression because in previous methodological steps, the results from the categorization of latent variables over

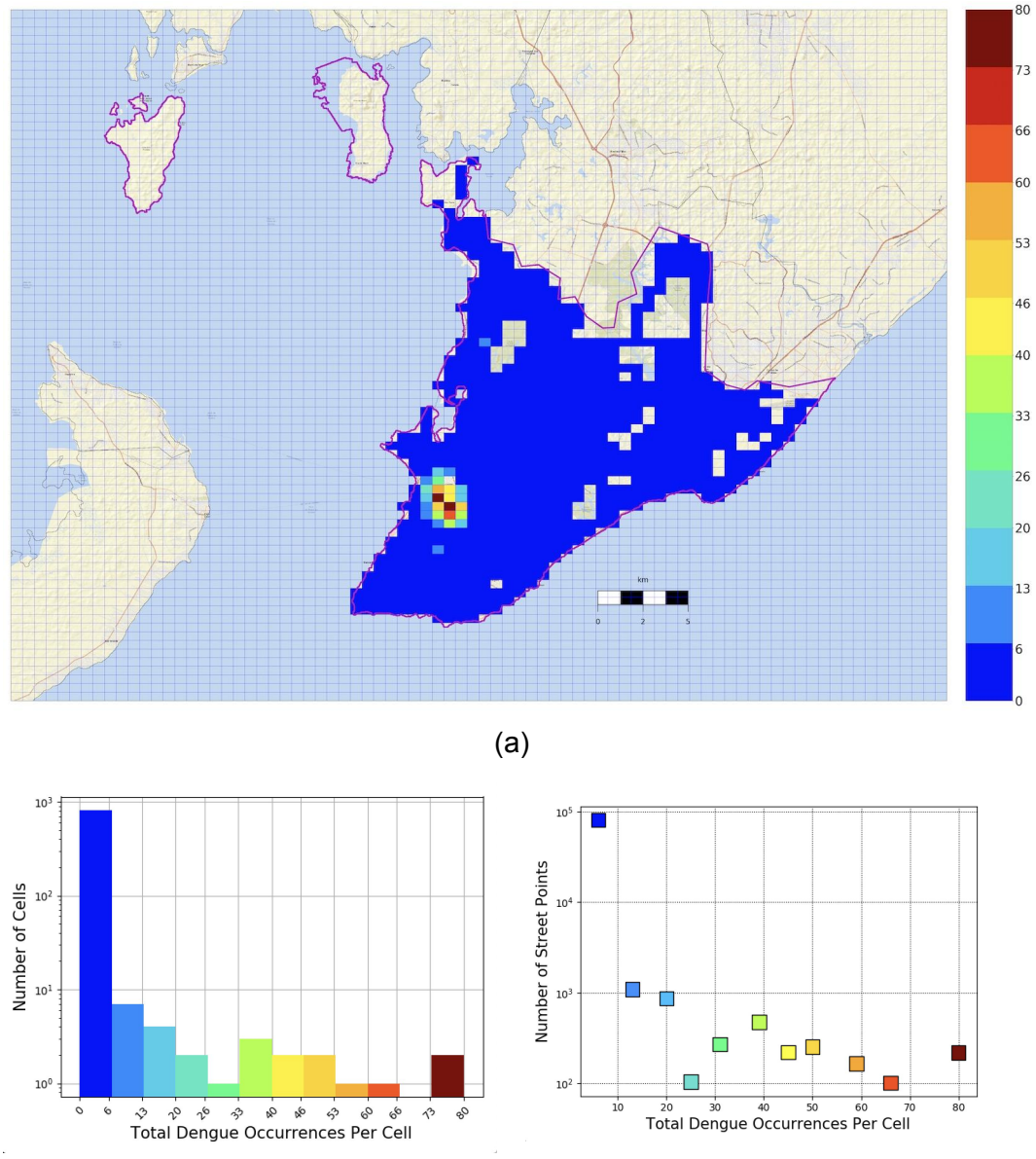


Figure 12 – (a) The 80 $\times$ 80 grid data structure with dengue fever notification cases in Salvador. (b) Histogram depicting the distribution of total occurrences by cell, in absolute values. (c) The total number of street location points regarding the total dengue occurrences distribution.

the territory, with “low” and “high” labels using classification, presented in ANDERSSON; BIRCK; ARAUJO (2017), does not yield significant gains to the model's performance. Similarly, the normalization of values was proposed in ANDERSSON; BIRCK; ARAUJO (2018), where it was concluded that using this approach presents a problem to infer values from cities whose dengue distribution is different or unknown.

From the presented regression problem, given a set of m location points $L = \{(lat, lon, Y)_k\}_{k=1}^{m}$ associated with a dengue incidence rate value $Y = \{y_k\}_{k=1}^{m}, y \in \mathbb{R} \mid y \geq 0$, being a set of aerial images $X^a = \{x^a(i,j)_k\}_{k=1}^{m}$ and a set of street-level images $X^s = \{[x^s(i,j)_1, ..., x^s(i,j)_4]_k\}_{k=1}^{m}$ correspondent to each point L, the proposed approaches takes the relation set $D = \{(X^a, X^s, L)_k\} = \{(X^a, X^s, Y)_k\}$ and learns the functions A, S and F in Equations 4, 5 and 6 respectively, parameterized by a set of weights $W_[a, s, f]$,

$$A(X^a) = A(W_a, X^a) \tag{4}$$

$$S(X^s) = S(W_s, X^s) \tag{5}$$

$$F(X^a, X^s) = F(W_f, A(W_a, X^a), S(W_s, X^s)) \tag{6}$$

where $A(X^a) = \hat{Y}^a$, $S(X^s) = \hat{Y}^s$ and $F(X^a, X^s) = \hat{Y}$, that satisfies $\mid Y - \hat{Y}^{[a,s]} \mid = 0$ and $\mid Y - \hat{Y} \mid = 0$ for the maximum number of location points in L.

Figure 13 shows the proposed conceptual models for A (Eq. 4), S (Eq. 5) and F (Eq. 6) functions to estimate dengue rates $\hat{Y}$ from location points L using aerial urban imagery X^a and street-level imagery X^s.

4.3 Proposed Models

In order to satisfy the problem of estimating dengue incidence values given a set of urban images from location points, presented in Section 4.2, Equations 4-6, it is proposed in the present work the general approach depicted in Figure 14.

The overall model is composed of a feature extraction layer and a regression layer. The feature extraction layer is responsible for obtaining attributes from image data to be processed as input in learning algorithms. The regression layer is a multiple regression estimator that uses the image features as predictors of dengue incidence rates. In this work, two types of models were investigated: (i) Deep Convolutional Neural Network (Deep ConvNet) as feature extractors, with fully-connected neural networks as multiple regression estimators, and (ii) designed descriptors in the feature extraction layer, together with fully-connected neural networks as multiple regression estimator. The models with designed descriptors represent a basic model concept for the proposed

problem and were built in order to compare the ability of dense features and designed descriptors as predictors for dengue rate estimates. The next Sections 4.3.1 and 4.3.2 describes the methodology involved in obtaining such models.

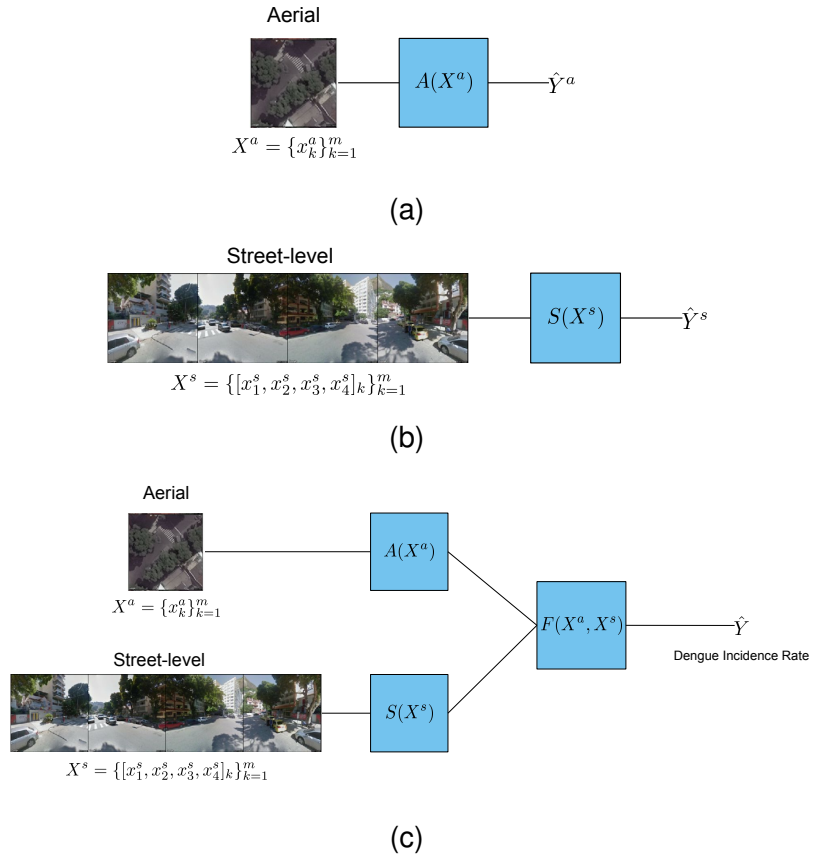


Figure 13 – Conceptual models for the presented problem of, given a set of urban images, learn a function capable of returning a real positive number or zero that satisfies $\mid Y - \hat{Y} \mid = 0$ for a maximum number of locations.

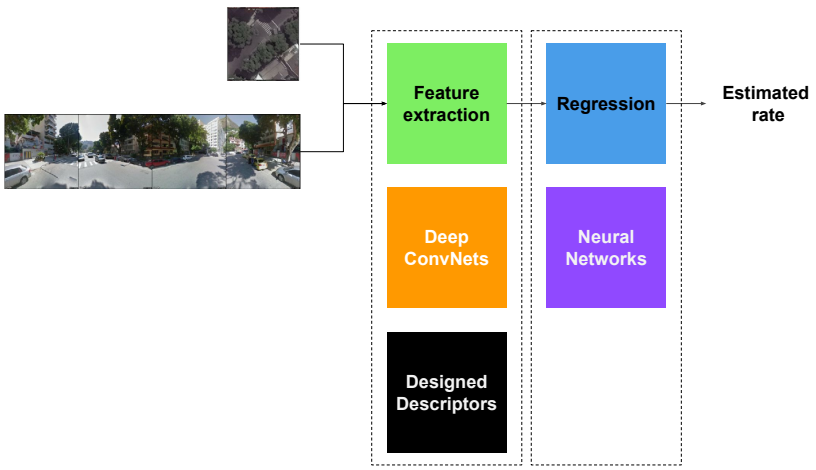


Figure 14 – Overall model for estimating dengue incidence values from urban images.

4.3.1 Baseline Models

Architectures composed of GISTs and HOGs descriptors (Sections 2.1.1 and 2.1.2) were proposed to evaluate the ability of designed descriptors in provide sufficient characteristics for the problem. These designed descriptors were placed in the feature extraction layer with fully-connected neural networks as multiple regressors. These models are presented to estimate dengue fever occurrences from images of a given geographic location point. Those models are a representation of a straight forward approach, where HOGs are used as global feature descriptors, providing a single feature vector from the original image to be input in the regression model.

The proposed baseline models for aerial images are: (I) a single layer neural network linear regressor with an aerial HOG descriptor as input, and alternatively a GIST descriptor, named "*Aerial HOG/GIST-LR*", and (II) a fully-connected neural network model named *"Aerial HOG"* as multiple regressors. For street-level images, it is proposed, as baseline models, a (III) single-layer neural network as a linear regressor for HOG feature inputs "*Street HOG-LR*" and (IV) a fully-connected neural network as multiple regressors named *Street HOG-NN*. The next paragraphs present each model for estimating dengue rates from aerial and street-level images.

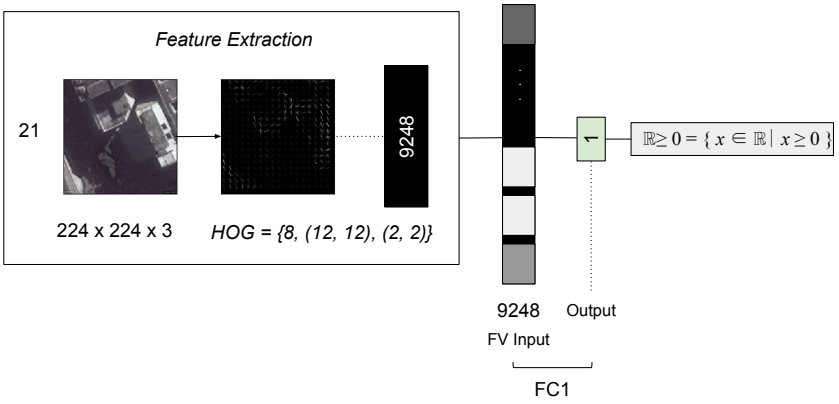
I. *Aerial-HOG/GIST-LR.* This model, depicted in Figure 15a, use the designed descriptor HOG and a fully-connected layer for multiple linear regression, composed of one fully-connected layer, with the input of same feature size and the output with one node, to retrieve a real positive number or zero. In this approach, HOG features were calculated using 8 orientation directions, a block of the size of 2×2, and resolution of cell equal to 12×12 in aerial images resized to $224 \times 224 \times 3$, resulting in a feature vector of 9248 positions. Alternatively, a multiple linear regression using a GIST descriptor, depicted in Figure 15b, is proposed, computed with 8×8 orientations per scale, sampled at 4×4 spatial locations, resulting in a feature vector of 960 positions.

II. *Aerial HOG-NN.* Figure 16 depicts the use of a designed descriptor HOG and a Neural Network composed of fully-connected layers for multiple regression. The HOG descriptors were calculated using 8 orientation directions, a block of size 2×2, and resolution of cell 12×12 in aerial images with resolution $224 \times 224 \times 3$, similar to *Aerial-HOG-LR* model. This network is built with four FC layers, including the output, being each layer output half the size of the previous layer, together with *ReLU* activation units, to output one real positive value or zero.

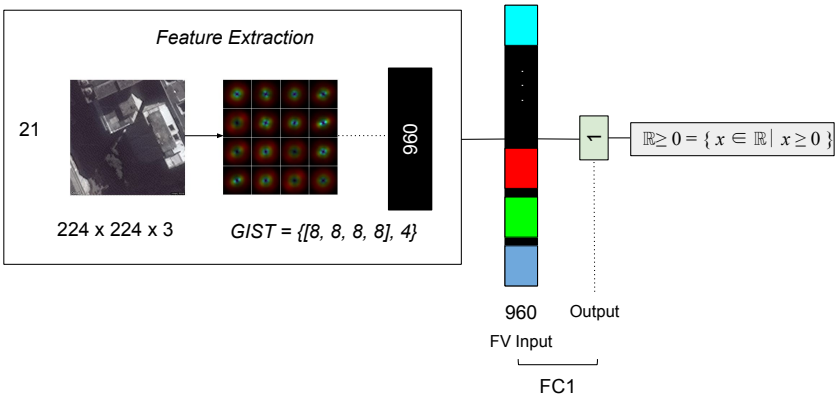
III. *Street HOG-LR.* Figure 17 presents the basic approach of HOG designed descriptors from street-level images as predictors of a linear regression model. The model is composed of the feature vectors calculated using 8 orientation directions, block size of 2×2 cells, and cell resolution equal 12×12 in images with resolution $112 \times 112 \times 3$. The features from camera directions 0 and 90 are combined through the mean average of each feature vector position, the same way that features from camera views of 180

and 270. Each combined feature results in the same size vector. Next, the feature vectors of $2,048$ positions are concatenated in a $4,096$ feature vector (FV Input) to be input in a single fully-connected layer (FC1) to predict a real positive number or zero as output.

IV. *Street HOG-NN.* Figure 18 depicts the second baseline model to apply in street-



(a) Aerial HOG-LR



(b) Aerial GIST-LR

Figure 15 – Baseline *Aerial-HOG/GIST-LR*: HOG (a) and GIST (b) features are input in a linear regressor built with a fully-connected (FC1) layer with one node for output real positive number or zero as output.

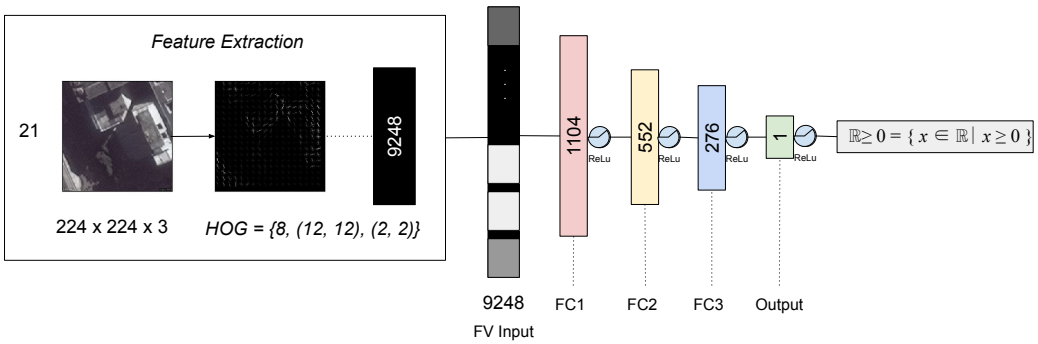


Figure 16 – Baseline *Aerial HOG-NN* proposed model for estimating dengue fever rates from aerial images using HOG designed descriptors and a fully-connected neural network.

level images, with a neural network as multiple regressor. This model follows the latter approach, in which the HOG feature vectors are calculated using 8 orientation directions, block size of 2×2 cells, and cell resolution equal 12×12 in images with increased resolution of $224 \times 224 \times 3$. Here, each feature vector of $9,248$ positions are averaged to compose a same size $9,248$ feature vector input (FV Input). Next, the neural network follows four fully-connected layers (FC) of $4,624$, $2,312$, $1,156$ and 1 nodes, respectively, and *ReLu* activation units to output a real positive number or zero.

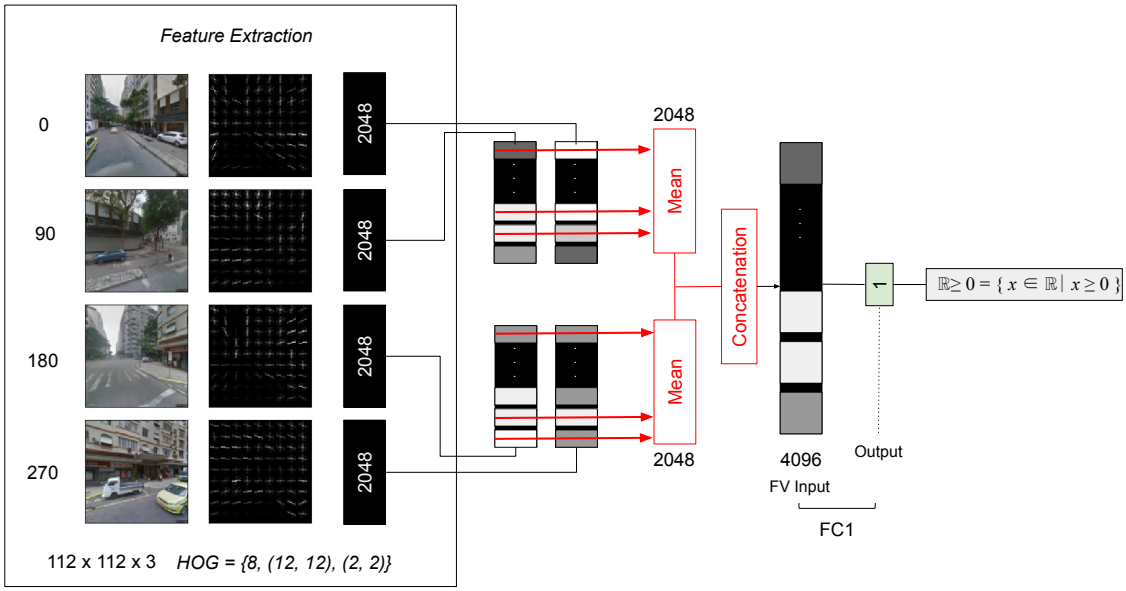


Figure 17 – Baseline *Street HOG-LR* basic model for estimating dengue fever rates from street-level images using HOG designed descriptors and a single fully-connected layer.

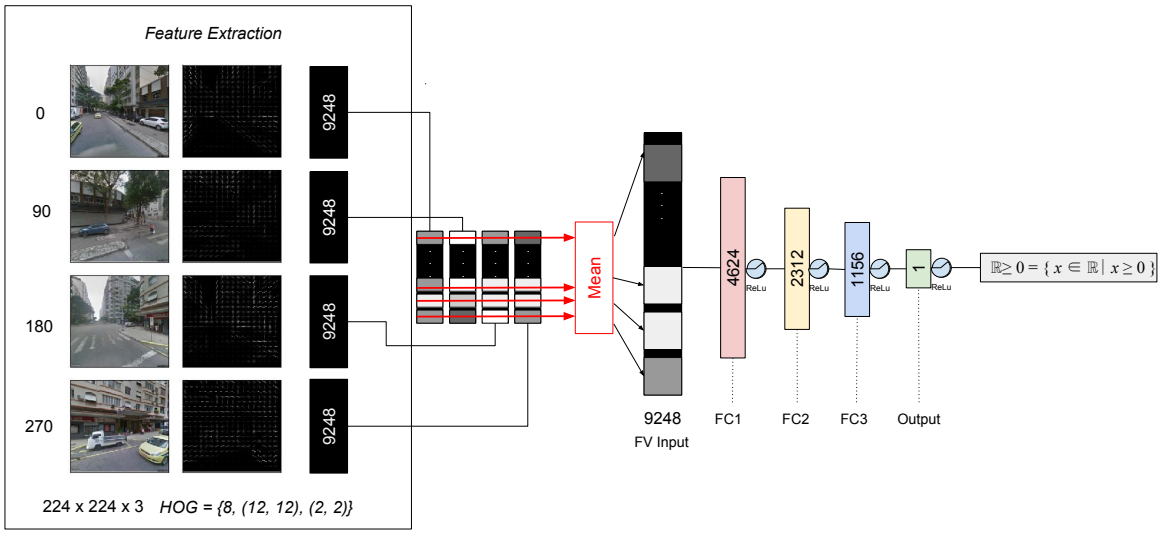


Figure 18 – Baseline *Street HOG-NN* proposed model for estimating dengue fever rates from street-level images using HOG designed descriptors and a fully-connected neural network.

4.3.2 Deep Convolutional Neural Network Models

To estimate dengue rates from aerial and street-level images using a Deep Convolutional Neural Networks (Deep ConvNets) (Section 2.2) in the feature extraction layer, it is proposed in this work the use of image features obtained from *DenseNet* (HUANG et al., 2017) architecture, specifically the *DenseNet 161* implementation. These features are extracted by the networks, using transfer-learning techniques or training without initialization, and their feature outputs are used in fully-connected neural networks as predictors to estimate dengue rates. In what follows, we detail each of the three models: (I) "*Sat*", which uses aerial images as input; (II) "*Street*", which uses four cardinal street-level images as input; (III) "*StreetSat*", which combines both aerial and street-level images.

I. *Sat.* The *Sat* Neural Network using *DenseNet-161* feature learning layers, replacing the original classifier block with an fully-connected neural network as multiple regression estimator, connecting the output feature vector of $2,208$ positions with four fully-connected (FC) layer including the output. Each FC1 to FC3 layers has its output with half the size of their input, and *ReLU* activation functions placed in the outputs of each layer. The last layer has one output node, to return zero or non-negative real number values, as depicted in Figure 19.

II. *Street.* The *Street* model, depicted in Figure 20, was built to estimate real positive numbers or zero from street-level images at a specific location, using as feature extractor the outputs of *DenseNet161* pretrained with *Places365* (ZHOU et al., 2018b) dataset, identified as *DenseNet161-Places*. The output feature vectors are obtained from the set of dense blocks followed by the last *BatchNorm* layer "Norm5" and an *Average Pooling 2D* operation, that compose the feature learning layers. The street-level dense features obtained from the *DenseNet161-Places* feature extractor, as depicted in 20, have $2,208$ positions, which are concatenated to form a $8,832$ feature vector.

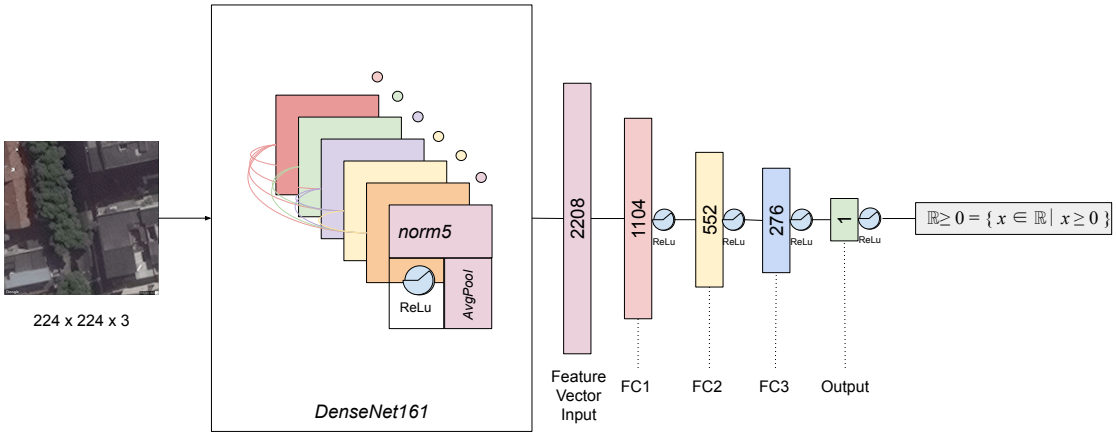


Figure 19 – ***Sat*** Neural Network to estimate real positive values or zero from aerial images.

Next, the input features (FV input) are feeded to a multiple regression estimator composed of four fully-connected layers (FCs and Output). Each FC1 to FC3 layer has the output half the size of their input, and the last layer with one node output, to return a real value positive number or zero. Activation functions of type *ReLU* are placed in each output layer.

III. *StreetSat.* To evaluate the use of aerial and street-level image features as complementary estimators, it is proposed the *StreetSat* model to estimate real pos-

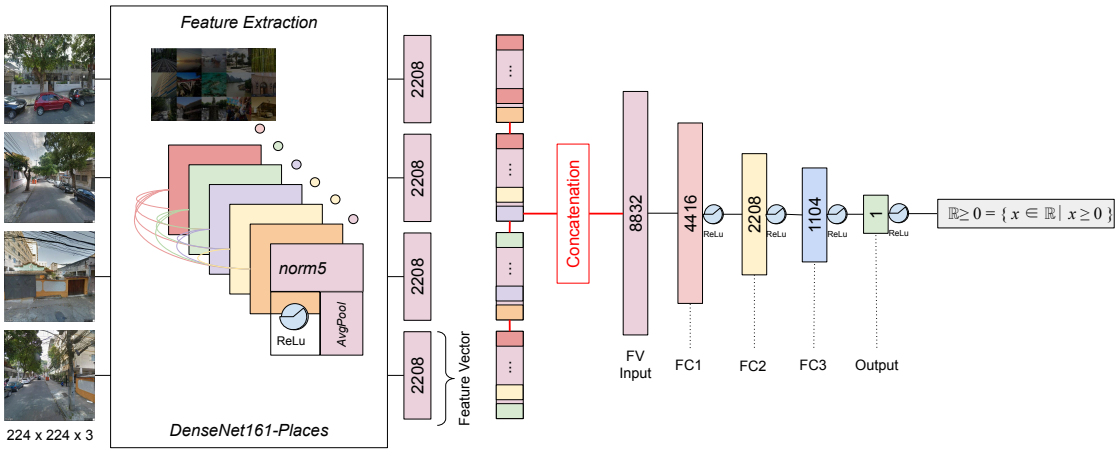


Figure 20 – ***Street*** Neural Network to estimate real positive values or zero from street-level images.

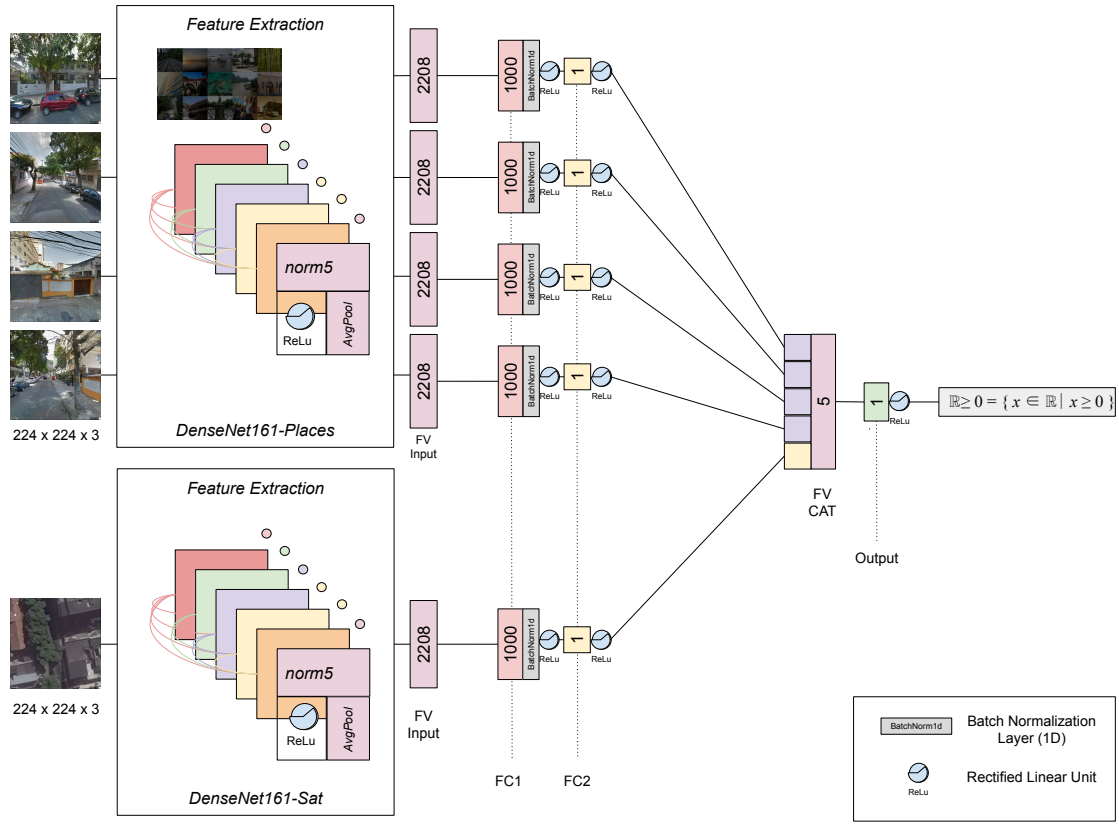


Figure 21 – ***StreetSat*** Neural Network.

itive value numbers or zero from aerial and street-level images, depicted in Figure 21. The network uses aerial dense features obtained from *DenseNet161* pretrained within the *Sat* model (Subsection 4.3.2.I.), identified as the feature extraction block *DenseNet161-Sat*, and street-level dense features, obtained from *DenseNet161* pretrained with pretrained with *Places365* dataset (*DenseNet161-Places*). All features are extracted from the *DenseNet* networks with the same feature extraction methodology explained in *Street* model (Subsection 4.3.2.II.), which of it having a vector size of $2,208$ positions. Each dense feature vector (FV Input) is input to a fully-connected (FC1) layer with output of 1000 nodes, equipped with a 1 dimension batch normalization (*BatchNorm*) layer and *ReLu* activation units, and an output fully-connected (FC2) with one output node. Next, the FC2 outputs for each image are concatenated in a 5 dimension feature vector (FV CAT) and then input to the final output node of the network, with one node to estimate zero or non-negative real number values.

4.4 Evaluation Methods

4.4.1 Training and Test methodologies

To compose train and test sets for evaluation of the proposed models, the *hold-out validation set* (HASTIE; TIBSHIRANI; FRIEDMAN, 2008; NG, 2018) technique was used, consisting of randomly dividing the available dataset in three parts: the training set, the validation (or development) set, and a test set with a specific portion related to whole dataset, e.g. $train = 50\%$, $validation = 25\%$, $test = 25\%$. After the split, the training set is used for learning weights and fit the model, while the validation set is used for evaluating the model performance at the learning process in each epoch. According to the authors, the test set should be kept aside and should not be used in any hyperparameter selection process for the model, and used only to evaluate the performance of the final model. The *hold-out* was applied in the works of KRIZHEVSKY; SUTSKEVER; HINTON (2012); SIMONYAN; ZISSERMAN (2015); HUANG et al. (2017), with datasets such as *CIFAR* (KRIZHEVSKY, 2009) with more than $60,000$ examples.

4.4.2 Train, validation and test sets distribution

Following the described *hold-out* methodology, the datasets for each city were split based on the distribution of the grid cells labeled with dengue incidence rates described in Section 4.1.3. In this approach, the *proportional stratified sampling* (HIRZEL; GUISAN, 2002) method was used to ensure the representativeness of the dengue values found in the territory in all train, validation, and test sets. Each city dengue cell distribution was arranged in bins, where samples from each bin are randomly drawn in each stratum, following a specific proportion set.

For Rio de Janeiro, from a distribution of $bins = 6$, the cell samples were drawn

Table 7 – Train gap calculated for train sets which uses aerial images with $zoom = \{20, 19\}$, with the result of $imageMeters$ distance and $testGap = 50m$. For $zoom = 21$, no necessary train gap to be set.

City	Zoom20		Zoom19	
	$imageMeters$	Train Gap	$imageMeters$	Train Gap
Rio de Janeiro	88.0 m	38.0 m	176.2 m	126.0 m
Sao Paulo	87.4 m	38.0 m	174.9 m	126.0 m
Salvador	93.0 m	43.0 m	186.2 m	136.0 m

from the bins and placed in the train, validation, and test strata following a proportion of $75\% - 05\% - 25\%$ split. For São Paulo and Salvador, a proportion of $70\% - 15\% - 15\%$ was used. For all cities, in validation and test cells, were discarded points where their location was less than 50 meters from the cell borders for test cells, to avoid possible overlap of points in training, validation and test datasets. Furthermore, different training sets were arranged for different aerial zoom levels in order to maintain the same validation and test set for all experiments and avoid the overlap of information between aerial images from train, validation, and test sets. The safe distances for different aerial zooms were calculated based on the Equations 7, 8,

$$metersPerPixel = 156543.03392 * \frac{cos(\varphi * \frac{\Pi}{180})}{2^{zoom}} \tag{7}$$

$$imageMeters = metersPerPixel * downloadResolution[height] \tag{8}$$

where φ is equal the latitude of the aerial image central point, $zoom$ the zoom level and $downloadResolution$ the height or width resolution of the Google Static Maps API requested image, in pixels, using the earth's radius $6378137m$ and Google's pixel per meters ratio $156,543.03392$ (BROADFOOT; GOOGLE, 2011; Answer from StackExchange, 2016).

Table 7 enumerates the resulted distances from the application of Equation 7 for different zoom levels in each city, and the train gap applied in each case using rounded values from $trainGap = imageMeters - testGap$, with $testGap = 50m$. Tables 8, 9 and 10 presents the total train, validation and test points available for each set after removing possible overlapping points, with the final resulting *train-validation-test* set proportion. In addition, for the cities of São Paulo and Salvador, a test set was constructed with valid points without the restriction of borders, in order to evaluate models trained with different cities. Table 11 shows the number of points used in the total test set for each city.

Table 8 – Train, test and validation datasets for use with aerial images $zoom = 21$.

City	Train points	Test points	Validation points	Total
Rio de Janeiro	151,458 ($\simeq 83\%$)	24,329 ($\simeq 13\%$)	7,115 ($\simeq 4\%$)	182,902
São Paulo	198,000 ($\simeq 77\%$)	31,174 ($\simeq 12\%$)	29,089 ($\simeq 11\%$)	258,263
Salvador	58,736 ($\simeq 79\%$)	7,739 ($\simeq 11\%$)	7,670 ($\simeq 10\%$)	74,145

Table 9 – Train, test and validation datasets for use with aerial images $zoom = 20$.

City	Train points	Test points	Validation points	Total
Rio de Janeiro	140,294 ($\simeq 82\%$)	24,329 ($\simeq 14\%$)	7,115 ($\simeq 4\%$)	171,738
São Paulo	184,381 ($\simeq 75\%$)	31,174 ($\simeq 13\%$)	29,089 ($\simeq 12\%$)	244,644
Salvador	52,051 ($\simeq 77\%$)	7,739 ($\simeq 12\%$)	7,670 ($\simeq 11\%$)	67,460

Table 10 – Train, test and validation datasets for use with aerial images $zoom = 19$.

City	Train points	Test points	Validation points	Total
Rio de Janeiro	117,600 ($\simeq 79\%$)	24,329 ($\simeq 16\%$)	7,115 ($\simeq 5\%$)	149,044
São Paulo	156,633 ($\simeq 72\%$)	31,174 ($\simeq 14\%$)	29,089 ($\simeq 14\%$)	216,896
Salvador	39,517 ($\simeq 72\%$)	7,739 ($\simeq 14\%$)	7,670 ($\simeq 14\%$)	54,926

Table 11 – Composition of location points sets and cells for testing with models from different cities.

City	Total test points	Total test cells
Sao Paulo	281,184	1,718
Salvador	84,192	834

4.4.3 Accuracy measurements

For all proposed models, the *Mean Squared Error* (MSE) loss function was used as main criterion for estimation of dengue fever values, as presented in Equation 9,

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n} \tag{9}$$

for each example $l(y, \hat{y})$ composed of a true value y and a value $\hat{y}$ predicted by the models, in a set of $L = \{l_1, ...l_n\}$ of a total of n examples. The *Root Mean Squared Error* (RMSE) and *Mean Absolute Error* (MAE), as presented in Equations 10 and 11,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n}} = \sqrt{MSE} \tag{10}$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}|}{n} \tag{11}$$

were also used to allow a better comparison of the actual and estimated values when evaluating the models, where y_i are actual values, and $\hat{y}$ estimated values.

In addition, the *Pearson Product Moment Correlation*, *Pearson* r, that measures how two variables are proportional to each other and can be predicted with linear regression, is used to measure the linear association between actual and estimated

values, given by the Equation 12,

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \tag{12}$$

where $\overline{x}$ is the mean average of actual values, and $\overline{y}$ the mean average of estimated values, with $-1 \leq r \leq 1$. When $r = -1$ there is total opposite correlation, $r = 1$ total positive correlation and $r = 0$ no correlation. In this work, the letter ρ is used to represent the Pearson r correlation, considering test dataset examples as a population.

4.5 Experiments

Baseline models were trained in Rio de Janeiro city, while Deep Convolutional Neural Networks models were trained in Rio de Janeiro and São Paulo cities, using the same proposed architectures without network topology modifications, sometimes with different hyper-parameters. Salvador was used only when evaluating the models for geographic portability. The next Section 4.5.1 presents the overall methodology for obtaining the architecture of proposed models, and in choosing the best hyper-parameters and final models. Particularities from baseline and Deep Convolutional Neural Network models and hyper-parameters selection are presented in Subsections 4.5.1.1 and 4.5.1.2.

4.5.1 Network, Hyper-parameter and Model Selection

Network Topology Setup. The overall topology of the proposed networks is based on works of ZAGORUYKO; KOMODAKIS (2015) for the siamese-like input of different domain images, and LIEMAN-SIFRY (2016) for the multiple camera views idea. Baselines models with Linear regression were built with a linear transformation of type $y = wA^T + b$ where A^T represents the input designed descriptor feature vector, w equal weights to be learned, and b an additive bias. For the topology selection for complex regressors with Neural Networks, in all models, the number of nodes in each hidden layer was set using a rule of the mean average between input size and output size layers, considering the output with 1 node. The hidden layers were iteratively added, using the same average rule for setting the number of nodes, while tracking the model's performance changes with the modifications.

Optimizer Algorithm. All models were trained using *Mini-Batch Gradient Descent* optimizer algorithm, a variant of *Stochastic Gradient Descent* (SGD), also referred as *Mini-Batch* SGD (BOTTOU, 2010; LIN et al., 2018) where weights are updated with a mini-batch of examples with size $1 < k < N$, where k is the mini-batch size and N is total number of examples. The optimizer takes, as hyper-parameters inputs the *learning rate* lr, *momentum* m, *Nesterov momentum* $nesterov$ and a regularization L2

weight decay wd.

Model Selection. To select the best model, for all proposed architectures, *early stopping* was used during the training stage, where the resulting MSE loss of the validation set in each training epoch is tracked, and the model selected when the lowest validation MSE is achieved. The training process is stopped when there is no progress in validation loss during the next epochs.

4.5.1.1 Baseline Models

The selection of HOG hyper-parameters was performed using *random search* technique (BERGSTRA; BENGIO, 2012), which consists of randomly drawing values for hyper-parameters from a uniform distribution. The hyper-parameters set that leads the model to the best result, during a number of epochs, is chosen.

The possibilities for HOG hyper-parameters sets were gathered from the best results of DALAL; TRIGGS (2005), with *block size* $= (2, 2)$ cells, varying the *cell size* $= \{(8, 8), (10, 10), (12, 12)\}$ pixels, two possible *orientations* $= \{8, 9\}$ and clipped block normalization $block_norm =$ *L2-Hys*. For GIST designed descriptor, it was used the parameters *number of blocks* $= 4$ and *orientations per scale* $= [8, 8, 8, 8]$ as proposed by DOUZE et al. (2009). The concatenation of descriptors using the mean average between feature vector positions is based on (HOFFMANN et al., 2019) proposed methodology.

The multiple linear regression models *Aerial HOG/GIST-LR* and *Street HOG-LR* were set as a linear transformation. The neural network approaches *Aerial HOG-NN* and *Street HOG-NN* neural networks where set up with similar architectures from the proposed deep neural network models (Subsection 4.5.1.2). The number of nodes in each hidden layer was chosen using the rule of the mean average between input and output number of nodes. For *Aerial HOG-NN*, the layer's nodes size was calculated considering the same input size as *Sat* Neural Network FC1 layer $(2, 208)$, and in *Street HOG-NN* the size of the HOG feature vector $(9, 248)$ used in this model.

The *Mini-batch* SGD optimizer hyper-parameters were selected using *random search* (BERGSTRA; BENGIO, 2012), with learning rates drawn from a uniform distribution of $lr = uniform(min = 1e - 10, max = 1e - 5, 10)$, *momentum* from $m = uniform([0.999, 0.995, 0.99, 0.98, 0.9, 0.0], 10)$, $nesterov = uniform([True, False], 10)$ and weight decay regularization from $wd = uniform(min = 1e^{-7}, max = 1e^{-4}, 10)$. Specifically, for *Aerial GIST-LR* and *Street HOG-LR* models, optimizer parameters were selected from KRIZHEVSKY; SUTSKEVER; HINTON (2012); HUANG et al. (2017) related works.

4.5.1.2 Deep Convolutional Neural Network Models

Architectures *Sat*, *Street* and *StreetSat* were constructed based on *transfer learning* methodologies, where a previous pretrained architecture, designed for a different task, is used for training another domain-related task (GOODFELLOW; BENGIO; COURVILLE, 2016). Transfer learning approaches used in the proposed models were the use of the architecture definition, specifically the feature extractor layers, and *fine-tuning* with previously trained weights from a related domain in all layers.

The *DenseNet* architecture (HUANG et al., 2017) was chosen as the model for feature extraction due to its state of the art performance in learning features for several different domain tasks. The specific version of the Deep ConvNet implementation *DenseNet-161* was used due to the availability of pretrained models with *Places* dataset (ZHOU et al., 2018b), a domain similar to urban street-level imagery data used in this work.

In the construction of the proposed *Sat* architecture, the transfer learning technique applied was to use the defined feature extraction layers from *DenseNet161*, composed of "dense blocks" of n dense layers, containing *BatchNorm2D*, *Convolution2D* and *ReLU* activations (see Section 2.2), finishing with a *global average pooling* operation. Then, the feature vector output is connected to the proposed neural network. In *Sat*, no previous weights were used to fine-tune, and the training process was made by initializing the network with random weights. The hidden layers were iteratively added to find the most appropriate setting for the task, using the same average rule for the number of nodes, while tracking the models' performance changes with the modifications.

For *Street* architecture, the transfer learning technique applied was to use the defined feature extraction layers from *DenseNet161*, similar to the previous model, but instead of training from scratch, the model was initialized with pretrained weights from *Places* dataset, and all dense layers were fixed ("frozen") to prevent *Places* weights from being updated during gradient descent. This is the standard methodology of a feature extractor with transfer learning. Each 4-camera view street-level image is propagated through the *DenseNet161-Places*. The outputs are concatenated to obtain a feature vector of $8,832$ positions to be connected to the proposed *Street* architecture. Similar to the previous model, hidden layers number of nodes were dimensioned using the mean average between input size and output size layers, that were iteratively added tracking the model's performance changes with the modifications.

Next, *StreetSat* architecture, similar to the previous model, uses feature extraction transfer learning technique, by propagating images in *DenseNet* models with frozen layers, i.e., without gradient update, for both aerial and street-level images. Street-level dense features were obtained from *DenseNet161-Places*, with *Places* dataset, and aerial dense features were obtained from *DenseNet161-Sat*, with pretrained weights from aerial images labeled with dengue fever rates, proposed in this work. The topology

of *StreetSat* was obtained by initially adopting the number of layers and nodes from the other proposed models and progressively reducing the number of units and layers, and tracking the obtained results. The *BatchNorm1D* layer was added to different FC outputs, tracking the model's performance, until a reasonably satisfactory position was found. Similarly, the CAT FV layer, for features concatenation, was also tested at different levels of the proposed network, until the best position was found.

The *Mini-batch* SGD optimizer hyper-parameter selection was performed using a *ad-hoc* grid search methodology, by tracking the models through training and validation using *MSE* loss as accuracy measure, for a range of initial learning rates $1e^{-9} \leq lr \leq 1e^{-3}$. The momentum $m = 0.9$, $nesterov = True$ and regularization L-2 weight decay $wd = 1e{-}4$ were set according to previous related work (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; HUANG et al., 2017).

4.5.2 Dengue Estimation with Baseline Experiments

For models *Aerial HOG/GIST-LR*, *Aerial HOG-NN* and *Street HOG-NN*, images were resized to $224 \times 224 \times 3$ resolution. In model *Street HOG-LR*, images were resized to $112 \times 112 \times 3$ to obtain a smaller feature vector size. No normalization was performed in aerial or street-level images. The chosen HOG parameters for both aerial and street-level baseline models were *block size* $= (2, 2)$, *cell size* $= (12, 12)$ and *orientations* $= 8$, resulting in a feature vector of $2,048$ positions for images with $112 \times 112 \times 3$ and $9,248$ positions for images with $224 \times 224 \times 3$. The GIST descriptor resulted in a feature vector of 960 positions.

In *Aerial GIST-LR* model, mini-batch size was set to $k = 32$ examples, initial learning rate was set to $lr = 1e-3$, reducing it by a factor of $f = 0.1$ when validation loss values reaches a plateau between 5 epochs, $momentum = 0.9$, $Nesterov = True$ and L-2 regularization to reduce overfitting with weight decay equal $wd = 1e-4$. For *Aerial HOG-LR* and *Aerial HOG-NN* models, mini-batch size was set to $k = 32$ examples, initial learning rate as $lr = 1.02878e^{-6}$ with exponential decay of $f = 0.2$ at every 16 epochs, $momentum = 0.99$, $Nesterov = True$ and L-2 regularization with weight decay equal $wd = 2.9358e^{-6}$. Model *Street HOG-LR* was trained with mini-batch size $k = 128$ examples, initial learning rate as $lr = 1e^{-4}$ decaying through inverse scaling by a factor of $f = 0.25$ and L-2 regularization with weight decay equal $wd = 1e-4$.

For *Aerial HOG-NN* and *Street HOG-NN* models, mini-batch size was set to $k = 32$ examples, initial learning rate as $lr = 1.02878e^{-6}$ with exponential decay of $f = 0.2$ at every 16 epochs, $momentum = 0.99$, $Nesterov = True$ and L-2 regularization with weight decay equal $wd = 2.9358e^{-6}$. All linear layers from baseline models were initialized with *LeCun* initialization method (LECUN et al., 1998).

4.5.3 Dengue Estimation with Deep ConvNets Experiments

In *Sat* models, aerial images were resized to $224 \times 224 \times 3$ and no normalization was made in the aerial before network input. Mini-batch size was set to $k = 32$ examples, initial learning rate $lr = 1e^{-6}$, reducing it by a factor of $f = 0.1$ when validation loss values reaches a plateau between 5 epochs, $momentum = 0.9$ and $Nesterov = True$. Regularization techniques were applied to reduce overfitting: L-2 regularization with weight decay of $1e^{-4}$; data augmentation with probability of execution $p = 0.5$ in each mini-batch, with random horizontal flip, random vertical flip and color jitter with $hue = 0.05$ and $saturation = 0.05$; and random dropout of nodes with probability of $[0.6, 0.5, 0.3]$ for FC1, FC2 and FC3 layers respectively. For *Sat* models trained in Rio de Janeiro, convolutional layers were initialized using *Kaiming Normal* initialization, the *DenseNet161* default initialization.

For *Street* models, image features were obtained by propagating street-level images with resized resolution of $224 \times 224 \times 3$ and normalization with values for average mean and standard deviation $mean = [0.485, 0.456, 0.406]$, $std = [0.229, 0.224, 0.225]$ following *DenseNet161-Places* methodology (CSAILVision, 2015). For *Street* model trained in Rio de Janeiro city (*RioStreet*), mini-batch size was set to $k = 128$ examples, linear layers initialized with *LeCun* initialization method and initial learning rate $lr = 1e^{-7}$, reducing it by a factor of $f = 0.1$ when validation loss values reaches a plateau between 5 epochs. For São Paulo *Street* model, linear layers were initialized with pretrained weights from *RioStreet*, and initial learning rate $lr = 4.13e^{-6}$.

In both models, it was applied $momentum = 0.9$, $Nesterov = True$ and regularization techniques used to reduce overfitting were L-2 regularization with weight decay $wd = 1e^{-4}$, shuffling the order of concatenation of feature vectors and random dropout of nodes with probability of $[0.5, 0.5, 0.5, None]$ in FV Input, FC1, FC2 and FC3 layers respectively.

Models *StreetSat* were trained using mini-batch $k = 32$, initial learning rate $lr = 1e^{-6}$, reducing it by a factor of $f = 0.1$ when validation loss values reaches a plateau between 5 epochs, $momentum = 0.9$ and $Nesterov = True$. Regularization techniques applied to avoid overfitting were L-2 penalty with weight decay $wd = 1e^{-4}$, random dropout of nodes with probability of $p = 0.8$ for feature vector concatenation layer (FV Cat) and F1 layers respectively.

4.5.4 Geographic Portability of Models

To verify the ability of trained models in a specific city to estimate dengue incidence in other cities, whose model has not been exposed to images examples, each proposed Deep Convolutional Neural Network trained and validated in Rio de Janeiro and São Paulo were tested for every other city, crossing the models between cities. Salvador is used exclusively to test the portability of models trained in other cities. Moreover, a

Table 12 – Combination of proposed models trained in each city and tested in their respective geographic limits (test set) and in the other studied cities, with each aerial and street features used.

Model	Aerial input	Street input	Rio de Janeiro	São Paulo	Salvador
Aerial HOG-LR	$zoom = 21$	-	Test set	-	-
Aerial GIST-LR	$zoom = 21$	-	Test set	-	-
Aerial HOG-NN	$zoom = 21$	-	Test set	-	-
Street HOG-LR	-	Street HOG	Test set	-	-
Street HOG-NN	-	Street HOG	Test set	-	-
RioSat21	$zoom = 21$	-	Test set	Whole city	Whole city
RioSat20	$zoom = 20$	-	Test set	Whole city	Whole city
RioSat19	$zoom = 19$	-	Test set	Whole city	Whole city
RioStreet	-	DenseNet-Places	Test set	Whole city	Whole city
RioStreetSat21	RioSat21	DenseNet-Places	Test set	Whole city	Whole city
RioStreetSat19	RioSat19	DenseNet-Places	Test set	Whole city	Whole city
RioStreetSat19	SpSat19	DenseNet-Places	Test set	Whole city	Whole city
SpSat19	$zoom = 19$	-	Test set	Test set	Whole city
SpStreet	-	DenseNet-Places	Test set	Test set	Whole city
SpStreetSat19	RioSat19	DenseNet-Places	Test set	Test set	Whole city
SpStreetSat19	SpSat19	DenseNet-Places	Test set	Test set	Whole city
MultiCity-StreetSat19	RioSat19	DenseNet-Places	Test set	Test set	Whole city
MultiCity-StreetSat19	SpSat19	DenseNet-Places	Test set	Test set	Whole city

StreetSat model trained with Rio de Janeiro and São Paulo dengue data all together is evaluated for each city test set and in Salvador, across all city location points.

To simplify the Deep ConvNet models identification, the format "*CityArchitecture*" and "*CityArchitectureZoom*" are used to denote models trained in specific cities using one of the proposed architectures, with certain zoom level, when applied. In *StreetSat* architecture that uses *DenseNet161-Sat* features and *DenseNet161-Places* features, specifically for aerial features, the *Sat* model that generates the features is also specified. The *StreetSat* model trained and validated using multiple cities dengue data is identified as *MultiCity-StreetSatZoom*.

All models trained in Rio de Janeiro were tested in the whole city of São Paulo and Salvador. Models trained and validated in São Paulo were tested in the whole city of Salvador and in the test set of Rio de Janeiro, due to the pretrained weights initialization with Rio de Janeiro models. Furthermore, *StreetSat* models, including the trained with multiple cities, were tested using as aerial input features resulted from *Sat* models from Rio de Janeiro and São Paulo, respectively.

Table 12 lists the possible combinations between each model trained in Rio de Janeiro (prefix *Rio*) and São Paulo (prefix *Sp*) when applied at their geographic limits test set, and in the other cities studied. All *Sat* models are capable of generating features for all cities, as well as street features are generated from the *DenseNet161-Places* model for all studied cities.

4.5.5 Model Evaluation Methods

The evaluation of each trained model for Rio de Janeiro and São Paulo cities was conducted over the location points of their test sets and in the test sets of other evaluated cities, in three different ways, presented in the next Subsections.

4.5.5.1 Overall Location Points

In overall location points, evaluation is conducted over individual location data points, i.e., Mean Absolute Error (MAE) is computed individually for each location point. The overall Pearson Correlation ρ is computed between actual values and estimated values for all evaluated test points. The number of location points that reached $MAE \leq 1.0$, $MAE \leq$ 10% maximum dengue value distribution, and $MAE \leq$ dengue distribution were set as auxiliary measures for the performance evaluation of the models.

4.5.5.2 Cell Region Estimation

Evaluation by cell region estimation aggregates location points into cells, with MAE being calculated at the cell level by taking the average estimation from overall points inside the cell. Similarly to the previous method, the Pearson Correlation ρ is computed between actual cell values and estimated cell values for all evaluated test cells. The number of cell regions that reached $MAE \leq 10.0$, $MAE \leq$ 10% maximum dengue value distribution, and $MAE \leq$ mean average dengue distribution were set as auxiliary measures for the performance evaluation of the models.

4.5.5.3 Risk Ranking Estimations

The dengue incidence risk ranking estimations are applied using *Risk-biased overlap* (RBO) (WEBBER; MOFFAT; ZOBEL, 2010) in ranked lists of actual values and predicted values over administrative sectors of each city, such as neighborhoods or districts. For the city of Rio de Janeiro, the *shapefiles* of neighborhoods' polygons were obtained from the city data portal (DATARIO, 2018). For São Paulo and Salvador, the administrative districts were used for RBO evaluation, obtained from the city's data portal and from *Database of Global Administrative Areas* (GADM) respectively (SMDU, 2018; GADM, 2018).

The RBO is a rank similarity measure that evaluates system retrieved lists at a specific top-k depth, i.e., at a k depth, with the possibility of lists being disjoint. The metric is based on agreement measure, which measures the proportion of overlap (the size of intersection) items, at a specific depth, and the average overlap that calculates the average between the agreement at k depth. Roughly speaking, the RBO adds weights to the agreement at each position, by geometric progression, regulated by a parameter p that indicates how abrupt the decline in weights is.

According to WEBBER; MOFFAT; ZOBEL (2010), RBO parameter p represents the user *persistance*, with a probability p of continuing visiting the next rank item. The RBO returns lower bound, and upper bound estimates, RBO_{MIN} and RBO_{MAX}, with the maximum bound being the minimal added to a residual $RBO_{MAX} = RBO_{MIN} + RBO_{RES}$. RBO also returns a score for observing the lists beyond the depth k, i.e., an extrapolated score, with RBO_{EXT}. All RBO scores values are $0 \leq RBO \leq 1.0$, with 0 indicating disjoint lists and 1.0 identical lists. Following the authors' methodology for the metric, the RBO could be reported as RBO_{EXT} followed by the $Residual$.

To compute the RBO in proposed models results, actual values are taken from the centroid of each cell's true values and distributed between the polygons of each administrative sectors, and predicted values are taken from the centroid of each cell's predicted values, totaling, in both cases, the dengue incidence values to get a single rate per administrative sector. Next, the RBO scores are computed between each ranked list from neighborhoods or districts with more dengue incidence risk to low incidence risk, using the parameters persistence $p = 0.9$ and depth $k = 10$ for Top-10 neighborhoods (Rio de Janeiro) or districts (São Paulo and Salvador). The scores for $Agreement$, $RBO_{EXT} = RBO$ and $Residual$ are reported for the best performance models.

5 RESULTS

This Chapter presents the results obtained during the training process of the proposed models, using the train and validation sets, in Section 5.1, and the results of their evaluations using test sets, in Section 5.2. Supplementary material from training and validation results can be found in Appendix B, for baseline models evaluation in Appendix C, and for Deep ConvNet models evaluation in Appendix D.

5.1 Train and Validation of Proposed Models

5.1.1 Baseline Models

The proposed linear regression models *Aerial HOG-LR* and *Aerial GIST-LR* trained over 50 epochs, reaching their best validation loss at $15th$ epoch. Next, *Street HOG-LR* model trained up to 48 epochs, reached its best validation loss at $43th$ epoch. The neural network baselines *Aerial HOG-NN* and *Street HOG-NN* models trained over 120 epochs in Rio de Janeiro train dataset, and reached the best validation loss at $17th$ and $10th$ epochs respectively. Table 13 summarizes the *Mean Square Error* (MSE) and *Root Mean Square Error* (RMSE) and Pearson Correlation ρ achieved for best validation loss during training process of all baseline models.

Figures 48, 49 and 50 in Appendix B presents the training process of baseline models with loss curves for each model.

Table 13 – Validation results during *Aerial HOG/GIST-LR* training in Rio de Janeiro. All *p-values* for correlation coefficient equal to ≤ 0.001

Model	MSE	RMSE	Pearson ρ	Best Loss Epoch
Aerial HOG-LR	16,216.44	127.34	0.20	15
Aerial GIST-LR	14,736.89	121.39	0.35	12
Street HOG-LR	**14,001.72**	**118.32**	**0.42**	**43**
Aerial HOG-NN	15,664.30	125.15	0.26	17
Street HOG-NN	14,216.18	119.23	0.41	10

5.1.2 Deep Convolutional Neural Network Models

5.1.2.1 Training and validating models in Rio de Janeiro

The Deep ConvNet models *Sat* with different zoom levels $zoom = \{21, 20, 19\}$ were trained in Rio de Janeiro over 60 epochs, reaching the best validation loss at $21st$ epoch for $zoom = 21$, at $17th$ epoch for $zoom = 20$ and at $13th$ epoch for $zoom = 19$. Training the *Street* model up to 100 epochs achieved the best validation loss of at $38th$ epoch. Figure 52 show the training, validation and pearson correlations resulting during training epochs. Table 14 enumerates all validation results from Deep ConvNet *Sat* and *Street* models in Rio de Janeiro.

The proposed *StreetSat* model was trained and tested with aerial features from *Sat* model with zoom levels $zoom = \{21, 19\}$ and street-level features from *DenseNet161-Places* over $50th$ epochs, achieving the best validation loss at the $8th$ epoch for *RioStreetSat21*. Table 14 enumerates validation results from Deep ConvNet *RioStreetSat21* and *RioStreetSat19* in Rio de Janeiro.

In Appendix B are presented Figures 51a, 51b and 51c that depicts train and validation loss curves during training epochs for each *Sat* model. Figures 53a and 53b depicts the training process, with Pearson correlations at best validation loss being $\rho = 0.68$ for both models.

5.1.2.2 Training and validating models in São Paulo

The *Sat* model in the city of São Paulo, using pretrained network weights from *RioSat19* in Rio de Janeiro as initializations, trained up to 40 epochs, whose continuity showed no improvement, and reached the best validation loss at $7th$ epoch. The best *Street* model for São Paulo, fine-tuned from Rio de Janeiro *RioStreet* model weights, trained over 100 epochs, achieved at early $9th$ epoch the best validation loss.

For *StreetSat* model, using aerial features from *Sat19* trained in São Paulo and street-level features from *Densenet161-Places*, training was over 50 epochs, achieving the best validation loss at $16th$ epoch. Table 15 enumerates enumerates all validation results from Deep ConvNet *Sat*, *Street* and *StreetSat* models in São Paulo.

Figures 54, 55 and 56 in Appendix B shows the loss curves in the training process

Table 14 – Validation results during training Deep ConvNet models in Rio de Janeiro. All *p-values* for correlation coefficient equal to ≤ 0.001

Model	MSE	RMSE	Pearson ρ	Best Loss Epoch
RioSat21	9,892.10	99.45	0.67	21
RioSat20	11,219.02	105.91	0.65	17
RioSat19	9,608.20	98.02	0.71	13
RioStreet	10,560.60	102.76	0.61	38
RioStreetSat21	**9,304.02**	**96.45**	**0.68**	**8**
RioStreetSat19	11,780.45	108.53	0.68	20

for each model trained and validated in São Paulo.

5.1.2.3 Training and validating a multiple city model

Aerial features for both cities used to train the multiple city model were extracted from pretrained *Sat* $zoom = 19$ (*Sat19*) models and street-level features from *DenseNet161-Places*. The model trained for over 50 epochs, achieving the best loss for validation set with data from both cities at $27th$ epoch. Table 16 presents the best validation results achieved while training the model.

The loss curves obtained during the process of training a *StreetSat* model using aerial and street-level features from different cities ("MultiCity") are depicted in Appendix B, Figure 57.

5.2 Evaluation of Proposed Models

The Subsection 5.2.1 presents the overall location points evaluation results, and for cell region estimation evaluation, obtained with baseline and Deep ConvNet models applied in Rio de Janeiro and São Paulo test sets. Next, Subsection 5.2.2 presents results from geographic portability of the models, i.e., results obtained with the application of the trained models in different cities, whose image examples were not seen during the training phase.

Additionally, the best resulting models were evaluated by dengue incidence risk ranking in districts, at Subsection 5.2.3. For the highlighted models, in each group, is presented the correlation plot between actual and estimated values, depicted with a tendency line, and the dengue incidence rates *hot-spot* map produced by the model.

5.2.1 Summary of Evaluations Results

Table 17 enumerates the evaluation of all trained and tested *Sat*, *Street* and *StreetSat* models with the city of Rio de Janeiro, from a total of $24,320$ test set location points,

Table 15 – Validation results during training Deep ConvNet models in São Paulo. All *p-values* for correlation coefficient equal to ≤ 0.001

Model	MSE	RMSE	Pearson ρ	Best Loss Epoch
SpSat19	1,058.00	32.52	0.46	7
SpStreet	**1,037.80**	**32.21**	**0.36**	**9**
SpStreetSat19	1,421.98	37.70	0.44	16

Table 16 – Validation results during training Deep ConvNet models in Rio de Janeiro and São Paulo with a "MultiCity" approach. All *p-values* for correlation coefficient equal to ≤ 0.001

Model	MSE	RMSE	Pearson ρ	Best Loss Epoch
MultiCity-StreetSat19	6,646.01	81.52	0.78	27

presenting the *mean squared error* (MSE) and their *root mean squared error* (RMSE), the *mean absolute error* (MAE), the estimations that had their absolute error ≤ 1.0 and Pearson Correlation coefficient for each model. Table 18 presents the evaluation of all trained and tested models for the city of São Paulo, from a total of $31,168$ test set location points.

Table 19 enumerates the evaluation of all trained and tested *Sat*, *Street* and *StreetSat* models with the city of Rio de Janeiro for cell region estimation, from a total of 352 test set cells, presenting the *mean squared error* (MSE) and their *root mean squared error* (RMSE), the *mean absolute error* (MAE), the estimations that had their absolute error ≤ 10.0 and Pearson Correlation coefficient for each model. Table 20 presents the evaluation of all trained and tested models for the city of São Paulo, from a total of 252 test set location points.

Figure 22 depicts the results obtained when estimating dengue rates for location points in test data set for the highlighted *StreetSat* models evaluated in Rio de Janeiro, with actual versus estimated plots. The correlation from overall location points between actual and estimated values for dengue incidence rates with *SpSat19* and *SpStreetSat19* for São Paulo test data set are depicted in Figure 23. The plots shows locations points that obtained absolute estimation errors ≤ 1.0 in red, and absolute errors > 1.0. Pearson correlation plots and *hot-spot* maps for baseline and other Deep ConvNet models evaluated in Rio de Janeiro are enumerated in Appendix C and D respectively.

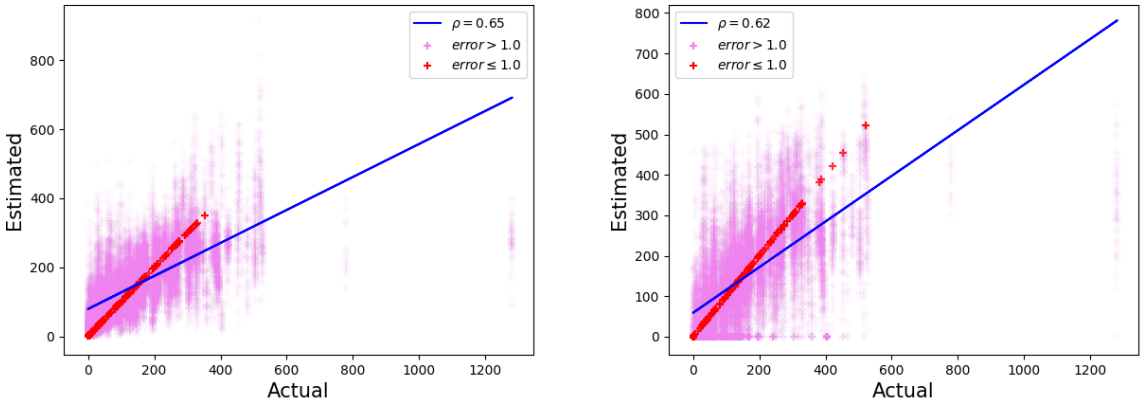
Figure 24 depicts the results obtained for *StreetSat* highlighted models when es-

Table 17 – Results for Rio de Janeiro model, for a total of $24,320$ test set location points. All Pearson Correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
Aerial HOG-LR	$zoom = 21$	19,378.88	139.20	106.12	124 (0.5%)	0.19
Aerial GIST-LR	$zoom = 21$	17,656.26	132.87	101.60	109 (0.4%)	0.36
Aerial HOG-NN	$zoom = 21$	19,176.39	138.47	107.70	110 (0.4%)	0.26
Street HOG-LR	-	16,225.00	127.37	93.56	120 (0.4%)	0.44
Street HOG-NN	-	15,987.72	126.44	92.58	142 (0.5%)	0.46
RioSat21	$zoom = 21$	12,204.40	110.47	71.58	252 (1%)	0.64
RioSat20	$zoom = 20$	12,714.50	112.75	72.13	285 (1%)	0.64
RioSat19	$zoom = 19$	13,524.16	116.29	72.72	285 (1%)	0.61
RioStreet	-	13,258.24	115.14	77.76	258 (1%)	0.58
RioStreetSat21	**RioSat21**	**11,677.62**	**108.06**	**69.16**	357 (1%)	**0.65**
RioStreetSat19	**RioSat19**	14,073.05	118.62	78.32	**1,252 (5%)**	0.62
MultiCity-StreetSat19	RioSat19	15,460.42	124.33	77.08	465 (1%)	0.59

Table 18 – Results for São Paulo models, for a total of $31,168$ test set location points. All Pearson Correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
SpSat19	$zoom = 19$	**2,261.35**	**47.55**	**22.70**	1,726 (5%)	0.51
SpStreet	-	2,739.88	52.34	25.83	921 (2%)	0.32
SpStreetSat19	**SpSat19**	2,293.74	47.89	24.43	**2,999 (9%)**	**0.52**
MultiCity-StreetSat19	SpSat19	5,523.55	74.32	44.14	869 (2%)	0.48



(a) *RioStreetSat21* (b) *RioStreetSat19*

Figure 22 – Overall individual correlation, from results of Rio de Janeiro highlighted models, between actual and estimated values for each location point in test set with aerial images with $zoom = 21$ and (b) for $zoom = 19$. All Pearson correlation ρ with $p - value < 0.001$

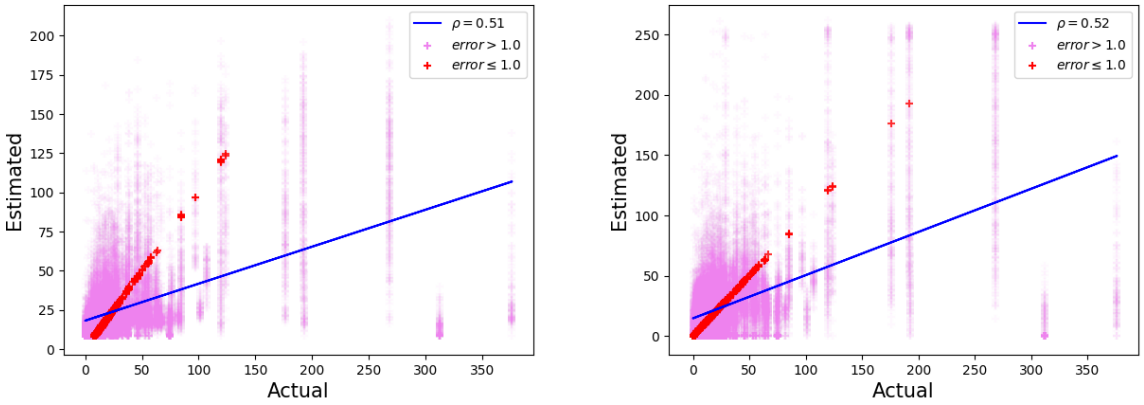
timating dengue rates for aggregated estimations in test data set cells for Rio de Janeiro, with actual versus estimated plots, with features from *RioSat* models trained with $zoom = \{21, 19\}$. The plot shows test cells that obtained absolute estimation errors ≤ 10.0 (highlighted) and absolute errors > 10.0. The standard deviation of each incidence average in the cell is shown as an error bar.

Table 19 – Results for Rio de Janeiro models, for a total of 352 test set cells. All Pearson Correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
Aerial HOG-LR	$zoom = 21$	22,604.94	150.34	123.34	12 (0.5%)	0.29
Aerial GIST-LR	$zoom = 21$	20,120.58	141.84	116.20	8 (0.2%)	0.46
Aerial HOG-NN	$zoom = 21$	23,386.37	152.92	128.23	10 (0.2%)	0.38
Street HOG-LR	-	17,027.95	130.49	101.93	11 (0.3%)	0.57
Street HOG-NN	-	16,302.85	127.68	98.59	14 (0.3%)	0.59
RioSat21	$zoom = 21$	9,342.54	96.65	60.80	34 (9%)	0.76
RioSat20	$zoom = 20$	9,183.5	95.83	56.3	49 (14%)	0.76
RioSat19	$zoom = 19$	10,343.67	101.70	57.72	75 (21%)	0.72
RioStreet	-	11,607.13	107.73	70.75	40 (11%)	0.70
RioStreetSat21	**RioSat21**	8,946.21	94.58	**51.21**	52 (15 %)	**0.78**
RioStreetSat19	**RioSat19**	**8,925.93**	**94.47**	51.25	**112 (32%)**	0.76
MultiCity-StreetSat19	RioSat19	10,519.52	102.56	52.97	104 (29%)	0.75

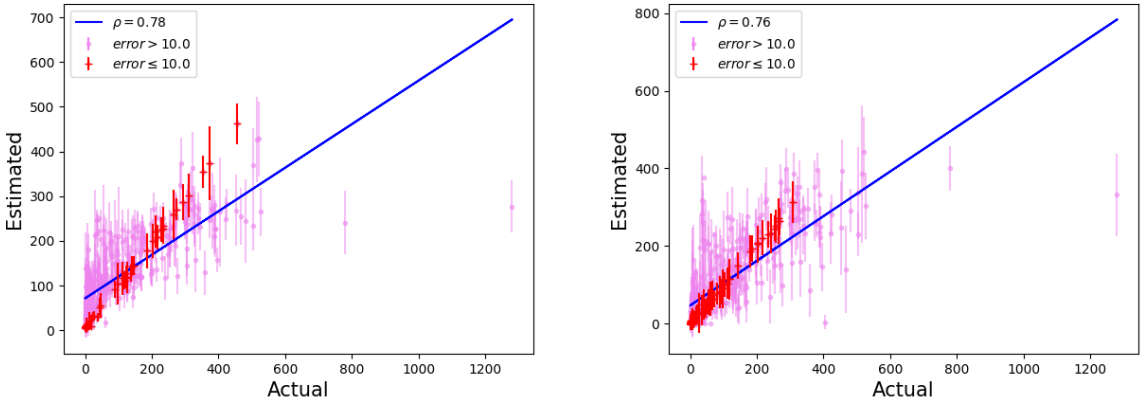
Table 20 – Results for São Paulo models, for a total of 252 test set cells. All Pearson Correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
SpSat19	$zoom = 19$	1,369.96	37.01	17.09	133 (52%)	0.57
SpStreet	-	1,672.77	40.89	20.00	87 (34%)	0.40
SpStreetSat19	**SpSat19**	**1,247.75**	**35.32**	**15.34**	**149 (59%)**	**0.61**
MultiCity-StreetSat19	SpSat19	2,613.49	51.12	34.61	41 (16%)	0.54



(a) *SpSat19* (b) *SpStreetSat19*

Figure 23 – Overall individual correlation, from results of São Paulo highlighted models, between actual and estimated values for each location point in test set. Pearson correlation with *p-value* < 0.001.

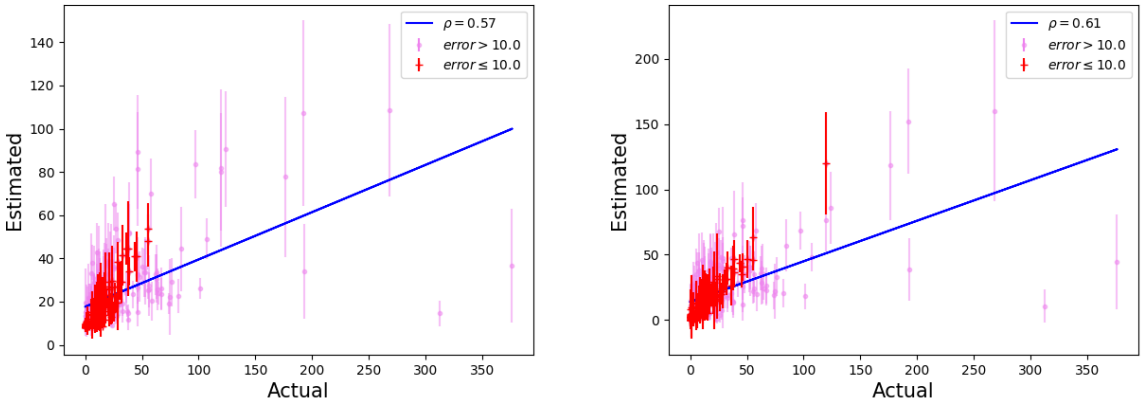


(a) *RioStreetSat21* (b) *RioStreetSat19*

Figure 24 – Cell estimation overall correlation for *StreetSat* models trained and tested in Rio de Janeiro. All Pearson correlation ρ with *p-value* < 0.001

Figure 25 depicts the results obtained from *SpSat19* and *SpStreetSat19* models trained and tested with São Paulo dengue fever data, for cell region estimates aggregated using the mean average of all predictions per cell, with actual versus estimated plots. For aerial imagery models, the zoom level applied was $zoom = \{19\}$. The plots highlight locations points that obtained the estimated absolute error of ≤ 10.0.

In Figure 26 it is presented the *hot-spots* maps of dengue fever estimations for cell region in Rio de Janeiro, computed with *RioStreetSat21* and *RioStreetSat21* models using *Sat* features from $zoom = \{21, 19\}$ as aerial inputs. Actual values of dengue are depicted in 26a for comparison. Figure 26b depicts the estimations for cell region with *RioStreetSat21*, with minimum estimation equal to 4.27 and maximum prediction equal to 461.87 dengue incidence rate. In *RioStreetSat19* (Figure 26c), minimum dengue



(a) *SpSat19* (b) *SpStreetSat19*

Figure 25 – Cell region estimation correlation, from results of São Paulo highlighted models, between actual and estimated values for each location point. Pearson correlation with *p-value* < 0.001.

estimated value equal 0 and maximum dengue estimated value equal 441.77.

Figure 27 presents the *hot-spots* maps of dengue fever estimations for cell region for all models trained and evaluated in São Paulo with *SpSat19* and *SpStreetSat19* models. Actual values of dengue are depicted in 27a for comparison. Figure 27b depicts the estimations for cell region with *SpSat19*, with minimum estimation equal to 8.29 and maximum prediction equal to 108.53 dengue incidence rate. In *SpStreetSat19* minimum dengue estimated value equal 0 and maximum dengue estimated value equal 160.10.

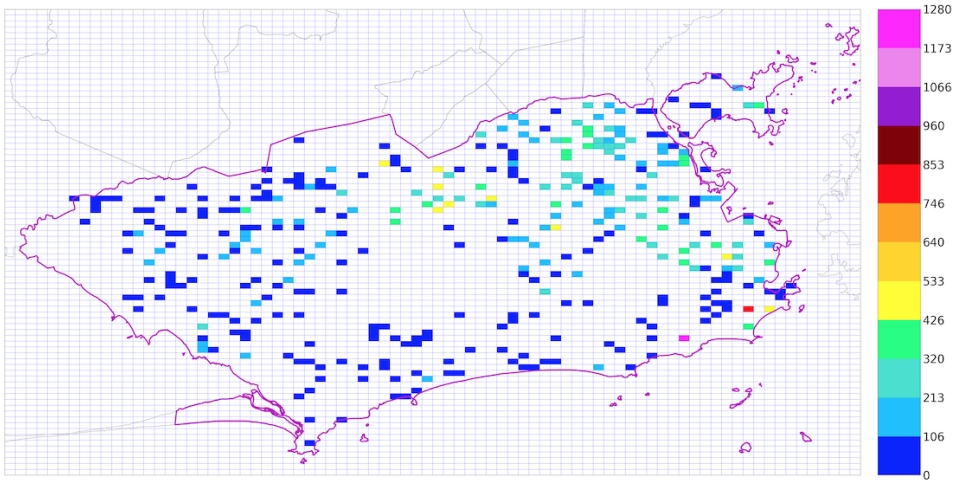
5.2.2 Geographic Portability Of Models

5.2.2.1 Rio de Janeiro Models applied in São Paulo and Salvador

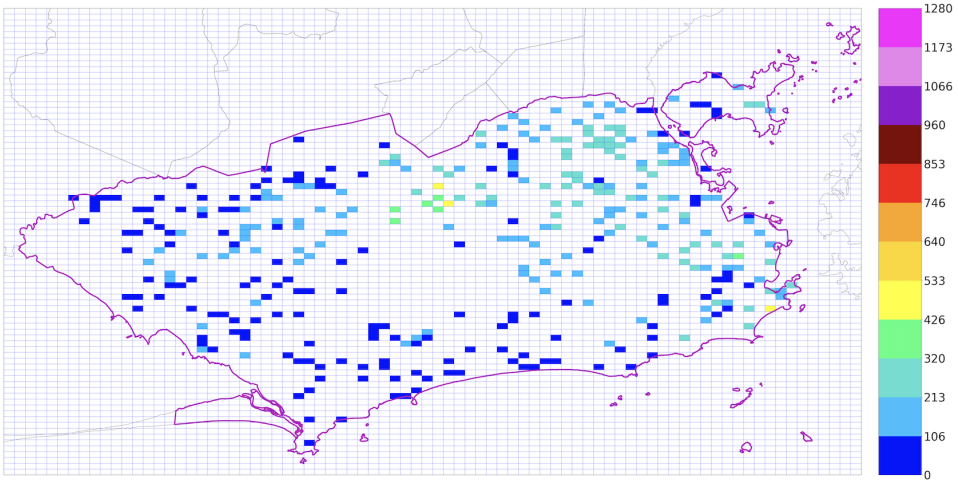
Table 21 presents the results while evaluating the models trained in Rio de Janeiro in $281,184$ location points from São Paulo city, and Table 22 presents the results by cell region estimation, for $1,718$ cells. For each model result, the column "Aerial Features" denote which network model produced the features to be used in the evaluation, when applicable.

For the highlighted *RioSat19*, *RioStreet* and *RioStreetSat19* models evaluated in São Paulo, Figure 28 depicts the results obtained when estimating dengue rates for all location points from São Paulo, in actual versus estimated plots. The plots shows locations points that obtained absolute estimation errors ≤ 1.0 in red, and absolute errors > 1.0.

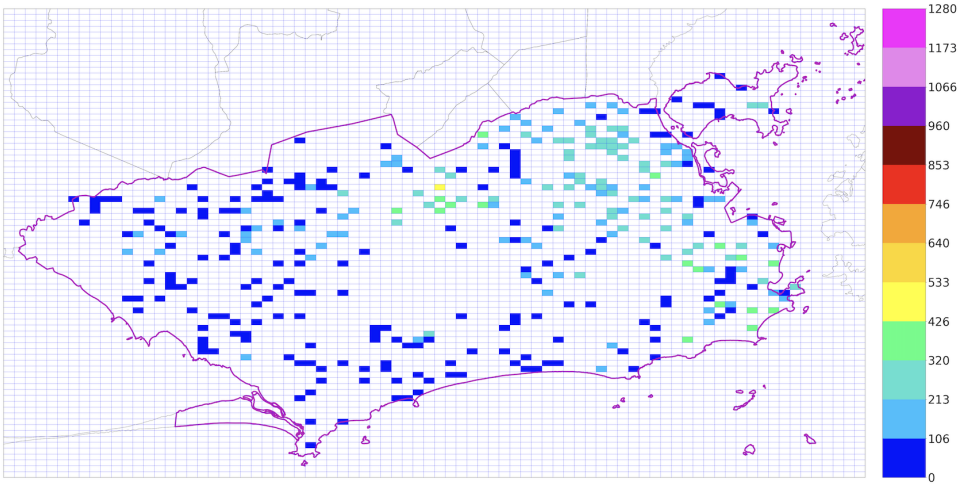
Figure 29 depicts the results obtained from *RioSat19*, *RioStreet* and *RioStreetSat19* models trained in Rio de Janeiro applied in whole city of São Paulo, considering the cell region estimations, with actual versus estimated plots. The plots highlight loca-



(a) Actual

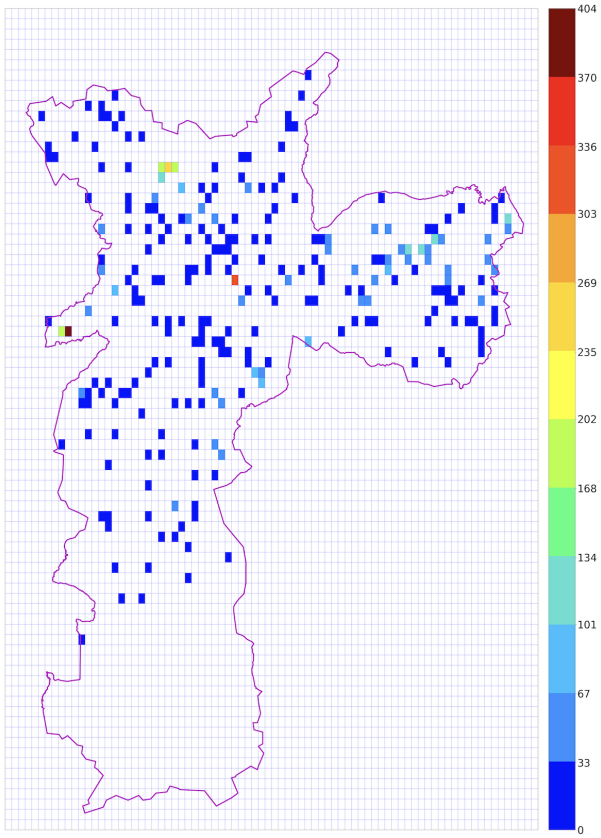


(b) *RioStreetSat21*

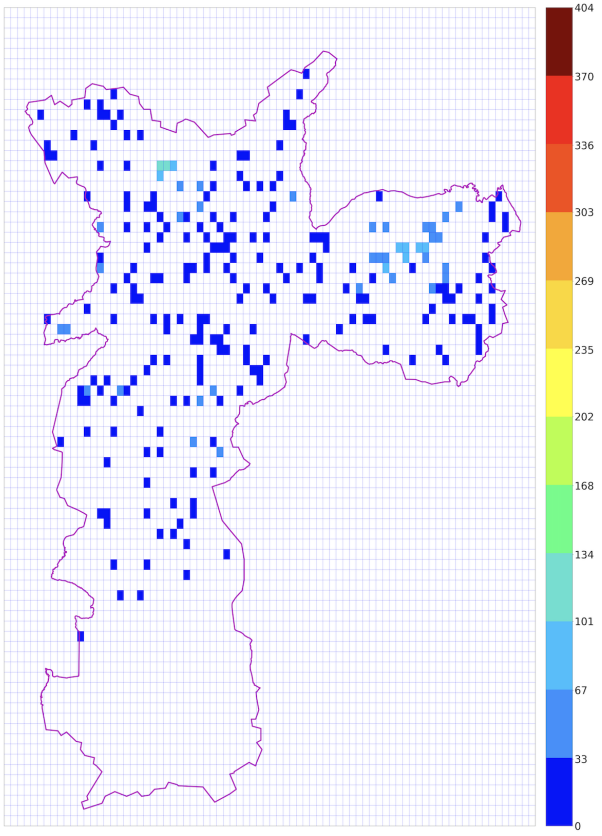


(c) *RioStreetSat19*

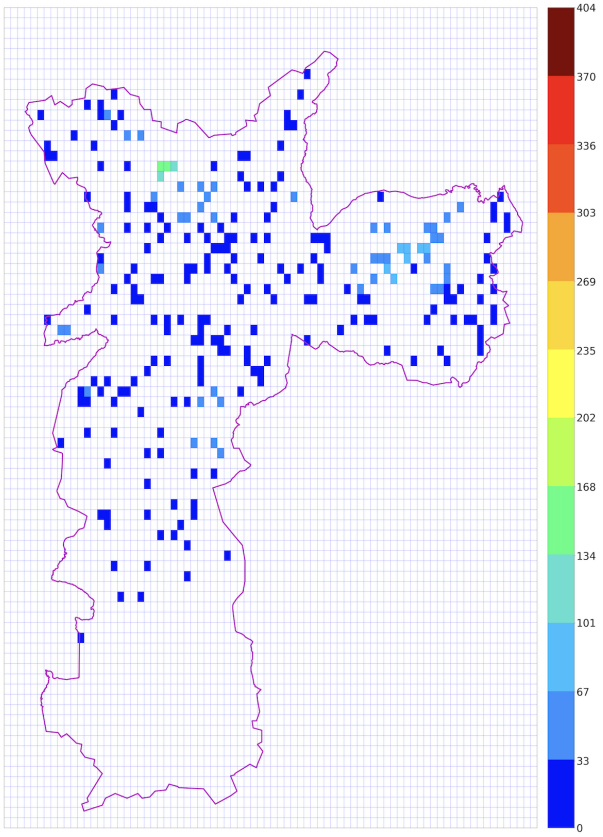
Figure 26 – Cell estimation hot-spot maps with *StreetSat* models (a) actual values for dengue fever rates and (b), (c) resulting estimated dengue values for Rio de Janeiro test set cells. Dengue rates ranges from 0 to a maximum of $1,280$, with maximum prediction range equal 461.87, depicted with shades of yellow.



(a) Actual dengue rates



(b) *SpSat19*



(c) *SpStreetSat19*

Figure 27 – Cell estimation maps with São Paulo models (a) actual values for dengue fever rates and (b), (c) resulting estimated dengue values for test set cells. Dengue rates ranges from 0 to a maximum of 404, with maximum prediction range equal 160, depicted with shades of green.

tions points that obtained the estimated absolute error of ≤ 10.0.

Figure 30 presents the *hot-spots* maps created with Rio de Janeiro models *RioSat19*, *RioStreet* and *RioStreetSat19* with *RioSat19* aerial features while evaluating dengue fever cell region estimations in whole city of São Paulo. Actual values of dengue are depicted in 30a for comparison. All models' estimated ranges are contained within the actual city dengue incidence rate spectrum.

Next, for Salvador city, Table 23 summarizes the results of applying Rio de Janeiro models for $84,192$ location points, and Table 24 shows the results by cell region estimation, for 834 cells over the territory. For each model result, the column "Aerial Features" denote which network model produced the features to be used in the evaluation, when applicable.

Table 21 – Results from Rio de Janeiro models applied in São Paulo, for $281,184$ location points. For *RioStreetSat19* with *SpSat19* aerial features, results are computed over São Paulo $31,168$ test set location points only. All Pearson correlation with *p-value* < 0.001.

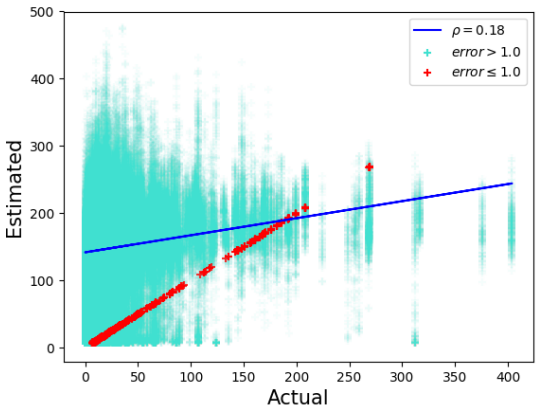
Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
RioSat21	$zoom = 21$	21,494.95	146.61	129.31	606 (0.2%)	0.07
RioSat20	$zoom = 20$	34,854.67	186.69	152.72	718 (0.2%)	0.14
RioSat19	$zoom = 19$	**18,788.87**	**137.07**	**123.08**	1,010 (0.3%)	0.18
RioStreet	-	28,662.16	169.29	157.42	234 (0.08%)	**0.24**
RioStreetSat21	RioSat21	24,184.25	155.51	125.8	889 (0.3%)	0.06
RioStreetSat19	**RioSat19**	26,780.5	163.65	131.32	**9,508 (3%)**	0.17
RioStreetSat19	SpSat19	22,344.84	149.48	113.57	1,558 (4%)	0.30

Table 22 – Results from Rio de Janeiro models applied in São Paulo, for $1,718$ cells. For *RioStreetSat19* with *SpSat19* aerial features, results are computed over São Paulo 252 test set cells only. All Pearson correlation with *p-value* < 0.001.

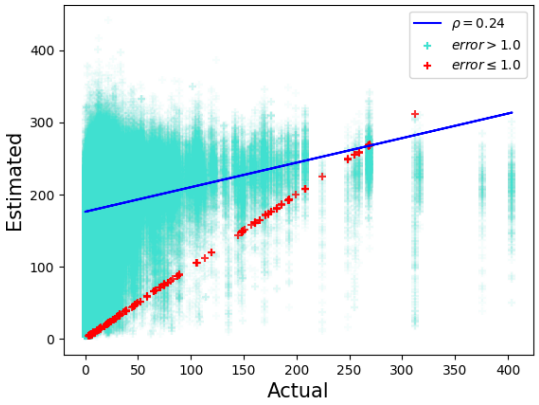
Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
RioSat21	$zoom = 21$	18,080.65	134.46	127.49	10 (0.5%)	0.18
RioSat20	$zoom = 20$	25,155.56	158.60	143.83	7 (0.4%)	0.19
RioSat19	$zoom = 19$	**15,321.93**	**123.78**	114.51	32 (1.8%)	0.31
RioStreet	-	24,321.66	155.95	146.45	38 (2%)	**0.38**
RioStreetSat21	RioSat21	17,145.21	130.93	121.06	16 (0.9%)	0.21
RioStreetSat19	**RioSat19**	16,765.59	129.48	**109.91**	**163 (9%)**	0.30
RioStreetSat19	SpSat19	14,501.37	120.42	99.15	16 (0.6%)	0.52

Table 23 – Results from Rio de Janeiro models applied in Salvador, for $84,192$ city points. All Pearson correlation with *p-value* < 0.001.

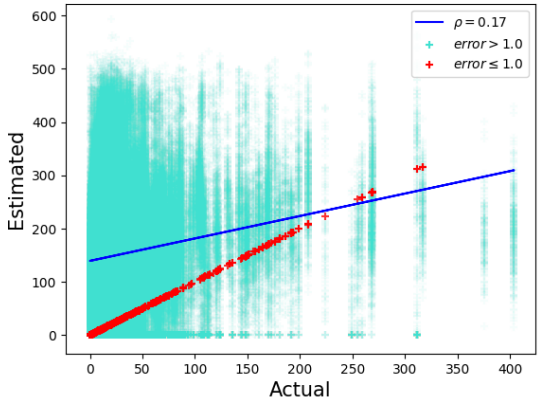
Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
RioSat21	$zoom = 21$	39,372.54	198.42	172.61	4 (0.004%)	**0.20**
RioSat20	$zoom = 20$	46,316.61	215.21	177.62	7 (0.008%)	0.11
RioSat19	$zoom = 19$	**17,028.74**	**130.49**	115.32	5 (0.005%)	0.19
RioStreet	-	31,631.38	177.85	163.08	0 (0%)	0.18
RioStreetSat21	RioSat21	32,323.36	179.79	151.065	16 (0.01%)	0.05
RioStreetSat19	**RioSat19**	32,757.88	180.99	143.48	**12,703 (15%)**	0.16
RioStreetSat19	SpSat19	32,907.00	181.40	142.42	10,886 (13%)	0.18
MultiCity-StreetSat19	**RioSat19**	19,146.15	138.37	**97.06**	111 (0.1%)	0.17



(a) *RioSat19*



(b) *RioStreet*



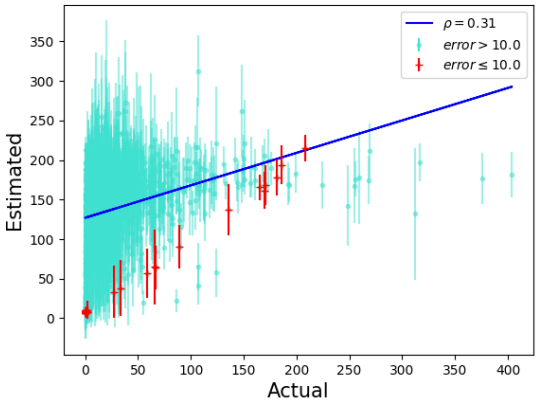
(c) *RioStreetSat19*, with *RioSat19* aerial features.

Figure 28 – Overall individual estimations for Rio de Janeiro models applied in the whole city of São Paulo. All Pearson correlation ρ with *p-value* < 0.001

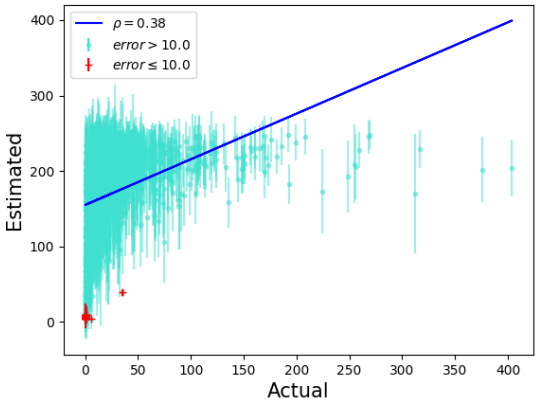
Figure 31 depicts the results obtained from the highlighted *RioSat21*, *RioSat19*, *RioStreetSat19* and *MultiCity-StreetSat19* models when evaluated in whole city of Salvador, for location points in all city examples, in actual versus estimated plots. The plots shows locations points that obtained absolute estimation errors ≤ 1.0 in red, and

Table 24 – Results from Rio de Janeiro models applied in Salvador, for 834 cells region estimations. All Pearson correlation with *p-value* < 0.001.

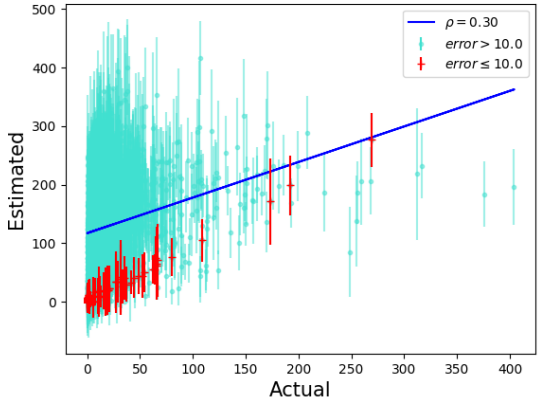
Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
RioSat21	$zoom = 21$	28,053.99	167.49	155.61	0 (0%)	**0.33**
RioSat20	$zoom = 20$	34,832.33	186.63	168.00	0 (0%)	0.17
RioSat19	$zoom = 19$	14,318.29	119.65	109.64	4 (0.4%)	0.25
RioStreet	-	24,750.95	157.32	143.76	7 (0.8%)	0.24
RioStreetSat21	RioSat21	28,267.46	168.13	155.18	0 (0%)	0.06
RioStreetSat19	RioSat19	23,940.75	154.73	130.62	55 (6%)	0.23
RioStreetSat19	SpSat19	20,198.89	142.12	117.69	**75 (8%)**	0.29
MultiCity-StreetSat19	RioSat19	**13,921.70**	**117.99**	**94.36**	32 (3%)	0.22



(a) *RioSat19*



(b) *RioStreet*



(c) *RioStreetSat19*, with *RioSat19* aerial features.

Figure 29 – Cell region estimations for Rio de Janeiro models applied in the whole city of São Paulo. All Pearson correlation ρ with *p-value* < 0.001

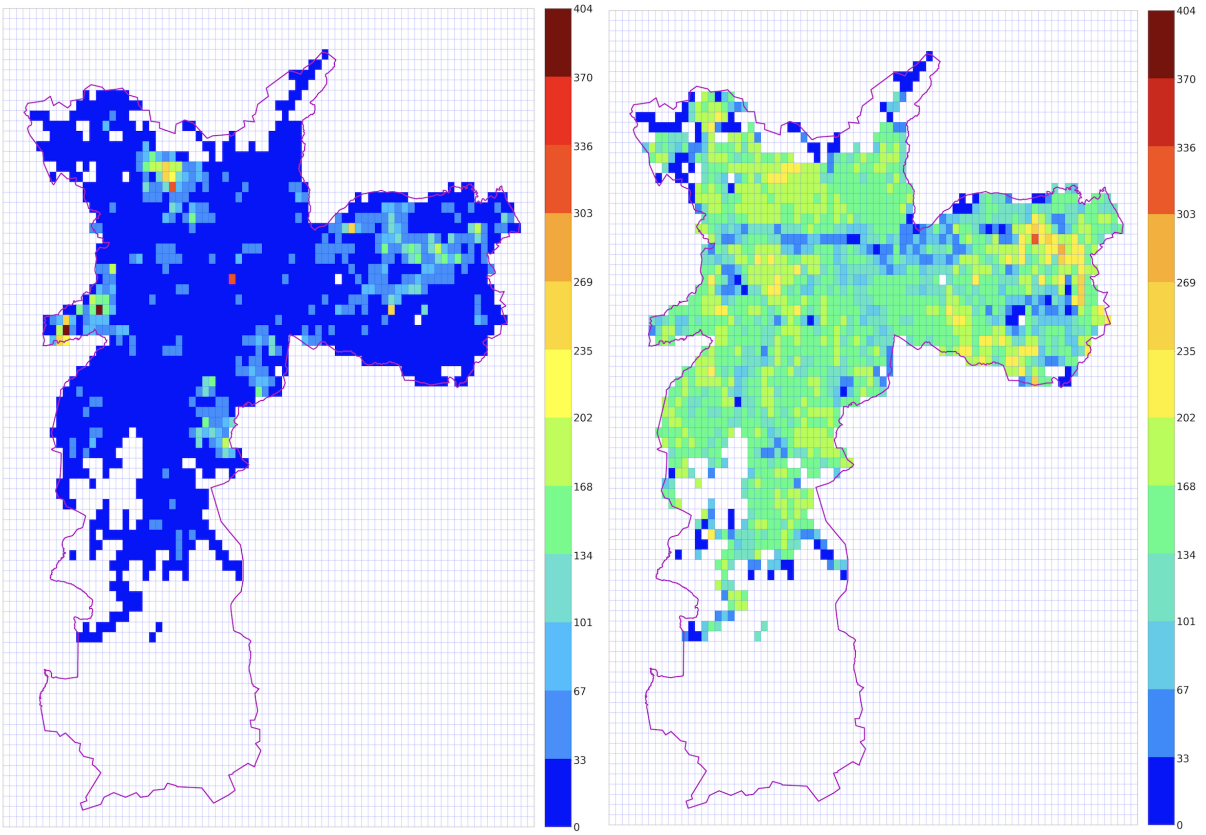
absolute errors > 1.0.

Figure 32 depicts the results obtained from *RioSat21*, *RioStreetSat19* using *SpSat19* aerial features and *MultiCity-StreetSat19* with *RioSat19* aerial features, the models trained in Rio de Janeiro with highlighted results for cell region estimations, when applied in whole city of Salvador. Plots depict actual versus estimated correlation, with locations points that obtained the estimated absolute error of ≤ 10.0 selected in red.

5.2.2.2 São Paulo Models applied in Salvador and Rio de Janeiro

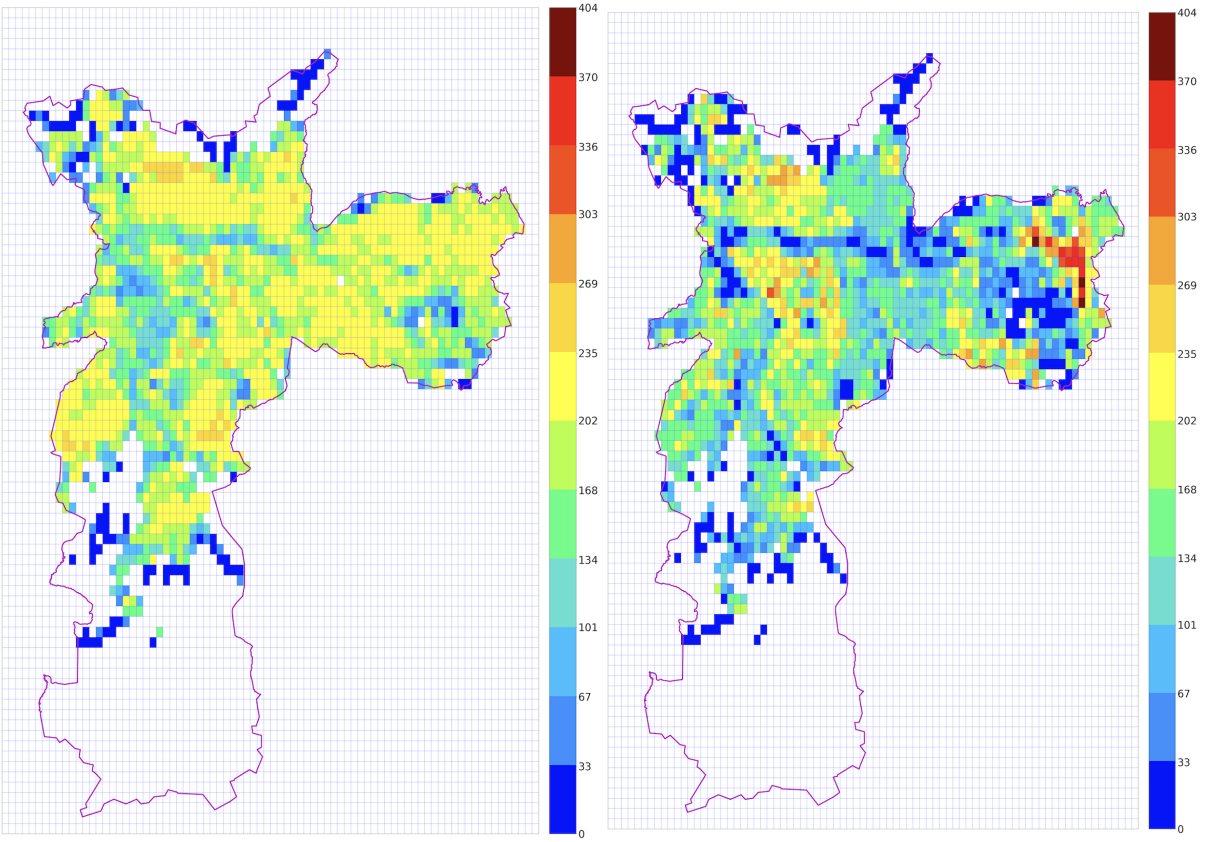
Table 25 presents the results while evaluating the models trained in São Paulo with $84,192$ location points from Salvador city, and Table 26 presents the results by cell region estimation, for the 834 cells of Salvador. Column “Aerial Features” denote which network model produced the aerial features to be used in the evaluation, when applicable.

Figure 33 depicts the correlation between actual and estimated overall location point



(a) Actual dengue rates

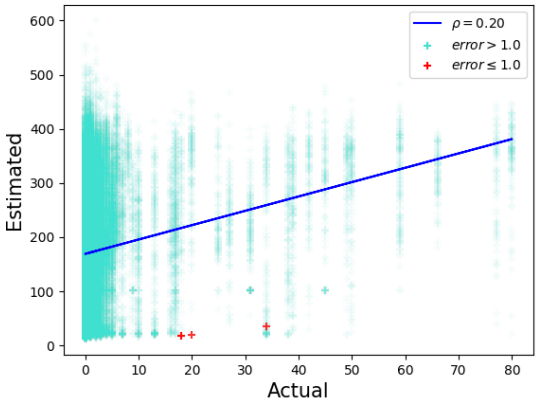
(b) *RioSat19*



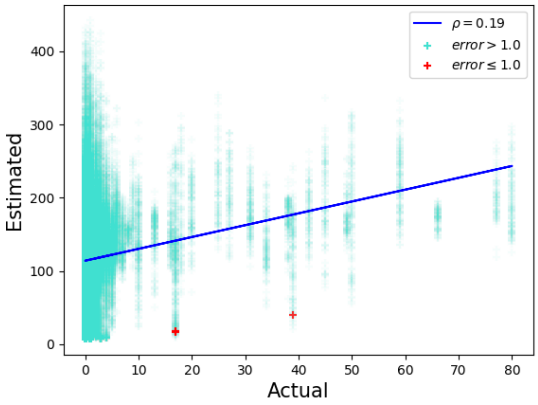
(c) *RioStreet*

(d) *RioStreetSat19*, with *RioSat19*.

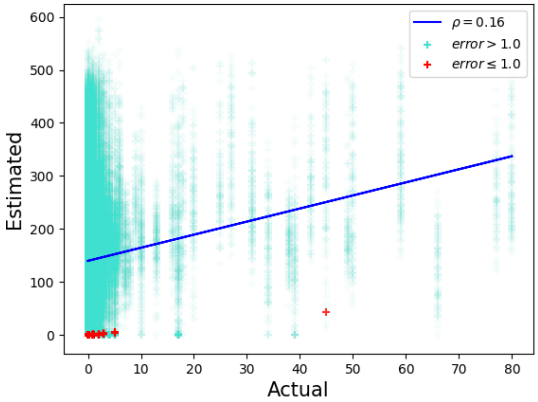
Figure 30 – Cell estimation maps with Rio de Janeiro models applied in whole city of São Paulo: (a) actual values for dengue fever rates and (b)-(d) resulting estimated dengue values for cells.



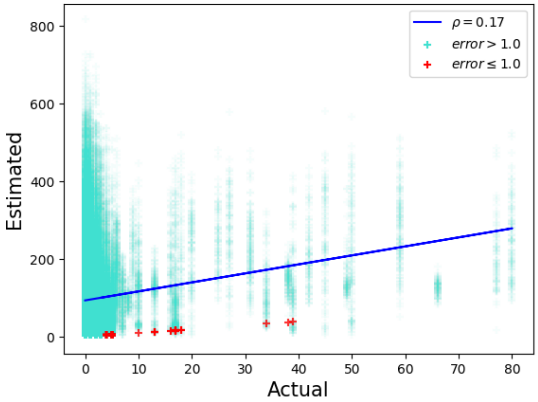
(a) *RioSat21*



(b) *RioSat19*



(c) *RioStreetSat19*, with *RioSat19* aerial features.



(d) *MultiCity-StreetSat19*, with *RioSat19* aerial features.

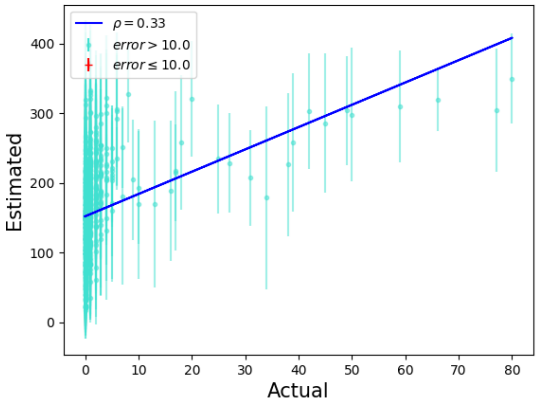
Figure 31 – Overall individual estimations with Rio de Janeiro models applied in the whole city of Salvador. All Pearson correlation ρ with *p-value* < 0.001

results, obtained from the highlighted models *SpStreet* and *SpStreetSat19* with *SpSat19* aerial features, when evaluated in whole city of Salvador. The plots shows locations points that obtained absolute estimation errors ≤ 1.0 in red, and absolute errors > 1.0.

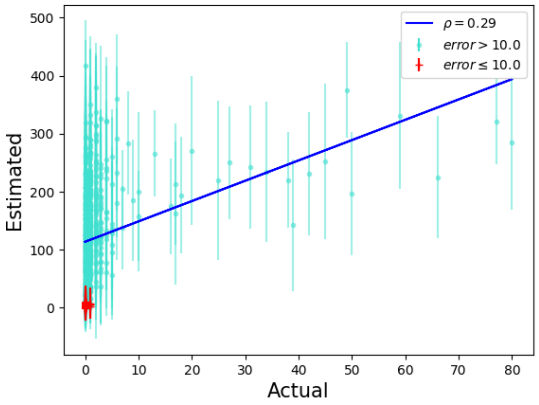
Figure 34 depicts the results obtained from *SpStreet* and *SpStreetSat19* with *SpSat19* aerial features, the models trained in São Paulo with highlighted results for cell

Table 25 – Results from São Paulo models applied in Salvador, for $84,192$ location points. All Pearson correlation with *p-value* < 0.001.

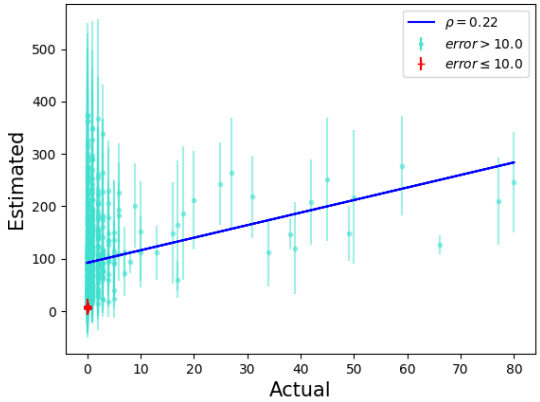
Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
SpSat19	$zoom = 19$	2,025.66	45.00	31.63	102 (0.1%)	0.20
SpStreet	-	**825.51**	**28.73**	25.39	87 (0.1%)	**0.22**
SpStreetSat19	SpSat19	1,489.57	38.59	24.98	18,539 (22%)	0.19
SpStreetSat19	**RioSat19**	1,146.91	33.87	**24.21**	**20,211 (24%)**	0.18
MultiCity-StreetSat19	SpSat19	8,071.25	89.84	60.69	391 (0.4%)	0.21



(a) *RioSat21*



(b) *RioStreetSat19*, with *SpSat19* aerial features.



(c) *MultiCity-StreetSat19*, with *RioSat19* aerial features.

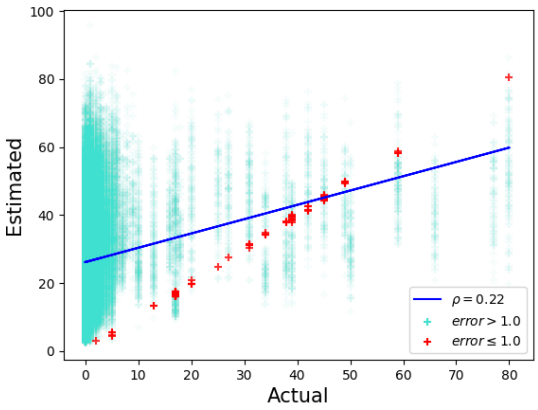
Figure 32 – Cell region estimations with Rio de Janeiro models applied in the whole city of Salvador. All Pearson correlation ρ with *p-value* < 0.001

region estimations, when applied in whole city of Salvador. Plots depict actual versus estimated correlation, with locations points that obtained the estimated absolute error of ≤ 10.0 selected in red.

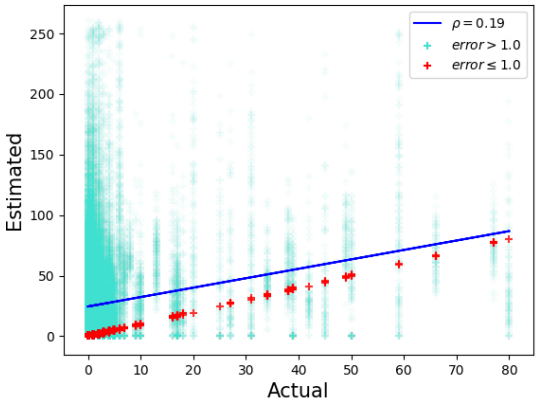
Figure 35 presents the *hot-spots* maps created with São Paulo models *SpStreet*, *SpStreetSat19* with *SpSat19* aerial features while evaluating dengue fever cell region estimations in whole city of Salvador. Actual values of dengue are depicted in 35a for

Table 26 – Results from São Paulo models applied in Salvador, for 834 cells region estimations. All Pearson correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
SpSat19	$zoom = 19$	1,098.04	33.13	26.86	146 (17%)	**0.33**
SpStreet	-	**683.53**	**26.14**	23.87	93 (11%)	0.28
SpStreetSat19	**SpSat19**	726.75	26.96	**20.52**	**277 (33%)**	0.31
SpStreetSat19	RioSat19	847.25	29.11	23.05	257 (30%)	0.23
MultiCity-StreetSat19	SpSat19	4,564.49	67.56	54.11	36 (0.4%)	**0.33**

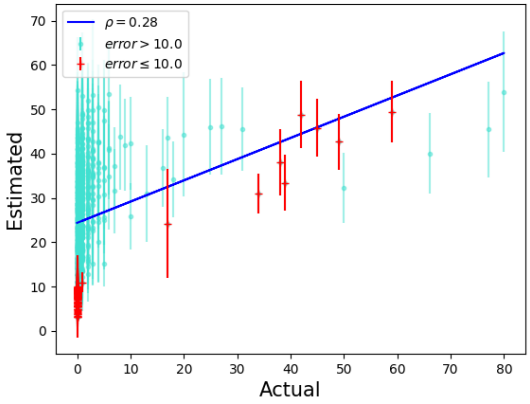


(a) *SpStreet*

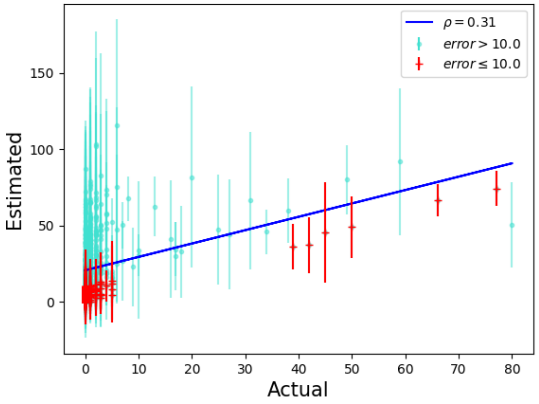


(b) *SpStreetSat19*, with *SpSat19* aerial features.

Figure 33 – Overall individual estimations with São Paulo models applied in the whole city of Salvador. All Pearson correlation ρ with *p-value* < 0.001



(a) *SpStreet*



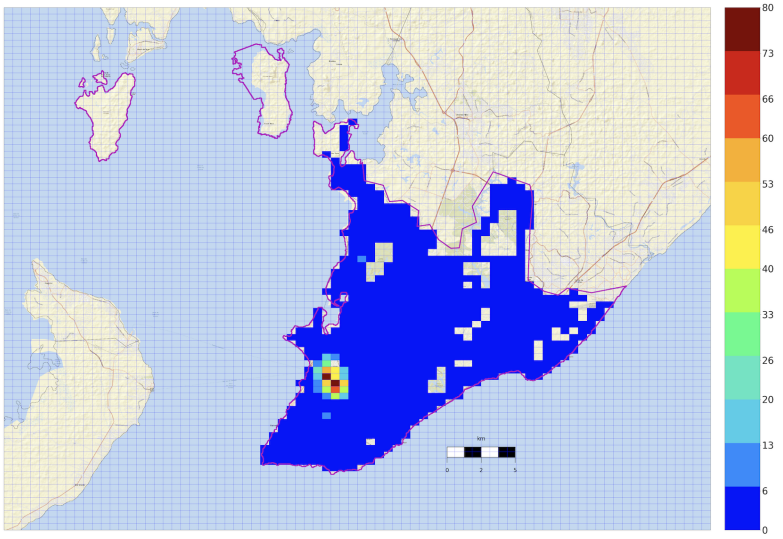
(b) *SpStreetSat19*, with *SpSat19* aerial features.

Figure 34 – Cell region estimations with São Paulo models applied in the whole city of Salvador. All Pearson correlation ρ with *p-value* < 0.001

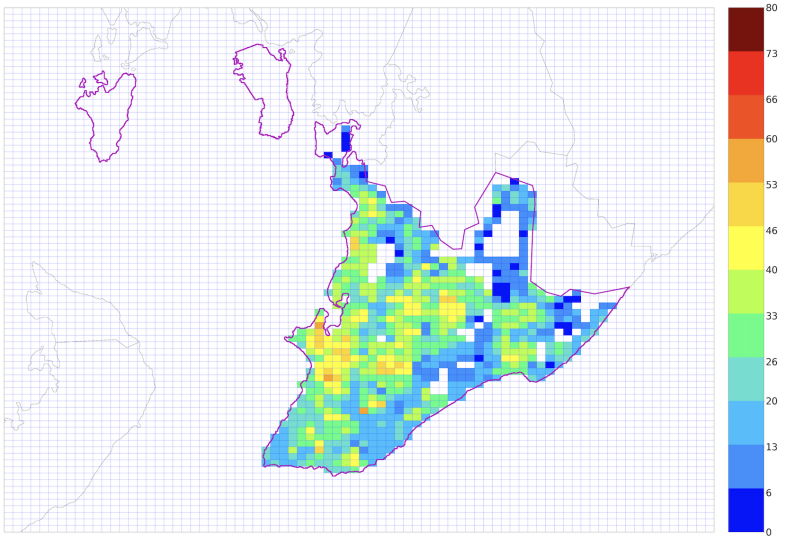
comparison.

Table 27 summarizes the results of applying São Paulo models in the test set from Rio de Janeiro, for $24,320$ location points, and Table 28 shows the results by cell region estimation, for 352 test cells over Rio de Janeiro territory. Column “Aerial Features” denote which network model produced the aerial features to be used in the evaluation, when applicable.

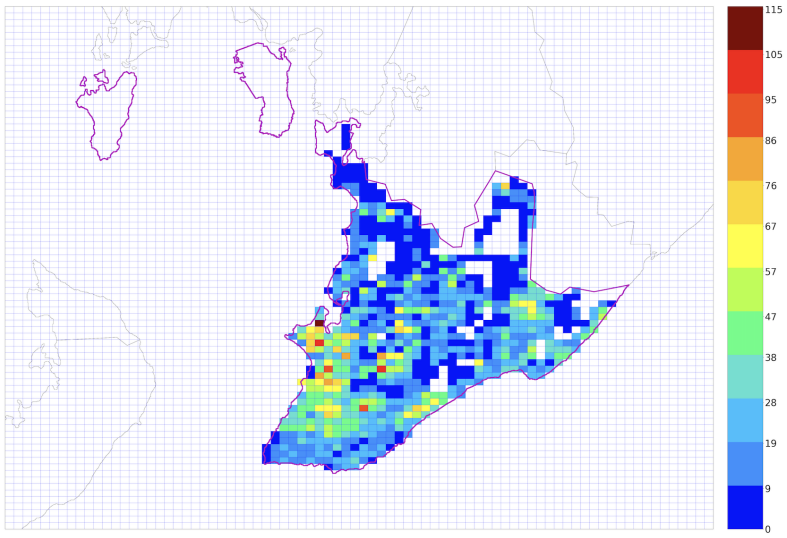
Figure 36 depicts the results obtained from *SpStreetSat19* with *RioSat19* aerial features, trained in São Paulo with highlighted results for overall location points and cell region estimations, when applied in the test set of Rio de Janeiro. Plots depict actual versus estimated correlation, with locations points that obtained the estimated



(a) Actual dengue rates

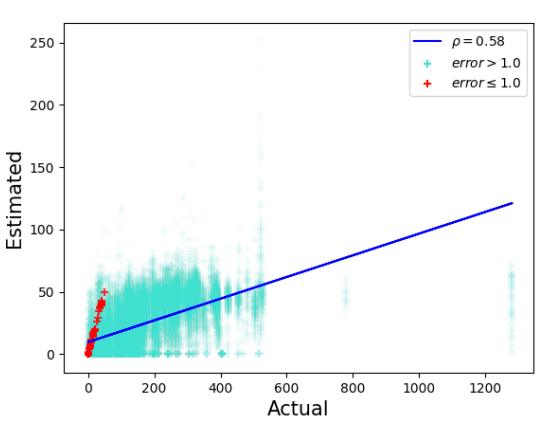


(b) *SpStreet*

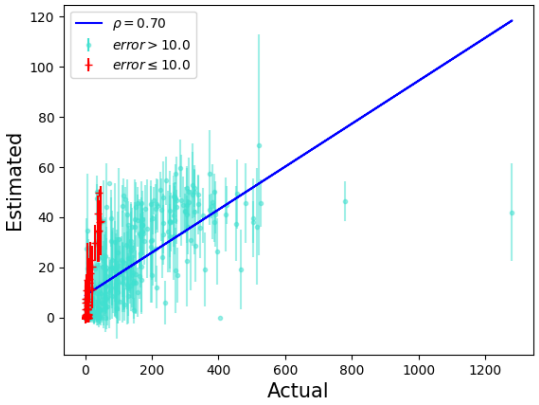


(c) *SpStreetSat19*, with *SpSat19* aerial features.

Figure 35 – Cell estimation maps with São Paulo models applied in whole city of Salvador. Original dengue incidence rates ranges varies from 0 to 80. *SpStreet* estimations are withing the original range, while estimations from *SpStreetSat19* extrapolate the range from 0 to 115.



(a) *SpStreet*



(b) *SpStreetSat19*, with *SpSat19* aerial features.

Figure 36 – Cell region estimations with São Paulo models applied in the test set of Rio de Janeiro. All Pearson correlation ρ with *p-value* < 0.001

absolute error of ≤ 1.0, for overall correlation, and absolute error of ≤ 10.0 for cell estimation selected in red.

5.2.3 Districts Risk Ranking Evaluation Results

The next Subsections present the resulting dengue risk ranking evaluation with Risk-biased Overlap (RBO) metric, for proposed models from Rio de Janeiro and São Paulo that presented better results in cell region estimation evaluation, from the application of models on same cities and the portability of the models to other cities. The next Subsections presents the risk ranking lists evaluations for each studied city.

Table 27 – Results from São Paulo models applied in Rio de Janeiro test set, for $24,320$ location points. All Pearson correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 1.0$	Pearson ρ
SpSat19	$zoom = 19$	39,581.79	198.95	143.33	192 (0.7%)	0.09
SpStreet	-	38,316.95	195.74	141.56	234 (0.9%)	0.45
SpStreetSat19	**RioSat19**	**36,894.51**	**192.08**	**141.55**	**1,133 (4.6%)**	**0.58**
SpStreetSat19	SpSat19	39,559.62	198.90	143.36	860 (3.5%)	0.10

Table 28 – Results from São Paulo models applied in Rio de Janeiro, for 352 test cells region estimations. All Pearson correlation with *p-value* < 0.001.

Model	Aerial Features	MSE	$RMSE$	MAE	$MAE \leq 10.0$	Pearson ρ
SpSat19	$zoom = 19$	29,786.61	172.59	104.43	85 (24%)	0.34
SpStreet	-	28,875.57	169.92	102.10	97 (27%)	0.55
SpStreetSat19	**RioSat19**	**27,964.97**	**167.23**	**101.87**	**102 (28%)**	**0.70**
SpStreetSat19	SpSat19	29,698.15	172.33	104.15	86 (24%)	0.31

5.2.3.1 Dengue Risk Ranking Estimations in Rio de Janeiro

Table 29 presents the output list of higher dengue estimations neighborhoods in Rio de Janeiro, from the models that presented the best results for this city. They are compared to their actual value risk neighborhoods rank, from higher incidence to lower incidence, including the resulting dengue risk neighborhood list from the application of São Paulo *StreetSat* model in Rio de Janeiro using aerial features from *RioSat19*. Table 30 enumerates the RBO scores between the actual value ranked lists and each estimated list, considering the weights of the positions of neighborhoods in the list.

5.2.3.2 Dengue Risk Ranking Estimations in São Paulo

Table 31 presents the output list of higher dengue estimations districts in São Paulo, from models that presented the best results for this city compared to their true value dengue risk district rank, from higher incidence to lower incidence. Table 32 enumerates the RBO scores between the two ranked lists. Next, the estimated risks ranked lists from $281,184$ city points of whole São Paulo, obtained with *StreetSat* models trained in Rio de Janeiro, are presented in Table 33 for comparison. The aerial features from *SpSat19* were also tested using *RioStreetSat19*, and evaluation was performed only in the test set. Table 34 enumerated the resulting RBO metric from the comparison of the actual ranked list from all São Paulo and estimated lists for the city by the proposed models.

Table 29 – Dengue risk neighborhood ranking for Rio de Janeiro city, applied in the test set of same city.

Risk Ranking	True Values	*RioStreetSat21*	*RioStreetSat19*	*SpStreetSat (RioSat19)*
1	Campo Grande	Campo Grande	Campo Grande	Campo Grande
2	Bangu	Santa Cruz	Santa Cruz	Santa Cruz
3	Copacabana	Bangu	Realengo	Bangu
4	Realengo	Realengo	Bangu	Olaria
5	Irajá	Guaratiba	Copacabana	Realengo
6	Rocinha	Padre Miguel	Irajá	Brás de Pina
7	Padre Miguel	Irajá	Olaria	Irajá
8	Santa Cruz	Cosmos	Padre Miguel	Guaratiba
9	Guaratiba	Copacabana	Brás de Pina	Cavalcanti
10	Cosmos	Olaria	Tijuca	Tijuca

Table 30 – Rank-biased Overlap (RBO) applied in true values dengue risk neighborhoods list and estimated dengue risk neighborhoods from Rio de Janeiro and São Paulo models, applied in Rio de Janeiro test set.

Model	Features	Agreement	RBO	Residual
RioStreetSat21	**RioSat21**	**0.90**	**0.74**	$2.68e^{-9}$
RioStreetSat19	RioSat19	0.70	0.72	$2.68e^{-9}$
SpStreetSat19	RioSat19	0.60	0.67	$2.68e^{-9}$

5.2.3.3 Dengue Risk Ranking Estimations in Salvador

Table 35 presents the estimated risk ranked lists from Salvador districts using Rio de Janeiro models that returned the best results for Salvador cell region estimation. Next, Table 36 shows the estimated dengue risk list from São Paulo models applied to Salvador. Risk rankings returned from *MultiCity-StreetSat19* model evaluated with Salvador using aerial features are enumerated in 37. Table 38 enumerates RBO metric results for actual dengue risk ranked list and each estimated list.

Table 31 – Dengue risk district ranking for São Paulo city, applied in the test set of same city.

Risk Ranking	True Values	*SpStreetSat19*
1	BRASILANDIA	BRASILANDIA
2	RAPOSO TAVARES	ITAQUERA
3	ITAQUERA	ARTUR ALVIM
4	SE	FREGUESIA DO O
5	ITAIM PAULISTA	JABAQUARA
6	FREGUESIA DO O	CACHOEIRINHA
7	ARTUR ALVIM	JOSE BONIFACIO
8	ARICANDUVA	GRAJAU
9	JABAQUARA	PEDREIRA
10	PENHA	CAMPO LIMPO

Table 32 – Rank-biased Overalap (RBO) applied in true values dengue risk neighborhoods list and estimated for São Paulo models, in São Paulo.

Model	Features	Agreement	RBO	Residual
SpStreetSat19	**SpSat19**	**0.50**	**0.62**	$3.99e^{-6}$

Table 33 – Dengue risk district ranking for whole São Paulo city, estimated from Rio de Janeiro models using aerial features from *RioSat21* and *RioSat19*. For *SpSat19* features, only test set is evaluated.

Risk Ranking	True Values	*RioSat19*	*RioStreet*	*RioStreetSat19*
1	BRASILANDIA	GRAJAU	GRAJAU	GRAJAU
2	ITAQUERA	PARELHEIROS	JARAGUA	JARAGUA
3	RAPOSO TAVARES	JARAGUA	JD ANGELA	SAPOPEMBA
4	RIO PEQUENO	TREMEMBE	JD SAO LUIS	PIRITUBA
5	CID ADEMAR	JD ANGELA	CID DUTRA	BRASILANDIA
6	CANGAIBA	CID DUTRA	TREMEMBE	JD ANGELA
7	CID LIDER	SAPOPEMBA	SAPOPEMBA	CID ADEMAR
8	SACOMA	PIRITUBA	PIRITUBA	CID DUTRA
9	JABAQUARA	ITAQUERA	ITAQUERA	TREMEMBE
10	FREGUESIA DO O	JD SAO LUIS	JABAQUARA	JABAQUARA

Table 34 – Rank-biased Overalap (RBO) applied in true values dengue risk neighborhoods list and estimated for whole São Paulo city, with Rio de Janeiro models.

Model	Features	Agreement	RBO	Residual
RioSat19	$zoom = 19$	0.10	0.14	$3.99e^{-6}$
RioStreet	-	0.20	0.17	$3.99e^{-6}$
RioStreetSat19	**RioSat19**	**0.30**	**0.24**	$3.99e^{-6}$

Table 35 – Dengue risk district ranking for whole Salvador city, estimated from Rio de Janeiro models.

Risk Ranking	True Values	*RioSat21*	*RioStreetSat19 (SpSat19)*
1	SANTO ANTÔNIO	PIRAJÁ	PIRAJÁ
2	SÃO CAETANO	SÃO CAETANO	ITAPOÃ
3	PIRAJÁ	ITAPOÃ	SÃO CAETANO
4	BROTAS	VALÉRIA	AMARALINA
5	S. TOMÉ/PARIPE	SÃO CRISTOVÃO	VALÉRIA
6	ITAPOÃ	AMARALINA	BROTAS
7	PENHA	S. TOMÉ/PARIPE	SÃO CRISTOVÃO
8	PERIPERI	BROTAS	SANTO ANTÔNIO
9	AMARALINA	VITÓRIA	VITÓRIA
10	VITÓRIA	SANTO ANTÔNIO	PENHA

Table 36 – Dengue risk district ranking for whole Salvador city, estimated from São Paulo models, with model *SpStreetSat19* using aerial features from *SpSat19*.

Risk Ranking	True Values	SpSat19	SpStreet	SpStreetSat19
1	SANTO ANTÔNIO	ITAPOÃ	PIRAJÁ	ITAPOÃ
2	SÃO CAETANO	SÃO CAETANO	VALÉRIA	SÃO CAETANO
3	PIRAJÁ	PIRAJÁ	SÃO CAETANO	PIRAJÁ
4	BROTAS	SÃO CRISTOVÃO	ITAPOÃ	SÃO CRISTOVÃO
5	S. TOMÉ/PARIPE	AMARALINA	SÃO CRISTOVÃO	BROTAS
6	ITAPOÃ	BROTAS	AMARALINA	VALÉRIA
7	PENHA	VALÉRIA	BROTAS	AMARALINA
8	PERIPERI	SANTO ANTÔNIO	S. TOMÉ/PARIPE	SANTO ANTÔNIO
9	AMARALINA	PENHA	SANTO ANTÔNIO	PENHA
10	VITÓRIA	VITÓRIA	VITÓRIA	VITÓRIA

Table 37 – Dengue risk district ranking for whole Salvador city, estimated from *MultiCity-StreetSat19* model using aerial features from *RioSat19* and *SpSat19*.

Risk Ranking	True Values	MultiCity (RioSat19)	MultiCity (SpSat19)
1	SANTO ANTÔNIO	ITAPOÃ	ITAPOÃ
2	SÃO CAETANO	PIRAJÁ	PIRAJÁ
3	PIRAJÁ	VALÉRIA	SÃO CAETANO
4	BROTAS	SÃO CAETANO	SÃO CRISTOVÃO
5	S. TOMÉ/PARIPE	SÃO CRISTOVÃO	VALÉRIA
6	ITAPOÃ	SANTO ANTÔNIO	BROTAS
7	PENHA	BROTAS	AMARALINA
8	PERIPERI	S. TOMÉ/PARIPE	SANTO ANTÔNIO
9	AMARALINA	PENHA	PENHA
10	VITÓRIA	AMARALINA	VITÓRIA

Table 38 – Rank-biased Overalap (RBO) between Salvador districts lists with Rio de Janeiro and São Paulo models.

Model	Features	Agreement	RBO	Residual
RioSat21	$zoom = 21$	0.80	0.64	0.02
RioStreetSat19	SpSat19	0.80	0.60	0.02
MultiCity-StreetSat19	RioSat19	0.80	0.59	0.02
SpSat19	$zoom = 19$	0.80	0.65	0.02
SpStreet	-	0.80	0.60	0.02
SpStreetSat19	**SpSat19**	**0.80**	**0.66**	0.02
MultiCity-StreetSat19	SpSat19	0.80	0.61	0.02

6 ANALYSIS AND DISCUSSION

This Chapter presents an analysis of the results obtained from the evaluation of the proposed models for dengue incidence rates estimation using urban images and computer vision models for Rio de Janeiro, São Paulo, and Salvador, for both modalities in the city itself or with the portability of the models. Section 6.1 discusses the results obtained between the models, reviewing the enrolled hypotheses presented in Chapter 3 and arguing the evidence by which the models presented better results among all the proposed ones. Section 6.2 presents a comparison of the performance of the proposed models and the results reported in of dengue prediction models related works, and for inferring latent variables using urban images within the area of visual computing sociology. The last Section 6.3 discuss the possibilities of using the proposed models within the context of public policies and methods to prevent and combat dengue fever epidemics in Brazilian urban centers.

6.1 Comparison of Proposed Models

The proposed models for Rio de Janeiro city, trained and evaluated with dengue autochthonous cases from 2010 to 2014 geographically distributed and aggregated by cell with a 80×80 grid data structure, using as input street-level and urban aerial images, presented as results Pearson Correlation varying from $\rho = 0.19$ to $\rho = 0.65$, including baseline and Deep ConvNet models, for overall location points evaluation and Pearson Correlation varying from $\rho = 0.29$ to $\rho = 0.78$ for cell region estimation evaluation methodology.

6.1.1 Baseline Models Comparison

In baseline aerial feature models, with results presented in Table 17, *Aerial HOG-LR*, *Aerial GIST-LR* and *Aerial HOG-NN* presented low Pearson Correlations $\rho = 0.19$, $\rho = 36$ and $\rho = 26$ for overall location points evaluation, and $\rho = 0.29$, $\rho = 0.46$ and $\rho = 0.38$ for cell region estimation. This results indicates low correlations between actual and estimated values using HOG designed descriptors, and with GIST, a near

noticeable correlation, for both evaluation methods.

The use of *Neural Network* slightly improves Pearson ρ when compared with a linear model in regressor block using HOG descriptors, although the number of evaluated location points that the model *Aerial HOG-LR* achieved $MAE \leq 1.0$ is 14 points ahead than the *Aerial HOG-NN*. The model *Aerial GIST-LR* with GIST designed descriptor presented better results than models with HOG descriptors for aerial approaches. The GISTs are global descriptors, i.e., they are computed over the entire input image and were designed for scene recognition problems, and their feature has fewer hyper-parameters to adjust when compared with HOG descriptors. On the other hand, HOG designed descriptors were designed for object recognition, specifically human detection, and can be classified as a region or local descriptor. Compared to GIST, HOG descriptors have more hyper-parameters to adjust and bigger feature size when used for a whole scene. The fact that HOGs have a higher number of hyper-parameters to adjust, and the different purposes of use may have influenced the obtained results.

Baseline street-level feature models *Street HOG-LR* and *Street HOG-NN* presented Pearson correlation $\rho = 0.44$ and $\rho = 0.46$ for overall location points evaluation and $\rho = 0.57$ and $\rho = 0.59$ for cell region estimation evaluation. The *Neural Network* multiple regressor approach presented better results when compared to linear regression, for both evaluation methods, including the number of location points evaluation that achieved $MAE \leq 1.0 = 142$, with a difference of 22 points ahead of linear model. The different methods of combining street-level descriptors resulted from different camera views may have influenced the obtained results and should be further investigated.

Between baseline proposed models, the use of street-level features as input entails better results when compared to models with aerial features. This indicates that street-level images, when combined with different camera views, leads to better results than single aerial imagery models. On the other hand, the results obtained from all baseline models, and the hot-spot maps produced by the baseline models with cell region estimations, depicted in Appendix C, Figures 58 and Figure 61, indicates that designed descriptors, specially HOG descriptors, are insufficient to estimate dengue incidence rates. Also, designed descriptors were unable to estimate a broader range of estimated values, unable to infer values close to 0 or ahead of 270 dengue occurrences.

6.1.2 Deep Convolutional Neural Network Models Comparison

Deep ConvNets for aerial images proposed models achieved Perason Correlation $\rho = 64$ for *RioSat21* and *RioSat20*, and $\rho = 0.61$ for *RioSat19* in overall location points evaluation (Table 17), and $\rho = 0.76$ for *RioSat21* and *RioSat20*, and $\rho = 0.72$ for for *RioSat19* in cell region estimations (Table 19). In *RioSat20* and *RioSat19*, location points evaluation with $MAE \leq 1.0$ are 33 points ahead of *RioSat21*. Between exclusively aerial models for Rio de Janeiro, *RioSat21* achieved the lowest absolute errors

$MAE = 71.58$, $RMSE = 110.47$ for overall location points evaluation, and *RioStreet20* achieved the lowest absolute errors $MAE = 56.3$, $RMSE = 95.83$ for cell region estimations. Overall, the *RioSat* with zoom levels $zoom = \{21, 20, 19\}$ models have roughly similar performance when applied to Rio de Janeiro, with $zoom = 20$ in *RioSat20* being able to estimate values with a wider range of dengue occurrences values, between all proposed models, as depicted in Appendix D, Figure 66c.

For São Paulo (Table 18), aerial *SpSat19* model with zoom $zoom = 19$ presented lower Pearson Correlation $\rho = 0.51$ when compared with Rio de Janeiro *Sat* models, albeit the number of location points evaluated with $MAE \leq 1.0$ are significant more, totaling $1,726$ examples. The overall absolute errors $MAE = 22.70$, $RMSE = 47.55$ are lower than Rio de Janeiro *Sat* models, which can be explained by the fact that dengue distribution values are lower than in Rio de Janeiro.

From all Rio de Janeiro Deep ConvNet models, *RioStreet* presented the lower Pearson $\rho = 0.58$ for location points estimations and $\rho = 0.70$ for cell region estimation when compared to aerial imagery models. When comparing the total location points that achieved $MAE \leq 1.0$, *RioStreet* presented 258 points with 6 points ahead of *RioSat21*, but 27 points below *RioStreet20* and *RioStreet19*. Since the street-level baseline models presented better results compared to aerial baseline models, there are evidences that, in the case of Deep ConvNet models, the feature concatenation approach for street-level images chosen for *Street* architecture may not be the best choice for this task, and should be further investigated.

The fusion *StreetSat* models that combine street-level and aerial features as inputs, for Rio de Janeiro, achieved Pearson Correlation from $\rho = 0.59$ to $\rho = 0.65$ for overall location points estimation, and $\rho = 0.75$ to $\rho = 0.78$ in cell region estimations. The lowest correlation $\rho = 59$, for overall location points, and $\rho = 0.75$ for cell region estimationis, is resulted from *MultiCity-StreetSat19* model, that uses features extracted from *RioSat19* model. The best *StreetSat* model for Rio de Janeiro, and between all proposed models, is *RioStreetSat21*, when considering absolute errors and Pearson Correlation as metrics, with $MAE = 69.16$ for overall location points estimation, $MAE = 51.21$ for cell region estimation and Pearson Correlation $\rho = 0.65, \rho = 0.78$ as metrics. Model *RioStreetSat19* obtained the highest number of location points and cell regions that achieved $MAE \leq 1.0 = 1,252$ and $MAE \leq 10.0 = 112$ respectively.

The use of aerial image features together with street-level features increased the correlation between the actual and estimated dengue occurrences values, as presented in Table 17 and 19, for both overall and cell region estimation. While mantaining correlation, *StreetSat* models were able to estimate lower values than their original *Sat* models, although the maximum dengue occurrence estimated values slightly reduced, for cell region estimations, presented in Figure 26.

With São Paulo, the *SpStreet* model reflects the same behaviour of *RioStreet* when

evaluated with Rio de Janeiro, whose Pearson Correlation $\rho = 0.32$ for overall location points evaluation and $\rho = 0.40$ for cell region estimation, are lower than *SpSat19* model for the same city, contradicting the results obtained in the experiments with baseline models. Again, when combining the two features - aerial and street-level - as inputs in *SpStreetSat19* model, the Pearson Correlation for overall location points evaluation and cell region estimation presents a small increase with $\rho = 0.52$ for location points evaluation and $\rho = 0.61$ for cell region evaluation. A somewhat significant rise in location points estimations is perceived, with $MAE \leq 1.0$, totaling $2,999(9\%)$ of all test examples, and $149(59\%)$ with $MAE \leq 10.0$ for cell estimation of all test cells in São Paulo.

When combining both imagery datasets labeled with dengue incidence rates, from Rio de Janeiro and São Paulo, obtaining the model *MultiCity-StreetSat19*, the performance of this model, in both cities, is somewhat smaller than their *StreetSat19* counterparts. For Rio de Janeiro, the multiple city model presents Pearson Correlation $\rho = 0.59$ for overall location points evaluation and $\rho = 0.54$ for cell region estimation, lower than *Sat* models, but higher than *Street* models for this city. In São Paulo, the model results in Pearson $\rho = 0.48$ and $\rho = 0.54$ for location points and cell region evaluations, respectively. In general, the *MultiCity-StreetSat19* does not present any performance addition to the proposed *Sat* and *StreetSat* models, only increases Pearson ρ when compared with *Street* models.

When applying the *Rank-Biased Overlap* (RBO) metric in best models of Rio de Janeiro *RioStreetSat21* and *RioStreetSat19* (Table 30), using test set cell estimations, the resulted agreement, i.e., the size of the intersection between the two ranked lists, between actual and estimated dengue risk ranking neighborhoods is $agreement = 0.90$ and $agreement = 0.70$ respectively, with 0 equal no intersection elements and 1 equal total intersection. The RBO weighs the position of each estimated neighborhood compared with the actual rank list, and returned $RBO = 0.74$, $residual = 2.68^{-9}$ and $RBO = 0.72$, $residual = 2.68e^{-9}$ for each model, meaning that the actual and estimated ranks for each model are similar in 73% and 72%, considering the relevance of position in which they were placed.

For São Paulo (Table 32), the best model *SpStreetSat19* presented an agreement of $agreement = 0.50$, meaning that there is an intersection in 50% of the actual and estimated districts, and $RBO = 0.62, residual = 3.99e^{-6}$, which indicates that the similarity between the two ranked list districts, considering the relevance of their positions, is 62%.

6.1.3 Portability of Models

When analyzing the results from Rio de Janeiro models applied in the whole city of São Paulo, presented in Tables 21 and 22, the model that presented better generaliza-

tion for this city was *RioStreet*, with Pearson Correlation $\rho = 0.24$ for overall location points estimation and Pearson $\rho = 0.38$ for cell region estimations. Moreover, the models *RioSat19* and *RioStreetSat19* with *RioSat19* aerial features resulted in $1,010(0.3\%)$ of location points estimations with $MAE \leq 1.0$ and $9,508(3\%)$ location points estimations with $MAE \leq 1.0$. From *hot-spots* maps produced with those models in São Paulo, depicted in Figure 30, it is possible to visualize that the all models extend the heat zones to adjacent regions, bearing little resemblance to the original areas, being *RioStreetSat19* with *RioSat19* the model closest to the original *hot-spots* map.

In Salvador (Table 23, 24), when considering the Pearson Correlation ρ, the Rio de Janeiro model that generalizes better is *RioSat21* with Pearson $\rho = 0.20$ for overall location points estimation and Pearson $\rho = 0.33$ for cell region estimations. From Figures 31 and 32 it is possible to visualize that estimations from Rio de Janeiro models in Salvador extrapolate the actual values to higher rates, including high values for the standard deviations of the estimated average incidence per cell region. The worst model performance is *RioStreetSat21* using *RioSat21* features, with Pearson $\rho = 0.05$, close to nonexistent correlation between actual and estimated results.

Models from São Paulo, when applied in the whole city of Salvador, enumerated in Tables 25 and 26, presented somewhat better results when compared with their counterparts in Rio de Janeiro models. For instance, *SpStreet* performed better than *RioStreet*, with Pearson $\rho = 0.22$ for *SpStreet*, against Pearson $\rho = 0.18$ for overall location points evaluation and $\rho = 0.28$ against $\rho = 0.24$. *SpSat19* performs slightly better than *RioSat19*, with Pearson $\rho = 0.20$ against Pearson $\rho = 0.19$. *StreetSat* models from São Paulo, including *MultiCity-StreetSat19* performed better than *StreetSat* Rio de Janeiro models when using aerial features from *SpSat19*, with Pearson $\rho = 0.21$ against $\rho = 0.17$ for overall location points evaluation and Pearson $\rho = 0.33$ against $\rho = 0.22$ for cell region estimation.

From Figures 33 and 34 it is possible to observe that the maximum overall estimations from São Paulo models applied in Salvador are lower than ranges estimated from Rio de Janeiro, although the standard deviations from average incidence estimations per cell region are similar to Rio de Janeiro models. From the *hot-spots* presented in Figure 35, *SpStreet* model's estimation maintain the actual dengue range from Salvador, while *SpStreetSat19*, with aerial features from *SpSat19*, extrapolates to a higher range. The produced *hot-spot* maps preserve somewhat the original heat area, although it has estimated other adjacent areas with higher dengue rates.

Overall, São Paulo models performed slightly better than Rio de Janeiro models when applied in Salvador, including the number of evaluations with $MAE \leq 1.0$ for location points and $MAE \leq 10.0$ for cell region estimations. This could be attributed to the transfer learning technique used, with São Paulo models initialized with Rio de Janeiro pretrained information. The fact that the model contains adjusted weights

from more than one city could generalize better for unseen city examples. When São Paulo models are applied back to the original city, Rio de Janeiro, the results maintain certain significative correlation, as seen in *SpStreet* when applied in Rio de Janeiro, with Pearson $\rho = 0.45$ and *SpStretSat19* with *RioSat19* aerial features, with Pearson $\rho = 0.58$ for overall location points correlation and $\rho = 0.55$ and $\rho = 0.70$ for cell region estimation.

The application of the RBO similarity metric between actual dengue risk ranking and estimated dengue risk ranking between districts for the best resulting models in Salvador (Table 38) resulted in an intersection agreement of $agreement = 0.80$ for all models. The best RBO similarity risk rank equal $RBO = 0.66$, $residual = 0.02$, meaning that the similarity between the two actual and estimated ranked lists are 66% similar, considering the position in which the districts appear and their original position.

In general, models trained in only one city generalize poorly to unseen cities. There are indications, considering all the scenarios of the presented results, that São Paulo models that use transfer learning with initialization from pretrained Rio de Janeiro weights generalize better to Salvador than Rio de Janeiro models.

6.1.4 Best Estimations and Attributes Discovering

The best estimations made by the proposed models with absolute error values ≤ 1.0, in each studied city, were grouped to investigate the possible visual attributes, i.e., urban elements, that contributed to the results. Figure 37 depicts some aerial and street-level images examples from estimations with $MAE \leq 1.0$ for low dengue incidence rate values, evaluated individually during test stage for the city of Rio de Janeiro, with the best-resulting model *RioStreetSat21*. Next, Figure 38 depicts examples of estimations with $MAE \leq 1.0$ with high dengue incidence rate values, among the best predictions from the model.

Figure 37 allows the visualization of typical scenarios with low dengue incidence rates, that can be predicted by the proposed models, in the city of Rio de Janeiro. From the depicted samples, it is possible to infer that typically green areas, with little presence of human urban elements, with possible low demographic concentration, are indicatives of low dengue incidence rates. In Figure 38, the presence of human urban elements is somewhat higher than in low estimates, with the presence of apartment buildings.

The same investigation was made in São Paulo and Salvador city. In São Paulo, results were obtained with the best-resulting model *SpStreetSat19*, and in Salvador with the test results from *SpStreetSat19* using aerial features from *SpSat19*. Figures 39 and 41 depicts the best predictions, with $MAE \leq 1.0$ among the lowest dengue incidence rates for that city. Figure 40 and 42 shows the best estimations performed with the proposed model for higher dengue incidence rates.

Figure 37 – Sample of best estimations among the lowest dengue incidence rates in Rio de Janeiro.

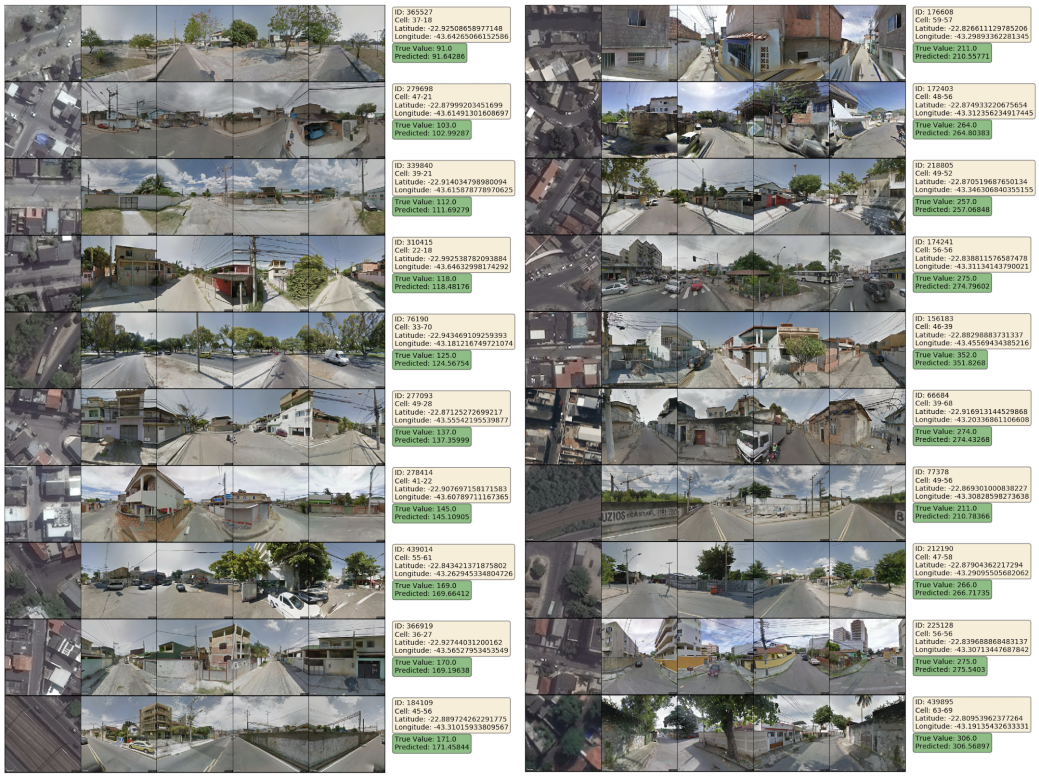


Figure 38 – Sample of best estimations among the highest dengue incidence rates in Rio de Janeiro.

Figure 39 – Sample of best estimations among the lowest dengue incidence rates in São Paulo.

Figure 40 – Sample of best estimations among the highest dengue incidence rates in São Paulo.

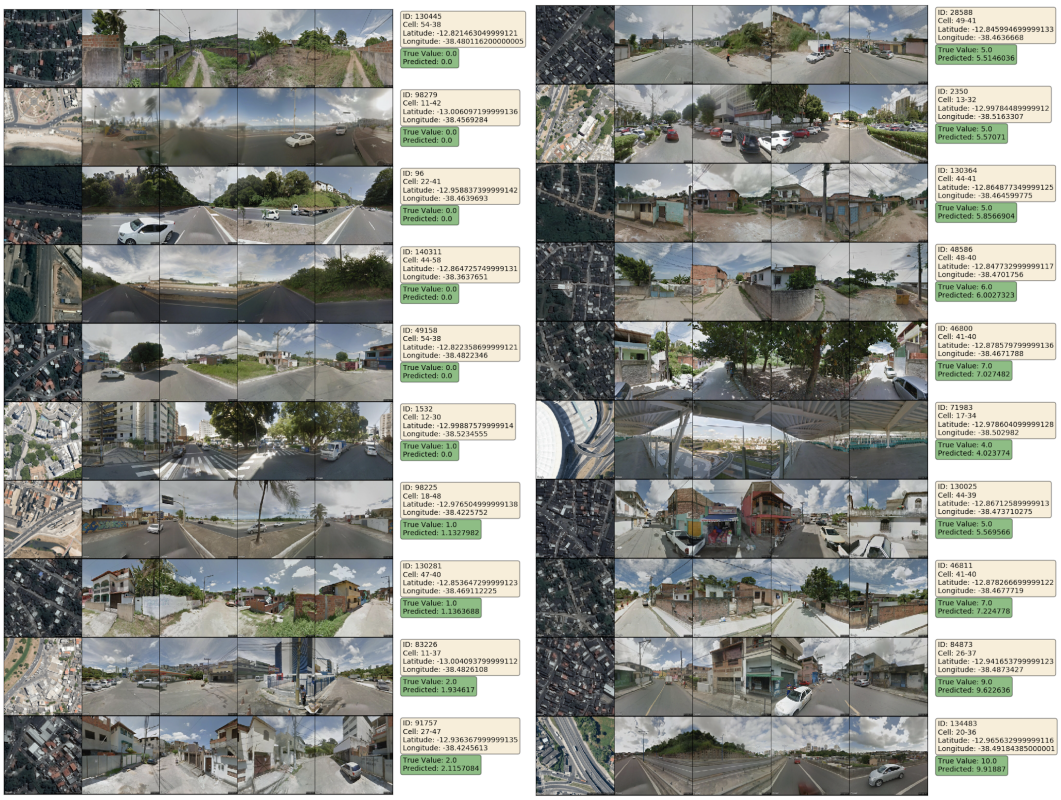


Figure 41 – Sample of best estimations among the lowest dengue incidence rates in Salvador.

Figure 42 – Sample of best estimations among the highest dengue incidence rates in Salvador.

São Paulo presents more urban elements than Rio de Janeiro in lower dengue incidence rates examples. In this city, “trees and roads” combination appears at street-level images among correct lower estimations, even when the aerial images indicate the presence of near habitations, indicating human activity. The number of dengue incidence rates gradually increases as the number of habitation grows near the analyzed location point.

In Salvador, the green nature elements, together with roads and a somewhat quantity of habitations near the location point, are the elements present in the sample images, from the best predictions among lower dengue incidence rates. As São Paulo, the highest correct estimated values present a significant concentration of houses and apartment buildings near the tested location point surroundings.

In order to quantify these assumptions, the classification labels from *Places* dataset from the best results of low and high dengue incidence rates were aggregated by labels that appear with a higher frequency in low or high dengue incidence rates estimates. This is a preliminary approach that explores the fact that the street-level features used to train the proposed models were obtained with *DenseNet161-Places* network, and their classification by this network might help understand the attributes that may contribute to the results.

For the city of Rio de Janeiro, Figure 43a presents *Places* labels that occoured with more frequency in correct low dengue estimations. Next, Figure 43b shows the *Places* labels that where more present in the classification of street-level images from high incidence rates correct estimations. Labels such as “*desert-road*”, “*field-road*” and “*forest-road*” appears among the first 10 labels for low dengue rates, while “*slum*”, “*loading-dock*”, “*promenade*”, “*street*”, “*beach-house*” and “*alley*” appears among the top labels for high dengue rates. Some labels appears in both distributions, e.g., “*residential-neighborhood*”, “*highway*”, “*industrial-area*”, with high occurrence.

For São Paulo, Figures 45a and 45b depicts the *Places* labels that appeared with more frequency in low and high dengue estimations, respectively. As presented in Rio de Janeiro, common labels such as “*residential-neighborhood*” and “*industrial-area*” appears in both high and low distributions. Also similar to Rio de Janeiro, “*field-road*”, “*desert-road*” and “*forest-road*” labels appears between the most frequent labels in low dengue incidence. The label “*slum*” appears in the highest rated labels for high dengue incidence, but also between the 10 most frequent labels in low dengue rates. The labels “*loading-dock*”, “*alley*” and “*street*” appears appears between the most frequent labels in high incidence rates.

In Salvador, low and high dengue incidence rate localities presented the class labels “*residential-neighborhood*”, “*industrial-area*”, “*slum*” and “*highway*”. Again, as in Rio de Janeiro and São Paulo, the labels “*desert-road*”, “*forest-road*” and “*field-road*” appears with high frequency among the top-10 *Places* classes in low dengue incidence rate

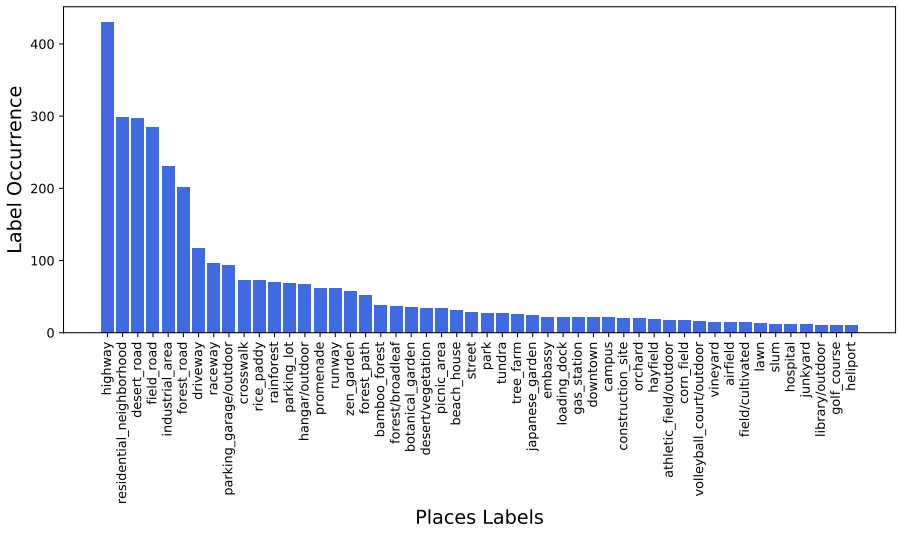
correct estimations. Other labels appear in both low and high estimations, without any conclusive significance for the different distributions.

6.1.5 Enrolled Hypotheses Discussion

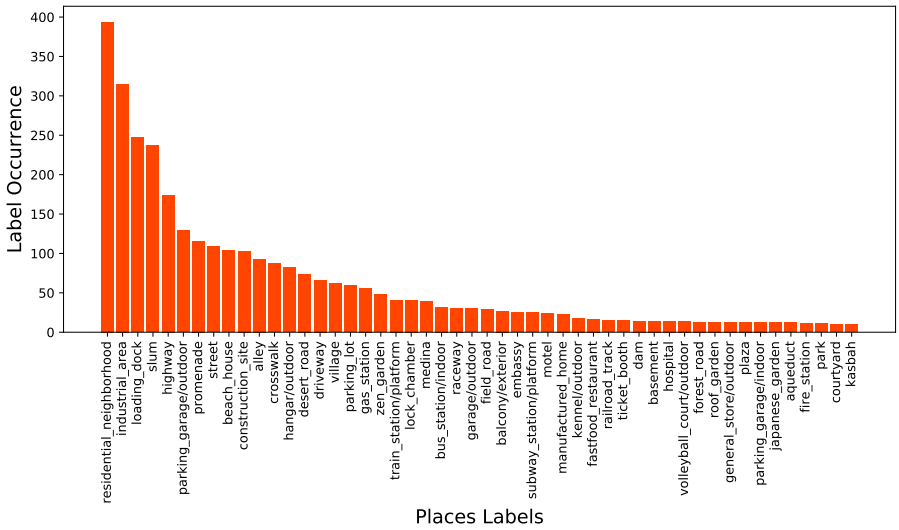
Following the specific enrolled hypotheses, presented in Chapter 3:

- **H. 1.** Using street-level urban imagery is as effective as using urban satellite imagery for estimating dengue-associated rates.

From baseline models obtained results, street-level images are more effective than aerial images, and, from Deep ConvNet models results, the use of street-level images



(a) Labels in low dengue incidence rates.



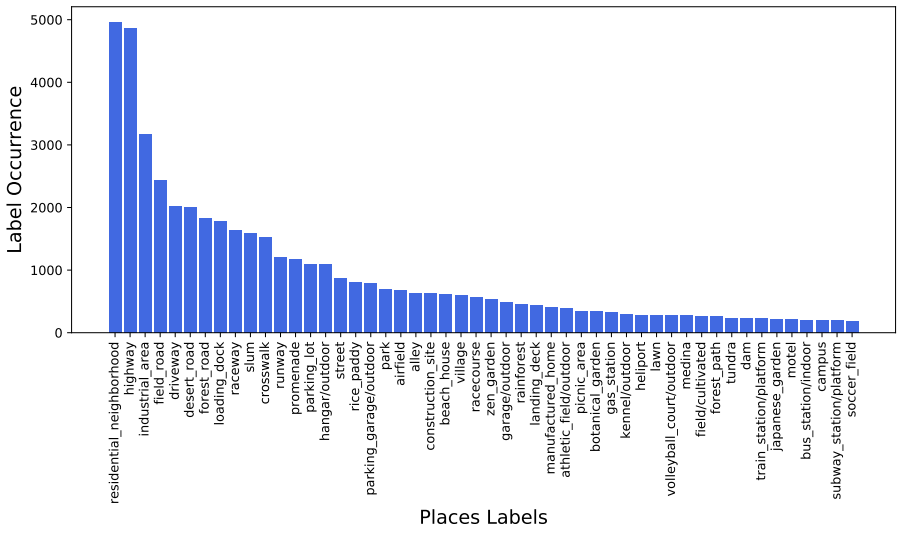
(b) Labels in high dengue incidence rates.

Figure 43 – Top-50 most frequent *Places* labels in (a) low and (b) high dengue incidence rates among correct estimations in Rio de Janeiro

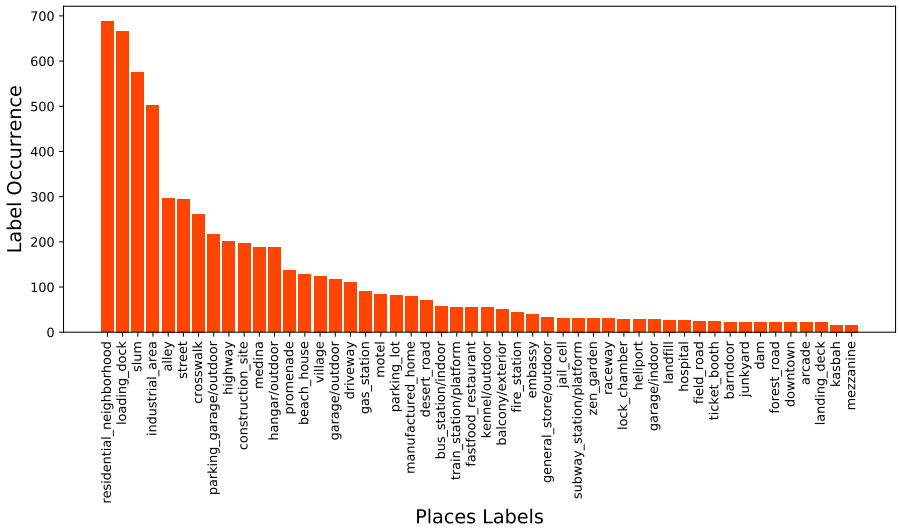
as inputs are useful as much as the use of aerial images, performing somewhat worse than models with aerial imagery only. This observed behavior from street-level models using deep features could be explained by choice of *Street* architecture topology, rather than the ability of urban images as predictors itself, and needs further investigation.

- **H. 2.** The combination of street-level and urban aerial imagery contributes to better results in estimating dengue-associated indices compared to the independent use of each image domain.

According to the Deep ConvNet models trained and evaluated in Rio de Janeiro and São Paulo, the use of both aerial and street-level features slightly increase the reported



(a) Labels in low dengue incidence rates.



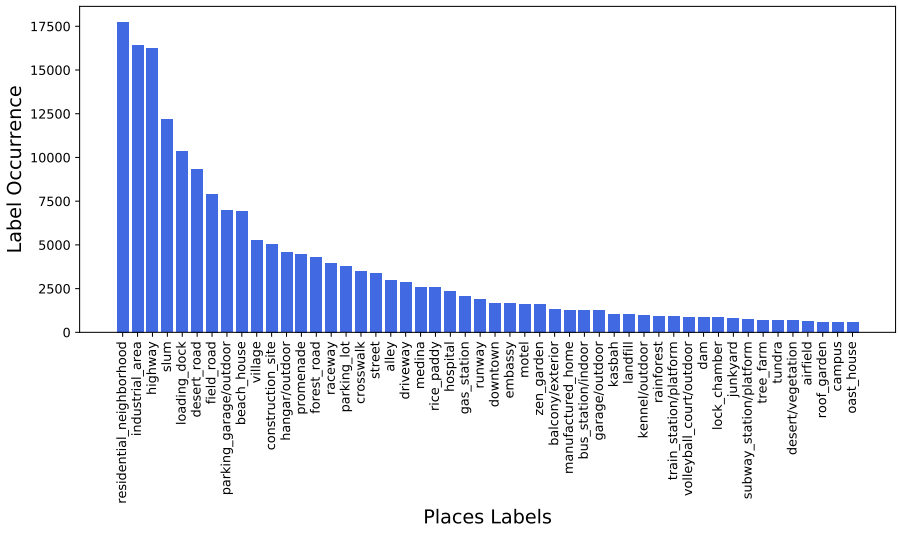
(b) Labels in high dengue incidence rates.

Figure 44 – Top-50 most frequent *Places* labels in (a) low and (b) high dengue incidence rates among correct estimations in São Paulo

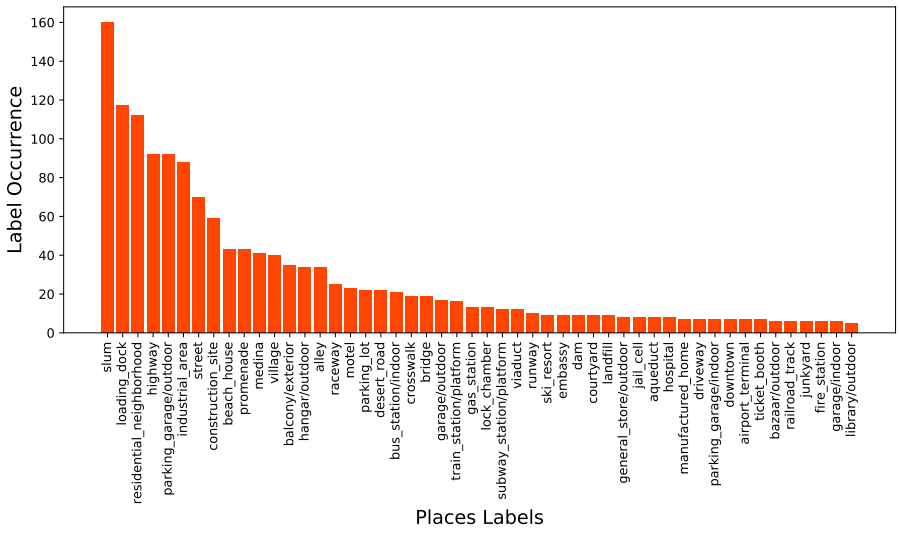
Pearson Correlation ρ, MAE and evaluations with $MAE \leq 1.0$ when compared with their counterparts using only aerial and street-level images.

- **H. 3.** Dengue estimation models trained in a single city are able to generalize estimation to other cities.

From the evaluations made with the proposed models, trained in a single city, and models using transfer learning and multiple cities image data, it was shown that models that have weights from more than one city generalize better to unseen cities. The Deep ConvNets were able to extract visual features, while training in one city, that could be transferred to a certain extent, for other unseen cities.



(a) Labels in low dengue incidence rates.



(b) Labels in high dengue incidence rates.

Figure 45 – Top-50 most frequent *Places* labels in (a) low and (b) high dengue incidence rates among correct estimations in Salvador

- **H. 4.** It is possible to extract, from the models that use street-level and aerial images data, informative attributes for dengue prevention not yet explored in urban areas that correlate with the occurrence of dengue.

Preliminary experiments on elements present in street-level images did not yield any significant results for the causality of correct low or high predictions. An investigation using methodologies for visualization and interpretation of deep learning models is needed, to understand which elements in urban images influence high or low dengue estimates. The interpretability of Deep Neural Networks models is difficult, mostly due to the high number of hyperparameters compared to simple linear approaches, giving these models a "black box" reputation. Methods for interpretability of deep learning models are still an open problem (MONTAVON; SAMEK; MÜLLER, 2018).

- **H. 5.** Deep Convolutional Neural Networks (Deep ConvNet) are suitable for use in the proposed models because they can learn and extract attributes from urban images, and estimate dengue-associated rates using these attributes.

According to the performed evaluations, Deep ConvNet models surpassed the results from baseline proposed models, indicating that deep features are more suitable for dengue estimation rates, considering the metrics of Pearson Correlation ρ, MAE and with evaluations with $MAE \leq 1.0$. This corroborates with related works results that showed the deep features advantages compared to designed descriptors in different tasks since the convolutional neural network layers can learn the best features for the specific problem.

6.2 Comparison with Related Works

After the evaluation of the models and the analysis of the results, *RioStreetSat21*, the model that presented the best Pearson Correlation ρ, applied in Rio de Janeiro for dengue incidence rates estimation, was compared with related works of the fields *Visual Computational Sociology* and dengue fever prediction models. From visual computational sociology models, only the best results from related regression works that reported Pearson ρ and ρ^2 (R^2) were compared with the present proposed model.

Table 39 presents the results from computer vision models in related works that infer socioeconomic variables and health-related indexes. Results are sorted by decreasing Pearson ρ. The proposed model *RioStreetSat21* performs somewhat worse than LIU et al. (2017); JEAN et al. (2016); NAIK (2017) proposed models. When compared with SUEL et al. (2019) works, the proposed *RioStreetSat21* model performs worse than the authors' model when compared to the Mean Income latent variable. However, the proposed model performs relatively equal to the authors' model when compared

Table 39 – Results from the best proposed model compared with *Visual Computational Sociology* related works reported Pearson ρ and R^2 values. All Pearson ρ with ***p-value*** ≤ 0.001

Author	Model	Goal	Attributes	Pearson ρ	R^2
(BENCY et al., 2017)	Inception V3 based	House Pricing	House Attributes, Points of interest, Aerial Images	0.95	0.91
(KANG; KANG, 2016)	Double-column CNN (AlexNet based)	Perceived Safety Score	Street-level images	0.90	0.81
(LIU et al., 2017)	HDMiR (AlexNet based)	Crime Rates	Street-level images	0.90	0.81
(JEAN et al., 2016)	AlexNet based	Local Per Capita	Aerial images	0.87	0.75
(SUEL et al., 2019)	VGG16 based	Mean Income	Street-level images	0.86	0.74
(NAIK, 2017)	StreetScore (AlexNet based)	Perceived Safety Score	Street-level images	0.85	0.72
(LAW; PAIGE; RUSSELL, 2018)	Linear Hedonic Model Deep ConvNet with 4 to 12 layers.	House Pricing	Hedonic Attributes, Aerial Images	0.84	0.70
(SUEL et al., 2019)	VGG16 based	Health deprivation and disability	Street level images	0.79	0.62
RioStreetSat21	**DenseNet161 based**	**Dengue Incidence Rates**	**Street-level and Aerial images**	**0.78**	**0.61**
(NAIK et al., 2014)	StreetScore (AlexNet based)	Perceived Safety Score	Bag of Features of Designed Descriptors	0.73	0.54
(ORDONEZ; BERG, 2014)	DeCaf	Human Judments Scores for safeness, uniqueness and whealth.	Street-level images	0.72	0.52
(SUEL et al., 2019)	VGG16 based	Self Reported Health	Street-level images	0.66	0.44
(LAW; PAIGE; RUSSELL, 2018)	Linear Hedonic Model Deep ConvNet with 4 to 12 layers.	House Pricing	Aerial images	0.40	0.16
(LAW; PAIGE; RUSSELL, 2018)	Linear Hedonic Model Deep ConvNet with 4 to 12 layers.	House Pricing	Street-level images	0.24	0.06

to health deprivation and disability, and better than SUEL et al. (2019) model, when compared to "self reported health" latent variable. Also, the proposed model *RioStreetSat21* performs worse than house pricing proposed that use house hedonic attributes together with street-level and aerial images, and better when the same models use only street-level or aerial images as inputs.

Table 40 enumerates the results from dengue incidence rates prediction and forecasting related works compared with the best result proposed model *RioStreetSat21*, with results are sorted by decreasing Pearson ρ. When comparing *RioSatStreet21* with dengue incidence rates week and year forecasting models, the proposed model *RioStreetSat21* performs worst than works from GUO et al. (2017); ASHBY et al. (2017); SCAVUZZO et al. (2017), that use satellite clime-related attributes. With models that use socioeconomic attributes, such as unemployment and household density, *RioStreetSat21* performs somewhat better than TEURLAI et al. (2015); ANGGRAENI et al. (2017); LAUREANO-ROSARIO et al. (2017) related works.

Table 40 – Results from the best proposed model compared with dengue incidence rates prediction and forecasting models from related works, with reported Pearson ρ and R^2 values. All Pearson ρ with *p-value* ≤ 0.001

Author	Model	Goal	Attributes	Pearson ρ	R^2
(GUO et al., 2017)	Support Vector Regression	Dengue Week Incidence Forecasting	Meteorological Data, Search Query Gov. Surveillance Data from Baidu	0.99	0.98
(ASHBY et al., 2017)	Poisson Model, Boosted Regression Trees	Dengue Annual Incidence Forecasting	Land Surface Temperature (LST), Surface Reflectance Daily, Enhanced Vegetation Index, Tropical Rainfall	0.99	0.97
(SCAVUZZO et al., 2017)	Neural Networks	Oviposition of Aedes Aegipty	Normalized Difference of Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Land Surface Temperature (LST)	0.98	0.97
RioStreetSat21	**DenseNet161 based**	**Dengue Incidence Rates**	**Street-level and Aerial images**	**0.78**	**0.61**
(TEURLAI et al., 2015)	Support Vector Regression	Dengue Incidence Rates	Activity Unenployment	0.76	0.58
(TEURLAI et al., 2015)	Support Vector Regression	Dengue Incidence Rates	Number of people per household.	0.74	0.55
(ANGGRAENI et al., 2017)	Linear Regression	Dengue Incidence Rates	Rainfall and TimeLag	0.66	0.44
(LAUREANO-ROSARIO et al., 2017)	Linear Regression	Dengue Incidence Rates	Precipitation, Minimum Air Temperature, Humidity, and SST (Sea Surface Temperature)	0.65	0.42
(TEURLAI et al., 2015)	Support Vector Regression	Dengue Incidence Rates	Mean Temperature	0.62	0.38
(ANGGRAENI et al., 2017)	Linear Regression	Dengue Incidence Rates	Temperature and TimeLag	0.54	0.29

6.3 Feasibility of the proposed model

The proposed models were designed to be used in dengue incidence estimation in specific locations, with location point predictions, and in micro-regions, using cell region estimations. The obtained results showed that, when training a model in a specific city, aerial images with $zoom = 21$ are more suitable for dengue estimation in the same city. Models trained in one city, with transfer learning from other trained cities, or trained in a multiple-city way, could estimate risk dengue incidence rates with reasonable correlation, for different Brazilian cities. Trained models also can help increase the estimation and the application of public policies. For the latter, aerial images with $zoom = 19$ generalize better to unseen cities.

Furthermore, using the proposed models with the cell region estimation methodology, it is possible to observe hot-spot estimated maps, as well as obtain a ranked list of neighborhoods or districts that need immediate attention from health agents for guidance and prevention of dengue epidemics. To illustrate this approach, Figure 46 presents a hot-spot map of dengue incidence rates within cell regions for the city of Salvador, created with a São Paulo model *SpStreetSat19* with aerial features from *SpSat19*. Although estimations are scaled with a higher maximum rate, it is possible to observe remnants of the original hot-spot at the same cell region locations. The high dengue rated cells could be targeted for local public policies in order to prevent a local outbreak of dengue fever.

Figure 47 exemplify a utility of the proposed models, with the use of resulted risk ranked lists in Salvador, with the application of São Paulo model *SpStreetSat19* with aerial features from *SpSat19*. On the left, the actual risk ranking of dengue incidence

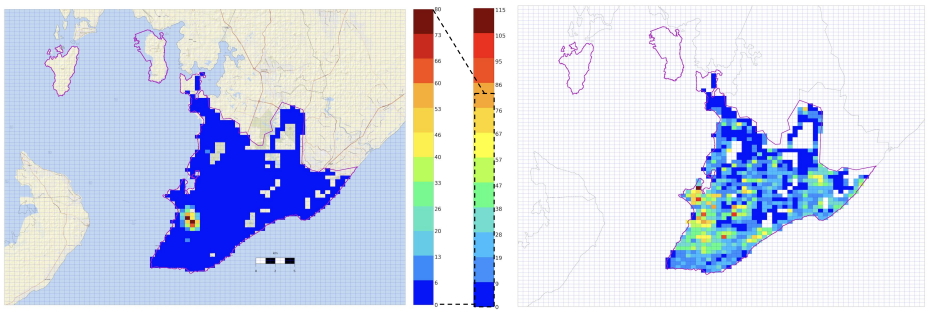
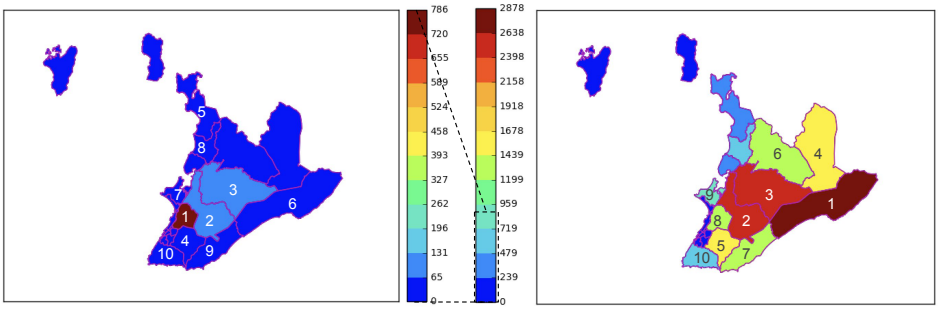


Figure 46 – Salvador original hot-spot map (left) of dengue incidence rates, and estimated values (right) with *SpStreetSat19* with aerial features from *SpSat19*.



	ACTUAL
1	SANTO ANTÔNIO
2	SÃO CAETANO
3	PIRAJÁ
4	BROTAS
5	SÃO TOMÉ/PARIPE
6	ITAPOÃ
7	PENHA
8	PERIPERI
9	AMARALINA
10	VITÓRIA

	ESTIMATED
1	ITAPOÃ
2	SÃO CAETANO
3	PIRAJÁ
4	SÃO CRISTOVÃO
5	BROTAS
6	VALÉRIA
7	AMARALINA
8	SANTO ANTÔNIO
9	PENHA
10	VITÓRIA

Figure 47 – Salvador actual risk ranked list (left) of districts with dengue incidence rates, and estimated values (right) risk ranked list, with *SpStreetSat19* with aerial features from *SpSat19*.

for districts of Salvador, in the right, the estimated risk ranking of dengue incidence rates, from the results presented in Table 38. Although the model cannot estimate the exact number of dengue incidence, two high-risk districts,“*São Caetano*” and “*Pirajá*” appears in the estimated list, which could be set as target regions by health professionals and public health policies.

7 CONCLUSION

Identifying places where intervention would be more productive is a central part of any public policy aiming at preventing dengue from spreading in urban centers. Motivated by this, and leveraged by the recent research area *Visual Computational Sociology*, this Thesis proposed Deep Neural Network models aimed at estimating dengue incidence rates automatically from urban images. Experiments were carried out with street-level and aerial images obtained from Google Street View and Maps API services, together with historical dengue fever data collected from the Brazilian capitals Rio de Janeiro (RJ), São Paulo (SP), and Salvador (BA).

The next sections enumerates the conclusions and considerations from evaluating the models, and possible future works related to this research field, as well as methods that may contribute to better results and topics that need further investigation.

7.1 Evaluation of Models

From the models' evaluation results, it was possible to conclude that there is information in street-level and aerial images that makes it possible to estimate dengue incidence rates in urban centers using these image categories solely as predictors. The use of combined information from both aerial and street-level domains as inputs adds more information to the model, allowing better results than their separated use.

It should be observed that models trained and validated in regions of a specific city perform better in other areas of the same city than in another distinct city. This leads to the conclusion that the potential of proposed models is restricted to the cities to which they were fitted. For better results when applied in other cities, one must expose the model to examples from more than one city or initializing it with weights from models trained in other cities.

The proposed models are limited to a few exact estimates, i.e., with the absolute error between actual and estimated values near zero. Even when the correlation between actual and estimated values is relatively strong, considering all tested examples, there are only a fraction of absolute errors near zero. Better results are obtained when

the evaluation is done considering regions larger than a single locality, through cell regions estimations or ranked lists of administrative districts with high to low dengue risk. In addition, the proposed models are also limited to estimations without considering periods of dengue occurrences, i.e., predictions and forecasting, because it is not possible to specify the date of the images used, nor to choose specific dates and years to compose blocks of images from a certain period.

Specifically, regarding aspects of the design of proposed models, this Thesis showed evidence that to extract information from urban imagery, the Deep Convolutional Neural Networks are more suitable than designed descriptors, since it produces more representative features than descriptors itself. Considering the *zoom* level of aerial images, for models trained and tested in the same city, input features from aerial images with $zoom = 21$ performed better than aerial images with other zoom values, and for generalization for other cities, $zoom = 19$ yielded the best results.

Overall, there is a relative advantage of using urban images when compared to demographic data, because images are a ubiquitous data domain that, in developing countries, can be less costly and straightforward to obtain than data collected directly from the inhabitants of the regions studied, such as in the census research. Street-level and aerial urban images can be used as a proxy for socioeconomic data, as demonstrated in related works, and specifically, with dengue incidence rates, with the results obtained in the experiments performed in the present work. Finally, this work adds to the *Visual Computational Sociology* research area by showing that it is possible to infer more than aesthetics, demographics, and socioeconomic factors from urban images.

7.2 Future Works

The following presents some possible future works related to the materials and methods used in this present Thesis.

7.2.1 Aerial and Street-view imagery

Google Street View and Static Maps imagery are continually being updated, and previously accessed images may become outdated from current ones. To this date, query for specific date images using the Street and Static APIs is not possible. While this functionality is not available, manually cataloging the date of the images and the impact that the use of images from different periods has on dengue estimated values should be considered as future work.

In addition, updating images that could not be accessed because they were unavailable at the time of download should be considered to increase the number of valid images available in the dataset.

7.2.2 Data distribution

This work presented the use of "*Quadrat Thematic Mapping*" in a grid data structure, forming a grid map. The following items show possibilities for the grid map exploitation related to the structure composition:

- Different grid data structures shapes (resolution), e.g. 100×100, 80×80, 40×40, resulting in different dengue incidence cell distribution, and their impact in estimation models.
- Grid maps structure is instantiated using a naive approach, where the resulting cell aspect ratio is a function of arbitrary lower and upper corners latitude and longitude coordinates. Composing a grid structure with a specific aspect ratio should be evaluated and compared with arbitrary grids, for the specific problem of dengue incidence rates and related problems.
- The results obtain using equal grid resolutions for different cities.
- The performance of grid data structure maps compared to different data distribution methods, i.e., by census sectors, districts or neighborhoods, and different ranges of covered areas, i.g., circle areas, ellipses, hexagons, perfect squares.

7.2.3 Designed Descriptors

A more extent investigation using the proposed baseline designed descriptors - *Histogram of Oriented Gradients* (HOGs) compared with *Spatial Envelopes* (GIST) - and the possible hyperparameter tuning involved in the optimization of the models should be considered, for street-level imagery, since the slightly better results achieved from GIST descriptors compared with HOGs for aerial feature models.

The different approaches in combining features in Street-level baseline models should be investigated, evaluating the impact of input image resolution and proposed an architecture in the obtained results. Furthermore, the GIST descriptor should be used in place of HOG descriptors, and both results compared.

7.2.4 Deep Convolutional Neural Networks Features

7.2.4.1 Transfer Learning from different architectures and datasets

In this work, it was proposed the use of *DenseNet161* using pretrained weights from *Places* dataset, as feature extractor for street-level images. The use of different state of the art architectures for fine-tuning models and using pretrained weights from different datasets, such as *ImageNet*, compared to the proposed models in this Thesis, should be evaluated.

7.2.4.2 Interpretability of proposed models

A preliminary investigation of which information in urban images is more present in location points estimations with low or high dengue incidence rates was performed in this work. Only street-level imagery was investigated due to the feature extraction technique from *DenseNet161* architecture pretrained with the Places dataset. From this approach, no significant information related to the causality of urban attributes in correct estimations were found while inspecting *Places* labels assigned to the street-level images.

Visualization and interpretation of deep learning models methods should be applied to extract information from street-level and aerial images according to the resulting estimated values, such as *Gradient-weighted Class Activation Mapping* (Grad-CAM) (SELVARAJU et al., 2017), or more specific for regression problems, the *Regression Activation Map* (RAM) (WANG; YANG, 2017).

7.2.5 Regression Neural Network Module

The proposed fully-connected neural networks placed at the regressor model should be compared with existents models from related works that propose regression problems using Deep ConvNets as feature extractors. The method applied should use the proposed models with related work latent variables data and comparing the results obtained from the author's proposed model with the present regression models in this Thesis.

7.2.6 Train and Test Metodology

Finally, a different approach in training and testing the proposed models, such as a k-fold cross-validation methodology, should be considered. In this method, the validation set is included in the training set, at each fold model, to increase models' generalization capacity for unseen images in the test set, leveraging the validation set images also in the model training. This may contribute to better results than those achieved in this present Thesis.

REFERENCES

ALBERT, A.; KAUR, J.; GONZALEZ, M. C. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 23., 2017, Halifax, NS, Canada. **Proceedings...** ACM Press, 2017. p.1357–1366.

AMARASINGHE, A. et al. Suppressing Dengue via a Drone System. In: INTERNATIONAL CONFERENCE ON ADVANCES IN ICT FOR EMERGING REGIONS (ICTER), 2017, Colombo, Sri Lanka. **Anais...** IEEE, 2017. p.202–208.

ANDERSSON, V. O.; BIRCK, M. A.; ARAUJO, R. M. Investigating Crime Rate Prediction Using Street-Level Images and Siamese Convolutional Neural Networks. In: LATIN AMERICAN WORKSHOP ON COMPUTATIONAL NEUROSCIENCE, 2017, Porto Alegre, Brazil. **Anais...** Springer, 2017. p.81–93.

ANDERSSON, V. O.; BIRCK, M. A. F.; ARAUJO, R. M. Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2018., 2018, Rio de Janeiro, Brazil. **Anais...** IEEE, 2018. p.1–8.

ANDERSSON, V. O.; CECHINEL, C.; ARAUJO, R. M. Combining Street-level and Aerial Images for Dengue Incidence Rate Estimation. In: TO APPEAR: 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, Budapest, Hungary. **Anais...** IEEE, 2019. p.1–8.

ANGGRAENI, W. et al. Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia. **Procedia Computer Science**, Elsevier B.V., Amsterdam, Netherlands, v.124, p.142–150, 2017.

ANGUELOV, D. et al. Google Street View: Capturing The World at Street Level. **Computer**, Washington, DC, v.43, n.June, p.32–38, 2010.

Answer from StackExchange. **Ratio of pixels per meter**. [Online; accessed 07-Oct-2019. Archived: http://archive.today/3TDWV], https://gis.stackexchange.com/a/178905.

ARIETTA, S.; AGRAWALA, M.; RAMAMOORTHI, R. **On relating visual elements to city statistics**. Berkeley, CA, USA: Technical Report UCB/EECS-2013-157, EECS Department, University of California (Sep 2013), 2013.

ARIETTA, S. M.; EFROS, A. A. City Forensics : Using Visual Elements to Predict Non-Visual City Attributes. **Transactions on Visualization and Computer Graphics**, IEEE, v.20, n.12, p.2624–2633, 2014.

ASHBY, J.; MORENO-MADRIÑÁN, M.; YIANNOUTSOS, C.; STANFORTH, A. Niche Modeling of Dengue Fever Using Remotely Sensed Environmental Factors and Boosted Regression Trees. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, Basel, Switzerland, v.9, n.4, p.328, 2017.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. **Transactions on pattern analysis and machine intelligence**, IEEE, v.39, n.12, p.2481–2495, 2017.

BAIDU. **Baidu Maps API**. China: Baidu, 2019. [Online; accessed 27-Jun-2019], http://developer.baidu.com/map/viewstatic.htm.

BASU, S. et al. Deepsat: a learning framework for satellite imagery. In: SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 23., 2015. **Proceedings...** ACM, 2015. p.37.

BENCY, A. J. et al. Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery. In: IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV), 2017, Santa Rosa, CA, USA. **Anais...** IEEE, 2017. p.320–329.

BENYISHAY, A. et al. Geocoding Afrobarometer rounds 1–6: methodology & data quality. **Geocoded Afrobarometer Surveys**, AidData, William & Mary, 2017.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, JMLR, v.13, n.Feb, p.281–305, 2012.

BOTTOU, L. Large-Scale Machine Learning with Stochastic Gradient Descent. **Proceedings of COMPSTAT'2010**, Springer, 2010.

BOWERS, K. J.; JOHNSON, S. D.; PEASE, K. Prospective hot-spotting: The future of crime mapping? **British Journal of Criminology**, Oxford, v.44, n.5, p.641–658, 2004.

BRACEWELL, R. N. **The Fourier transform and its applications**. McGraw-Hill New York: McGraw-Hill New York, 1986. v.31999.

BROADFOOT, C.; GOOGLE. **We need a Map.getScale() method - Google Groups**. [Online; accessed 07-Oct-2019. Archived: http://archive.today/xge58], https://groups.google.com/d/msg/google-maps-js-api-v3/hDRO4oHVSeM/osOYQYXg2oUJ.

BUCZAK, A. L. et al. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. **BMC Medical Informatics and Decision Making**, London, UK, v.12, n.1, p.124, dec 2012.

BUCZAK, A. L. et al. Prediction of High Incidence of Dengue in the Philippines. **PLoS Neglected Tropical Diseases**, San Francisco, CA, USA, v.8, n.4, p.e2771, apr 2014.

CAO, R. et al. Integrating Aerial and Street View Images for Urban Land Use Classification. **Remote Sensing**, Basel, Switzerland, v.10, n.10, p.1553, 2018.

CASE, E. H. **MosquitoNet**: Investigating the use of unmanned aerial vehicles and neural networks in integrated mosquito management. 2017. Dissertação (Mestrado em Ciência da Computação) — Cornell University, Ithaca, NY, USA.

CGU. **Electronic System Of Citizen Information Service**. [Online; accessed 15-Aug-2019], https://esic.cgu.gov.br/sistema/site/index.aspx.

CHAINEY, S.; TOMPSON, L.; UHLIG, S. The utility of hotspot mapping for predicting spatial patterns of crime. **Security journal**, Springer, v.21, n.1-2, p.4–28, 2008.

CHENG, G.; HAN, J. A survey on object detection in optical remote sensing images. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v.117, p.11–28, 2016.

CHEONG, Y. L.; LEITÃO, P. J.; LAKES, T. Assessment of land use factors associated with dengue cases in Malaysia using boosted regression trees. **Spatial and Spatio-temporal Epidemiology**, Amsterdam, Netherlands, v.10, p.75–84, 2014.

CHOPRA, S.; HADSELL, R.; LECUN, Y. Learning a similiarty metric discriminatively, with application to face verification. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2005, San Diego, CA, USA. **Proceedings...** IEEE, 2005. p.349–356.

COELHO, F. C.; DE CARVALHO, L. M. Estimating the attack ratio of dengue epidemics under time-varying force of infection using aggregated notification data. **Scientific reports**, Nature Publishing Group, v.5, p.18455, 2015.

COHEN, D. A. et al. Neighborhood Physical Conditions and Health. **American Journal of Public Health**, Washington, DC, v.93, n.3, p.467–471, 2003.

COHEN, D. et al. “Broken Windows” and the Risk of Gonorrhea. **American Journal of Public Health**, Washington, DC, v.90, n.2, p.230–236, 2000.

CSAILVision. **CSAIL Computer Vision Places365 GitHub Project**. MIT: GitHub, 2015. https://github.com/CSAILVision/places365.

CURTIS, A.; BLACKBURN, J. K.; WIDMER, J. M.; MORRIS, G. A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Petit-Goave, Haiti. **International Journal of Health Geographics**, BioMed Central, Springer Nature, London, UK, v.12, n.1, p.21, 2013.

CURTIS, A.; QUINN, M.; OBENAUER, J.; RENK, B. M. Supporting local health decision making with spatial video: Dengue, Chikungunya and Zika risks in a data poor, informal community in Nicaragua. **Applied Geography**, Elsevier Ltd, Amsterdam, Netherlands, v.87, p.197–206, 2017.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2005, San Diego, CA, USA. **Anais...** IEEE, 2005. v.I, p.886–893.

DATARIO. **Notified and Georeferenced Dengue Cases**. [Offline; accessed 10-Jan-2018], http://dadosabertos.rio/dataset/casos_notificados_de_dengue.

DATARIO. **Neighborhoods limits from Rio de Janeiro**. [Online; accessed 12-Jan-2018], http://www.data.rio/datasets/limite-bairro.

De Nadai, M. et al. Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life. In: ACM ON MULTIMEDIA CONFERENCE, 2016, Amsterdam, Netherlands. **Anais...** ACM, 2016. p.1127–1135.

DEB, S. et al. An ensemble prediction approach to weekly Dengue cases forecasting based on climatic and terrain conditions. **Journal of Health and Social Sciences**, SIPISS FerrariSinibaldi, Milano, Italy, v.23, p.257–272, 2017.

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2009., 2009, Miami Beach, FL, USA. **Anais...** IEEE, 2009. p.248–255.

DOERSCH, C. et al. What makes Paris look like Paris? **ACM Transactions on Graphics**, ACM, New York, NY, USA, v.31, n.4, p.1–9, 2012.

DOUZE, M. et al. Evaluation of GIST descriptors for web-scale image search. In: CIVR 2009 - International Conference on Image and Video Retrieval, 2009, Santorini, Greece. **Anais...** ACM, 2009. p.19:1–8.

DRUCKER, H. et al. Support Vector Regression Machines. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 1997, NIPS, Denver, Colorado, USA. **Anais...** NIPS, 1997. v.1, p.155–161.

DUBEY, A. et al. Deep learning the city: Quantifying urban perception at a global scale. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2016, Amsterdam, Netherlands. **Anais...** Springer, 2016. p.196–212.

ECK, J.; CHAINEY, S.; CAMERON, J.; WILSON, R. Mapping crime: Understanding hotspots. **National Institute of Justice Special Report**, Washington, DC, USA, 2005.

ESRI. **ESRI Shapefile Technical Description**. Redlands, CA, USA: [s.n.], 1998. (July).

FANG, Q.; SANG, J.; XU, C. Giant: Geo-informative attributes for location recognition and exploration. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 21., 2013. **Proceedings...** ACM, 2013. p.13–22.

FU, K.; CHEN, Z.; LU, C.-T. StreetNet: preference learning with convolutional neural network on urban crime perception. In: ACM SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 26., 2018, ACM. **Proceedings...** ACM, 2018. p.269–278.

GADM. **Administrative limits from Salvador**. [Online; accessed 12-Jan-2018], `https://www.gadm.org/maps/BRA/bahia/salvador.html`.

GANGULI, S.; DUNNMON, J.; HAU, D. Predicting Food Security Outcomes Using Convolutional Neural Networks (CNNs) for Satellite Tasking. **arXiv preprint arXiv:1902.05433**, ArXiv, 2019.

GEBRU, T. et al. Fine-grained car detection for visual census estimation. In: THIRTY-FIRST AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2017, AAAI. **Anais...** AAAI, 2017.

GEBRU, T. et al. Using deep learning and google street view to estimate the demographic makeup of the us. **Preprint arXiv:1702.06683**, arXiv, 2017.

GEIPEL, J.; LINK, J.; CLAUPEIN, W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v.6, n.11, p.10335–10355, 2014.

GEONET, T. E. C. **Distance on a sphere**: The Haversine Formula. [Online; accessed 21-Aug-2019], https://community.esri.com/groups/coordinate-reference-systems/blog/2017/10/05/haversine-formula.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: THIRTEENTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 2010, Sardinia, Italy. **Proceedings...** JMLR, 2010. p.249–256.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. London, UK: MIT Press, 2016. http://www.deeplearningbook.org.

GOOGLE. **Google Street View**. [Online; accessed 09-May-2017], https://www.google.com/streetview/.

GOOGLE. **Google Static Maps**. [Online; accessed 09-May-2018], https://developers.google.com/maps/documentation/maps-static/dev-guide.

GOOGLE. **Google Earth**. [Online; accessed 30-Mar-2018], https://www.google.com/earth/.

GOOGLE. **Google Geocoding API**. [Online; accessed 15-Aug-2019], https://developers.google.com/maps/documentation/geocoding/start.

GOOGLE. **Google Maps**. [Online; accessed 07-Oct-2019], https://www.google.com/maps.

GUO, P. et al. Developing a dengue forecast model using machine learning: A case study in China. **PLoS Neglected Tropical Diseases**, Public Library of Science, San Francisco, CA, USA, v.11, n.10, 2017.

GUZMAN, A.; ISTÚRIZ, R. E. Update on the global spread of dengue. **International Journal of Antimicrobial Agents**, Elsevier B.V., v.36, n.SUPPL. 1, p.S40–S42, 2010.

HARIHARAN, B.; ARBELAEZ, P.; GIRSHICK, R.; MALIK, J. Hypercolumns for Object Segmentation and Fine-Grained Localization. In: THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2015, Boston, MA, US. **Anais...** IEEE, 2015.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**: Data Mining, Inference, and Prediction. 2.ed. New York, NY, USA: Springer New York Inc., 2008. (Springer Series in Statistics).

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recogniton. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016, Las Vegas, NV, USA. **Anais...** IEEE, 2016. p.770–778.

HIRZEL, A. H.; GUISAN, A. Which is the optimal sampling strategy for habitat suitability modelling. **Ecological Modelling**, Amsterdam, Netherlands, v.157, n.2-3, p.331–341, 2002.

HOFFMANN, E. J. et al. Model Fusion for Building Type Classification from Aerial and Street View Images. **Remote Sensing**, Elsevier, v.11, n.11, p.1259, 2019.

HOSSACK, W. **The Fourier Transform (What you need to know)**. The University of Edinburgh: The School of Physics and Astronomy, 2016.

HUANG, G.; LIU, Z.; WEINBERGER, K. Q.; MAATEN, L. van der. Densely connected convolutional networks. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2017, Puerto Rico. **Anais...** IEEE, 2017. v.1, n.2, p.3.

HUBEL, D. H.; WIESEL, T. N. Receptive Fields, Binocular interaction and functional architecture in the cats visual cortex. **Neurophysiolojy Pharmacology**, Printed in Great Britain, p.106–154, 1962.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, IEEE, v.9, n.3, p.90–95, 2007.

IANDOLA, F. N. et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. **arXiv preprint arXiv:1602.07360**, Cornell University Library, 2016.

IBGE. **Síntese de indicadores sociais: uma análise das condições de vida da população brasileira**. Instituto Brasileiro de Geografia e Estatística: Instituto Brasileiro de Geografia e Estatística, 2016. 146p. v.39.

JEAN, N. et al. Combining satellite imagery and machine learning to predict poverty. **Science**, New York, N.Y., v.353, n.6301, p.790–4, 2016.

KANG, H.-W.; KANG, H.-B. Urban Safety Prediction Using Context and Object Information via Double-Column Convolutional Neural Network. In: CONFERENCE ON COMPUTER AND ROBOT VISION (CRV), 2016., 2016, IEEE. **Anais...** IEEE, 2016. p.399–405.

KARPATHY, A. **"CS231n Convolutional Neural Networks for Visual Recognition"**. [Online; accessed 30-Mar-2018], `http://cs231n.github.io/convolutional-networks/`.

KHORMI, H. M.; KUMAR, L. Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing based case study. **Science of the Total Environment**, Elsevier B.V., Amsterdam, Netherlands, v.409, n.22, p.4713–4719, 2011.

KHOSLA, A.; AN, B.; LIM, J. J.; TORRALBA, A. Looking beyond the visible scene. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2014, IEEE. **Proceedings...** IEEE, 2014. p.3710–3717.

KIKUTI, M. et al. Spatial distribution of dengue in a Brazilian Urban slum setting: Role of socioeconomic gradient in disease risk. **PLoS Neglected Tropical Diseases**, PLOS, v.9, n.7, p.1–18, 2015.

KITA, K.; KIDZIŃSKI, Ł. Google Street View image of a house predicts car accident risk of its resident. **Preprint arXiv:1904.05270**, arXiv, 2019.

KRASIN, I. et al. **"OpenImages: A public dataset for large-scale multi-label and multi-class image classification."**. [Online; accessed 30-Mar-2018], `https://storage.googleapis.com/openimages/web/index.html`.

KRIZHEVSKY, A. **Learning multiple layers of features from tiny images**. MIT: Tech Report, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 2012, Lake Tahoe, NV, USA. **Anais...** Neural Information Processing Systems Foundation Inc., 2012. p.1097—-1105.

KRYSTOSIK, A. R. et al. Community context and sub-neighborhood scale detail to explain dengue, chikungunya and Zika patterns in Cali, Colombia. **PLoS ONE**, Public Library of Science, San Francisco, CA, USA, v.12, n.8, 2017.

LATKIN, C. A.; CURRY, A. D. Stressful Neighborhoods and Depression: A Prospective Study of the Impact of Neighborhood Disorder. **Journal of Health and Social Behavior**, Washington, DC, v.44, n.1, p.34–44, 2003.

LAUREANO-ROSARIO, A. E. et al. Modelling dengue fever risk in the State of Yucatan, Mexico using regional-scale satellite-derived sea surface temperature. **Acta Tropica**, Elsevier B.V., Amsterdam, Netherlands, v.172, n.January, p.50–57, 2017.

LAUREANO-ROSARIO, A. et al. Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico. **Tropical Medicine and Infectious Disease**, Multidisciplinary Digital Publishing Institute, Basel, Switzerland, v.3, n.1, p.5, jan 2018.

LAW, S.; PAIGE, B.; RUSSELL, C. Take a look around: using street view and satellite images to estimate house prices. **Preprint arXiv:1807.07155**, arXiv, 2018.

LE, H. T. et al. Study of CNN Based Classification for Small Specific Datasets. In: **Studies in Computational Intelligence**. International Publishing AG: Springer Nature, 2018. v.769.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-Based Learning Applied to Document Recognition. **Proceedings of the IEEE**, IEEE, p.2278—-2324, 1998.

LECUN, Y.; BOTTOU, L.; Orr, B, G.; MULLER, K.-R. Efficient BackProp. **Neural Networks: tricks of the trade**, Springer, p.9–48, 1998.

LEFÈVRE, S. et al. Toward seamless multiview scene analysis from satellite to street level. **Proceedings of the IEEE**, New Jersey, US, v.105, n.10, p.1884–1899, 2017.

LI, Q. et al. Ecological niche modeling identifies fine-scale areas at high risk of dengue fever in the pearl river delta, China. **International Journal of Environmental Research and Public Health**, Multidisciplinary Digital Publishing Institute, Basel, Switzerland, v.14, n.6, p.619, jun 2017.

LIEMAN-SIFRY, J. **Convolutional Neural Networks to Predict Location from Colorado Google Street View Images**. Galvanize Capstone Project: GitHub, 2016. `https://github.com/jliemansifry/streetview/`.

LIN, C. H.; WEN, T. H. Using geographically weighted regression (GWR) to explorespatial varying relationships of immature mosquitoes andhuman densities with the incidence of dengue. **International Journal of Environmental Research and Public Health**, Multidisciplinary Digital Publishing Institute, Basel, Switzerland, v.8, n.7, p.2798–2815, 2011.

LIN, T.; STICH, S. U.; PATEL, K. K.; JAGGI, M. Don't Use Large Mini-Batches, Use Local SGD. **Preprint**, arXiv, n.1, p.1–35, 2018.

LIN, T.-Y. et al. Microsoft COCO: Common objects in context. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2014, Zurich, Switzerland. **Anais...** Springer, 2014. p.740—-755.

LIU, X.; CHEN, Q.; XU, Y.; LIN, L. Place-centric Visual Urban Perception with Deep Multi-instance Regression. In: ACM CONFERENCE ON MULTIMEDIA, 2017, Mountain View, CA. **Proceedings...** ACM Press, 2017. n.October, p.1–9.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, New York, NY, USA, v.60, n.2, p.91–110, 2004.

MESINA, J.; ISANAN, J.; MADERAZO, C. Poverty incidence identification of cities and municipalities using convolutional Neural Network as applied to satellite imagery. **IOP Conference Series: Materials Science and Engineering**, Bristol, England, v.482, n.1, p.012044, 2019.

MEYNBERG, O.; CUI, S.; REINARTZ, P. Detection of high-density crowds in aerial images using texture classification. **Remote Sensing**, Elsevier, v.8, n.6, p.470, 2016.

MICROSOFT. **Bing Maps REST Toolkit**. San Francisco, California, US: Github, 2019. [Online; accessed 28-Jun-2019], https://github.com/Microsoft/BingMapsRESTToolkit/.

Ministry of Health. **LIRAa appoints 357 at-risk municipalities for dengue, zika and chikungunya.** [Online; accessed 19-Aug-2019], http://www.saude.gov.br/noticias/agencia-saude/42041-liraa-aponta-357-municipios-em-situacao-de-risco-para-dengue-zika-e-chikungunya.

MNIH, V.; HINTON, G. E. Learning to detect roads in high-resolution aerial images. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2010, Springer. **Anais...** Springer, 2010. p.210–223.

MONTAVON, G.; SAMEK, W.; MÜLLER, K. R. Methods for interpreting and understanding deep neural networks. **Digital Signal Processing: A Review Journal**, Elsevier Inc., v.73, p.1–15, 2018.

MUJAHID, M. S. et al. Relation between neighborhood environments and obesity in the multi-ethnic study of atherosclerosis. **American Journal of Epidemiology**, Baltimore, MD, v.167, n.11, p.1349–1357, 2008.

NAIK, N. **Visual Urban Sensing: Understanding Cities throught Computer Vision**. 2017. 1–172p. Tese (Doutorado em Ciência da Computação) — Massachusets Insitute of Technology.

NAIK, N.; PHILIPOOM, J.; RASKAR, R.; HIDALGO, C. Streetscore - Predicting the Perceived Safety of One Million Streetscapes. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPR), 2014, Columbus, Ohio, USA. **Anais...** IEEE, 2014.

NAJJAR, A.; KANEKO, S.; MIYANAGA, Y. Crime Mapping from Satellite Imagery via Deep Learning. **arXiv preprint arXiv:1812.06764**, ArXiv, 2018.

NASA.GOV. **Earth Fact Sheet**. [Online; accessed 21-Aug-2019], https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html.

NG, A. **Machine learning yearning**: Technical strategy for ai engineers in the era of deep learning. [Online; accessed 18-Oct-2019], https://www.deeplearning.ai/machine-learning-yearning/.

O'HAVER, T. A pragmatic introduction to signal processing. **University of Maryland**, College Park, MD, 1997.

OLIVA, A.; TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. **International Journal of Computer Vision**, [S.l.], v.42, n.3, p.145–175, 2001.

OpenStreetMap. **Hot Export Tool**: Download OpenStreetMap Data. [Online; accessed 10-Apr-2018], https://export.hotosm.org/en/v3/.

OpenStreetMap contributors. **Planet dump retrieved from https://planet.osm.org** . https://www.openstreetmap.org.

ORDONEZ, V.; BERG, T. L. Learning High-level Judgments of Urban Perception. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, Springer, New York, NY, USA, v.8694 LNCS, n.PART 6, p.494–510, 2014.

OSHRI, B. et al. Infrastructure quality assessment in africa using satellite imagery and deep learning. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY & DATA MINING, 24., 2018, London, United Kingdom. **Proceedings...** ACM, 2018. p.616–625.

PERRONNIN, F.; SANCHEZ, J.; MENSINK, T. Improving the Fisher Kernel for Large-Scale Image Classification. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2010, Heraklion, Crete, Greece. **Anais...** Springer, 2010. p.143–156.

PORZI, L.; ROTA BULÒ, S.; LEPRI, B.; RICCI, E. Predicting and understanding urban perception with convolutional neural networks. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 23., 2015, Brisbane, Australia. **Proceedings...** ACM, 2015. p.139–148.

QGIS Development Team. **QGIS Geographic Information System**. OSGeo: Open Source Geospatial Foundation, 2009. http://qgis.osgeo.org.

ROSSER, G. et al. Predictive Crime Mapping: Arbitrary Grids or Street Networks? **Journal of Quantitative Criminology**, Springer US, v.33, n.3, p.569–594, 2016.

RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision**, Springer, v.115, n.3, p.211–252, 2015.

SALESSES, P.; SCHECHTNER, K.; HIDALGO, C. A. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. **PLoS ONE**, San Francisco, CA, USA, v.8, n.7, p.e68400, jul 2013.

SARFRAZ, M. S.; TRIPATHI, N. K.; KITAMOTO, A. Near real-time characterisation of urban environments: a holistic approach for monitoring dengue fever risk areas. **International Journal of Digital Earth**, Abingdon, UK, v.7, n.11, p.916–934, 2014.

SCAVUZZO, J. M. et al. Modeling the temporal pattern of Dengue, Chicungunya and Zika vector using satellite data and Neural Networks. In: XVII WORKSHOP ON INFORMATION PROCESSING AND CONTROL (RPIC), 2017, Mar de Plata, Argentina. **Anais...** IEEE, 2017. p.1–6.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, IEEE. **Proceedings...** IEEE, 2017. p.618–626.

SHAPIRO, A. Street-level: Google Street View's abstraction by datafication. **New Media & Society**, SAGE, p.146144481668729, 2017.

SHERMAN, L. W.; GARTIN, P. R.; BURGER, M. E. Hot spots of predatory crime : routine activitices and the criminology of place. **Criminology**, Washington, DC, USA, v.27, n.June, p.27—-55, 1989.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **arXiv preprint arXiv:1409.1556**, ArXiv, p.1–10, 2015.

SMDU, S. M. d. D. U. **Administrative limits from São Paulo**. [Online; accessed 12-Jan-2018], http://dados.prefeitura.sp.gov.br/pt_PT/dataset/distritos.

SMS-RIO, S. M. d. S. d. C. d. R. d. J. **Notified and Georeferenced Dengue Cases**. [Online; accessed 29-Apr-2019], http://dadosabertos.rio.rj.gov.br/apiSaude/Apresentacao/pdf/documentacao_dengue.pdf.

SRINIVAS, S. et al. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision. **Vision Systems Theory, Tools and Applications, Frontiers in Robotics and AI**, University of Bath, UK, v.2, n.January, p.1–13, 2016.

STOLERMAN, L.; MAIA, P.; KUTZ, J. N. Data-Driven Forecast of Dengue Outbreaks in Brazil: A Critical Assessment of Climate Conditions for Different Capitals. In: XIV, 2016. **Anais...** ArXiv.org, 2016.

SUDUWELLA, C. et al. Identifying Mosquito Breeding Sites via Drone Images. In: WORKSHOP ON MICRO AERIAL VEHICLE NETWORKS, SYSTEMS, AND APPLICATIONS (DRONET-17), 3., 2017, Niagara Falls, NY, USA. **Anais...** ACM Press, 2017. n.3, p.27–30.

SUEL, E.; POLAK, J. W.; BENNETT, J. E.; EZZATI, M. Measuring social, environmental and health inequalities using deep learning and street imagery. **Scientific reports**, Nature Publishing Group, v.9, n.1, p.6229, 2019.

SUS. **SINAN - Sistema de Informação de Agravos de Notificação**. Ministério da Saúde: Sistema Único de Saúde, 2018. [Online; accessed 29-Apr-2019], `http://portalsinan.saude.gov.br/notificacoes`.

SZEGEDY, C. et al. Going Deeper with Convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2015, Boston, MA, USA. **Proceedings...** IEEE, 2015. p.1–9.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016, Las Vegas, Nevada, USA. **Anais...** IEEE, 2016. p.2818–2826.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI), 2017, San Francisco, California, USA. **Anais...** AAAI, 2017. v.4, p.12.

TEURLAI, M. et al. Socio-economic and Climate Factors Associated with Dengue Fever Spatial Heterogeneity: A Worked Example in New Caledonia. **PLoS Neglected Tropical Diseases**, San Francisco, CA, USA, v.9, n.12, p.1–31, 2015.

THAMMAPALO, S.; CHONGSUVIVATWONG, V.; GEATER, A.; DUERAVEE, M. Environmental factors and incidence of dengue fever and dengue haemorrhagic fever in an urban area, Southern Thailand. **Epidemiology and Infection**, Cambridge University Press, v.136, n.1, p.135–143, 2008.

WANG, W. et al. Urban perception of commercial activeness from satellite images and streetscapes. In: COMPANION OF THE THE WEB CONFERENCE, 2018, ACM. **Anais...** ACM, 2018. p.647–654.

WANG, Z.; YANG, J. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. **Preprint arXiv:1703.10757**, arXiv, 2017.

WEBBER, W.; MOFFAT, A.; ZOBEL, J. A similarity measure for indefinite rankings. **ACM Transactions on Information Systems (TOIS)**, ACM, v.28, n.4, p.20, 2010.

WHO. **Dengue and Severe Dengue**. World Health Organization: Geneva, Switzerland, 2018. [Online; accessed 1-Feb-2018], http://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

WILDER-SMITH, A.; GUBLER, D. J. Geographic Expansion of Dengue: The Impact of International Travel. **Medical Clinics of North America**, Elsevier, v.92, n.6, p.1377–1390, 2008.

WILSON, J. Q.; KELLING, G. L. Broken Windows. **Atlantic Monthly**, Boston, MA, USA, v.249, n.3, p.29, 1982.

WORKMAN, S.; ZHAI, M.; CRANDALL, D. J.; JACOBS, N. A unified model for near and remote sensing. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, IEEE. **Proceedings...** IEEE, 2017. p.2688–2697.

XIAO, J.; HAYS, J.; EHINGER, K. A.; TORRALBA, A. SUN database: Large-scale scene recognition from abbey to zoo. In: COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2010, San Francisco, CA, USA. **Anais...** IEEE, 2010.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? **Advances in Neural Information Processing Systems 27 (Proceedings of NIPS)**, Neural Information Processing Systems Foundation, Inc., v.27, p.1–9, nov 2014.

ZAGORUYKO, S.; KOMODAKIS, N. Learning to compare image patches via convolutional neural networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, IEEE. **Proceedings...** IEEE, 2015. p.4353–4361.

ZELLWEGER, R. M. et al. Socioeconomic and environmental determinants of dengue transmission in an urban setting: An ecological study in Nouméa, New Caledonia. **PLoS Neglected Tropical Diseases**, Public Library of Science, San Francisco, CA, USA, v.11, n.4, p.1–18, 2017.

ZHOU, B. et al. Learning Deep Features for Scene Recognition using Places Database. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 2014, Montreal, Canada. **Anais...** NIPS, 2014. p.487–495.

ZHOU, B. et al. Places: A 10 Million Image Database for Scene Recognition. **Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, USA, v.40, n.6, p.1452–1464, 2018.

ZHOU, B. et al. Places: A 10 Million Image Database for Scene Recognition. **Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, USA, v.40, n.6, p.1452–1464, 2018.

Appendix A GRID DATA STRUCTURE PARAMETERS

The developed grid data structure API, named *GridMapAPI*, to map street location points to urban latent variables is available for download and use at `https://github.com/Vortander/GridMapAPI`. The API also includes routines for distributing and creating train, validation and test datasets.

Tables 41, 42 and 43 enumerate the parameters used for creating the 80 $\times$ 80 grid structures for Rio de Janeiro, São Paulo and Salvador respectively, and the geographic characteristics for each city grid data structure. The territory dimensions were calculated using the grid data structure map by multiplying the cell area by the number of cells within the territory, and values from the *Brazilian Institute of Geography and Statistics* - IBGE[1] are provided to verify the calculated dimensions.

Table 41 – Parameters used for creating the 80 $\times$ 80 grid structure in Rio de Janeiro and resulting geographic dimensions including, the territory dimension calculated using the grid structure compared with IBGE values.

Initial Coordinates	Latitude	Longitude
Lower Left Corner	-23.091585650000013	-43.80890349999994
Upper Right Corner	-22.73487665000009	-43.100445999999764
	Cells in Territory	Used Cells
Number of Cells	2,673	1,855
	Width	Height
Cell Dimension	907.52 m	496.36 m
Total Grid Dimension	72.60 km	39.71 km
	Cell Area	Total Grid Area
Grid Areas	450,456.63 m^2	2,882.95 km^2
	Grid Map	IBGE
Territory Dimensions	1,204.07 km^2	1,200.26 km^2

[1] `https://www.ibge.gov.br/cidades-e-estados/`

Table 42 – Parameters used for creating the 80 $\times$ 80 grid structure in São Paulo and resulting geographic dimensions, including the territory dimension calculated using the grid structure compared with IBGE values.

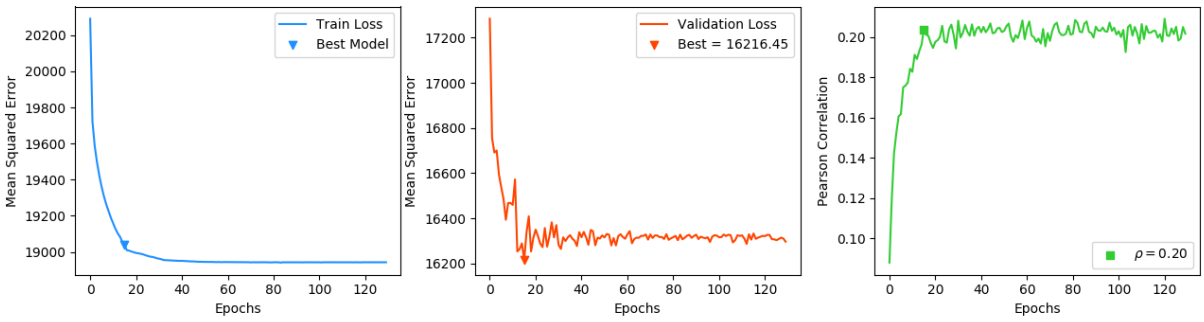
Initial Coordinates	Latitude	Longitude
Lower Left Corner	-24.021281	-46.8460877
Upper Right Corner	-23.319459	-46.3517027
	Cells in Territory	Used Cells
Number of Cells	2,495	1,718
	Width	Height
Cell Dimension	630.48 m	976.58 m
Total Grid Dimension	50.44 km	78.13 km
	Cell Area	Total Grid Area
Grid Areas	615,714.16 m^2	3,940.88 km^2
	Grid Map	IBGE
Territory Dimensions	1,536.21 km^2	1.521,11 km^2

Table 43 – Parameters used for creating the 80 $\times$ 80 grid structure in Salvador and resulting geographic dimensions, including the territory dimension calculated using the grid structure, not including maritime areas, compared with IBGE values, with maritime areas.

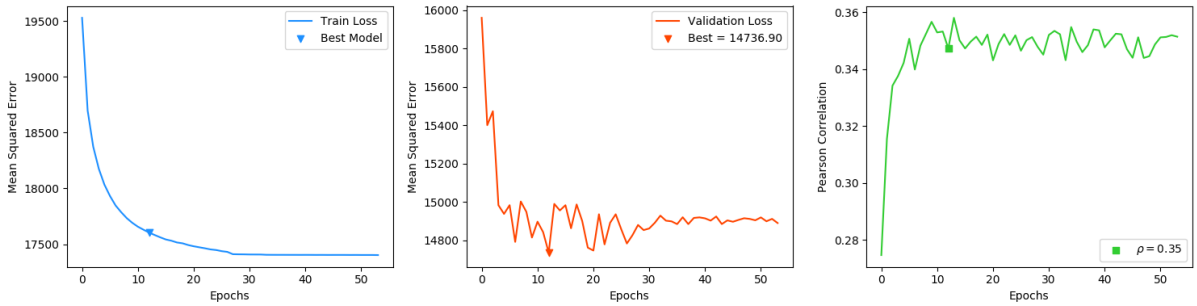
Initial Coordinates	Latitude	Longitude
Lower Left Corner	-13.053447	-38.704311
Upper Right Corner	-12.715086	-38.237315
	Cells in Territory	Used Cells
Number of Cells	1,057	834
	Width	Height
Cell Dimension	633.28 m	470.83 m
Total Grid Dimension	50.66 km	37.67 km
	Cell Area	Total Grid Area
Grid Areas	298,167.22 m^2	1,908.36 km^2
	Grid Map	IBGE
Territory Dimensions	315.16 km^2	693.83 km^2

Appendix B TRAIN AND VALIDATION LOSS CURVES

B.1 Baseline Models



(a) *HOG descriptor*



(b) *GIST descriptor*

Figure 48 – *Aerial-HOG/GIST-LR* model trained and validated in Rio de Janeiro. Pearson *p-value* ≤ 0.001.

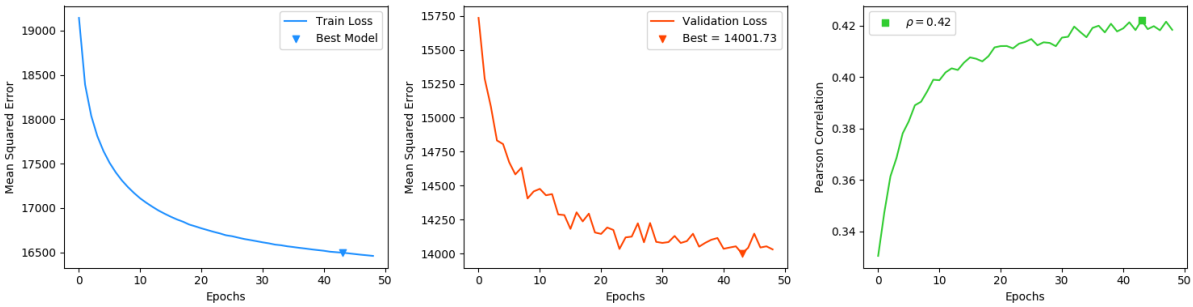
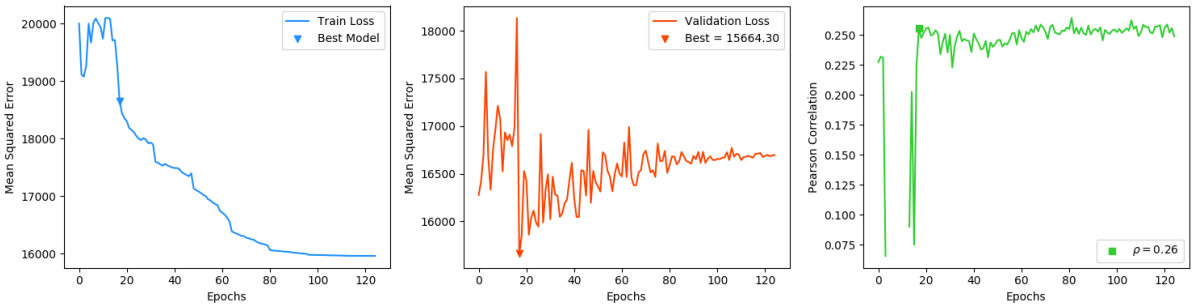
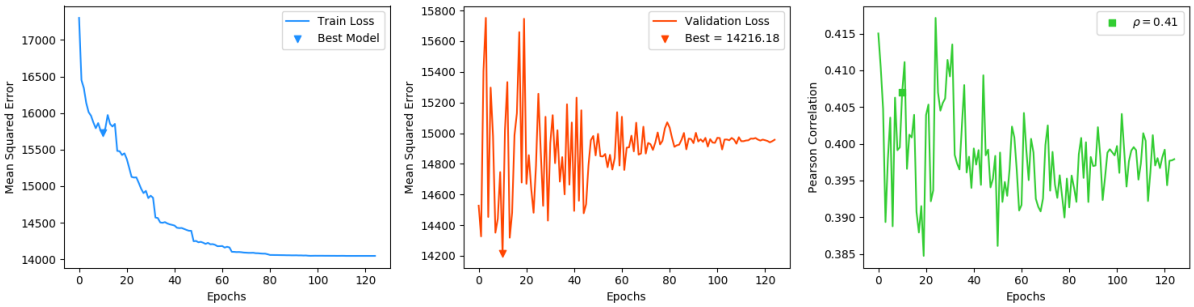


Figure 49 – *Street HOG-LR* train and validation loss for curves in Rio de Janeiro. All Pearson *p-value* < 0.001.



(a) *Aerial HOG-NN*

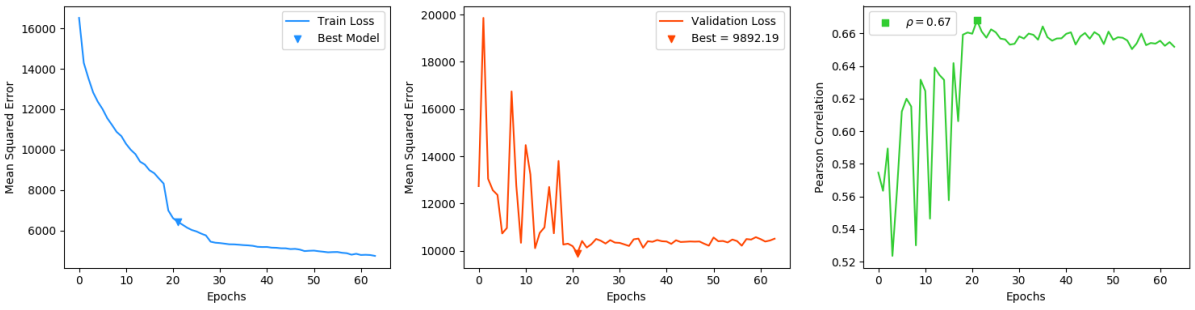


(b) *Street HOG-NN*

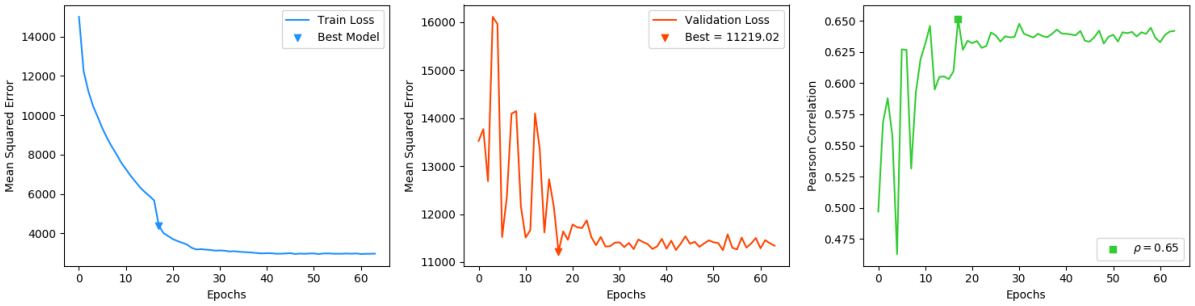
Figure 50 – *Aerial HOG-NN* and *Street HOG-NN* neural networks train and validation loss for Rio de Janeiro. All Pearson correlation with *p-value* < 0.001.

B.2 Deep ConvNet Models

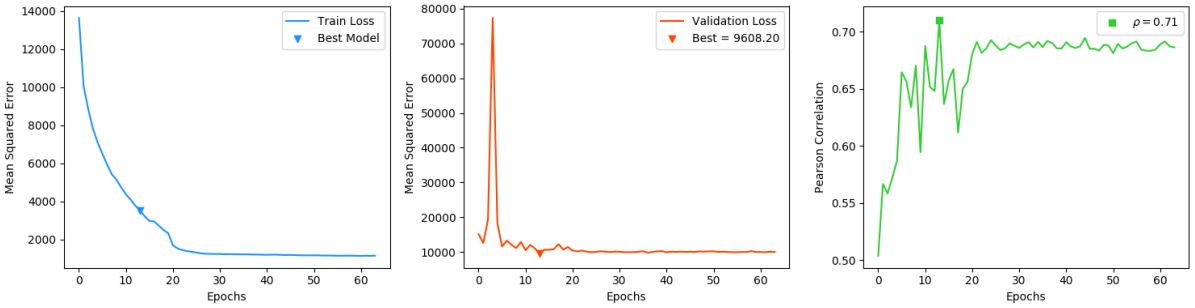
B.2.1 *Sat* models in Rio de Janeiro



(a) *Sat21*



(b) *Sat20*



(c) *Sat19*

Figure 51 – Train and validation loss curves for *Sat* models with $zoom = \{21, 20, 19\}$ in Rio de Janeiro. Pearson correlations in best validation epoch with *p-value* < 0.001.

B.2.2 *Street* Model in Rio de Janeiro

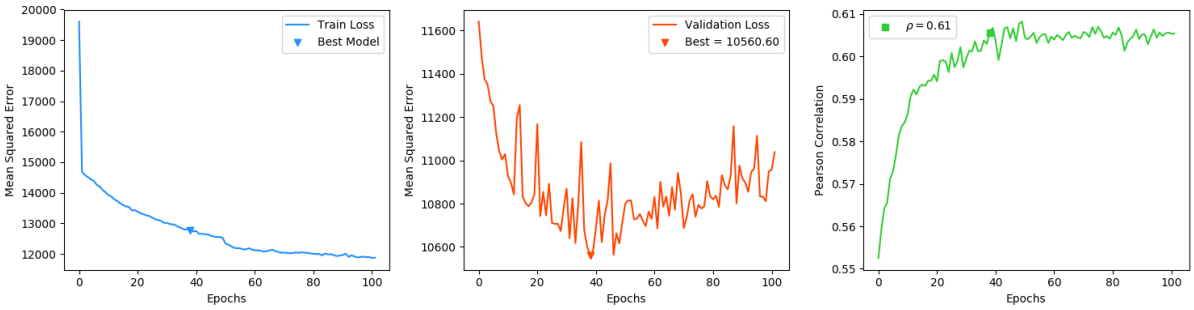
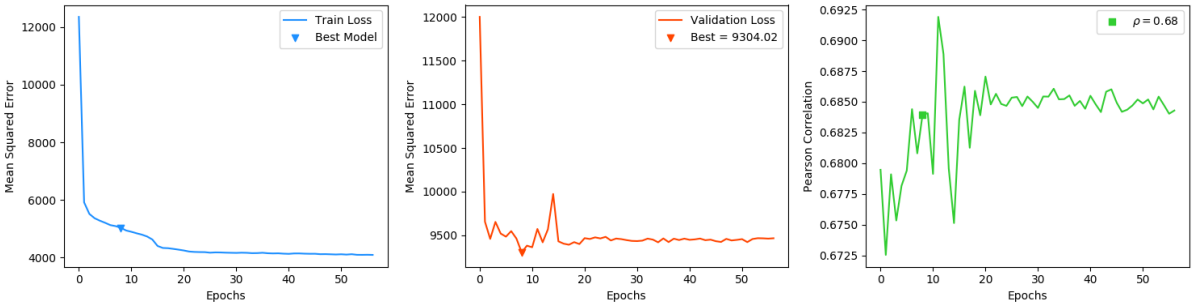
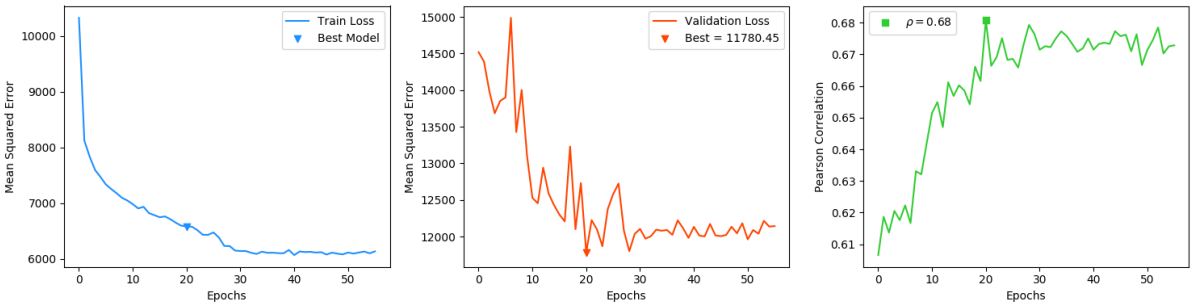


Figure 52 – Train and Validation loss for *Street* model in Rio de Janeiro. Pearson correlation in best validation epoch with *p-value* < 0.001.

B.2.3 *StreetSat* Models in Rio de Janeiro



(a) *StreetSat21*



(b) *StreetSat19*

Figure 53 – Train and validation loss curves for *StreetSat* models with $zoom = \{21, 19\}$ in Rio de Janeiro. All Pearson correlations *p-value* < 0.001.

B.2.4 *Sat* Model in São Paulo

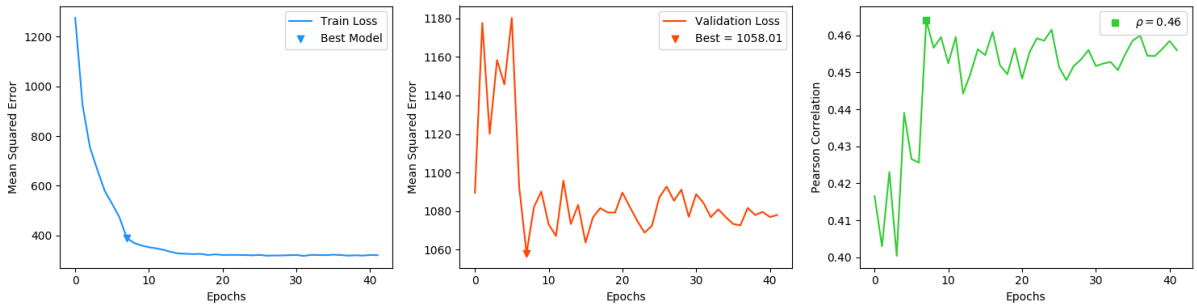


Figure 54 – *Sat* model train and validation loss curves in São Paulo, with $zoom = 19$. Pearson correlation in best validation epoch with *p-value* < 0.001.

B.2.5 *Street* Model in São Paulo

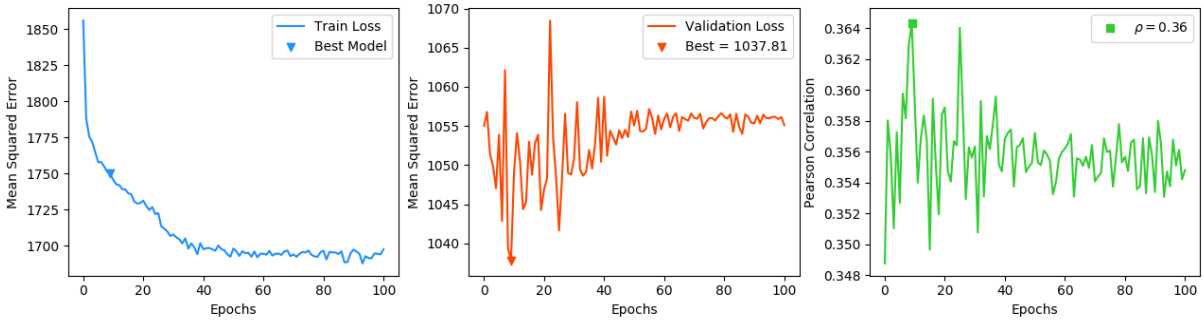


Figure 55 – *Street* model training process for São Paulo with loss curves. Pearson correlation in best validation epoch with *p-value* < 0.001.

B.2.6 *StreetSat* Model in São Paulo

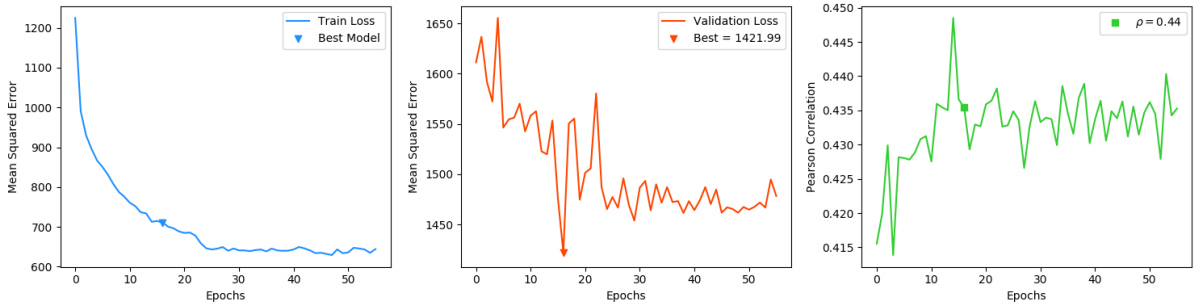


Figure 56 – *StreetSat* train and validation loss curves for São Paulo, using aerial features from *Sat19* and street-level features from *DenseNet161-Places*. Pearson correlation in best validation epoch with *p-value* < 0.001.

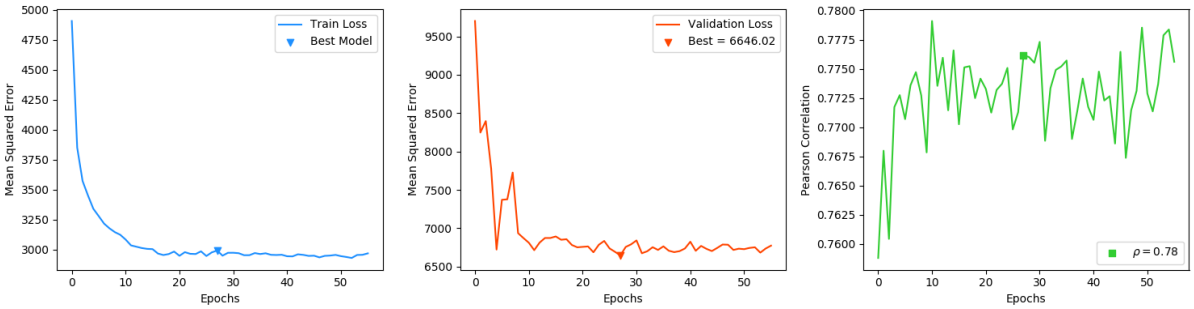
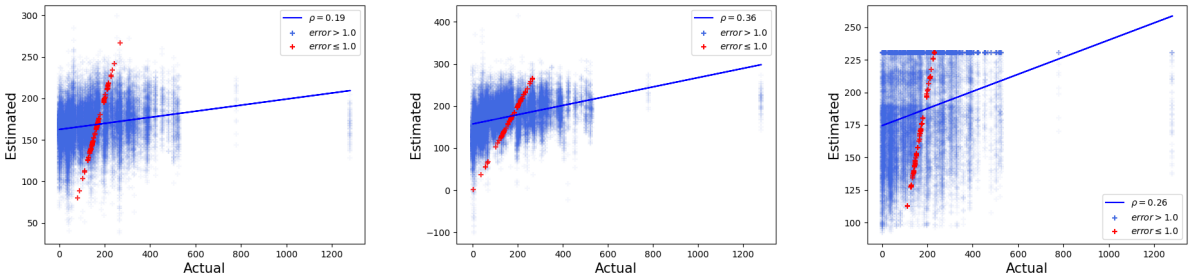


Figure 57 – *StreetSat* "MultiCity" model train and validation loss curves using aerial and street-level features from Rio de Janeiro and São Paulo. Pearson correlation in best validation epoch with *p-value* < 0.001.

Appendix C SUPPLEMENTARY RESULTS FROM BASELINE MODELS

C.1 Aerial imagery models

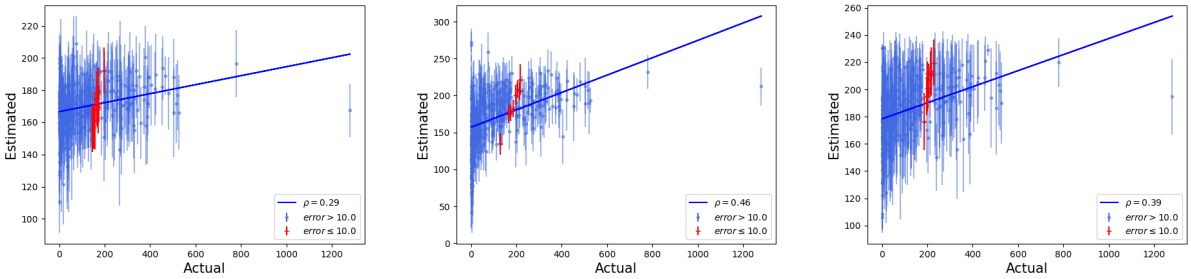
Figure 58 depicts the results obtained for the baseline models when estimating dengue real value rates from aerial images for location points in test data set, in actual versus estimated plots, for *Aerial HOG-LR*, *Aerial GIST-LR* and *Aerial HOG* neural network models, respectively. The plot shows locations points that obtained $MAE \leq 1.0$ and $MAE > 1.0$.



(a) *Aerial HOG-LR* (b) *Aerial GIST-LR* (c) *Aerial HOG-NN*

Figure 58 – Overall location points correlation results for baseline *Aerial HOG/GIST* models. All Pearson correlation ρ with $p - value < 0.001$.

Figure 59 depicts the results obtained when aggregating estimates at the cell level, in actual versus estimated plots, for *Aerial HOG-LR*, *Aerial GIST-LR* and *Aerial HOG-NN* models, respectively. The plot shows test cells that obtained $MAE \leq 10.0$ and $MAE > 10.0$, with the standard deviation for each cell depicted in error bars for each cell. Pearson correlation was computed between actual and estimated dengue cell values, and is depicted with the tendency line for all 352 test cells in Rio de Janeiro.

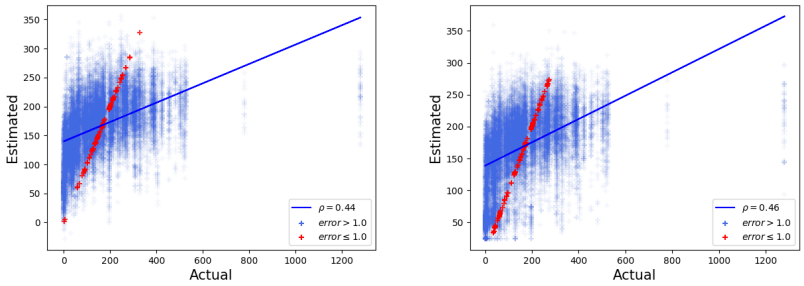


(a) *Aerial HOG-LR* (b) *Aerial GIST-LR* (c) *Aerial HOG-NN*

Figure 59 – Cell region estimation correlations for baseline *Aerial HOG/GIST* models. All Pearson correlation ρ with $p - value < 0.001$.

C.2 Street-level imagery models

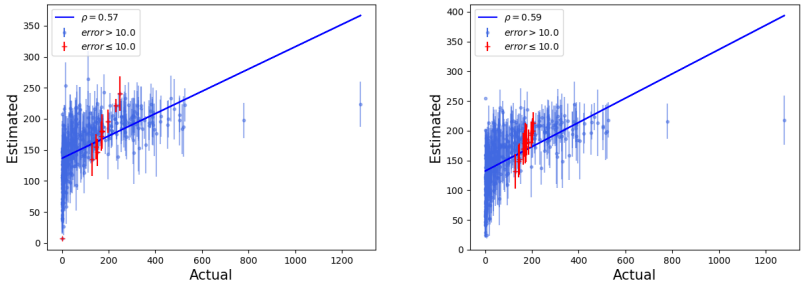
Figure 60 depicts the results obtained for the baseline models when estimating dengue real value rates from street-level images for location points in test data set, in actual versus estimated plots, for *Street HOG-LR* linear regression model and *Street HOG-NN* with a fully-connected neural network regressor. The plot shows locations points that obtained $MAE \leq 1.0$ and $MAE > 1.0$.



(a) *Street HOG-LR* (b) *Street HOG-NN*

Figure 60 – Overall location points correlation results for baseline *Street HOG* models. All Pearson correlation ρ with $p - value < 0.001$.

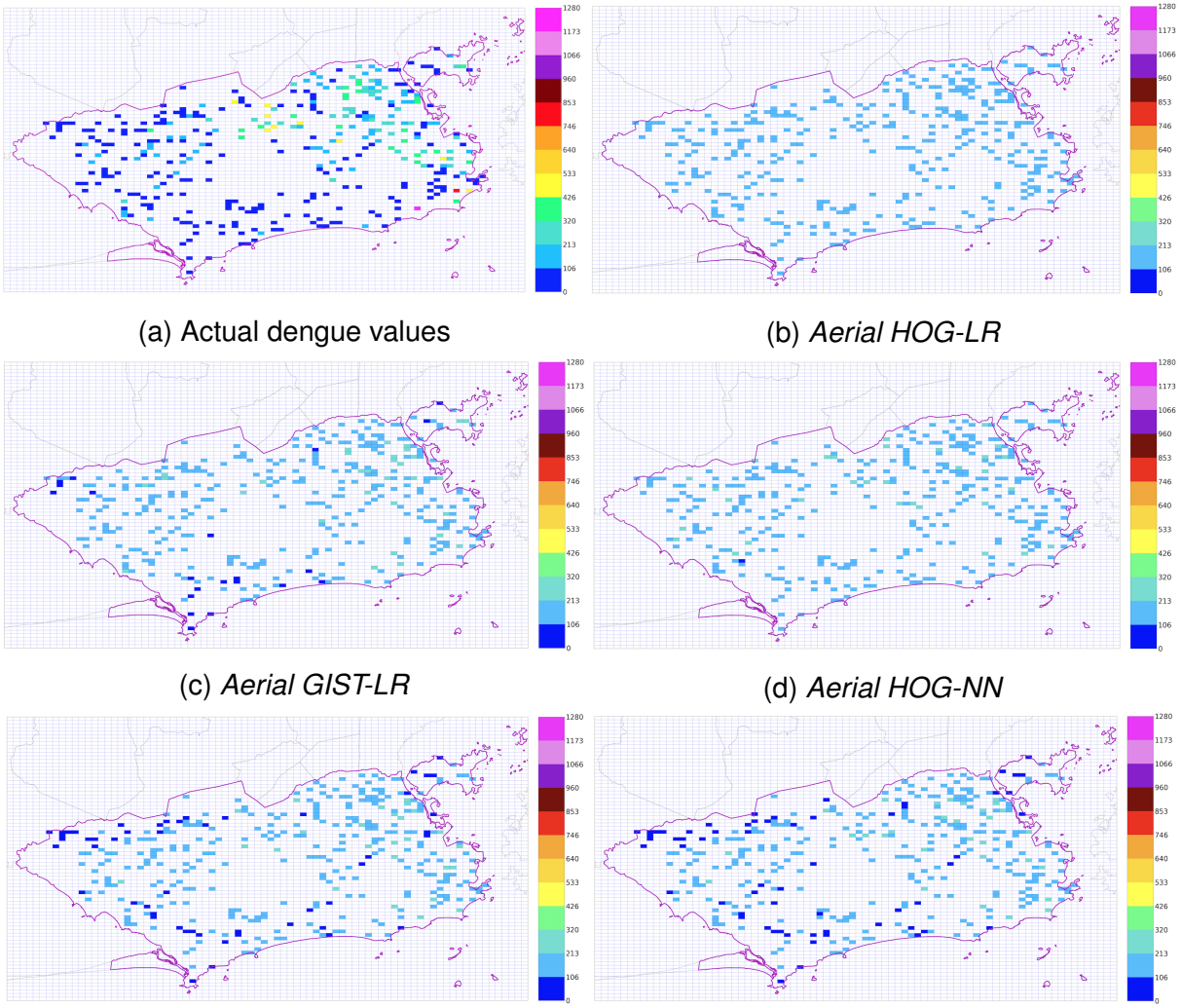
Figure 61 depicts the results obtained when aggregating estimates at the cell level, in actual versus estimated plots, for baseline *Street HOG-LR* and *Street GIST-LR* proposed models. The plots show test cells that obtained $MAE \leq 10.0$ and $MAE > 10.0$, with the standard deviation for each cell depicted in error bars for each cell. Pearson correlation was computed between actual and estimated dengue cell values, and is depicted with the tendency line for all 352 test cells in Rio de Janeiro.



(a) *Street HOG-LR* (b) *Street HOG-NN*

Figure 61 – Cell region estimation correlation results for baseline *Street HOG* models. All Pearson correlation ρ with $p - value < 0.001$.

C.3 Baseline Hot-Spots



(a) Actual dengue values (b) *Aerial HOG-LR*

(c) *Aerial GIST-LR* (d) *Aerial HOG-NN*

(e) *Street HOG-LR* (f) *Street HOG-NN*

Figure 62 – Cell region estimation for baseline aerial and street-level proposed models. Actual dengue values presented in (a) for comparison.

Figure 62 illustrate the hot-spots maps of dengue estimations for cell region made by the baseline proposed models. Actual dengue values are depicted in 62a for comparison with aerial and street-level models. Figure 62b depicts the estimations for cell region with *Aerial HOG-LR*, with minimum estimation equal to 110.44 and maximum prediction equal to 209.03 dengue incidence rate. In *Aerial GIST-LR* (Figure 62c), minimum dengue estimated value equal 42.07 and maximum dengue estimated value equal 272.27.

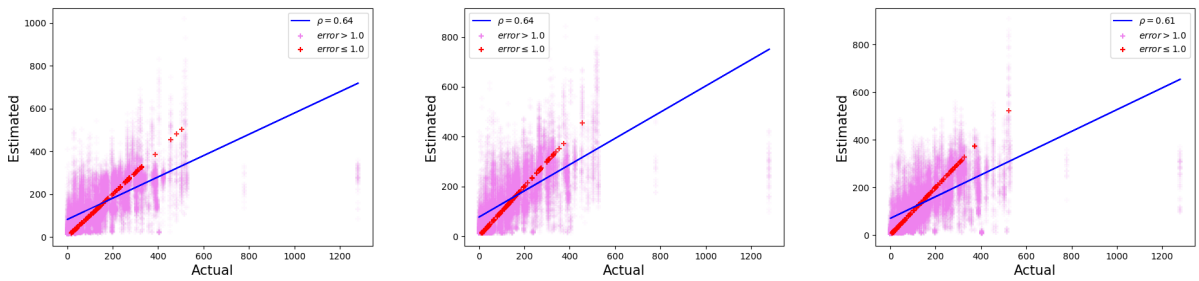
Figure 62d presents the estimations using model *Aerial HOG-NN*, with minimum dengue value estimated equal 105.86 and maximum equal 230.29. Cell estimation maps with baseline street-level models are presented in Figures 62e and 62f, depicting ranges from minimum dengue estimated value equal 6.9 and maximum equal 264.04 for *Street HOG-LR* and a minimum dengue estimated value equal 23.87 and maximum equal 254.43 for *Street HOG-NN*.

Appendix D SUPPLEMENTARY RESULTS FROM DEEP CONVNET MODELS

D.1 Models for Rio de Janeiro

Figure 63 depicts the results obtained from *Sat* model when estimating dengue rates for location points in Rio de Janeiro test data set, with actual versus estimated plots, for different zoom levels $zoom = 21, 20, 19$ of aerial imagery. The plots highlight locations points that obtained $MAE \leq 1.0$ and $MAE > 1.0$. Figure 64 depicts the results obtained when aggregating estimates at the cell levels. The plots highlight test cells that obtained $MAE \leq 10.0$ and $MAE > 10.0$, with the standard deviation for each cell depicted in error bars for each cell.

Figure 65a depicts the results obtained for the *Street* proposed model in Rio de Janeiro (*RioStreet*), when estimating dengue real value rates for location points in test data set, in actual versus estimated points. In overall results for locations points, the plot shows test examples that obtained the estimations absolute error ≤ 1.0 (highlighted) and other points that presented error > 1.0. Figure 65b depicts the results obtained for the *Street* proposed model when aggregating estimations of dengue rates per cells, in actual versus estimated points, for Rio de Janeiro (*RioStreet*). The plot



(a) *RioSat21*, $zoom = 21$ (b) *RioSat20*, $zoom = 20$ (c) *RioSat19*, $zoom = 19$

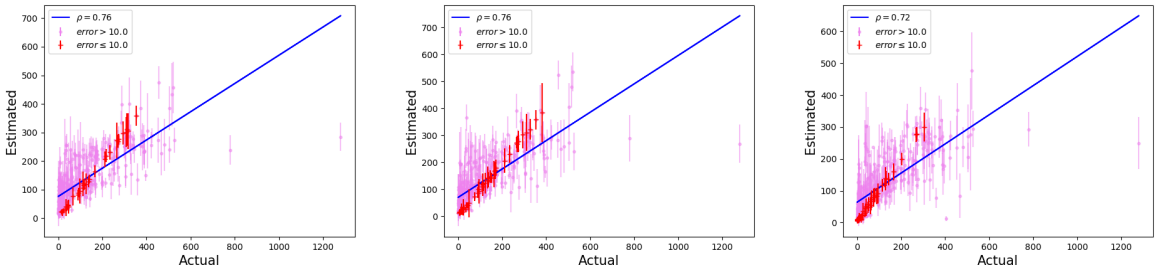
Figure 63 – *Sat* models overall individual correlation between actual and estimated values, with different zoom levels, for Rio de Janeiro location points test set. All Pearson correlation ρ with $p - value < 0.001$.

highlights test cells that obtained aggregated estimations with absolute error of ≤ 10.0. Pearson correlation was computed between actual cell values and aggregated estimated dengue values, and is depicted with the tendency line.

D.1.1 Hot-Spot Maps

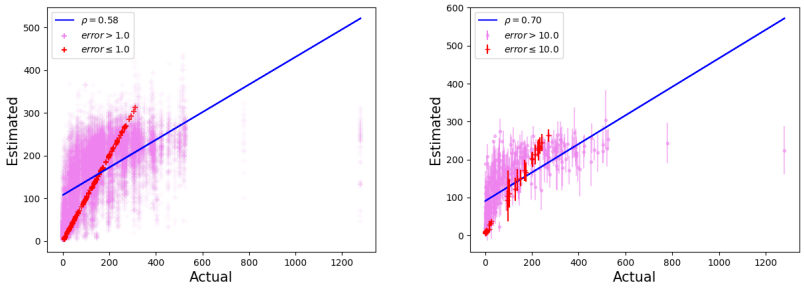
Figure 66 illustrate the *hot-spots* maps of dengue estimations for cell regions, resulting from the *Sat* proposed models with different zoom levels for Rio de Janeiro. Actual values of dengue are depicted in 66a for comparison. Figure 66b depicts the estimations for cell region with *RioSat21*, with minimum estimation equal to 15.26 and maximum prediction equal to 475.13 dengue incidence rate. In *RioSat20* (Figure 66b), minimum dengue estimated value equal 12.24 and maximum dengue estimated value equal 535.72. Figure 66d presents the estimations using model *RioSat19*, with minimum dengue value estimated equal 7.14 and maximum equal 475.92.

Figure 67 illustrate the *hot-spots* maps of dengue estimations for cell regions, resulting from the *Street* proposed models trained and evaluated in Rio de Janeiro. Actual



(a) *RioSat21*, $zoom = 21$ (b) *RioSat20*, $zoom = 20$ (c) *RioSat19*, $zoom = 19$

Figure 64 – *Sat* models correlation between actual and estimated values for each test cell in Rio de Janeiro test set, calculated using the mean average between all test points inside a cell, for different zoom levels. All Pearson correlation ρ with $p - value < 0.001$.



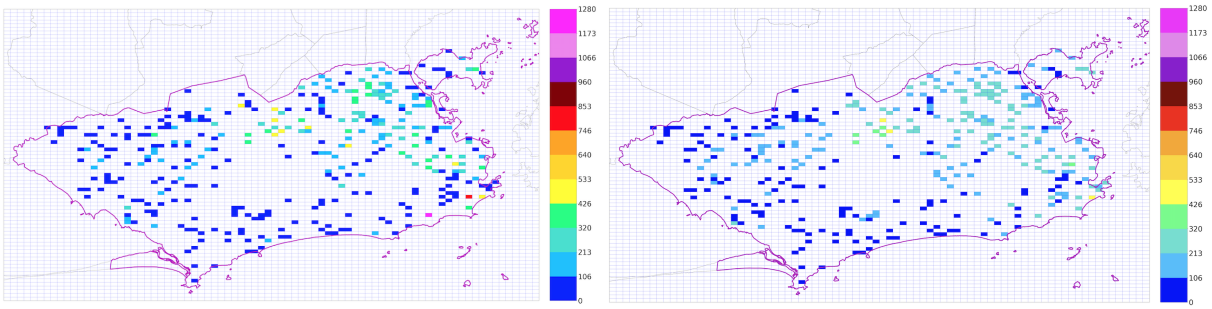
(a) *Overall Locations* (b) *Cell Region*

Figure 65 – (a) Overall correlation, from results of *RioStreet* model, between actual and estimated values for each location point. (b) Cell region estimation correlation, from results of *RioStreet* model, between actual and estimated values for each location point. All Pearson correlation ρ with $p - value < 0.001$.

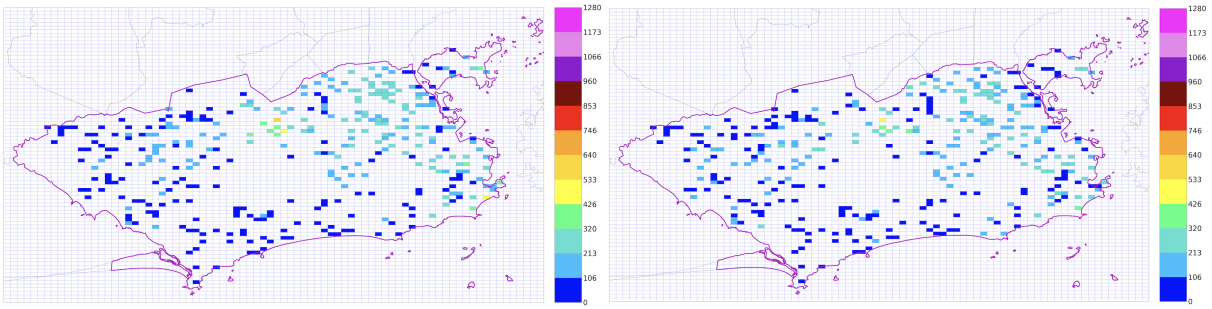
values of dengue are depicted in 67a for comparison. Figure 67b depicts the estimations for cell region with *RioStreet*, with minimum estimation equal to 5.27 and maximum prediction equal to 304.81 dengue incidences.

D.2 Models for São Paulo

Figure 68 depicts the results obtained from *Street* model trained and tested with São Paulo dengue fever data, when estimating rates for location points in test data set, and cell region estimation, in Figure 68b, with actual versus estimated plots. The plots highlight locations points that obtained the estimated absolute error of ≤ 1.0 and ≤ 10.0

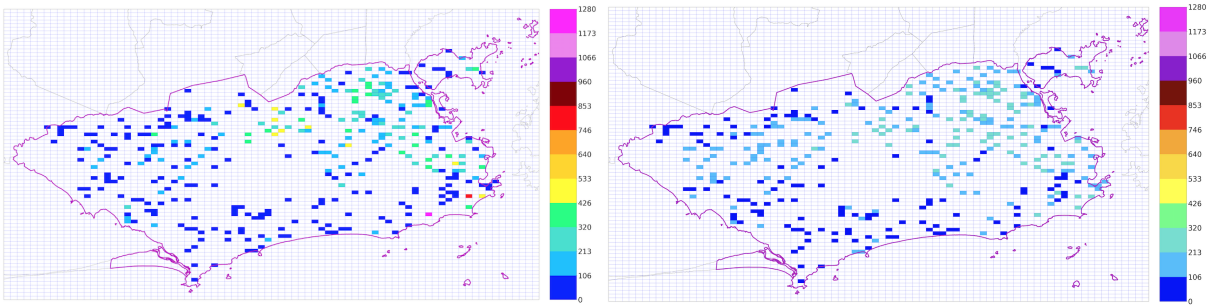


(a) Actual dengue rates (b) *RioSat21*



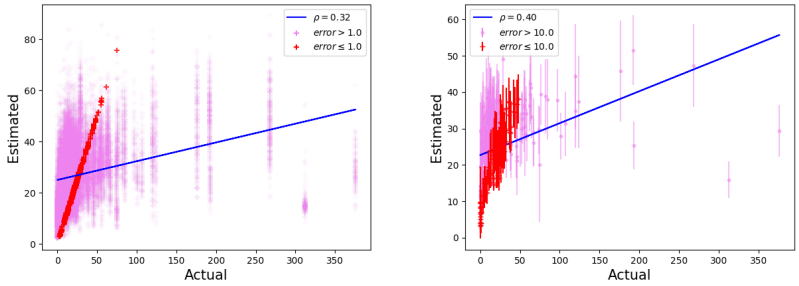
(c) *RioSat20* (d) *RioSat19*

Figure 66 – Cell region estimation from *Sat* models trained and tested over Rio de Janeiro: (a) actual dengue values for each test cell, and (b)-(d) the resulting maps obtained with different zooms levels.



(a) Actual (b) *RioStreet*

Figure 67 – Cell region estimation with *Street* model for Rio de Janeiro test set cells (a) actual dengue values and (b) predicted.



(a) *Overall Locations* (b) *Cell Region*

Figure 68 – (a) Overall correlation, from results of São Paulo models, between actual and estimated values for each location point. (b) Cell region estimation correlation, from results of São Paulo models, between actual and estimated values for each location point. Pearson correlation with *p-value* < 0.001.

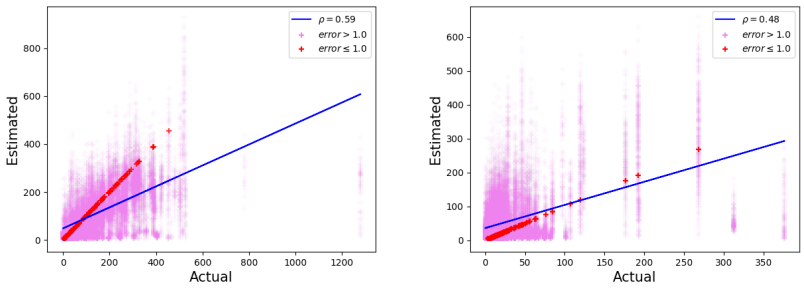
for cell estimation. Pearson correlation was computed between actual and estimated dengue values, and is depicted with the tendency line.

D.2.1 Hot-Spot Maps

Figure 70 presents the *hot-spots* maps of dengue fever estimations for cell region for all models trained and evaluated in São Paulo *Street* model. Actual values of dengue are depicted in 70a for comparison.

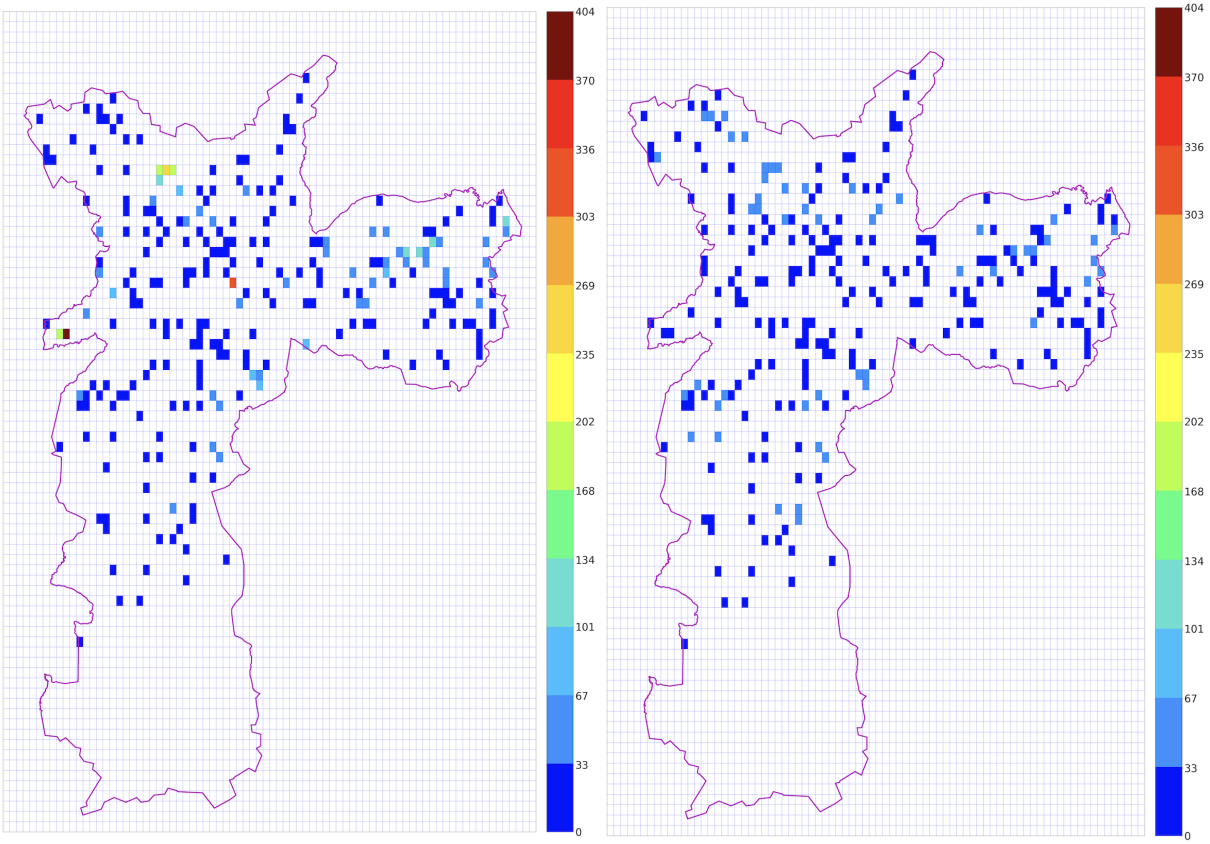
D.3 Multiple Cities Model

Figure 69 and 71 depicts the results obtained from *StreetSat* model trained and tested with Rio de Janeiro and São Paulo dengue fever data (*MultiCity-StreetSat19*), when estimating rates for location points and cell region estimations in test data set, with actual versus estimated plots. The plots highlight locations points that obtained the



(a) *MultiCity-StreetSat19* results in Rio de Janeiro. (b) *MultiCity-StreetSat19* results in São Paulo.

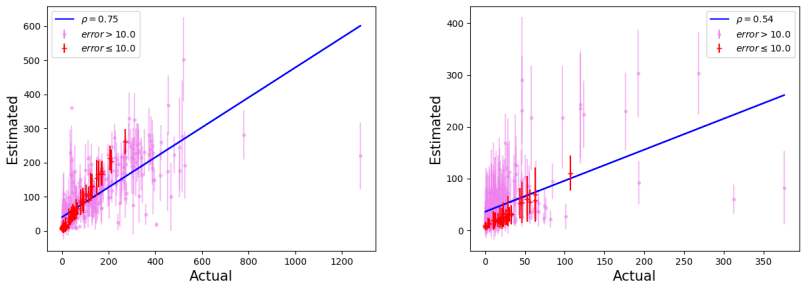
Figure 69 – Overall correlation from results of *MultiCity-StreetSat19*, between actual and estimated values for each location point in Rio de Janeiro and São Paulo. Pearson correlation with *p-value* < 0.001.



(a) Actual dengue rates (b) *SpStreet*

Figure 70 – Cell estimation maps with São Paulo *Street* model (a) actual values for dengue fever rates and (b) resulting estimated dengue values for test set cells.

estimated absolute error of ≤ 1.0 and cell estimation regions with absolute error ≤ 10.0. Pearson correlation was computed between actual and estimated dengue values, and is depicted with the tendency line.

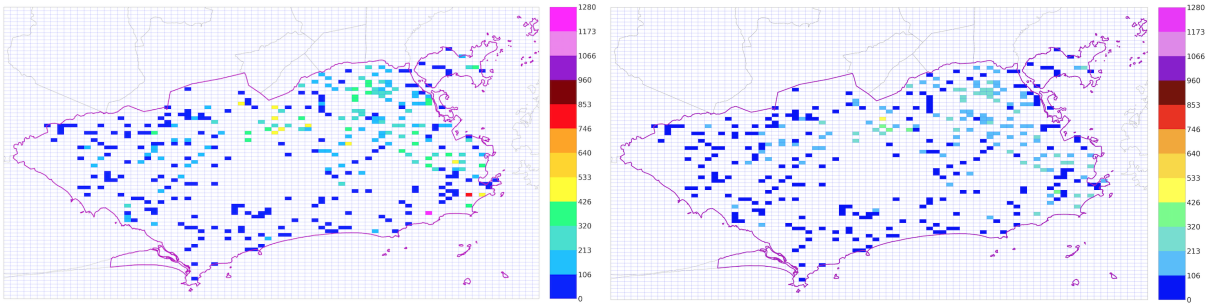


(a) *MultiCity-StreetSat19* in Rio de Janeiro. (b) *MultiCity-StreetSat19* in São Paulo.

Figure 71 – Overall correlation from results of *MultiCity-StreetSat19*, between actual and estimated values for each location point in Rio de Janeiro and São Paulo. Pearson correlation with *p-value* < 0.001.

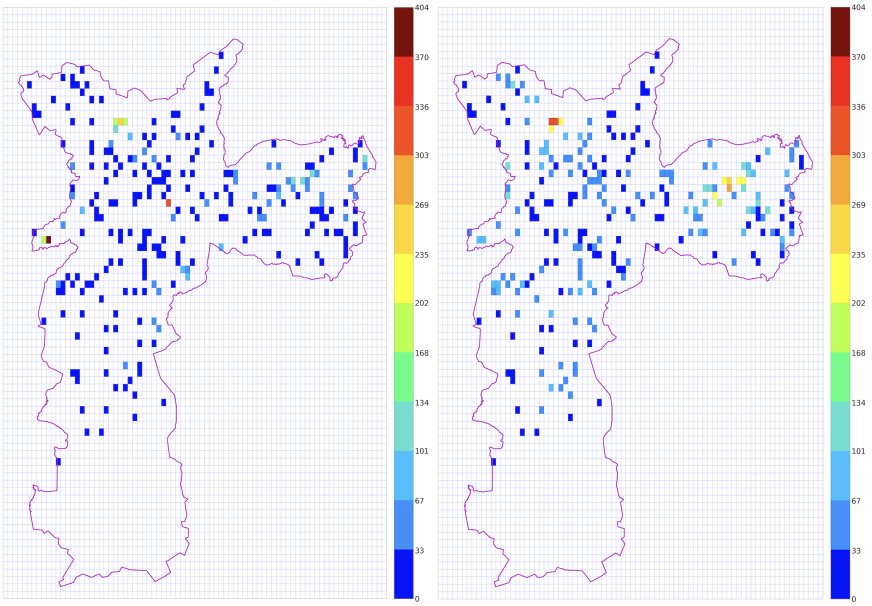
D.3.1 Hot-Spot Maps

Figures 72 and 73 presents the *hot-spots* maps of dengue fever estimations for cell region for *MultiCity-StreetSat19* model evaluated in Rio de Janeiro and São Paulo. Actual values of dengue for Rio de Janeiro are depicted in Figure 72a, and for São Paulo in Figure 73a for comparisons. Figure 72b depicts the estimations for cell region with *MultiCity-StreetSat19* model in Rio de Janeiro, with minimum estimation equal to 4.7 and maximum prediction equal to 503.56 dengue incidence rate. In *MultiCity-StreetSat19* model in São Paulo (Figure 73b), minimum dengue estimated value equal 5.49 and maximum dengue estimated value equal 303.47.



(a) Actual dengue in Rio de Janeiro (b) *MultiCity-StretSat19*

Figure 72 – Cell estimation maps with MultiCity model in Rio de Janeiro (a) actual values for dengue fever rates and (b) resulting estimated dengue values for test set cells. Actual dengue rates ranges from 0 to $1,280$, and estimated values are within the actual range, reaching at maximum shades of yellow in the color scale.



(a) *Actual dengue in São Paulo* (b) *MultiCity-StretSat19*

Figure 73 – Cell estimation maps with MultiCity model in São Paulo (a) actual values for dengue fever rates and (b) resulting estimated dengue values for test set cells. Actual dengue rates ranges from 0 to 404, and estimated values are within the actual range, reaching at maximum shades of red in the color scale.

Appendix E SOURCE CODES

The developed grid data structure API, named *GridMapAPI*, to map street location points to urban latent variables, including routines for train, validation and test datasets distribution, is available at:

- `https://github.com/Vortander/GridMapAPI`

Baseline and Deep Convolutional Neural Networks architectures proposed in this work, including dataloader classes for different set of images and features, are available at:

- `https://github.com/Vortander/VisualRegressionAPI`

Datasets and trained models in Rio de Janeiro and São Paulo for dengue incidence estimation will be avaliable at:

- `https://github.com/Vortander/DeepDengue`