



Cassio Freitas Pereira de Almeida

**Mapeamento da Distribuição Populacional
Através da Detecção de Áreas Edificadas em
Imagens de Regiões Heterogêneas do Google
Earth Usando Deep Learning**

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática da PUC-Rio.

Orientador: Prof. Hélio Côrtes Vieira Lopes

Rio de Janeiro
Outubro de 2017



Cassio Freitas Pereira de Almeida

**Mapeamento da Distribuição Populacional
Através da Detecção de Áreas Edificadas em
Imagens de Regiões Heterogêneas do Google
Earth Usando Deep Learning**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Hélio Côrtes Vieira Lopes

Orientador

Departamento de Informática – PUC-Rio

Prof. Simone Diniz Junqueira Barbosa

Departamento de Informática – PUC-Rio

Prof. Bruno Feijó

Departamento de Informática – PUC-Rio

Prof. Alex Laier Bordignon

Departamento de Geometria – UFF

Prof. Esteban Walter Gonzalez Clua

Instituto de Computação – UFF

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 26 de Outubro de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Cassio Freitas Pereira de Almeida

Bacharel em Estatística pela UFPR. Fez mestrado em Estatística no Departamento de Estatística de UFRJ. Atualmente é Pesquisador em Informações Geográficas na Escola Nacional de Ciências Estatísticas - ENCE/IBGE.

Ficha Catalográfica

Almeida, Cassio Freitas Pereira de

Mapeamento da Distribuição Populacional Através da Detecção de Áreas Edificadas em Imagens de Regiões Heterogêneas do Google Earth Usando Deep Learning / Cassio Freitas Pereira de Almeida; orientador: Hélio Côrtes Vieira Lopes. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2017.

v., 72 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Informática – Teses. 2. Rede Neural Convolucional. 3. Segmentação de Imagens. 4. Detecção de Objetos. 5. Aprendizado de Máquina. 6. Classificação de Imagens. 7. Dasimetria.
I. Lopes, Hélio Côrtes Vieira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Nestes anos dedicados ao doutorado muitas pessoas foram importantes e de alguma forma contribuíram para a realização deste trabalho.

Agradeço especialmente ao meu orientador Professor Hélio Lopes, pela disponibilidade, atenção dispensada, paciência, dedicação e profissionalismo.

À minha família, em especial à Sonia, Sara e Julia, pela compreensão e incentivo, durante todo este período.

Aos professores Bruno Feijó e Denise Britz, pelo apoio e orientações que foram fundamentais para a realização deste projeto.

À amiga Sandra Canton, pela amizade e incentivo.

À Professora Simone Barbosa, pelas orientações e por ter aberto as portas do Laboratório Ideias, que junto com o Professor Hélio Lopes, possibilitou a criação de um grupo de trabalho de altíssimo nível em um ambiente agradável e estimulante para trabalhar.

Neste grupo de trabalho formidável, agradeço a colaboração dos colegas Jefry Sastre, Ariane Bueno, Guilherme Schardong, Luiz Shirmer, Jonatas Grosman, Pedro Thompson e Felipe Baldino.

Em particular, agradeço a Sonia Fiol pela amizade, apoio e colaboração e ao William Ducca pelas discussões e colaboração intensa.

Por fim, agradeço ao IBGE pela oportunidade proporcionada e à PUC-Rio pela bolsa de isenção.

Resumo

Almeida, Cassio Freitas Pereira de; Lopes, Hélio Côrtes Vieira.
Mapeamento da Distribuição Populacional Através da Detecção de Áreas Edificadas em Imagens de Regiões Heterogêneas do Google Earth Usando Deep Learning.
Rio de Janeiro, 2017. 72p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Informações precisas sobre a distribuição da população são reconhecidamente importantes. A fonte de informação mais completa sobre a população é o censo, cujos os dados são disponibilizados de forma agregada em setores censitários. Esses setores são unidades operacionais de tamanho e formas irregulares, que dificulta a análise espacial dos dados associados. Assim, a mudança de setores censitários para um conjunto de células regulares com estimativas adequadas facilitaria a análise. Uma metodologia a ser utilizada para essa mudança poderia ser baseada na classificação de imagens de sensoriamento remoto para a identificação de domicílios, que é a base das pesquisas envolvendo a população. A detecção de áreas edificadas é uma tarefa complexa devido a grande variabilidade de características de construção e de imagens. Os métodos usuais são complexos e muito dependentes de especialistas. Os processos automáticos dependem de grandes bases de imagens para treinamento e são sensíveis à variação de qualidade de imagens e características das construções e de ambiente. Nesta tese propomos a utilização de um método automatizado para detecção de edificações em imagens Google Earth que mostrou bons resultados utilizando um conjunto de imagens relativamente pequeno e com grande variabilidade, superando as limitações dos processos existentes. Este resultado foi obtido com uma aplicação prática. Foi construído um conjunto de imagens com anotação de áreas construídas para 12 regiões do Brasil. Estas imagens, além de diferentes na qualidade, apresentam grande variabilidade nas características das edificações e no ambiente geográfico. Uma prova de conceito será feita na utilização da classificação de área construída nos métodos dasimétrico para a estimação de população em gride. Ela mostrou um resultado promissor quando comparado com o método usual, possibilitando a melhoria da qualidade das estimativas.

Palavras-chave

Rede Neural Convolutacional; Segmentação de Imagens; Detecção de Objetos; Aprendizado de Máquina; Classificação de Imagens; Dasimetria.

Abstract

Almeida, Cassio Freitas Pereira de; Lopes, Hélio Côrtes Vieira (Advisor). **Population Distribution Mapping Through the Detection of Building Areas in Google Earth Images of Heterogeneous Regions Using Deep Learning**. Rio de Janeiro, 2017. 72p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The importance of precise information about the population distribution is widely acknowledged. The census is considered the most reliable and complete source of this information, and its data are delivered in an aggregated form in sectors. These sectors are operational units with irregular shapes, which hinder the spatial analysis of the data. Thus, the transformation of sectors onto a regular grid would facilitate such analysis. A methodology to achieve this transformation could be based on remote sensing image classification to identify building where the population lives. The building detection is considered a complex task since there is a great variability of building characteristics and on the images quality themselves. The majority of methods are complex and very specialist dependent. The automatic methods require a large annotated dataset for training and they are sensitive to the image quality, to the building characteristics, and to the environment. In this thesis, we propose an automatic method for building detection based on a deep learning architecture that uses a relative small dataset with a large variability. The proposed method shows good results when compared to the state of the art. An annotated dataset has been built that covers 12 cities distributed in different regions of Brazil. Such images not only have different qualities, but also shows a large variability on the building characteristics and geographic environments. A very important application of this method is the use of the building area classification in the dasimetric methods for the population estimation into grid. The concept proof in this application showed a promising result when compared to the usual method allowing the improvement of the quality of the estimates.

Keywords

Convolutional Neural Network; Image Segmentation; Object Detection; Machine Learning; Image Classification; Dasimetry.

Sumário

1	Introdução	11
2	Detecção de Objetos	16
2.1	Métodos de comparação de modelos	16
2.2	Métodos baseados no conhecimento	17
2.3	Métodos de análise de objetos em imagem	17
2.4	Métodos de aprendizado de máquina	18
2.5	Extração de Atributos	19
3	Deep Learning em Imagens	21
3.1	Camada de convolução	22
3.2	Camada de não-linearidade	23
3.3	Camadas de Agrupamento	24
3.4	A Arquitetura U-Net	24
4	Trabalhos relacionados à detecção de edificações	27
5	Aquisição e Anotação das Imagens	30
5.1	Obtenção das Imagens	30
5.2	Anotação das Áreas Construídas	32
6	Resultados	37
6.1	Resultados com a U-Net	37
6.2	Comparação U-Net e <i>Patch-based Classification Model</i>	40
7	Aplicação em Dasimetria	48
8	Conclusão	56
	Referências bibliográficas	59

Lista de figuras

1.1	Imagens com delimitações de setores censitários, linhas amarelas, e delimitações do gride em vermelho. (<i>Imagens Google Earth</i>)	14
1.2	Imagens de varias regiões do Brasil. Grandes variação na qualidade de imagens e nas características geográficas e urbanísticas (<i>Imagens Google Earth</i>)	14
3.1	Rede Neural totalmente conectada com uma camada escondida	22
3.2	Convolução 2D [1]	23
3.3	Convolução 2D [2]	24
3.4	Exemplo de Agrupamento 2D - 2×2 . Redução da altura e largura do objeto mantendo a profundidade [1]	25
3.5	Exemplo de Agrupamento 2D, função $max()$, em uma única camada [1]	26
3.6	The U-net architecture.	26
5.1	Locais onde foram coletadas as imagens GE (pontos em vermelho).	31
5.2	Formação do mosaico com imagens <i>Google Earth</i> . Em vermelho a delimitação geográfica da região de interesse. Em azul, o gride georreferenciado.	32
5.3	Zonas metropolitanas. 5.3(a) região com edifícios, casas e grandes avenidas. Imagem nítida, com pouca aclusão e muitas sombras. 5.3(b) região com edifícios, casas, áreas verdes e rua estreitas. Imagem com menor nitidez, pouca aclusão e poucas sombras. (<i>Imagens Google Earth</i>).	33
5.4	Diversidade urbanística. 5.4(a) região com grande densidade de construções de pequena dimensão e estrutura urbanística não padronizada. 5.4(b) região com grande densidade de construções de grande porte. (<i>Imagens Google Earth</i>).	34
5.5	Cidades do interior. 5.5(a) imagem com alto brilho e pouca nitidez. Área residencial com variação nas cores de telhado, areas verdes com tonalidade variável. 5.5(b) imagem mais escura com boa nitidez. Área residencial com áreas verdes. (<i>Imagens do Google Earth</i>).	34
5.6	Cidades do interior. 5.6(a) adensamento de casa com muita área verde. 5.6(b) região árida com residências dispersas quase sem variação de cor. <i>Imagens Google Earth</i> .	35
5.7	Anotação de áreas construídas: 5.7(a)) Imagem com baixo adensamento. 5.7(b)) Imagem com as as construções anotadas (preto). <i>Imagens Google Earth</i> .	35
5.8	Anotação de áreas construídas. 5.8(a) Imagem original. 5.8(b) Imagem anotada, blocos de áreas construídas. 5.8(c) Imagem P & B da classificação de áreas construída. <i>Imagens Google Earth</i>	36
6.1	Distribuição das estatísticas de teste obtidos com a U-Net no conjunto de teste, 56 imagens.	38

6.2	Imagens de Teste. 6.2(a) Imagem RGB de referência. 6.2(b) Imagem da camada de predição sobre a imagem de referência. Em verde, <i>True Positives</i> , em amarelo <i>False Negatives</i> e em vermelho <i>False Positives</i> . <i>Imagens Google Earth</i> .	39
6.3	Imagens de Teste. 6.3(a) Imagem RGB de referência. 6.3(b) Imagem da camada de predição sobre a imagem de referência. Em verde, <i>True Positives</i> , em amarelo <i>False Negatives</i> e em vermelho <i>False Positives</i> . <i>Imagens Google Earth</i> .	40
6.4	Imagens de Teste. 6.4(a) Imagem RGB de referência. 6.4(b) Imagem da camada de predição sobre a imagem de referência. Em verde, <i>True Positives</i> , em amarelo <i>False Negatives</i> e em vermelho <i>False Positives</i> . <i>Imagens Google Earth</i> .	41
6.5	Imagens de Teste. 6.5(a) Imagem RGB de referência. 6.5(b) Imagem da camada de predição sobre a imagem de referência. Em verde, <i>True Positives</i> , em amarelo <i>False Negatives</i> e em vermelho <i>False Positives</i> . <i>Imagens Google Earth</i> .	42
6.6	Modelo <i>Patch-based</i> . Arquitetura CNN utilizada em [3].	43
6.7	Imagem RGB e imagens classificadas pelo <i>Patch-based</i> e U-Net em duas regiões: N.Mutum - MT e Trairi - CE. Em verde - (<i>True Positives</i>), em vermelho - <i>False Positives</i> e em amarelo - <i>False Negatives</i>	45
6.8	Imagem RGB e imagens classificadas pelo <i>Patch-based</i> e U-Net em duas regiões: RJ (B. Sucesso) e RJ (Tijuca). Em verde - (<i>True Positives</i>), em vermelho - <i>False Positives</i> e em amarelo - <i>False Negatives</i>	46
6.9	Imagem RGB e imagens classificadas pelo <i>Patch-based</i> e U-Net em duas regiões: Rio Bonito - RJ e RJ (Botafogo). Em verde - (<i>True Positives</i>), em vermelho - <i>False Positives</i> e em amarelo - <i>False Negatives</i>	47
7.1	Dispersão do arruamento e de áreas construídas. Áreas construídas anotadas em azul. Rua anotadas em preto. Imagens subdivididas em quatro regiões. <i>Imagens Google Earth</i> .	52
7.2	Dispersão das proporções obtidas nas imagens de teste pelo arruamento (Rua) e pela área de construções anotada (Mask)	53
7.3	Boxplot das diferenças entre: áreas construídas preditas (U-Net) e áreas construídas anotadas (Pred-Mask); proporções de arruamento e áreas construídas anotadas (Rua-Mask); proporções de arruamento e áreas construídas preditas com U-Net(Rua-Pred).	54
7.4	Divergências entre a proporção de vias e área construída. Áreas construídas anotadas(azul). Vias anotadas em preto. Imagens subdivididas em quatro regiões. <i>Imagens Google Earth</i> .	54
7.5	Área rural. Difícil identificação de vias públicas e vias particulares para mapeamento digital. Áreas construídas anotadas (azul). <i>Imagens Google Earth</i> .	55
8.1	Mapa do Índice de Desenvolvimento Humano Municipal da Cidade do Rio de Janeiro, fornecido pela ONG Atlas do Desenvolvimento Humano do Brasil (http://www.atlasbrasil.org.br/).	58

Lista de tabelas

6.1	Rede U-Net: Resultados do Treinamento, Validação e Teste	38
6.2	Resultados do treinamento e validação para os modelos <i>Patch-based</i> e U-Net	42
6.3	Resultados da predição dos modelos <i>Patch-based</i> e U-net. Os resultados correspondem a média e desvio padrão (<i>mean (standard deviation)</i>) sob o mesmo conjunto de imagens de Teste.	43
6.4	Resultados da predição dos modelos <i>Patch-based</i> e U-net. Os resultados correspondem a média em cada uma das 12 localidades nas imagens de Teste.	44
7.1	Média e desvio padrão das diferenças absolutas entre as proporções de arruamento, áreas construídas preditas(U-Net) e áreas construídas anotadas.	52

1

Introdução

Informações precisas sobre a distribuição da população são importantes para a gestão de problemas sociais, econômicos e ambientais relacionados à população. A fonte de informação mais completa sobre a população é o censo. Em muitos países, particularmente nos países em desenvolvimento, os censos são a cada dez anos e os dados são disponibilizados ao público de forma agregada como anuários estatísticos, geralmente divididos em áreas político-administrativas. Estes agrupamentos, os setores censitários, são unidades operacionais homogêneas de tamanho e formas irregulares, e tem o objetivo de facilitar o planejamento e o gerenciamento da operação censitária, mas não foram pensadas objetivando a divulgação dos dados. De um modo geral, as propriedades dos dados agregados baseados em setores são limitadas para a análise espacial de problemas socioeconômicos e ambientais relacionados com a população [4]. Assim, o gride de população surgiu como um formato alternativo para o mapeamento de dados populacionais. O gride de populações é uma rede geograficamente referenciada de células quadradas com cada célula carregando uma contagem de população ou o valor da densidade populacional em sua localização. A Figura 1.1 contém exemplos de setores censitários limitados pelas linhas em amarelo e de um gride cujas células estão marcadas em vermelho. A geração de um sistema de grades com objetivos estatísticos consiste em uma operação de mudança de suporte espacial de dados e têm as vantagens de que todas as células têm o mesmo tamanho, são estáveis ao longo do tempo e propicia uma maior facilidade de obter recortes espaciais mais adequados aos objetivos de estudo. Recentemente no Brasil, a Fundação Instituto Brasileiro de Geografia e Estatística - IBGE iniciou um projeto para a divulgação de dados estatísticos em gride que teve a primeira divulgação em 2016. O objetivo é converter os dados estatísticos do censo em gride. Na metodologia desenvolvida há a necessidade do emprego de métodos que utilizam informações correlacionadas à população provenientes da classificação de imagens de sensoriamento remoto. A localização dos domicílios é a informação mais relevante, mas considerando o custo para obtenção, as estimativas são feitas com informações menos eficientes, como o arruamento. Assim, métodos de baixo custo para geração de informações

de imagens de sensoriamento remoto podem melhorar o processo existente e inclusive possibilitar a criação de métodos mais eficientes. Neste caso, a detecção de áreas construídas traria um incremento na metodologia, pois estão diretamente relacionadas à identificação de domicílios, que é a base das pesquisas envolvendo a população.

As imagens de alta resolução (HSR) [5] de sensoriamento remoto estão cada vez mais acessíveis e são fontes de dados relevantes para análises em sistemas de informação geográfica (SIG). O desenvolvimento tecnológico permitiu a coleta e distribuição destas imagens em grande escala. Por exemplo, a empresa Planet Labs [6, 7] tem um grande número de mini satélites em órbita que enviam imagens diárias de quase toda a superfície terrestre. A popularização dos veículos aéreos não tripulados também permite coletar imagens aéreas de altíssima qualidade com baixo custo [8]. Por fim, empresas como DigitalGlobe e Google disponibilizam as recentes imagens com alta qualidade de praticamente toda a superfície da terra, isso sem custo para o usuário e com muita facilidade de acesso. Esta facilidade na obtenção de imagens RGB de alta qualidade tem gerado uma demanda de processos automatizados para extração de informações desta grande massa de imagens. Estas imagens possibilitam o mapeamento com um alto nível de detalhe, o que possibilita a criação de novas aplicações como a detecção de objetos. No entanto, estas novas aplicações exigem a transformação destes dados em informações tangíveis para que a quantificação ou análise possa ser feita conjuntamente com outras bases de dados nos Sistemas de Informação Geográfica (SIG) [9]. O núcleo dessa transformação é a interpretação de objetos dentro de cada imagem. Este processo de extração de objeto é demorado e dependente quase que exclusivamente da interpretação humana. Grandes esforços têm sido empregados no aprimoramento e desenvolvimento de novos métodos de classificação [10, 11].

A pesquisa em sensoriamento remoto focada na classificação de imagens vem atraindo a atenção da comunidade porque os resultados de classificações são a base de muitas aplicações socioeconômicas e ambientais.

A detecção de objetos [11] em imagens de sensoriamento remoto (RSI) geralmente refere-se à localização e delimitação de determinadas formas generalizadas, classificando-as como objetos de paisagem (ex. rios) ou objetos artificiais (ex. veículos; edificações). Este é o problema fundamental no campo da análise de imagens e desempenha um papel importante para varias aplicações, como monitoramento ambiental, detecção de perigo geológico, mapeamento de uso da terra, agricultura de precisão e planejamento urbano.

A tarefa de detecção de objetos é desafiadora por causa da complexidade

das imagens. Os principais passos da classificação de imagens incluem a determinação de um sistema de classificação adequado, seleção de amostras de treinamento, pré-processamento de imagem, extração de atributos (*features*), processamento, pós-classificação e avaliação da qualidade. As necessidades do usuário, a escala da área de estudo, as restrições econômicas e as habilidades dos analistas, são fatores importantes que influenciam a seleção de dados, a definição do procedimento de classificação e a qualidade dos resultados de classificação.

Em particular, a detecção de edificações é considerada uma tarefa difícil [12, 13, 14, 3, 15]. Com o aumento da resolução das imagens, as características internas dos objetos ficaram discerníveis. Assim, grandes variações podem ser observadas dentro da mesma categoria de objeto, conhecida como variação intra-classe. Por exemplo, as edificações aparecem com diferentes cores, formas, tamanhos e orientações, conforme mostrado na Figura 1.2. Outro fator relevante é a qualidade da imagem, como por exemplo a presença de sombras e oclusões que afetam as aparências de objetos. Nesses casos, as edificações podem ficar parcialmente encobertas pelas sombras de outras edificações ou, no caso da oclusão, ficam encobertas por outras edificações devido ao ângulo de obtenção da imagem. Quando se trata de aplicações em grande escala, há uma grande quantidade de imagens a serem processadas. Estas aplicações são de grande abrangência territorial, gerando grandes variações nas imagens devido às características geográficas. Vale citar que o tipo de solo e vegetação mudam de uma região para outra. Variam também as características urbanísticas, como o adensamento das construções, a forma e tamanho das quadras, as características do arruamento e a existência de áreas com finalidades diversas como parques e reservas. A qualidade da imagem é muitas vezes afetada devido às variações atmosféricas. Nestes casos, o método de detecção de objetos utilizado deve ser robusto o suficiente para não só ter uma boa performance na classificação com todas essas características, mas também ser eficiente em tempo de processamento para que seu uso seja viável.

Muitos avanços foram feitos na área de sensoriamento remoto a fim de aperfeiçoar os métodos para classificação de objetos e especificamente identificação de edificações. No entanto, estudos existentes geralmente são limitados a alguns conjuntos de dados públicos com cenas relativamente simples, o que facilita a comparabilidade entre as propostas, mas limita a diversificação de soluções para o mundo real. Não há relato de algum método que tenha funcionado de forma confiável para cenários complexos do mundo real em grande escala. Ainda, grande parte desses métodos é fortemente dependente da interação humana.

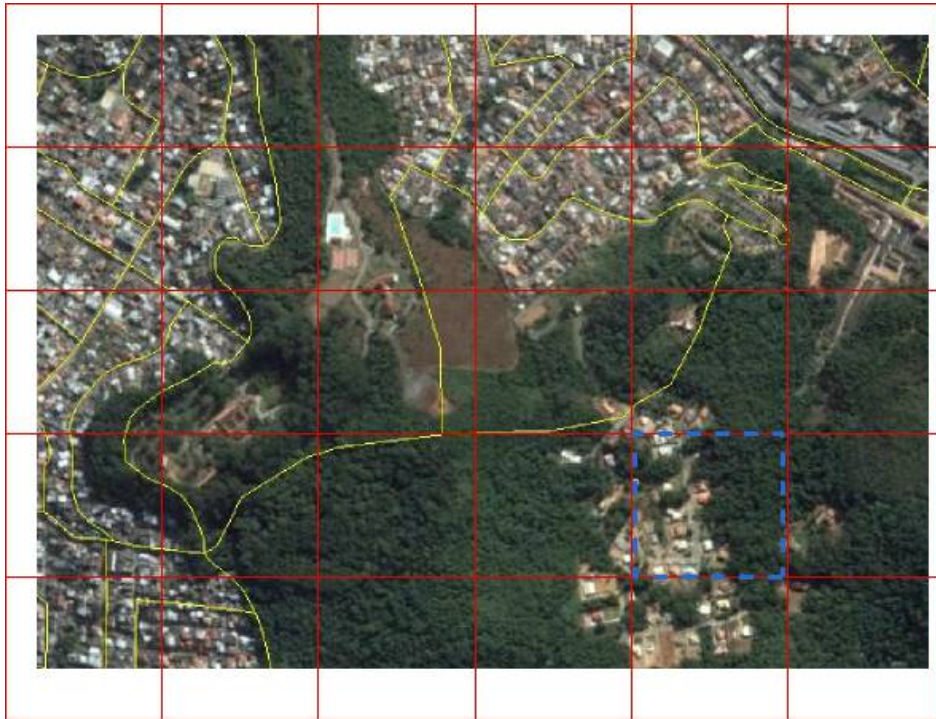


Figura 1.1: Imagens com delimitações de setores censitários, linhas amarelas, e delimitações do gride em vermelho. (*Imagens Google Earth*)



Figura 1.2: Imagens de varias regiões do Brasil. Grandes variação na qualidade de imagens e nas características geográficas e urbanísticas (*Imagens Google Earth*)

Os métodos de *Deep Learning*, um desenvolvimento recente na área de aprendizado de máquinas, trouxe novas possibilidades. Estes métodos possibilitam o aprendizado automático de atributos que aumentam a performance de classificação [16, 17]. Várias aplicações em diversas áreas mostraram melhoria de precisão significativas [18, 19].

Esta tese propõe um método para solucionar algumas destas dificuldades na detecção de edificações, possibilitando a aplicação prática em grande escala. As principais contribuições desta tese são os seguintes dois aspectos:

- 1- Utilização da área construída no Método Dasimétrico para melhoria da performance nas estimativas do Gride.
- 2- Criação de um único modelo com uma rede neural convolucional com boa performance na tarefa de detecção e segmentação automática de edificações a partir de um dataset pequeno de imagens RGB com grande variabilidade de qualidade, iluminação, oclusão e características geográficas e urbanísticas.
- 3- A boa performance com um conjunto de imagens relativamente pequeno torna possível, e com baixo custo, criar o próprio dataset na região de interesse e treinar inteiramente o modelo. O conjunto de imagens utilizado, composto por 126 imagens RGB (Google Earth) de 12 regiões do Brasil anotadas, está disponível em [20].

O restante deste trabalho está organizado da seguinte forma. O Capítulo 2 apresenta uma revisão dos trabalhos em detecção de objetos. O Capítulo 3 apresenta brevemente as duas arquiteturas de Deep Learning que foram utilizadas nesse trabalho. O Capítulo 4 apresenta uma discussão sobre os trabalhos relacionados que utilizam Deep Learning na tarefa de detecção de edificações. O Capítulo 5 apresenta o conjunto de dados. O Capítulo 6 descreve o procedimento de teste e os resultados. O Capítulo 7 apresenta uma prova de conceito do uso do método de detecção de edificações para a aplicação em dasimetria no cálculo da população em gride. Finalmente, o Capítulo 8 conclui esta tese fazendo algumas observações finais e sugerindo trabalhos futuros.

2

Detecção de Objetos

A detecção de objetos em imagens de alta resolução não visa somente determinar a existência de um objeto em uma imagem, mas também, localizá-lo na imagem. O termo "objeto" refere-se à forma generalizada e é dividido em objetos artificiais e naturais. Os objetos artificiais (ex. veículos, navios e edificações) têm limites bem definidos e não fazem parte do ambiente de fundo. Os objetos naturais (ex. paisagem, rios e parcelas de uso da terra (LULC)) possuem limites menos definidos e são parte do ambiente de fundo.

Com os avanços da tecnologia em sensoriamento remoto, as imagens têm fornecido informações espaciais mais detalhadas. Além das propriedades da região, uma maior variedade de objetos artificiais tornam-se reconhecíveis e até mesmo podem ser identificados separadamente devido ao aumento da resolução. Isso abriu novas perspectivas no campo da detecção automática de objetos geoespaciais e diversos métodos foram desenvolvidos. Estes métodos podem ser divididos em quatro categorias principais [11]: métodos de comparação de modelos (*Template matching-based*), métodos baseados no conhecimento (*Knowledge-based*), métodos de análise de objetos em imagens (*Object-Based Image Analysis - OBIA*), e métodos de aprendizado de máquina (*Machine Learning - ML*). Essas quatro categorias não são necessariamente independentes e às vezes o mesmo método integra diferentes categorias.

2.1

Métodos de comparação de modelos

Os métodos de comparação de modelos são os mais simples. Consistem na criação de um modelo (*template*) para um determinado objeto de interesse. Então é feita uma busca de compatibilidade deste *template* sobre a imagem, considerando uma medida de similaridade. Esta busca leva em consideração a rotação, translação e variações de escala do objeto. São duas as categorias de *templates*: o rígido e o deformável. O *template* rígido é mais adequado para objetos de aparência simples e com pouca variação, como rodovias em [21] e [22]. No caso de detecção de edificações, as aplicações de [23, 24, 25], mostraram que o método é preciso, mas muito sensível a variações de formato e densidade, o que limita as aplicações em RS. Os *templates*

deformáveis de forma livre representam um objeto de forma arbitrária com base em algumas regularizações gerais de continuidade ou aplicando funções paramétricas que representes os objetos. Os métodos mais populares são os modelos de contorno ativos, também conhecidos como modelos de *snake*, [26, 27]. Os *templates* deformáveis se mostraram mais poderosos e flexíveis, adaptando-se melhor às variações intra-classe [28, 29]. Contudo, estes métodos necessitam de informações e parâmetros a priori para a definição do *template*.

2.2

Métodos baseados no conhecimento

Os métodos de detecção de objetos baseados no conhecimento (*knowledge-based*) [30] usam parâmetros e regras para determinar objetos por meio de testes de hipóteses. Os métodos mais conhecidos usam os parâmetros geométricos da forma do objeto e o contexto da imagem como uma restrição espacial a presença do objeto. A dificuldade na aplicação destes métodos é devida à subjetividade na definição das informações a priori dependente do conhecimento de um analista.

2.3

Métodos de análise de objetos em imagem

Os métodos de análise de objetos em imagem (OBIA) para extração de objetos a partir de imagens HSR [31],[32] e [33] são realizados em duas etapas. Primeiro, os *pixels* de uma imagem são agrupados em regiões homogêneas com base em seus atributos de espectros e textura, formando as segmentações da imagem. Então, um classificador utiliza as características extraídas da região e classifica cada região com uma classe exclusiva referente a um objeto. Essas regiões segmentadas são consideradas objetos potenciais e, portanto, esse tipo de métodos é referido como "análise de imagem baseada em objeto"(OBIA). Mais especificamente, a análise de objetos geográficos ficou denominada GEOBIA [34]. Este método pode incorporar recursos de alto nível, como contexto espacial ou da geometria do objeto na classificação e com isto superar as limitações dos métodos convencionais de classificação de imagem baseada em pixel [11]. Métodos OBIA são atualmente os métodos mais utilizados para a tarefa de extração automática de objetos em imagens HSR. O desenvolvimento desses métodos foi considerado um avanço na literatura de sensoriamento remoto e tem moldado, em grande parte, o conhecimento da comunidade de sensoriamento remoto em torno dessa tarefa [11].

A qualidade de segmentação dos objetos alvo tem uma influência direta sobre a precisão da classificação posterior. Várias técnicas de segmentação

de imagens foram desenvolvidas e aplicadas na análise de imagens de sensoriamento remoto [35],[36],[37] ,[38],[39],[40] e [41]. Quando os *pixels* são agrupados em regiões homogêneas, utilizando apenas as características de espectro e textura para medir a homogeneidade sem incluir recursos de alto nível, não há garantia de que as regiões geradas correspondam a objetos reais ou partes de objetos devido à ambiguidade de recursos de baixo nível. Isso acontece mesmo quando são utilizados os algoritmos de segmentação de imagem de última geração [42] e [43]. Por exemplo, estradas e entradas de estacionamentos podem ser agrupadas em uma mesma região devido à sua semelhança em termos de características de espectros. No OBIA é necessário definir parâmetros de segmentação, geralmente com base na homogeneidade espectral e tamanho, para os diferentes objetos da imagem dimensionados, moldados e espacialmente distribuídos em uma cena. Isso é feito de modo que a segmentação seja feita satisfazendo os requisitos dos usuários [44]. O algoritmo de segmentação multi-resolução (MRS) [31] é provavelmente o mais popular para a delimitação de objetos relativamente homogêneos e significativos. A questão é que a performance do MSR depende de parâmetros, que são geralmente selecionados subjetivamente a partir da avaliação visual da adequação da segmentação. Embora este processo subjetivo permita a incorporação mais flexível de forma, textura, geometria, recursos semânticos contextuais e conhecimento especializado, este procedimento dificilmente será reproduzível e levanta algumas limitações em relação à robustez da abordagem. Um conjunto de técnicas para automação do processo de segmentação em OBIA foram desenvolvidos [38], [39] e [45]. Contudo, a definição das regras de classificação ainda é subjetiva e não há uma solução genérica para a automação completa do processo de segmentação [11]. Uma revisão mais detalhada pode ser encontrada em [9] ou [33].

2.4

Métodos de aprendizado de máquina

As técnicas de aprendizado de máquina criaram poderosos classificadores. Muitas abordagens recentes consideraram a detecção de objetos como um problema de classificação e alcançaram melhorias significativas. A detecção de objeto pode ser realizada treinando um classificador que captura as variações de aparências e visualizações de objetos a partir de um conjunto de dados. A entrada usual destes classificadores é um conjunto de *pixels*, que são selecionados na imagem utilizando um sistema de janelas móveis ou proveniente de modelos de objetos com seus atributos. A saída são seus respectivos rótulos, ou seja, se é um determinado objeto ou não. A etapa mais relevante neste

processo é a extração de atributos que sejam relevantes para a classificação do objeto. Este é um processo fundamental na potencialização da performance do classificador [46, 47, 48, 49]. Alguns dos métodos de aprendizagem de máquina mais comumente utilizados em RS são: *Support Vector Machine* (SVM) [50], AdaBoost [51, 52], k-nearest-neighbors (kNN) [53] e *Artificial Neural Network* (ANN) [54]. Na avaliação feita em [55] que comparou os métodos SVM, ANN e árvore de decisão para o mapeamento LULC em imagens de satélite, mostrou que a precisão alcançada pela SVM foi maior do que outros métodos. Estes classificadores apresentaram melhoras significativas no processo de classificação em RS. No entanto, requerem grandes amostras, de objetos e não-objetos, para treinamento. A performance e capacidade de detecção dependem diretamente da qualidade desta amostra.

2.5

Extração de Atributos

Os métodos citados são bons classificadores, mas a alta dependência de processos adjacentes para extração de atributos dificulta a aplicação em larga escala. A extração de objetos em imagens HSR não é complicada para um analista humano. Contudo, considerando a dificuldade em obter bons resultados de classificação em processo automatizados utilizando apenas imagens de HSR, é natural pensar na inclusão de recursos mais discriminatórios utilizando dados auxiliares. Os métodos mais comuns para extração de atributos são: Histogram of oriented gradients (HOG)[56]; Bag-of-words (BoW), [57]; Texture features, [58]; Sparse representation (SR)-based, [59, 60, 61, 62, 63]; Haar-like, [64]. Contudo, muitos outros foram propostos, como em [65], [66]. Em [67], os autores utilizaram dados de elevação, como o modelo de superfície digital (DSM) extraído da fotogrametria estéreo ou diretamente das imagens LiDAR, o que aumentou significativamente a precisão da extração de objetos. No entanto, a aquisição destes atributos tem custo elevado. Por exemplo, um modelo de superfície usando LiDAR pode custar até quatro vezes mais que a coleta de imagens RGB aéreas [5], o que pode tornar inviável a aplicação em grande escala. Neste sentido, vários esforços foram feitos na tentativa da extração de atributos diretamente das imagens HSR. Por exemplo, [68] propuseram extrair características espectrais, de textura e geométricas da segmentação em várias escalas e demonstraram que esses recursos ajudaram a melhorar a precisão do mapeamento urbano LULC. No trabalho de [69] foi desenvolvido um índice morfológico de construção e sombra utilizado para extrair edificações de imagens HSR. Em [70], uma medida espacial foi criada para aumentar a precisão da extração de objeto a partir de imagens de HSR.

Um método para extrair atributos na forma de pixel foi proposto por [71], mostrando que esses recursos eram mais discriminatórios do que os recursos espectrais usados tradicionalmente.

3

Deep Learning em Imagens

O método de Deep Learning (DL) é hoje um dos mais promissores métodos em aprendizagem de máquinas para classificação de imagens. Este método permite que o computador aprenda automaticamente atributos complexos e relevantes a partir de um conjunto de treinamento [72, 73, 74].

Dentre os modelos supervisionados de DL, a rede neural convolutiva (Convolutional Neural Network) (CNN)[54, 75, 76, 73] é a que superou a maioria dos algoritmos no reconhecimento visual. A estrutura da CNN permite que o modelo aprenda atributos altamente abstratos e mapeie estes atributos de entrada em representações que podem aumentar o desempenho dos classificadores.

A CNN tornou-se popular e com grande sucesso em aplicações de reconhecimento visual de larga escala. Esta repercussão teve início com o trabalho de LeCun et al. [54], com a *Le-Net*, a primeira aplicação em CNN bem sucedida. Esta aplicação foi uma rede para reconhecimento de dígitos em imagens de 32×32 bits. Em seguida, Krizhevsky et al. [18], com *AlexNET*, tornou a CNN mais popular na área de visão computacional. Esta rede mostrou ótimo desempenho na classificação de imagens do *ImageNet* [77]. O *ImageNet* é um conjunto de imagens organizado de acordo com a hierarquia do WordNet [78]. Existem em média 1000 imagens representativas de cada conceito, todas controladas pela qualidade e anotadas por humanos. Atualmente, tem mais de 14 milhões de imagens. Este banco de imagens virou uma referência para comparação de classificadores de imagem utilizando a CNN. Outros modelos importantes, todos usando o *ImageNet*, foram o *ZF NET* [79], *GoogleLeNet* [80], *VGGNet* [81] e o *ResNet* [82].

A CNN é projetada para aproveitar a estrutura bidimensional da imagem de entrada. Tem uma arquitetura de multicamada composta de múltiplos estágios de extração de atributos. Cada estágio geralmente é composto por três camadas: uma camada convolucional, uma camada de não-linearidade e uma camada de agrupamento. Uma CNN típica, geralmente, é composta de até três desses estágios de extração de atributos, seguidos por uma ou mais camadas de redes neurais, totalmente conectadas e uma camada classificadora final.

3.1

Camada de convolução

Considerando que a entrada da rede é uma imagem, as Redes Neurais convencionais são limitadas devido à sua estrutura com grande número de parâmetros. Nestas redes, a camada de entrada é um vetor com n neurônios seguido por camadas escondidas (*Hidden*) inteiramente conectadas (*full connected*) (veja Figura 3.1). Se a entrada é uma imagem RGB de 32×32 *pixel*, isto implica em um vetor de entrada de tamanho $32 \times 32 \times 3 = 3072$. Isto significa que cada neurônio da primeira camada escondida terá 3072 parâmetros (*weights*). Em uma aplicação utilizando a *ImageNet*, por exemplo, serão $224 \times 225 \times 3 = 150528$ parâmetros por neurônio, um número bastante elevado, o que gera modelos muito complexos e dificulta a aplicação desta configuração em imagens.

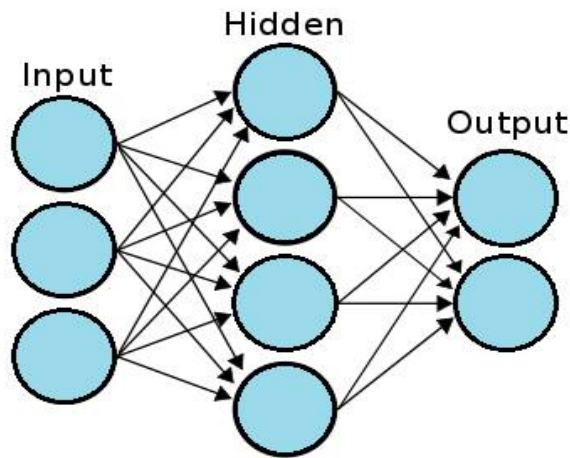


Figura 3.1: Rede Neural totalmente conectada com uma camada escondida

A CNN possui uma estrutura que se adapta bem em imagens. Seus neurônios, ou filtros (*kernels*), têm uma estrutura de três dimensões: altura, largura e profundidade. Este filtro, está conectado a uma pequena região da camada anterior, o que reduz o número de parâmetros. Nesta configuração, as transformações ocorrem em volumes de três dimensões. Cada camada de convolução é formada por um conjunto de filtros tridimensionais de pequena dimensão espacial (altura e largura) e profundidade fixa, igual ao número de camadas da imagens de entrada ou da camada anterior. Estes filtros são compostos por parâmetros dinâmicos que assumem valores aprendidos durante o processo de treinamento. O filtro percorre a imagem em um processo de janelas móveis (veja Figura 3.3), gerando novas imagens ou mapas de ativação com as estimativas em cada posição espacial. De forma geral, funciona como

se o conjunto de filtros em cada camada aprendesse a extrair os atributos significativos.

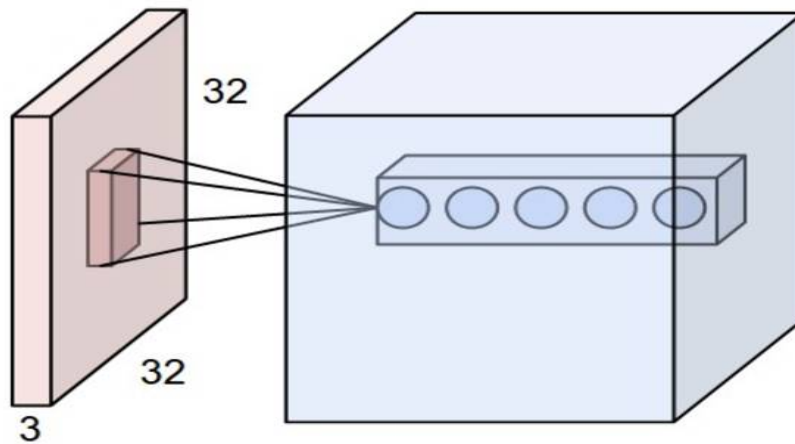


Figura 3.2: Convolução 2D [1]

Para construção de uma CNN é preciso definir as características denominadas hiperparâmetros. Os principais são: o tamanho dos filtros (altura e largura), a profundidade (*depth*) ou número de filtros e o passo de deslocamento (*stride*) dos filtros sobre a camada anterior. Estes hiperparâmetros definem a quantidade de variáveis a serem estimadas e o volume de saída. A estratégia para a redução do número de parâmetros considera a suposição de que se um atributo é importante para um determinado local da imagem, este atributo também é importante nas outras posições da imagem. Isto significa que um filtro em uma determinada profundidade deve usar os mesmos parâmetros para toda a área da camada de entrada.

3.2

Camada de não-linearidade

Esta camada é uma transformação do valor de saída. Consiste em uma função não linear. As mais comuns são: Sigmoid, Tanh e ReLu.

A ReLu, uma função simples $f(x) = \max(0, x)$. Funciona como um limitador para valores negativos. Tem sido muito utilizada por apresentar vantagens em relação às outras medidas [18]: facilidade de cálculo e aceleração da convergência do otimizador - *Stochastic Gradient Descent*- SGD.

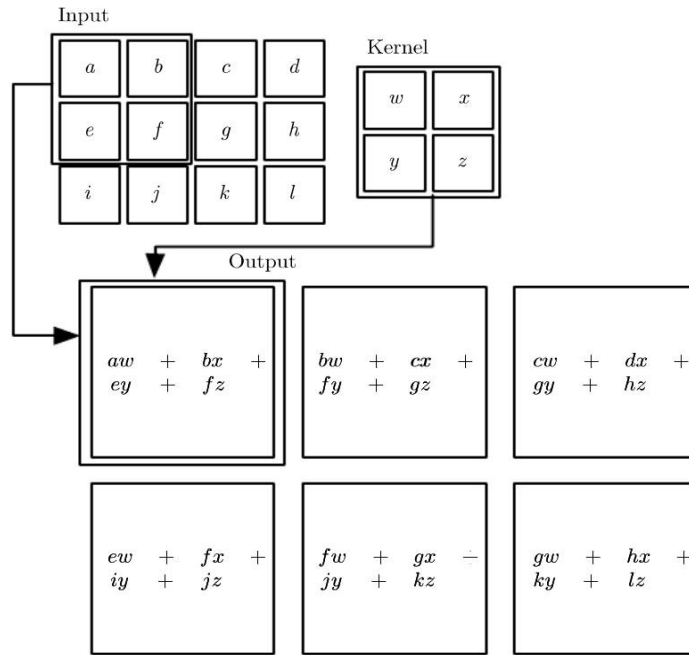


Figura 3.3: Convolução 2D [2]

3.3

Camadas de Agrupamento

Outra estratégia na redução de parâmetros é a inserção de camadas de agrupamento (*Pooling Layers*) entre camadas de convolução. Estas camadas tem a finalidade de reduzir a dimensão espacial (altura e largura) dos blocos intermediários. Esta camada também tem a estrutura bidimensional, geralmente de 2×2 *pixels* que desliza por cada uma das camadas de forma independente. A função geralmente utilizada é $Max()$.

3.4

A Arquitetura U-Net

O uso típico das CNN na tarefa de classificação de imagens é atribuir um rótulo a uma imagem de entrada. Para a segmentação de uma imagem, a estratégia é a classificação em janelas móveis (*patches*) variando *pixel* a *pixel*. Outra estratégia é a classificação do *pixel* considerando o contexto de toda a imagem como na U-net proposta por [83].

A U-Net é uma rede só com camadas de convolução e não tem camadas *fully connected*, conforme ilustrado na Figura 3.6. É composta por camadas de contração (lado esquerdo) e camadas expansivas (lado direito), apresentando uma simetria entre as camadas dos dois lados, no formato da letra "U". A modificação mais relevante está na camada expansiva, o *upsampling*,

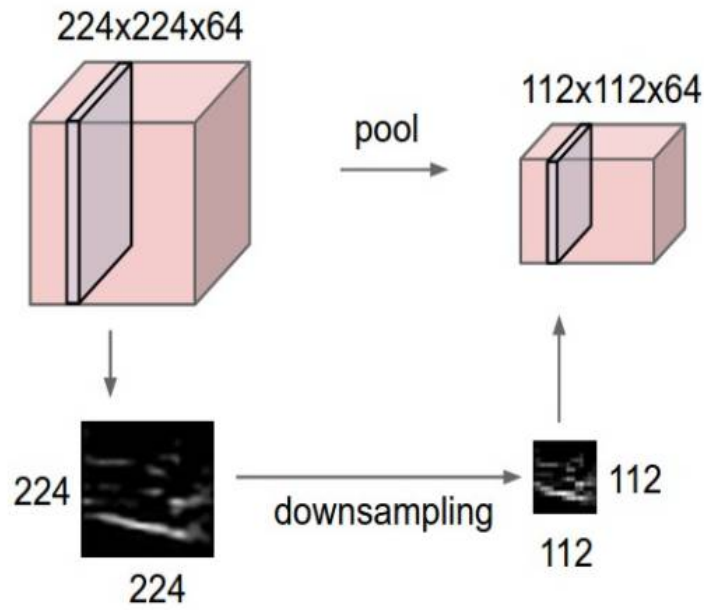


Figura 3.4: Exemplo de Agrupamento 2D - 2×2 . Redução da altura e largura do objeto mantendo a profundidade [1]

que permite a propagação das informações de contexto para camadas com maior resolução. Além disso, esta camada expansiva é incrementada com a concatenação das camadas do nível equivalente da camada de contração. Na aplicação de Ronneberger et al. [83] as imagens não eram coloridas, a entrada é de 572×572 *pixels* e uma camada. Na implementação para este trabalho a entrada tem o mesmo tamanho, mas com três camadas (RGB). A saída tem 388×388 *pixels*. A redução é porque só são utilizadas a parte válida de cada convolução, ou seja, a saída ou o mapa de segmentação só tem os *pixels* que tenham informação de contexto completa na imagem de entrada.

No lado esquerdo, cada etapa do caminho de contração consiste em duas camadas sequenciais de: filtros de convolução 3×3 , uma camada ReLU e uma max-pooling de 2×2 para *downsampling*. Do outro lado da U-net, cada passo do caminho expansivo é composto por uma camada de *upsampling*, uma expansão das linhas e colunas da camada de origem, seguida por uma convolução 2×2 , uma concatenação com as camadas de contração correspondente e duas camadas de convolução 3×3 . Ao final, a camada final, é uma convolução 1×1 . Esta camada é usada para mapear cada vetor de recursos de 64 componentes para o número desejado de classes. Para mais detalhes sobre o U-net, veja [83].

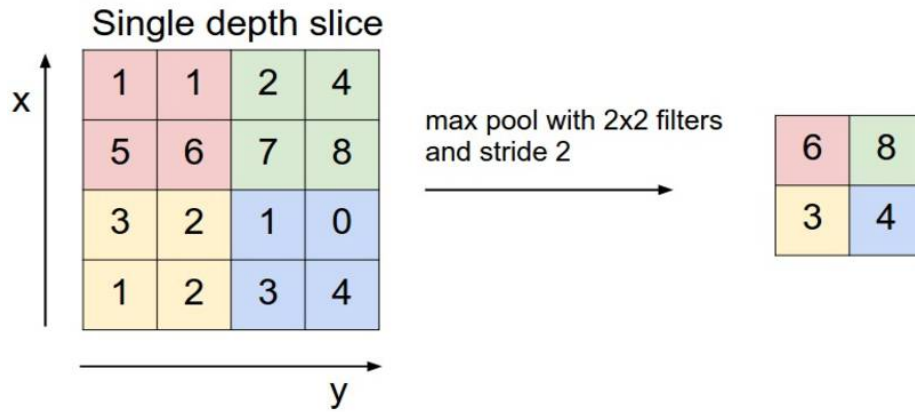


Figura 3.5: Exemplo de Agrupamento 2D, função $\max()$, em uma única camada [1]

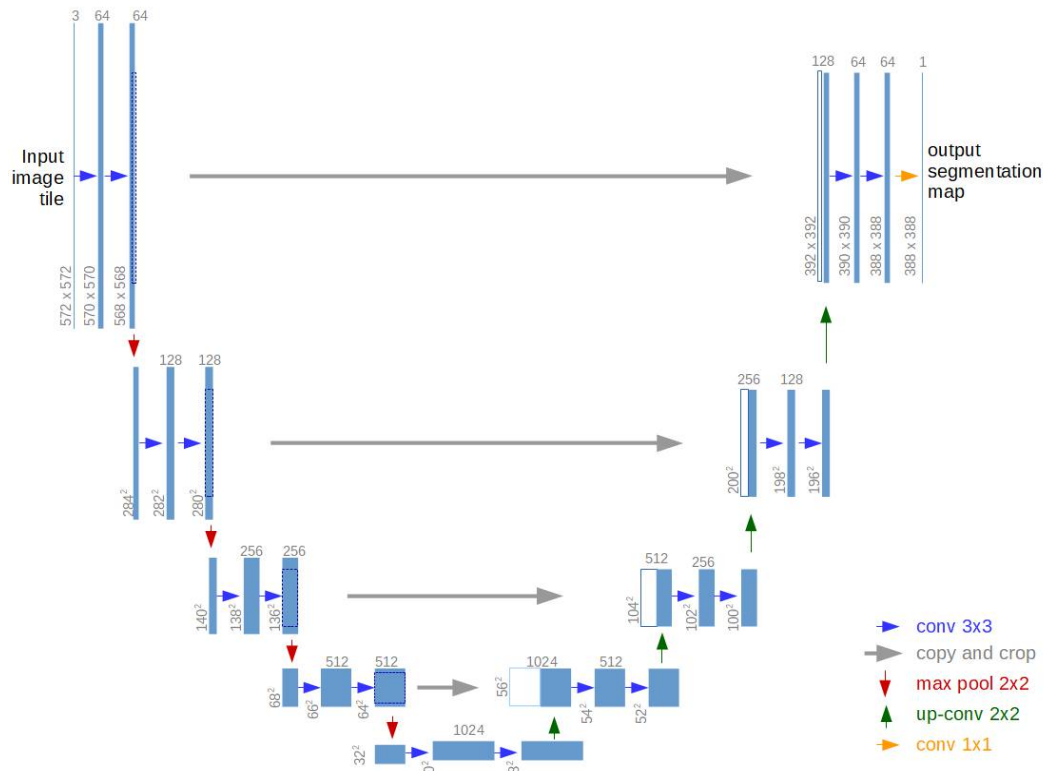


Figura 3.6: The U-net architecture.

Nos últimos anos, as Redes Neurais Convolucionais (CNNs) têm sido amplamente utilizadas para resolver diferentes tarefas de sensoriamento remoto [84].

De acordo com Kampffmeyer et al. [85], existem duas abordagens principais para segmentação de imagens usando CNNs. A primeira, chamado de abordagem *patch-based*, faz a classificação de cada pixel na imagem, observando uma sub-região que contém o pixel. O segundo, chamado de abordagem *baseada em pixel*, classifica cada pixel na imagem usando uma abordagem *downscaling-upscaling*. O trabalho de Guo et al. [3] é um exemplo da abordagem baseada em *patch*, e o trabalho de Kampffmeyer et al. [85] é um exemplo da abordagem baseada em *pixel*.

Na prática, existem atualmente duas abordagens principais para a realização de segmentação de imagem usando CNNs[85]. A primeira, conhecida por *patch-based*, foi inspirada no trabalho de reconhecimento de dígitos [54] e no de classificação de imagens, o Imagenet [77, 18]. Neste método, a classificação utiliza pequenas janelas, *patches*, como informação para classificação da sua região central que pode ser um *pixel* ou conjunto de *pixels*, um gride, gerando pequenas imagens para serem classificadas [3, 86, 48, 87]. Posteriormente a classificação em uma imagem é feita utilizando uma janela móvel, de mesma dimensão do patch treinado, que desloca sobre toda a imagem fazendo a classificação. Os trabalhos recentes em SR mostram vários empregos desta abordagem em imagens RGB como em [88, 89] que propõem métodos de ajuste fino para uma rede pré-treinada, a *GoogleNet*, e melhoram a acurácia de classificação no conjunto de dados *UCMerced* [90]. Em Penatti et al. [91], também é utilizada a CNN em associação a um método clássico de extração de atributos, o *Bag of Visual Word Classification*. Nos trabalhos de Zhang F. et al. [92] e Zhang L. et al. [84], foi utilizado o conjunto de imagens *UCMerced* para treinar uma rede CNN e posteriormente fizeram um teste em imagens com atributos semelhantes numa região de Sidney, Austrália, coletadas do *Google Earth*. No trabalho de Guo et al. [3], os autores usam a CNN para identificar construções com a finalidade de localizar aldeias em uma região do Laos. Este trabalho se diferencia por ser o único a criar o próprio conjunto de imagens

RGB de alta resolução do *Google Earth* e utilizar uma configuração própria de CNN realizando o treinamento e teste.

Em todas estas aplicações o uso da CNN é para a classificação e posteriormente é utilizadas para segmentação. Neste processo, toda a segmentação esta condicionada ao tamanho do *patch* e ao passo de deslocamento (*stride*) [93]. A variação de tamanho do *patch* influencia a precisão, pois um *patch* pequeno traz menos informação para o modelo, o que impacta negativamente na acurácia. O passo de deslocamento do *patch*, quanto menor, melhor a acurácia de predição, contudo, aumenta a demanda de processamento devido ao incremento do número de *patches*.

A segunda abordagem é baseada na ideia da segmentação semântica. Classifica pixel a pixel considerando o contexto da localização deste pixel [94]. Este trabalho foi o primeiro a utilizar uma (*Fully Convolutional Network* - FCN) supervisionada para a predição de *pixel* (pixelwise). Neste método a rede é totalmente convolucional, dividida em duas partes, codificadora e decodificadora. A codificadora é uma série de camadas de convolução responsáveis por mapear a imagem para uma representação de baixa resolução. Em seguida, a parte decodificadora, formada por camadas de interpolação (*upsampling*) e deconvolução, recriam uma imagem em que cada pixel terá uma classificação. Esta abordagem permitiu uma melhoria na acurácia em segmentação de imagens. Inspirado por estes resultados, Kampffmeyer et al. [85] utiliza para uma aplicação em SR. O autor aplica esta metodologia associada ao método *patch-based* no conjunto de dados *ISPRS Vaihingen 2D* [95] para criar a segmentação associada a um mapa de incerteza de predição.

De forma geral, os trabalhos citados na sua maioria empregam redes reconhecidamente eficientes como geradora automática de atributos e então empregadas em um classificador. Outro fato é a utilização quase que unânime do conjunto de imagens públicas, o UCMerced [90], que permite a comparação com outros trabalhos, mas não proporciona a avaliação dos modelos em situações diferentes e mais complexas a aquelas apresentadas nas referidas imagens. Como observa Zhang L. et al. [84] e Karpatne et al. [13], ainda há alguns desafios para o uso do *Deep Learning* nas tarefas em sensoriamento remoto:

- Como ter boa performance no aprendizado em métodos de *Deep Learning* com pequenas bases de dados;
- Como lidar com a complexidade das imagens de sensoriamento remoto. Imagens com grande variabilidade tanto no objetos quanto no contexto de fundo (*background*). Isto torna difícil o aprendizado de representações robustas e discriminativas dos objetos com *Deep Learning*;

Neste contexto, para enfrentar estes desafios propomos a utilização de uma arquitetura CNN, a U-Net, apresentada por Ronneberger et al. em 2015 [83]. Inspirado no trabalho de [94], [83] modificou e estendeu a arquitetura original criando uma rede capaz de prover uma melhor segmentação com uma quantidade relativamente menor de imagens para treinamento. No seu trabalho original, a U-net foi aplicada à segmentação baseada em *pixels* em imagens biomédicas com excelente desempenho.

5

Aquisição e Anotação das Imagens

Para a localização de edificações em larga escala, é primordial que o método seja robusto quanto às variações da qualidade de imagens, como variações de iluminação (brilho e sombras) e oclusão. Deve ser robusto a alta variância intra-classe, ou seja, deve reconhecer edificações com diferentes tamanhos, formatos e colorações. Também tem que ter alta capacidade de discriminação das edificações dentre a complexidade geográfica, como a variação de características de solo, vegetação e urbanísticas. Neste sentido, a solução foi a construção de um conjunto de imagens com esta complexidade.

5.1

Obtenção das Imagens

Para a construção de conjunto de imagens com as características desejadas, foi decidido utilizar imagens RGB *Google Earth* de regiões brasileiras. Ao pegar imagens de diversas regiões do país garantimos a diversidade geográfica, urbanística. Com as imagens provenientes do GE, temos a certeza de que tem boa qualidade para extração de objetos, mas, devido à extensão do país e das regiões selecionadas, as imagens têm característica de qualidade diversa.

Neste trabalho, coletamos e anotamos 126 imagens RGB, cada uma com aproximadamente 900×900 *pixels*. Essas imagens foram coletadas do GE em 12 locais diferentes no Brasil, pelo uso da API do GE. A Figura 5.1 mostra essas localizações no mapa. Na Tabela 6.4 tem a identificação destas regiões.

A API do GE limita o tamanho da imagem e o número de downloads por dia. O processo de coleta de imagens começa com a seleção de municípios, ou uma região de interesse. Nesta região é construído um gride georreferenciado com células do tamanho das imagens provenientes da API do GE. Então em um processo iterativo este gride é preenchido formando um mosaico com imagens da área. A Figura 5.2 mostra o processo de criação do mosaico.

As Figuras 5.3, 5.4, 5.5 e 5.6 apresentam algumas imagens coletadas. A Figura 5.35.3(a) e 5.35.3(b) mostram uma zona metropolitana com edifícios e casas. A Figura 5.35.3(a) é uma imagem bem nítida e mostra avenidas largas, um pouco de oclusão e muita sombra. Na Figura 5.35.3(b), apresenta



Figura 5.1: Locais onde foram coletadas as imagens GE (pontos em vermelho).

imagem com qualidade diferente, com menos contraste e brilho. Tem maior adensamento de construções, ruas estreitas, mais áreas verdes e menos sombra.

A Figura 5.45.4(a), tem um único tom. Apresenta um grande adensamento de construções de pequeno porte com arruamento estreito e quadras em formatos variados. A Figura 5.45.4(b) apresenta o mesmo tom, mas a urbanização muda. Apresenta quadras grandes com grandes construções bem adensadas.

As Figuras 5.55.5(a) e 5.5(b) têm semelhança na composição urbanística, contudo as imagens têm características bem distintas quanto a nitidez, e brilho.

As Figuras 5.65.6(a) e 5.6(b) apresentam grande contraste. A Figura 5.65.6(a), composta por casa adensadas em uma região bastante verde à margem de um rio. Na Figura 5.65.6(b) contrasta por ser uma região árida quase com um tom único de cor. tem semelhança na composição urbanística, contudo as imagens tem características bem distintas quanto a nitidez, e brilho.

Como mostrado nas Figuras 5.3, 5.4, 5.5 e 5.6, o conjunto de imagens é desafiador e apresenta os principais desafios levantados na bibliografia.

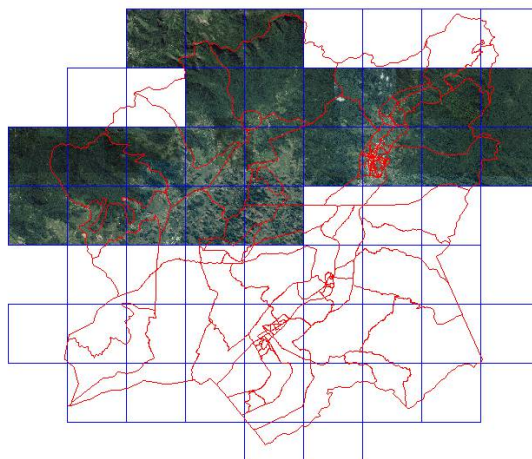


Figura 5.2: Formação do mosaico com imagens *Google Earth*. Em vermelho a delimitação geográfica da região de interesse. Em azul, o gride georreferenciado.

5.2

Anotação das Áreas Construídas

Para o processo de aprendizado supervisionado foi necessária a identificação das áreas construídas. Para cada imagem coletada adicionamos manualmente uma nova camada que indica, em forma binária, onde existe ou não uma edificação. Este é um processo manual que depende da interpretação humana. Ele é geralmente feito desenhando um polígono que coincida com os limites da edificação. Uma tarefa delicada e demorada e que se torna muito complicada nos grandes adensamentos de edificações pela dificuldade de identificação dos limites das construções. No caso das imagens obtidas, na escala fornecida, este processo se mostrou inviável, em muitos casos foi difícil determinar os limites corretamente e em muitos outros, por causa do adensamento, as anotações ficavam emendadas. A alternativa adotada foi marcar as regiões construídas em blocos. Foram utilizados *tablets* com canetas para a marcação das áreas. Assim, nas regiões pouco densas as construções ficaram anotadas de modo similar ao gerado pelos polígonos como se observa nas Figuras 5.75.7(a) e 5.7(b). Nas regiões mais densas, ficaram representados os blocos de edificações, 5.85.8(a) e 5.8(b). Ficando assim definido que o que se pretende classificar são áreas construídas, ou seja blocos com uma ou mais construções.

Posteriormente, as imagens anotadas foram convertidas em uma imagem *P&B* identificando as áreas construídas para serem utilizadas como *labels* no processo de aprendizado supervisionado (veja Figura 5.85.8(c)). Este conjunto de dados anotado está disponível em [20].

Foram anotadas 126 imagens de aproximadamente 900×900 *pixel*. Para a



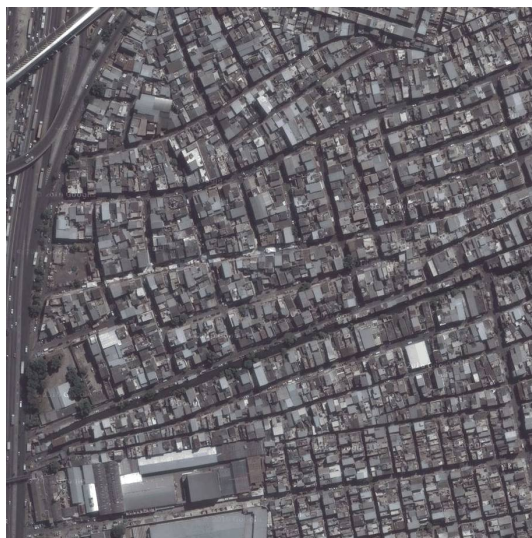
5.3(a):

5.3(b):

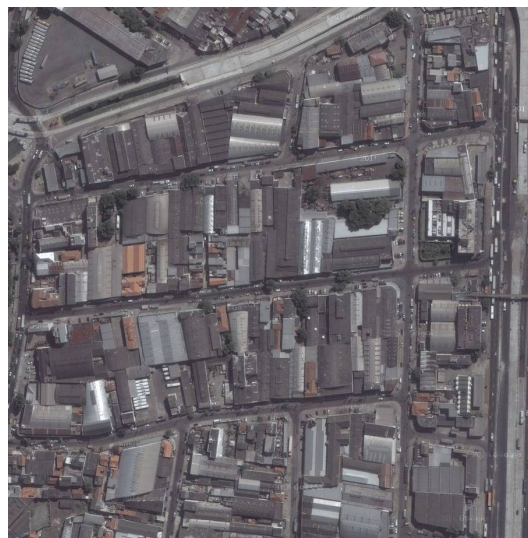
Figura 5.3: Zonas metropolitanas. 5.3(a) região com edifícios, casas e grandes avenidas. Imagem nítida, com pouca aclusão e muitas sombras. 5.3(b) região com edifícios, casas, áreas verdes e rua estreitas. Imagem com menor nitidez, pouca aclusão e poucas sombras. (*Imagens Google Earth*).

realização dos experimentos foram separadas 12 imagens, uma imagem de cada região para teste e o restante, 114 imagens, foram utilizadas no treinamento e validação.

Na aplicação da U-net, usamos uma estratégia de janela móvel extraindo imagens de tamanho 572×572 a cada 150 pixels (*stride*) nas 114 imagens anotadas. Este processo é idêntico tanto nas imagens RGB quanto na respectiva imagem de classificação, gerando 772 conjuntos de imagens (imagem RGB e de classificação). Devido às características da U-Net, as imagens de classificação devem ser ajustadas. A imagem de referência na U-Net tem tamanho 388×388 , então é feito o recorte das imagens, retirando uma borda de 93 pixels .



5.4(a):

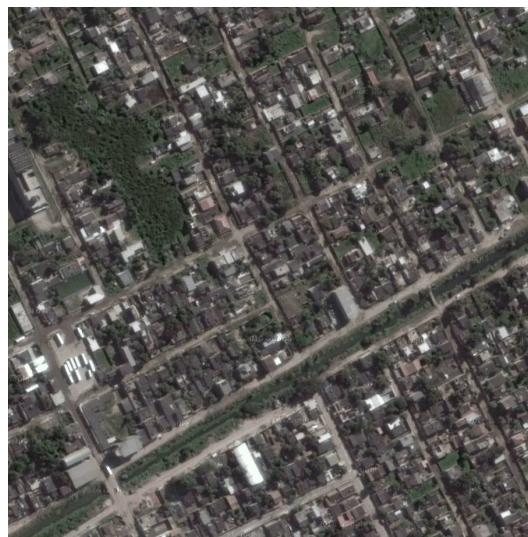


5.4(b):

Figura 5.4: Diversidade urbanística. 5.4(a) região com grande densidade de construções de pequena dimensão e estrutura urbanística não padronizada. 5.4(b) região com grande densidade de construções de grande porte. (*Imagens Google Earth*).



5.5(a):



5.5(b):

Figura 5.5: Cidades do interior. 5.5(a) imagem com alto brilho e pouca nitidez. Área residencial com variação nas cores de telhado, áreas verdes com tonalidade variável. 5.5(b) imagem mais escura com boa nitidez. Área residencial com áreas verdes. (*Imagens do Google Earth*).



5.6(a):



5.6(b):

Figura 5.6: Cidades do interior. 5.6(a) adensamento de casa com muita área verde. 5.6(b) região árida com residências dispersas quase sem variação de cor. *Imagens Google Earth.*



5.7(a): x



5.7(b): x

Figura 5.7: Anotação de áreas construídas: 5.7(a)) Imagem com baixo adensamento. 5.7(b)) Imagem com as as construções anotadas (preto). *Imagens Google Earth.*



5.8(a): Imagem Original

5.8(b): Imagem Anotada



5.8(c): Imagem de Classificação

Figura 5.8: Anotação de áreas construídas. 5.8(a) Imagem original. 5.8(b) Imagem anotada, blocos de áreas construídas. 5.8(c) Imagem P & B da classificação de áreas construída. *Imagens Google Earth*

6

Resultados

Neste capítulo, analisamos os resultados em duas etapas. Na primeira, falamos da aplicação da U-Net e analisamos o desempenho de classificação, com base nas estatísticas e visualização das imagens. Na segunda parte, comparamos os resultados da U-Net com os resultados da aplicação do modelo descrito por [3], que é uma aplicação clássica da CNN em *patch*. O modelo foi escolhido por ter sido utilizado em condição semelhante ao experimento proposto. No experimento descrito em [3], também utilizam imagens *Google Earth*.

6.1

Resultados com a U-Net

Implementamos o U-net em python com Keras [96]. O processamento foi feito em um computador com CPU Intel i7, 64 GB de RAM e com uma GPU - NVidia TitanX. No treinamento do modelo U-net, foram executadas 450 épocas em cerca de 45 horas. Para aumentar o número de exemplos de treinamento e minimizar a chance de *overfitting*, usamos a funcionalidade do gerador de imagens da Keras. Aplicamos as transformações aleatórias de *Flips* verticais, *Flips* horizontais e *scaling ranging* de 0,9 a 1. Outras transformações usuais, como a rotação, não foram empregadas. Pois elas causam, nos cantos das imagens, uma deformação. Esta alteração é inadequada no processo de segmentação da U-Net. Um conjunto de *pixel* deformados gerariam imagens com edificações de formatos irreais causando um ruído na performance de classificação. O conjunto de teste, 56 imagens das 12 regiões, teve a mesma preparação, exceto as deformações.

Os resultados de treinamento, validação e teste estão na Tabela 6.1 onde,

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN},$$

e

$$F_1score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}.$$

Na Tabela 6.1 os resultados de Teste correspondem a média e desvio padrão relativos às 56 imagens. A precisão média de 80,4% é expressiva. O trabalho de Zhang Q. et al. [73], também usando imagens do Google Earth, obteve precisão máxima de 89% e abaixo de 78% em situações mais complexas onde as propriedades espectrais das construções e do terreno são similares. Considerando que [73] tem imagens de uma única província chinesa e faz a classificação em *patches*, o desafio apresentado com a U-Net é mais complexo. Na Figura 6.1 observamos que a precisão variou entre 60% e 100%. A precisão mínima foi de 56,8% e a máxima 95,7%. Uma das imagens com baixa precisão, a da Figura 6.46.4(b), teve precisão de 68,1%. Observou-se que em imagens com grande adensamento de construções, muitas sombras e construções com propriedades espectrais similares à do terreno, a incidência de falsos positivos pode ser mais alta.

Tabela 6.1: Rede U-Net: Resultados do Treinamento, Validação e Teste

Model	Task	Precision	Recall	F_1 -score
U-net	Treinamento	95.9%	95.9%	95.9%
	Validação	93.8%	84.1%	93.9%
	Teste	80,4% (7,7%)	85,6% (9,1%)	82,6% (6.5%)

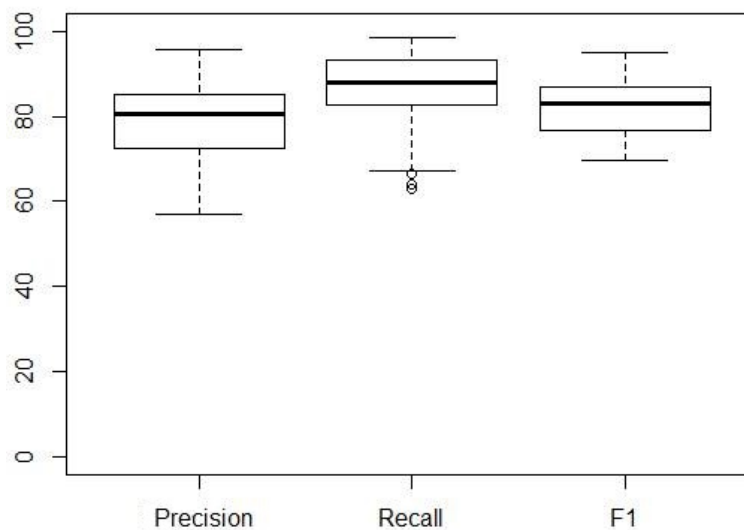


Figura 6.1: Distribuição das estatísticas de teste obtidos com a U-Net no conjunto de teste, 56 imagens.

A Figura 6.26.2(b) mostra regiões com falso positivos geralmente nas bordas das edificações. Há duas grandes áreas de estacionamento que foram consideradas edificações. Alguns pontos considerados falsos positivos ou falsos

negativos, se observados na imagem de referência, são difíceis de serem definidos como parte ou não de uma edificação. Estes pontos podem ser erros de anotação. De forma geral, os grandes blocos de áreas construídas ficaram bem definidos.



Precision	Recall	F1
79,5	92,1	85,4

6.2(c):

Figura 6.2: Imagens de Teste. 6.2(a) Imagem RGB de referência. 6.2(b) Imagem da camada de predição sobre a imagem de referência. Em verde, *True Positives*, em amarelo *False Negatives* e em vermelho *False Positives*. *Imagens Google Earth*.

A Figura 6.3, diferentemente das outras, tem brilho e nitidez diferenciadas. Com precisão de 85,8%, o resultado também apresentou divergências nas bordas das edificações. Porém, o modelo conseguiu classificar bem mesmo nas ruas de terra com cor semelhante aos telhados.

Uma das imagens mais desafiadoras é a Figura 6.5, que tem construções difíceis de serem identificadas. O resultado do modelo teve precisão de 70,2%, com algumas regiões de falsos positivos. Estes erros ocorreram no entorno das construções. Alguns desses erros nestas regiões podem ter sido causados por imprecisões de anotação. Pelas características da imagem, muitas vezes é difícil definir claramente os limites das construções. Isto ocorre devido à cobertura por vegetação, tipo de calçamento e a modificações de características do solo devido ao uso intenso.

De forma geral, o modelo obtido foi capaz de localizar as áreas de construção. Conseguiu discriminar bem o arruamento e outras áreas fora do contexto urbano. Os pontos críticos foram as áreas adjacentes ou entre



6.3(a):

6.3(b):

Precision	Recall	F1
85,8	86,6	86,2

6.3(c):

Figura 6.3: Imagens de Teste. 6.3(a) Imagem RGB de referência. 6.3(b) Imagem da camada de predição sobre a imagem de referência. Em verde, *True Positives*, em amarelo *False Negatives* e em vermelho *False Positives*. *Imagens Google Earth*.

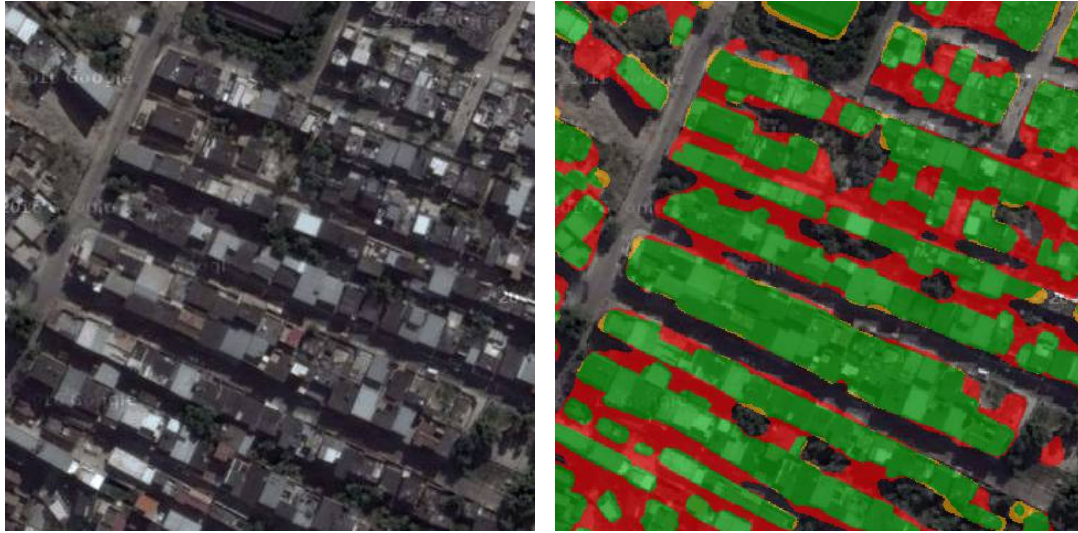
construções que dependendo do tipo de cobertura foi selecionada erroneamente. Uma parte dos erros também são devidos ao erro de anotação. Nas áreas de grande adensamento alguns limites e áreas entre construções são difíceis de serem identificados. Este tipo de situação varia conforme a qualidade da imagem.

Selecionamos o modelo CNN proposto em Guo et al. [3] para comparar com a U-net porque é um trabalho bem recente que apresenta uma solução para um problema semelhante, fazer a detecção de construção em imagens da GE. Neste trabalho os autores comparam um método tradicional, o AbaBoost com o Haar-like e informação de cor, com uma rede CNN. Mostram que a CNN teve um desempenho 25% superior na classificação.

6.2

Comparação U-Net e *Patch-based Classification Model*

No trabalho de [3] foi proposto um modelo para resolver um problema semelhante. Mostram uma boa performance da CNN na identificação de vilas numa região do Laos utilizando imagens do Google Earth. A CNN proposta para predição em *patch*, a qual vamos nos referir como o *Patch-based Classification Model* ou só *Patch-based*, tem a forma tradicional, como



6.4(a):

6.4(b):

Precision	Recall	F1
68,1	96,6	80,0

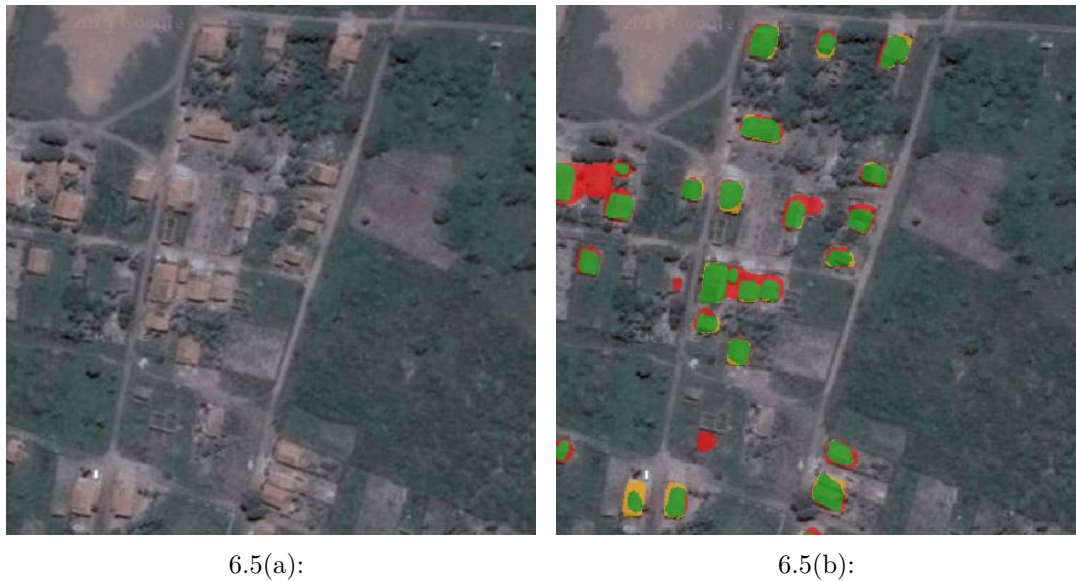
6.4(c):

Figura 6.4: Imagens de Teste. 6.4(a) Imagem RGB de referência. 6.4(b) Imagem da camada de predição sobre a imagem de referência. Em verde, *True Positives*, em amarelo *False Negatives* e em vermelho *False Positives*. *Imagens Google Earth*.

mostardo na Figura 6.6. É composta por duas camadas convolucionais, cada uma seguida por uma camada *pooling layer* com *stride 2* para *downsampling*. A primeira camada convolucional contém seis filtros 5×5 e o segundo contém 12 filtros 4×4 e uma camada densa com um neurônio no final para realizar a classificação. Considera como uma entrada, um *patch* RGB de tamanho 18×18 *pixel*.

Com o conjunto de imagens de treinamento, 114 imagens RGB, foram gerados aproximadamente 600.000 *patches* de 18×18 com as respectivas classificações. Para cada *patch*, um rótulo binário foi atribuído indicando se há uma edificação no *patch* ou não.

O *Patch-based* foi implementado em Python com o Keras [96]. O processo de treinamento foi executado na mesma máquina que U-Net, CPU Intel i7 com 64 GB de RAM e com uma GPU NVidia Titan X. Foram executadas 250 épocas em 2,7 horas. Neste processo, também foram empregadas as funcionalidades do gerador de imagens da Keras com transformações aleatórias de *Flips* verticais, e horizontais e *scaling ranging* de 0,9 a 1. O resultado de treinamento e validação está na Tabela 6.2, onde o resultado da U-Net, na Tabela 6.1, foi repetido para facilitar a comparação.



Precision	Recall	F1
70,2	88,1	78,2

6.5(c):

Figura 6.5: Imagens de Teste. 6.5(a) Imagem RGB de referência. 6.5(b) Imagem da camada de predição sobre a imagem de referência. Em verde, *True Positives*, em amarelo *False Negatives* e em vermelho *False Positives*. Imagens Google Earth.

Tabela 6.2: Resultados do treinamento e validação para os modelos *Patch-based* e U-Net

Modelo	Tarefa	Precisão	Recall	F1-score
<i>Patch-based</i>	Training	83.1%	87.7%	85.2%
	Validation	81.6%	92.2%	86.6%
U-net	Training	95.9%	95.9%	95.9%
	Validation	93.8%	84.1%	93.9%

No procedimento de teste foram utilizadas as imagens de teste, as mesmas utilizadas no U-Net. A Tabela 6.3 mostra os resultados do teste para o *Patch-based* e repete os resultados da U-Net que estão na Tabela 6.1, para facilitar a comparação.

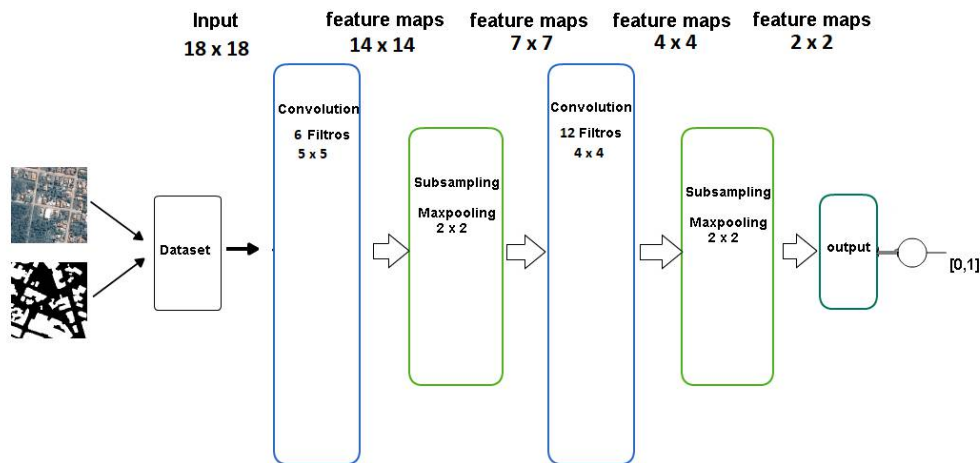


Figura 6.6: Modelo *Patch-based*. Arquitetura CNN utilizada em [3].

Tabela 6.3: Resultados da predição dos modelos *Patch-based* e U-net. Os resultados correspondem a média e desvio padrão (*mean (standard deviation)*) sob o mesmo conjunto de imagens de Teste.

Modelo	Precisão	Recall	F1-score
<i>Patch-based</i> - Patch	68.1% (15.1%)	87.6% (11.0%)	75.1% (11.7%)
U-net	80.5% (7.7%)	85.6% (9.1%)	82.6% (6.6%)

Nas fases de treinamento e validação, o modelo U-net produziu melhores resultados, com F1-score de 93,9% na validação contra 86,6% para o modelo *Patch-based*. Na fase de teste, conforme a Tabela 6.3, o modelo *Patch-based* produziu piores resultados do que o modelo U-net. A precisão média com o *Patch-based* foi de 68,1% contra 80,5% do U-Net. Também, a incerteza associada ao *Patch-based* é maior, pois o desvio padrão é praticamente o dobro com relação ao observado no U-Net.

Na Tabela 6.4 é possível verificar com mais detalhes a variação das estatísticas de predição. O F_1 – score varia de 53,4% a 88% com o *Patch-based* enquanto a variação é de 73,4% a 93,9%.

Nesta comparação de performance é importante entender que a unidade de predição de cada modelo é distinta. No U-Net, a unidade é o *pixel*, e no *Patch-based* é o *patch*. A predição em *patch* é mais suavizada, tem menor variabilidade, quanto maior for o tamanho do *patch*. Assim na comparação dos dois modelos, valores iguais não significa desempenho semelhante. De forma geral, espera-se que com performance semelhante o *Patch-based* tenha indicadores, como precisão, com valores maiores. De qualquer forma a comparação dos dois modelos é válida com o intuito de comparar a qualidade

de predição destes modelos quando comparados aos respectivos valores de referência.

Tabela 6.4: Resultados da predição dos modelos *Patch-based* e U-net. Os resultados correspondem a média em cada uma das 12 localidades nas imagens de Teste.

Imagem	Modelo	Precisão	Recall	F1-score
Cáceres - MT	<i>Patch-based</i>	39.6	84.4	53.8
	U-net	69.8	79.6	74.3
Curitiba - PR	<i>Patch-based</i>	66.7	81.7	73.1
	U-net	78.0	92.3	84.5
Miranorte - TO	<i>Patch-based</i>	70.5	93.7	80.3
	U-net	88.8	85.6	87.2
N. Mutum - MT	<i>Patch-based</i>	59.9	86.9	70.6
	U-net	83.8	66.8	74.2
Nilópolis - RJ	<i>Patch-based</i>	74.0	92.8	82.2
	U-net	71.9	91.0	80.1
RJ (Botafogo)	<i>Patch-based</i>	71.6	96.6	82.0
	U-net	81.9	94.9	87.9
RJ (B.Sucesso)	<i>Patch-based</i>	82.7	94.1	88.0
	U-net	93.7	93.0	93.3
RJ (Tijuca)	<i>Patch-based</i>	75.5	96.8	84.8
	U-net	82.6	98.2	89.7
Rio Bonito - RJ	<i>Patch-based</i>	71.8	80.3	75.6
	U-net	75.7	87.8	81.2
Trairí - CE	<i>Patch-based</i>	38.3	91.8	53.4
	U-net	73.2	80.8	76.6
Tucuruí - AM	<i>Patch-based</i>	72.0	80.9	75.7
	U-net	83.4	75.9	79.4
Xique - BA	<i>Patch-based</i>	71.9	96.4	82.3
	U-net	84.2	86.1	85.1

O *Patch-based* tem alta incidência de falsos positivos. As Figuras 6.76.7(b) e 6.7(e) exemplificam as previsões *Patch-based* com várias áreas falsas positivas. Nessas duas imagens, a precisão do modelo *Patch-based* é 59,9% e 38,3% (avaliação por *patch*), respectivamente (como na Tabela 6.4). A precisão do modelo *Patch-based* para a imagem na Figura 6.86.8(b) foi de 82,7% usando o *patch* na avaliação e 71,3% usando *pixel*. Nesta imagem, pode-se observar que o modelo previu todas as ruas como áreas de construção. A alta precisão nesta imagem é justificada pelo fato de a imagem ter uma alta densidade de construção. Nessas imagens de teste, podemos observar que

os erros provavelmente foram devidos ao modelo *Patch-based* que classifica um *patch* 18×18 *pixels*. Como um *patch* poderia incluir parcialmente áreas de construção, o modelo *Patch-based* geraria falsos positivos, classificando erroneamente ruas e áreas verdes como edificações.

As Figuras 6.76.7(b) e 6.7(e) exemplificam as previsões *Patch-based* com várias áreas falsas positivas. Nessas duas imagens, a precisão do modelo *Patch-based* é 47,3% e 48,5%, respectivamente (como na Tabela 6.4). Estas são áreas de baixa densidade de construções e os erros observados inviabilizam o modelo para uso prático. As Figuras 6.76.7(b) e 6.7(e) ilustram a predição com o U-Net, que tiveram precisão de 86,4% e 77,6%, respectivamente. É possível observar que elas tem uma baixa incidência de falsos positivos e nestes casos os erros ficaram concentrados em poucos *pixel*, sempre nas bordas das construções.



Figura 6.7: Imagem RGB e imagens classificadas pelo *Patch-based* e U-Net em duas regiões: N.Mutum - MT e Trairi - CE. Em verde - (True Positives), em vermelho - *False Positives* e em amarelo - *False Negatives*

Em outra situação, como a da Figura 6.86.8(b), a precisão do modelo *Patch-based* foi de 82,8%. Nesta imagem, pode-se observar que também há alta incidência de falsos positivos. O modelo previu todas as ruas como áreas de construção. A alta precisão nesta imagem é justificada pelo fato de a imagem

ter uma alta densidade de construção. Na predição do U-Net, mostrado no Figura 6.86.8(c), com precisão de 95,6%, as ruas estão bem discriminadas.

As imagens da Figura 6.8 trazem regiões urbanas densas com áreas não construídas, espaços como praças e vegetação. Estes espaço dificultam a classificação pois muitas vezes tem forma e cor semelhante a construções. Na Figura 6.76.8(e), o *Patch-based* teve teve precisão de 77,55%. Também com muitos erros, falsos positivos, tomando todas as ruas. Na predição com a U-Net, na Figura 6.86.8(f) (precisão de 82,8%), os erros ficaram mais uma vez concentrados nas bordas das construções e em áreas internas dos blocos de construções.

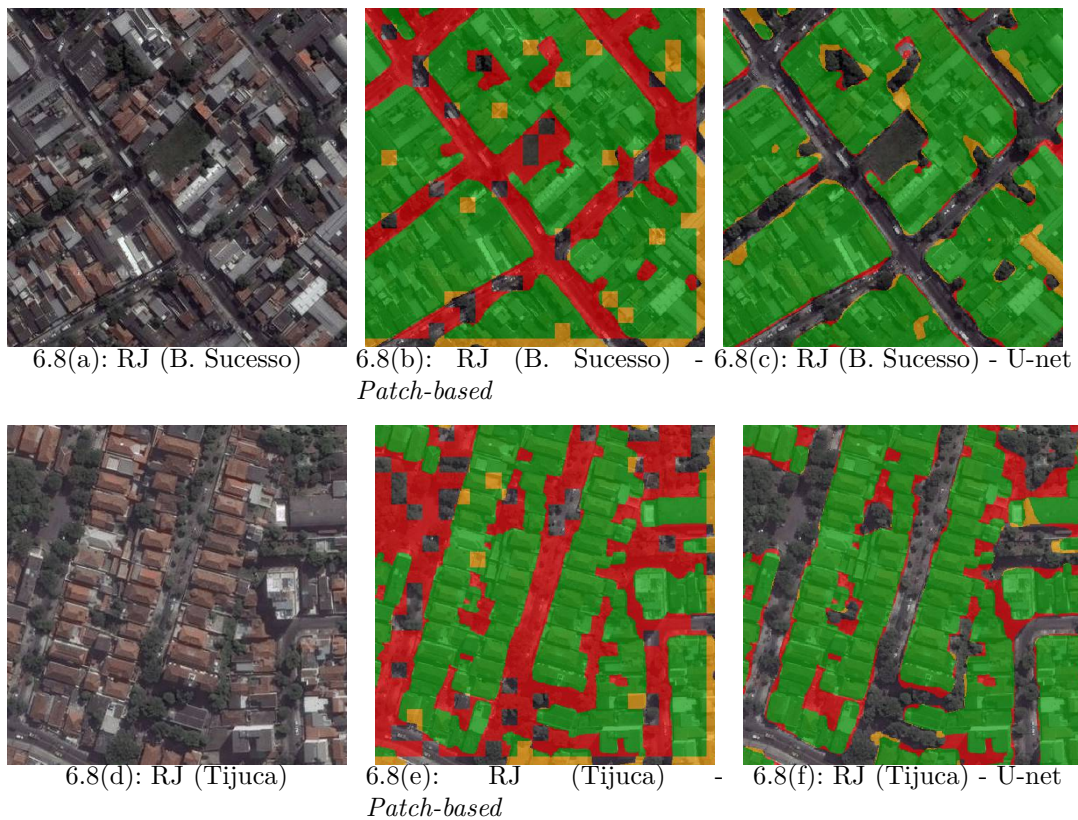


Figura 6.8: Imagem RGB e imagens classificadas pelo *Patch-based* e U-Net em duas regiões: RJ (B. Sucesso) e RJ (Tijuca). Em verde - (True Positives), em vermelho - *False Positives* e em amarelo - *False Negatives*

O modelo *Patch-based* proposto por Guo et al. [3], não conseguiu classificar adequadamente as imagens. Na aplicação original do artigo, haviam apenas imagens contíguas que foram utilizadas para treinamento e também pelo teste. Isto reduz a variabilidade e consequentemente facilita o aprendizado e a classificação, porém deixa bem fraca a possibilidade de generalização do modelo. No problema proposto, apesar do grande número de amostras para treinamento, a performance não foi boa. Quase a totalidade das imagens



Figura 6.9: Imagem RGB e imagens classificadas pelo *Patch-based* e U-Net em duas regiões: Rio Bonito - RJ e RJ (Botafogo). Em verde - (True Positives), em vermelho - *False Positives* e em amarelo - *False Negatives*

foram preditas como área construída, com grande número de falsos positivos. Em casos onde há uma grande densidade de edificações, a precisão não reflete adequadamente o desempenho do classificador, pois a proporção de área sem edificação é pequena e a variação de falsos positivos pouco influencia na precisão. Como exemplo, no caso da Figura 6.96.8(b), que tem aproximadamente 80% de área construída, a precisão foi de 82,2% e a proporção de falsos negativos foi de 76,7%. Para a mesma área o U-Net teve, 9,6%, de falsos positivos e precisão de 95,7%, qualidade bem superior como pode ser visto na Figura 6.96.8(c). Como observado nos resultados apresentados, o modelo *Patch-based*, de classificação em *patches*, não performou bem. A configuração simples, com três camadas de convolução, número pequeno de filtros e os tamanho dos *patches*, não foi suficiente para extrair atributos eficientes para a correta predição. Outras implementações, com variação de tamanho dos *patches* e com modelos mais complexos, poderiam trazer resultados melhores. O que fica evidenciado é que o problema de classificação propostos é mais complexo que o enfrentado em [3] e a U-Net teve boa performance.

Informações precisas sobre a distribuição espacial da população são importantes para a gestão de problemas sociais, econômicos e ambientais relacionados à população. Os governos precisam de dados precisos sobre a população para apoiar o planejamento de projetos de infra-estrutura [97], para a preparação de respostas a desastres naturais, emergências e avaliação de risco e vulnerabilidade [98, 99, 100, 101]. Da mesma forma, o setor privado necessita de dados populacionais para planejar a localização de suas instalações [102], para a otimização de sistemas de prestação de serviços e para avaliações de risco [103]. Informações confiáveis sobre a distribuição da população também são essenciais para avaliar a pressão humana sobre o meio ambiente [104], para quantificar o impacto ambiental na população [105] e para aplicações de saúde pública [44].

A fonte de informação mais completa sobre a população é o censo. Os censos de População e Habitação podem ser definidos como o processo de coletar, compilar, avaliar, analisar e publicar os dados estatísticos, econômicos e sociais, relativos a todas as habitações e seus residentes, em em uma região ou um país. Em muitos países, particularmente nos países em desenvolvimento, os censos são a cada dez anos. Os dados populacionais do censo são disponibilizados ao público de forma agregada como anuários estatísticos, geralmente divididos em áreas político-administrativas [106]. Tomando como base o modelo de censo adotado pela maioria dos países, entre eles, o Brasil, a unidade básica de interesse é o indivíduo e, em um nível de agregação superior, a família, tendo como referência o domicílio. Logo, o domicílio é a menor unidade para localização espacial. A disponibilização da geolocalização destes domicílios possibilitaria que os dados fossem trabalhados em todas as escalas e em qualquer tipo de compartimentação espacial [107]. Contudo, existem restrições à divulgação de dados desta forma devido a questões legais relacionadas com a proteção da identidade dos informantes ou confidencialidade. Então a divulgação é disponibilizada em agrupamentos de domicílios denominados setores censitários. Estes agrupamentos são unidades operacionais homogêneas de tamanho e formas irregulares, e têm o objetivo de facilitar o planejamento e o gerenciamento da operação censitária, mas não

são adequadas para a divulgação dos dados.

Conforme descrito em [4], os problemas na utilização de setores são diversos. A resolução espacial dos dados agregados é variável e baixa, exceto nas áreas urbanas mais densamente povoadas. Os setores e as unidades político-administrativas podem ter seus limites alterados a qualquer momento, o que afeta a comparação dos dados ao longo do tempo [108, 109, 110, 111]. Por fim, pode haver a inadequação entre a unidade geográfica em que os dados demográficos são divulgados e a unidade geográfica de interesse dos usuários [112, 107, 113, 114, 115].

De modo geral, as propriedades dos dados agregados baseados em setores são limitadas para a análise espacial em problemas socioeconômicos e ambientais relacionados com a população. Assim, uma possibilidade é a utilização de dados sobre a população na forma de uma grade espacial. Este é um formato alternativo para fornecer dados populacionais. A grade de populações é uma rede georreferenciada de células quadradas. Para cada célula há o valor da contagem de população ou o valor da densidade populacional para aquela localização. A geração de um sistema de grades com objetivos estatísticos consiste em uma operação de mudança de suporte espacial de dados e tem as vantagens de que todas as células têm o mesmo tamanho, são estáveis ao longo do tempo e propiciam uma maior facilidade de obter recortes espaciais mais adequados aos objetivos de estudo.

As técnicas para esta transferência de suporte podem ser classificadas de acordo com três diferentes abordagens [116, 117]: agregação, desagregação e interpolação. A agregação acumula os dados das unidades de origem contidas nas unidades de destino. Na interpolação [114, 113], uma superfície de densidade é estimada usando dados auxiliares. Na desagregação os dados são realocados em subespaços contidos no espaço ou suporte original. O método mais conhecido para desagregação é a modelagem dasimétrica [118, 119, 120]. Neste método, a estimativa da variável, no novo suporte (destino), é dada por uma função que relaciona a variável entres os suportes de origem e destino. Na sua versão mais simples, conhecida como ponderação zonal simples ou método *overlay* [121] é calculada de acordo com a equação abaixo:

$$\hat{P}_o = \sum_{o=1}^n P_t \frac{A_{o \cap d}}{A_o}. \quad (7-1)$$

Cada área de origem (A_o) contribui com uma fração correspondente ao percentual da área de interseção com as regiões de destino correspondentes($A_{o \cap d}$)[114].

Com relação à construção de uma grade populacional, a modelagem pode ser descrita como uma técnica de desagregação de dados de população para reagregação em células de grade em uma resolução espacial mais alta (*downscaling*) utilizando dados auxiliares que se correlacionam com a densidade populacional.

A modelagem dasimétrica tem sido extensivamente estudada [122], incorporando vários dados auxiliares e variando na sofisticação do uso da ponderação de áreas [123, 124], a fim de aumentar a precisão do modelo. Os dados ancilares usados com maior frequência são os dados de cobertura da terra / uso da terra [118, 125, 98].

Imagens de satélite de alta resolução foram recentemente utilizadas para gerar dados auxiliares [126, 127]. No trabalho de Lung et al. [127], o autor utilizou imagens VHR (*Very High Resolution*) e OBIA para identificar edificações mostrando que esta é uma alternativa com bons resultados principalmente para locais onde os dados auxiliares usuais (produtos cartográficos) não estão disponíveis. Também já foram empregadas informações de infraestrutura local, como a densidade de ruas [128], ou a densidade de pontos de interesse extraídos do OpenStreetMap [129]. Dados de arrecadação de impostos por [130, 131, 132, 133, 134, 4], a emissão de luz em [135] e cadastros de endereços em [136, 137]. Apesar do rápido progresso no desenvolvimento dos métodos dasimétricos, a adoção prática das grades de população é baixa. Isso ocorre porque a maioria dos potenciais usuários só são capazes de utilizar o produto pronto. Então, para aumentar o uso de dados demográficos para análise espacial, é preciso que estas estimativas, para grandes regiões e em alta resolução, estejam disponíveis em domínio público [4].

No Brasil, o IBGE já faz a divulgação dos dados de população em gride. Na metodologia adotada, [138, 117], utiliza de vários métodos dependendo do tipo e qualidade das informações disponíveis. Apesar do emprego de GPS para coleta dos dados, ainda não há registro de geolocalização ou identificação de endereço para muitas regiões do país, inviabilizando uso da agregação para composição do gride. Então, foi adotada uma metodologia híbrida que mescla métodos de agregação e desagregação conforme a disponibilidade de informação. No método de desagregação, conforme os estudos empíricos realizados por Bueno [117], foi constatado que o método dasimétrico com dados auxiliares de vias apresenta o melhor desempenho. O segundo melhor foi o método dasimétrico com dados auxiliares de classificação de uso/cobertura das terras e como última opção, quando não há disponibilidade de dados auxiliares, foi o método de ponderação zonal. Desta forma, na construção do gride de

população o método de modelagem dasimétrico tem papel fundamental e os dados auxiliares influenciam diretamente na qualidade dos resultados obtidos. O uso de métodos computacionais para prover estes dados auxiliares podem trazer ganhos significativos, baixando os custos e viabilizando o emprego em grande escala. Considerando que o domicílio é a unidade básica para a contagem da população, o método apresentado para localização de edificações com imagens RGB do GE tem emprego promissor como dado auxiliar no cálculo dasimétrico.

Para mostrar o desempenho do mapeamento de edificações desenvolvido com a U-Net, será feita uma comparação das ponderações obtidas considerando o arruamento.

Ambos os métodos têm restrição semelhante. Por exemplo, pode haver ruas ou edificações em áreas não residenciais. Contudo, onde há população há edificações, o que não é sempre verdade com o arruamento. Há ruas em áreas sem edificações e há edificações em áreas sem ruas.

Nesta comparação, não usaremos o cômputo final da densidade populacional porque não temos o valor de referência da população para comparação e aferição da qualidade. Isto não afeta a validade da comparação, pois estaremos comparando a fração atribuída a cada parcela por cada um dos dados auxiliares, vias e área construída.

Utilizando as imagens de teste já com a classificação de as áreas construídas, foram acrescentadas anotações do arruamento. Estas imagens foram subdivididas em quatro partes e para cada uma foi calculado a proporção de arruamento na sub-região com relação ao total de arruamento na imagem. De forma semelhante foram calculadas a proporção de área edificada predita e a proporção de área edificada anotada. A Figura 7.1 mostra duas imagens como exemplo. Em azul estão as áreas com edificações preditas pelo U-Net. Em preto, foram anotadas manualmente as ruas. A marcação em vermelho mostra a divisão das sub-regiões.

O experimento foi executado com 42 imagens. A comparação foi feita observando a diferença entre as proporções calculadas nas diferentes formas em cada sub-região. O principal objetivo é comparar as diferenças entre o as proporções obtidas pelo arruamento e a área construída predita. Os resultados estão na Tabela 7.1 e na Figura 7.3.

Na Tabela 7.1 e na Figura 7.3 estão três comparações. A primeira compara os resultados obtidos com a diferença das área construída predita ("Pred") e anotada ("Mask"). Tem valor baixo, o que era de se esperar considerando a qualidade da predição, já discutida em capítulos anteriores. Ainda, devido a proximidades destas medidas, as outras duas medidas que

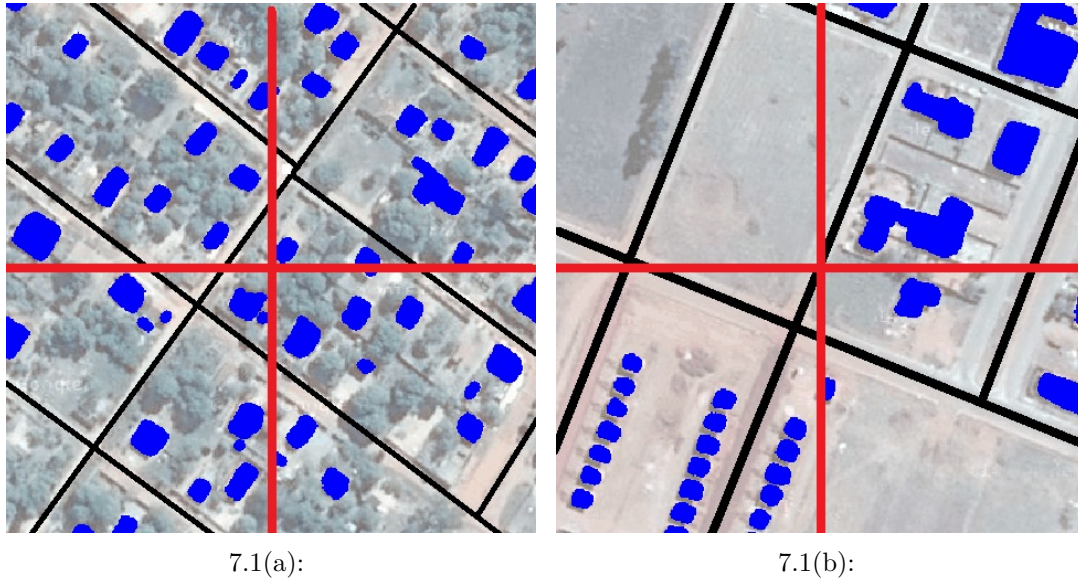


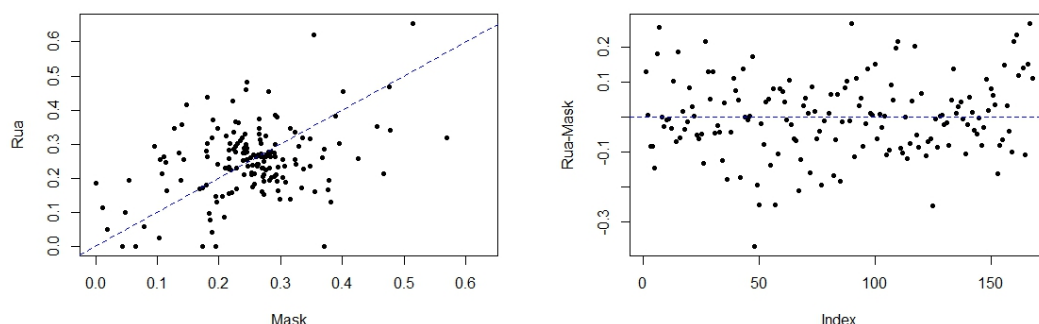
Figura 7.1: Dispersão do arruamento e de áreas construídas. Áreas construídas anotadas em azul. Rua anotadas em preto. Imagens subdivididas em quatro regiões. *Imagens Google Earth.*

Comparação	Média (sd)
Pred-Mask	2,3% (2,5%)
Rua-Mask	8,4% (7,0%)
Rua-Pred	8,2% (7,0%)

Tabela 7.1: Média e desvio padrão das diferenças absolutas entre as proporções de arruamento, áreas construídas preditas(U-Net) e áreas construídas anotadas.

comparam os resultados obtidos com proporção de arruamento ("Rua") ficam com distribuição semelhante. Neste caso, vamos nos ater a análise da comparação das diferenças entre as proporções obtidas pelo arruamento e a área construída anotada, considerada como referência. Na Figura 7.27.2(a) observamos a baixa correlação ente as medidas, indicando que para as diversas sub-regiões as proporções obtidas pelos dois métodos não coincidem. Na Figura 7.27.2(b) a dispersão da diferença entre as proporções calculadas se distribui aleatoriamente em torno de zero. Pela Figura 7.3, observa-se que tem distribuição simétrica, com muita dispersão e poucos valores extremos. Isto indica a ausência de de qualquer padrão sistemático entre as medidas. A diferença observada entre os dois procedimento mostra que as medidas não se correlacionam. Embora, o método com o arruamento seja regulamente utilizado e tido como eficiente, há diversas situações onde falha. Em locais onde haja arruamento e não tenha população o método irá atribuir valor, como mostra a Figura 7.17.1(b). No quadrante superior esquerdo, pelo método de arruamento

se obteve 19%, enquanto que o de área construída, o Mask, foi 0%.



7.2(a): Corelação entre Rua e Mask

7.2(b): Dispersão da diferença entre Rua e Mask

Figura 7.2: Dispersão das proporções obtidas nas imagens de teste pelo arruamento (Rua) e pela área de construções anotada (Mask)

Outro caso, pode ser visto na Figura 7.47.4(a), onde no quadrante superior esquerdo há pouco arruamento, 11,4%, e praticamente nenhuma construção. O Mask foi de 1,1%. No quadrante inferior, há mais vias e a proporção foi de 19%. Neste mesmo quadrante o Mask foi de 5%. Na Figura 7.47.4(b), no quadrante superior esquerdo, onde praticamente não há vias, a proporção obtido pelo Rua foi de 4%. Visivelmente, há varias construções nesta área e o Mask foi mais coerente com valor de 18%.

Outra questão relevante é a existência de arruamento digitalizado para áreas não urbanizadas. A inexistência destes produtos é devido à baixa demanda e à dificuldade de identificação das vias. Por exemplo, na Figura 7.57.5(a) e 7.5(b), é difícil distinguir as vias públicas de outros caminhos privados ou que interligam pequenas propriedades. Isto muitas vezes inviabiliza o mapeamento deixando, muitas áreas descobertas.

Essa prova de conceito deu um resultado bem promissor. Foi mostrado que a classificação de áreas construídas pode gerar estimativas dasimétricas mais precisas que o método por ponderação de vias. Foram mostrados alguns exemplos onde as divergências entre os ponderadores foram grandes. Presumindo que a existência de população fixa implica na ocorrência de construções, as estimativas por área construída é mais coerente. Outra vantagem é a possibilidade da construção do próprio mapeamento em áreas de maior necessidade evitando a utilização de métodos menos eficientes.

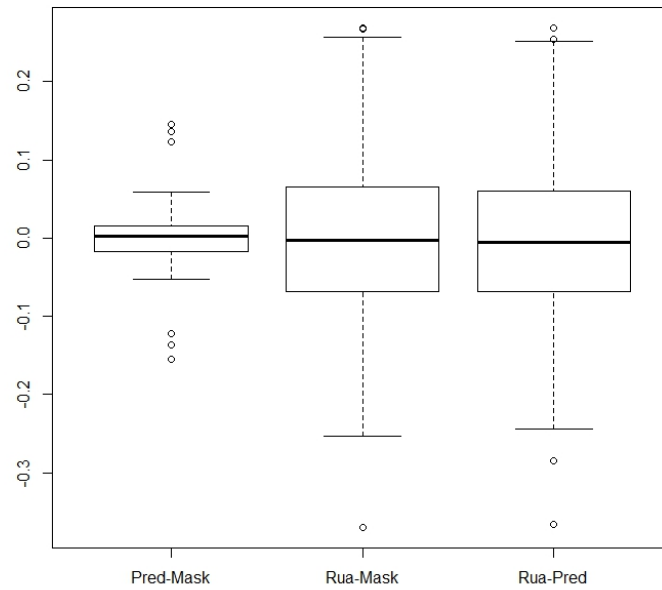
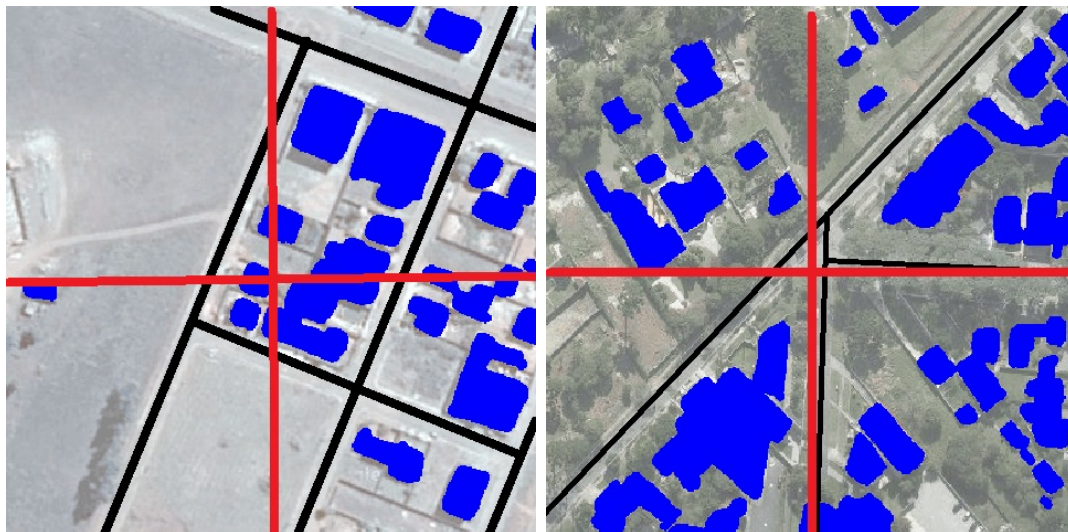


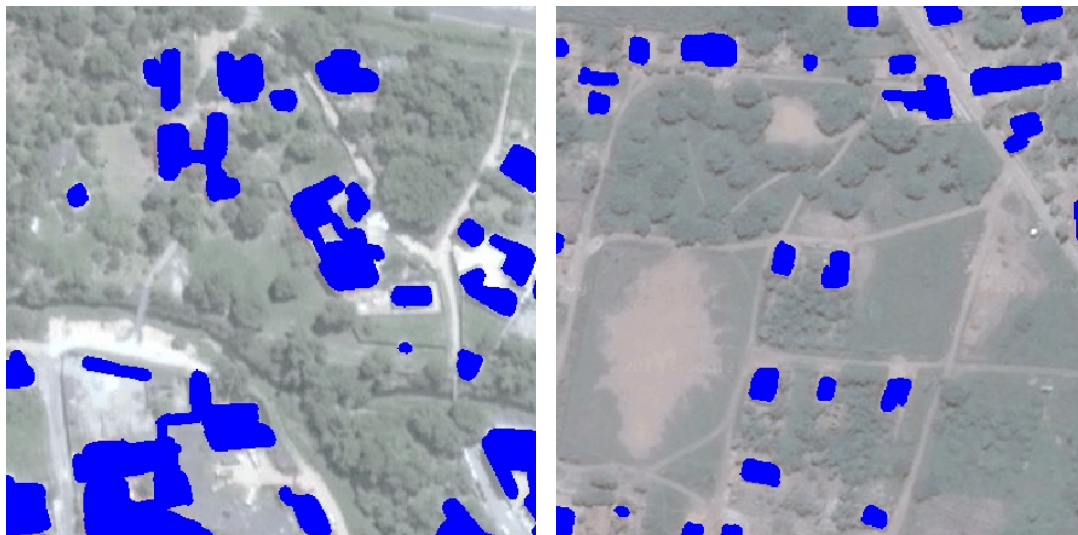
Figura 7.3: Boxplot das diferenças entre: áreas construídas preditas (U-Net) e áreas construídas anotadas (Pred-Mask); proporções de arruamento e áreas construídas anotadas (Rua-Mask); proporções de arruamento e áreas construídas preditas com U-Net(Rua-Pred).



7.4(a):

7.4(b):

Figura 7.4: Divergências entre a proporção de vias e área construída. Áreas construídas anotadas(azul). Vias anotadas em preto. Imagens subdivididas em quatro regiões. *Imagens Google Earth.*



7.5(a):

7.5(b):

Figura 7.5: Área rural. Dificil identificação de vias públicas e vias particulares para mapeamento digital. Áreas construídas anotadas (azul). *Imagens Google Earth.*

Neste trabalho, investigamos o uso da U-net para identificar a área de construção em imagens RGB do Google Earth no Brasil. A U-net mostra um resultado muito promissor, considerando que usamos um modelo único para executar a tarefa em um conjunto de imagens com grande variação intra-classe, de *background* e também em termos de qualidade de imagem. Não se tem registro, pela literatura investigada, de experimento semelhante. Além disso, os resultados foram obtidos usando um conjunto relativamente pequeno de imagens para treinamento. A U-Net foi mais versátil e poderosa com relação ao método que utiliza classificador em *patch*. A estrutura do U-Net, com imagem de entrada de 572×572 e com as camadas de *downscaling* associadas às de *upscaling*, traz mais possibilidade no aprendizado de atributos abstratos. A saída de 388×388 *pixel*, já segmentada pela classificação, e com a rapidez de processamento, menos de 1s por imagem nos testes realizados utilizando GPU, permite aplicações eficientes em larga escala. A possibilidade de bons resultados com conjuntos de dados relativamente pequenos e com a baixa exigência de padronização das imagens, facilita a construção de bancos de imagens anotadas para diversas aplicações.

Isto mostra a viabilidade prática de aplicações em larga escala em países como o Brasil, que tem grande extensão e muitas diferenças regionais.

Apesar dos bons resultados, há problemas nas classificações que são relevantes e comuns entre outros classificadores. As duas principais são a classificação incorreta de áreas pavimentadas e os erros causados pelo sombreamento. No primeiro caso, pátios e estacionamentos são classificados como área construída. Estas áreas são difíceis de classificar pois têm valores espectrais e textura semelhantes às coberturas de edificações. No segundo caso, a dificuldade é que o objeto, ou parte dele, pode estar oculto devido a sombra. Neste caso não há informação espectral ou textura para a classificação. Em ambos os casos há a necessidade de captura do contexto local, um atributo abstrato que a U-Net deveria aprender. Para melhorar estas classificações difíceis e obter melhor generalização há a necessidade de aumentar o conjunto de imagens anotadas, principalmente nos casos citados. Na implementação original do U-Net [83] foi aplicada a deformação elástica

sobre as imagens a fim de aumentar artificialmente o número de imagens para treinamento. Na aplicação para classificação de áreas edificadas estes processos que causam deformação geométrica não são aplicáveis, pois a forma dos objetos é importante na caracterização. Outro fato é que comparado as imagens biomédicas utilizadas originalmente, as imagens de sensoriamento remoto são mais complexas. Com isto, utilizamos uma amostra de treinamento bem maior que a do artigo original.

O método tradicional, o *Patch Based*, comparado ao U-Net, tem desvantagem. Está diretamente relacionado com o tamanho do *patch*. Quanto maior o *patch*, maior é a informação de contexto. Contudo torna-se mais lento, pois deve computar separadamente cada *patch* e existe muita redundância devido a sobreposição dos *patches* na estratégia de segmentação.

Considerando a complexidade da U-Net, pretende-se testar algumas simplificações na rede, como a redução do números de filtros e camadas. Isto poderia trazer ganhos no tempo de treinamento e qualidade da classificação considerando amostras pequenas.

No conjunto de imagens anotadas, não foi utilizado nenhum critério específico para avaliação dos erro de anotação. Isto pode impactar no processo de treinamento do modelo e na avaliação da precisão de classificação. Um método de crítica e avaliação das anotações em imagens traria maior confiabilidade nos resultados

Esse trabalho também mostrou uma aplicação, prova de conceito, em métodos dasimétricos para estimação da população. A utilização das áreas de construídas permite o cálculo de pesos mais precisos, consequentemente uma melhor precisão das estimativas nas sub-regiões. Nesta aplicação, a dasimetria foi sempre relacionada com a estimativa da população, contudo o conceito é válido para outras variáveis que tenham os domicílios como unidade de referência. Estes resultados bastante promissores com a possibilidades de obtenção do mapeamento de áreas edificada em larga escala, possibilita o aprimoramento do método de estimativas da população em gride feita pelo IBGE.

O mapeamento de áreas construídas serve a muitas outras aplicações como a localização de novos assentamentos, avaliação de expansão de assentamentos e áreas urbanas. No caso de adaptação para a detecção de outro objetos(ex. ruas, veículos, áreas de cultivo), possibilita outros empregos como no mapeamento de usos da terra.

Em outro contexto, pretende-se usar a U-net para resolver outra tarefa mais abstrata. O objetivo e associar a imagem das construções a uma classe do Índice de Desenvolvimento Humano, a fim de obtenção do mapeamento

como mostra a Figura 8.1 para o caso da Cidade do Rio Janeiro. Com isso, o problema em questão seria um classificação multi-classe ao invés de uma classificação binária.

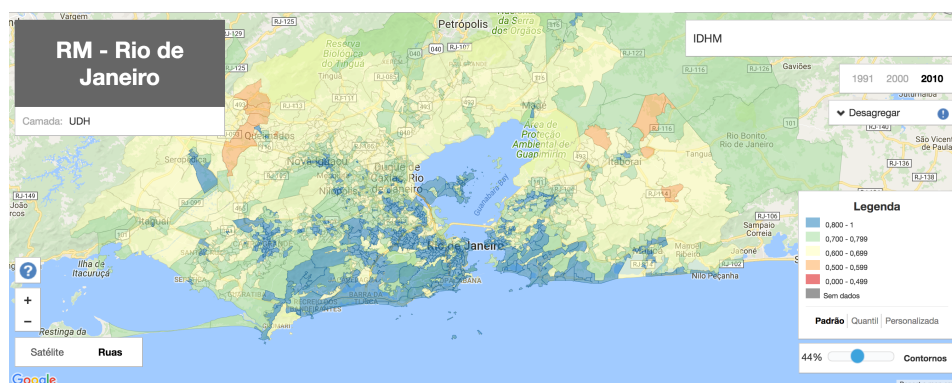


Figura 8.1: Mapa do Índice de Desenvolvimento Humano Municipal da Cidade do Rio de Janeiro, fornecido pela ONG Atlas do Desenvolvimento Humano do Brasil (<http://www.atlasbrasil.org.br/>).

Referências bibliográficas

- [1] KARPATY, A.. Stanford cs231n: Convolutional neural networks for visual recognition. <http://cs231n.github.io/>, 2015.
- [2] GOODFELLOW, I.; BENGIO, Y. ; COURVILLE, A.. Deep Learning. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [3] GUO, Z.; SHAO, X.; XU, Y.; MIYAZAKI, H.; OHIRA, W. ; SHIBASAKI, R.. Identification of village building via google earth images and supervised machine learning methods. Remote Sensing, 8(4):271, 2016.
- [4] DMOWSKA, A.; STEPINSKI, T. F.. A high resolution population grid for the conterminous United States: The 2010 edition. Computers, Environment and Urban Systems, 61:13–23, 2017.
- [5] SHU, Y.. Deep Convolutional Neural Networks for Object Extraction from High Spatial Resolution Remotely Sensed Imagery, 2014.
- [6] BUTLER, D.. Many eyes on earth. Nature, 505(7482):143–144, 2014.
- [7] SIERAWSKI, B. D.; REED, R. A.; WARREN, K. M.; STERNBERG, A. L.; AUSTIN, R. A.; TRIPPE, J. M.; WELLER, R. A.; ALLES, M. L.; SCHRIMPF, R. D.; MASSENGILL, L. W. ; OTHERS. Cubesat: Real-time soft error measurements at low earth orbits. In: RELIABILITY PHYSICS SYMPOSIUM (IRPS), 2017 IEEE INTERNATIONAL, p. 3D–1. IEEE, 2017.
- [8] COLOMINA, I.; MOLINA, P.. ISPRS Journal of Photogrammetry and Remote Sensing Unmanned aerial systems for photogrammetry and remote sensing : A review. ISPRS Journal of Photogrammetry and Remote Sensing, 92:79–97, 2014.
- [9] BLASCHKE, T.. Object based image analysis for remote sensing. ISPRS journal of photogrammetry and remote sensing, 65(1):2–16, 2010.

- [10] LU, D.; WENG, Q.. **A survey of image classification methods and techniques for improving classification performance.** International Journal of Remote Sensing, 28(5):823–870, mar 2007.
- [11] CHENG, G.; HAN, J.. **A survey on object detection in optical remote sensing images.** ISPRS Journal of Photogrammetry and Remote Sensing, 117:11–28, 2016.
- [12] MISHRA, A.; PANDEY, A. ; BAGHEL, A. S.. **Building detection and extraction techniques: A review.** In: COMPUTING FOR SUSTAINABLE GLOBAL DEVELOPMENT (INDIACOM), 2016 3RD INTERNATIONAL CONFERENCE ON, p. 3816–3821. IEEE, 2016.
- [13] KARPATNE, A.; JIANG, Z.; VATSAVAI, R. R.; SHEKHAR, S. ; KUMAR, V.. **Monitoring land-cover changes: A machine-learning perspective.** IEEE Geoscience and Remote Sensing Magazine, 4(2):8–21, 2016.
- [14] DORNAIKA, F.; MOUJAHID, A.; EL, Y. ; RUICHEK, Y.. **Building detection from orthophotos using a machine learning approach : An empirical study on image segmentation and descriptors.** Expert Systems With Applications, 58:130–142, 2016.
- [15] LIASIS, G.; STAVROU, S.. **Building extraction in satellite images using active contours and colour features.** International Journal of Remote Sensing, 37(5):1127–1153, 2016.
- [16] HINTON, G. E.; OSINDERO, S. ; TEH, Y.-W.. **A fast learning algorithm for deep belief nets.** Neural computation, 18(7):1527–1554, 2006.
- [17] BENGIO, Y.; LAMBLIN, P.; POPOVICI, D. ; LAROCHELLE, H.. **Greedy layer-wise training of deep networks.** In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 153–160, 2007.
- [18] KRIZHEVSKY, A.; SUTSKEVER, I. ; HINTON, G. E.. **Imagenet classification with deep convolutional neural networks.** In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 1097–1105, 2012.
- [19] GIRSHICK, R.; DONAHUE, J.; DARRELL, T. ; MALIK, J.. **Rich feature hierarchies for accurate object detection and semantic segmentation.** In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 580–587, 2014.

- [20] ALMEIDA, C. F. P.; FERNANDES, W. P. D.; BARBOSA, S. D. J. ; LOPES, H.. Using u-net to identify building areas on google earth images from brazil - annotated images. <http://www.ideias.inf.puc-rio.br/GEimages.html>, 2017.
- [21] CHAUDHURI, D.; KUSHWAHA, N. ; SAMAL, A.. Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques. IEEE journal of selected topics in applied earth observations and remote sensing, 5(5):1538–1544, 2012.
- [22] ZHANG, J.; LIN, X.; LIU, Z. ; SHEN, J.. Semi-automatic road tracking by template matching and distance transformation in urban areas. International journal of remote sensing, 32(23):8331–8347, 2011.
- [23] LEFÈVRE, S.; WEBER, J.. Automatic building extraction in vhr images using advanced morphological operators. In: URBAN REMOTE SENSING JOINT EVENT, 2007, p. 1–5. IEEE, 2007.
- [24] STANKOV, K.; HE, D.-C.. Building detection in very high spatial resolution multispectral images using the hit-or-miss transform. IEEE Geoscience and Remote Sensing Letters, 10(1):86–90, 2013.
- [25] STANKOV, K.; HE, D.-C.. Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(10):4069–4080, 2014.
- [26] JAIN, A. K.; ZHONG, Y. ; DUBUISSON-JOLLY, M.-P.. Deformable template models: A review. Signal processing, 71(2):109–129, 1998.
- [27] PENG, J.; ZHANG, D. ; LIU, Y.. An improved snake model for building detection from urban aerial images. Pattern Recognition Letters, 26(5):587–595, 2005.
- [28] AHMADI, S.; ZOEJ, M. V.; EBADI, H.; MOGHADDAM, H. A. ; MOHAMMADZADEH, A.. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. International Journal of Applied Earth Observation and Geoinformation, 12(3):150–157, 2010.
- [29] SIRMACEK, B.; UNSALAN, C.. Urban-area and building detection using sift keypoints and graph theory. IEEE Transactions on Geoscience and Remote Sensing, 47(4):1156–1167, 2009.

- [30] BALTSAVIAS, E.. **Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems.** ISPRS journal of photogrammetry and remote sensing, 58(3):129–151, 2004.
- [31] BAATZ, M.. **Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation.** *Angewandte geographische informationsverarbeitung*, p. 12–23, 2000.
- [32] BLASCHKE, T.. **Object-based contextual image classification built on image segmentation.** In: *ADVANCES IN TECHNIQUES FOR ANALYSIS OF REMOTELY SENSED DATA*, 2003 IEEE WORKSHOP ON, p. 113–119. IEEE, 2003.
- [33] BLASCHKE, T.; HAY, G. J.; KELLY, M.; LANG, S.; HOFMANN, P.; ADDINK, E.; FEITOSA, R. Q.; VAN DER MEER, F.; VAN DER WERFF, H.; VAN COILLIE, F. ; OTHERS. **Geographic object-based image analysis—towards a new paradigm.** *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191, 2014.
- [34] HAY, G. J.; CASTILLA, G.. **Geographic object-based image analysis (geobia): A new name for a new discipline.** *Object-based image analysis*, p. 75–89, 2008.
- [35] ESCH, T.; THIEL, M.; BOCK, M.; ROTH, A. ; DECH, S.. **Improvement of image segmentation accuracy based on multiscale optimization procedure.** *IEEE Geoscience and Remote Sensing Letters*, 5(3):463–467, 2008.
- [36] KIM, M.; WARNER, T. A.; MADDEN, M. ; ATKINSON, D. S.. **Multi-scale geobia with very high spatial resolution digital aerial imagery: scale, texture and image objects.** *International Journal of Remote Sensing*, 32(10):2825–2850, 2011.
- [37] GAO, Y.; MAS, J. F.; KERLE, N. ; NAVARRETE PACHECO, J. A.. **Optimal region growing segmentation and its effect on classification accuracy.** *International journal of remote sensing*, 32(13):3747–3763, 2011.
- [38] ARDILA, J. P.; BIJKER, W.; TOLPEKIN, V. A. ; STEIN, A.. **Context-sensitive extraction of tree crown objects in urban areas using vhr satellite images.** *International journal of applied earth observation and geoinformation*, 15:57–69, 2012.

- [39] DRĂGUȚ, L.; CSILLIK, O.; EISANK, C. ; TIEDE, D.. **Automated parameterisation for multi-scale image segmentation on multiple layers**. ISPRS Journal of Photogrammetry and Remote Sensing, 88:119–127, 2014.
- [40] IM, J.; QUACKENBUSH, L. J.; LI, M. ; FANG, F.. **Optimum scale in object-based image analysis**. Scale Issues in Remote Sensing, p. 197–214, 2014.
- [41] MING, D.; LI, J.; WANG, J. ; ZHANG, M.. **Scale parameter selection by spatial statistics for geobia: Using mean-shift based multi-scale segmentation as an example**. ISPRS Journal of Photogrammetry and Remote Sensing, 106:28–41, 2015.
- [42] ARBELAEZ, P.; MAIRE, M.; FOWLKES, C. ; MALIK, J.. **Contour detection and hierarchical image segmentation**. IEEE transactions on pattern analysis and machine intelligence, 33(5):898–916, 2011.
- [43] ARBELAEZ, P.; PONT-TUSET, J.; BARRON, J. T.; MARQUES, F. ; MALIK, J.. **Multiscale combinatorial grouping**. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 328–335, 2014.
- [44] HAY, S.; NOOR, A.; NELSON, A. ; TATEM, A.. **The accuracy of human population maps for public health application**. Tropical Medicine & International Health, 10(10):1073–1086, 2005.
- [45] DRĂGUȚ, L.; EISANK, C.. **Automated object-based classification of topography from srtm data**. Geomorphology, 141:21–33, 2012.
- [46] ANDREOPOULOS, A.; TSOTSOS, J. K.. **50 years of object recognition: Directions forward**. Computer Vision and Image Understanding, 117(8):827–891, 2013.
- [47] HUANG, Y.; WU, Z.; WANG, L. ; TAN, T.. **Feature coding in image classification: A comprehensive study**. IEEE transactions on pattern analysis and machine intelligence, 36(3):493–506, 2014.
- [48] VAKALOPOULOU, M.; KARANTZALOS, K.; KOMODAKIS, N. ; PARAGIOS, N.. **Building detection in very high resolution multispectral data with deep learning features**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2015 IEEE INTERNATIONAL, p. 1873–1876. IEEE, 2015.

- [49] LI, Y.; WANG, S.; TIAN, Q. ; DING, X.. **Feature representation for statistical-learning-based object detection: A review**. Pattern Recognition, 48(11):3542–3559, 2015.
- [50] BOSER, B. E.; GUYON, I. M. ; VAPNIK, V. N.. **A training algorithm for optimal margin classifiers**. In: PROCEEDINGS OF THE FIFTH ANNUAL WORKSHOP ON COMPUTATIONAL LEARNING THEORY, p. 144–152. ACM, 1992.
- [51] FREUND, Y.; SCHAPIRE, R. E.. **A decision-theoretic generalization of on-line learning and an application to boosting**. In: EUROPEAN CONFERENCE ON COMPUTATIONAL LEARNING THEORY, p. 23–37. Springer, 1995.
- [52] FREUND, Y.; SCHAPIRE, R. ; ABE, N.. **A short introduction to boosting**. Journal-Japanese Society For Artificial Intelligence, 14(771-780):1612, 1999.
- [53] COVER, T.; HART, P.. **Nearest neighbor pattern classification**. IEEE transactions on information theory, 13(1):21–27, 1967.
- [54] LECUN, Y.; BOTTOU, L.; BENGIO, Y. ; HAFFNER, P.. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, 86(11):2278–2324, 1998.
- [55] HUANG, C.; DAVIS, L. ; TOWNSHEND, J.. **An assessment of support vector machines for land cover classification**. International Journal of remote sensing, 23(4):725–749, 2002.
- [56] DALAL, N.; TRIGGS, B.. **Histograms of oriented gradients for human detection**. In: COMPUTER VISION AND PATTERN RECOGNITION, 2005. CVPR 2005. IEEE COMPUTER SOCIETY CONFERENCE ON, volumen 1, p. 886–893. IEEE, 2005.
- [57] FEI-FEI, L.; PERONA, P.. **A bayesian hierarchical model for learning natural scene categories**. In: COMPUTER VISION AND PATTERN RECOGNITION, 2005. CVPR 2005. IEEE COMPUTER SOCIETY CONFERENCE ON, volumen 2, p. 524–531. IEEE, 2005.
- [58] SENARAS, C.; OZAY, M. ; VURAL, F. T. Y.. **Building detection with decision fusion**. IEEE journal of selected topics in applied earth observations and remote sensing, 6(3):1295–1304, 2013.

- [59] CHEN, Y.; NASRABADI, N. M. ; TRAN, T. D.. **Hyperspectral image classification using dictionary-based sparse representation**. IEEE Transactions on Geoscience and Remote Sensing, 49(10):3973–3985, 2011.
- [60] CHEN, L.; ZHAO, S.; HAN, W. ; LI, Y.. **Building detection in an urban area using lidar data and quickbird imagery**. International Journal of Remote Sensing, 33(16):5135–5148, 2012.
- [61] ZHANG, L.; ZHANG, L.; TAO, D. ; HUANG, X.. **Sparse transfer manifold embedding for hyperspectral target detection**. IEEE Transactions on Geoscience and Remote Sensing, 52(2):1030–1043, 2014.
- [62] ZHANG, Y.; DU, B. ; ZHANG, L.. **A sparse representation-based binary hypothesis model for target detection in hyperspectral images**. IEEE Transactions on Geoscience and Remote Sensing, 53(3):1346–1354, 2015.
- [63] ZHANG, Y.; ZHANG, L.; DU, B. ; WANG, S.. **A nonlinear sparse representation-based binary hypothesis model for hyperspectral target detection**. IEEE Journal of selected topics in applied earth observations and remote sensing, 8(6):2513–2522, 2015.
- [64] VIOLA, P.; JONES, M.. **Rapid object detection using a boosted cascade of simple features**. In: COMPUTER VISION AND PATTERN RECOGNITION, 2001. CVPR 2001. PROCEEDINGS OF THE 2001 IEEE COMPUTER SOCIETY CONFERENCE ON, volumen 1, p. I–I. IEEE, 2001.
- [65] KOSAKA, N.; AKIYAMA, T.; TSAI, B. ; KOJIMA, T.. **Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2005. IGARSS'05. PROCEEDINGS. 2005 IEEE INTERNATIONAL, volumen 4, p. 2980–2983. IEEE, 2005.
- [66] SOHN, G.; DOWMAN, I.. **Data fusion of high-resolution satellite imagery and lidar data for automatic building extraction**. ISPRS Journal of Photogrammetry and Remote Sensing, 62(1):43–63, 2007.
- [67] GONG, B.; IM, J. ; MOUNTRAKIS, G.. **An artificial immune network approach to multi-sensor land use/land cover classification**. Remote Sensing of Environment, 115(2):600–614, 2011.

- [68] BRUZZONE, L.; CARLIN, L.. **A multilevel context-based system for classification of very high spatial resolution images.** IEEE transactions on Geoscience and Remote Sensing, 44(9):2587–2600, 2006.
- [69] HUANG, X.; ZHANG, L.. **Morphological building/shadow index for building extraction from high-resolution imagery over urban areas.** IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1):161–172, 2012.
- [70] ZHANG, P.; LV, Z. ; SHI, W.. **Object-based spatial feature for classification of very high resolution remote sensing images.** IEEE Geoscience and Remote Sensing Letters, 10(6):1572–1576, 2013.
- [71] ZHANG, H.; SHI, W.; WANG, Y.; HAO, M. ; MIAO, Z.. **Classification of very high spatial resolution imagery based on a new pixel shape feature set.** IEEE Geoscience and Remote Sensing Letters, 11(5):940–944, 2014.
- [72] BENGIO, Y.. **Learning Deep Architectures for AI.** Foundations and Trends® in Machine Learning, 2(1):1–127, 2009.
- [73] ZHANG, Q.; WANG, Y.; LIU, Q.; LIU, X. ; WANG, W.. **Cnn based suburban building detection using monocular high resolution google earth images.** In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2016 IEEE INTERNATIONAL, p. 661–664. IEEE, 2016.
- [74] DENG, L.; YU, D.. **Deep Learning Methods and Applications.** Foundations and Trends® in Signal Processing, 7(3-4), 2013.
- [75] LECUN, Y.; KAVUKCUOGLU, K. ; FARABET, C.. **Convolutional networks and applications in vision.** ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems, p. 253–256, 2010.
- [76] CIREŞAN, D. C.; MEIER, U.; MASCI, J.; GAMBARDELLA, L. M. ; SCHMIDHUBER, J.. **Flexible, high performance convolutional neural networks for image classification.** IJCAI International Joint Conference on Artificial Intelligence, p. 1237–1242, 2011.
- [77] DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K. ; FEI-FEI, L.. **Imagenet: A large-scale hierarchical image database.** In: COMPUTER VISION AND PATTERN RECOGNITION, 2009. CVPR 2009. IEEE CONFERENCE ON, p. 248–255. IEEE, 2009.

- [78] FELLBAUM, C.. **Wordnet: An electronic lexical database:** Bradford book, 1998.
- [79] ZEILER, M. D.; FERGUS, R.. **Visualizing and Understanding Convolutional Networks.** In: EUROPEAN CONFERENCE ON COMPUTER VISION, p. 818–833. Springer, 2014.
- [80] SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V. ; RABINOVICH, A.. **Going deeper with convolutions.** In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 1–9, 2015.
- [81] SIMONYAN, K.; ZISSERMAN, A.. **Very Deep Convolutional Networks for Large-Scale Image Recognition.** International Conference on Learning Representations (ICRL), p. 1–14, 2015.
- [82] HE, K.; ZHANG, X.; REN, S. ; SUN, J.. **Deep residual learning for image recognition.** In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 770–778, 2016.
- [83] RONNEBERGER, O.; FISCHER, P. ; BROX, T.. **U-net: Convolutional networks for biomedical image segmentation.** In: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION, p. 234–241. Springer, 2015.
- [84] ZHANG, L.; ZHANG, L. ; KUMAR, V.. **Deep learning for remote sensing data.** IEEE Geoscience and Remote Sensing Magazine, (6):18, 2016.
- [85] KAMPFFMEYER, M.; SALBERG, A.-B. ; JENSSEN, R.. **Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks.** In: COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW), 2016 IEEE CONFERENCE ON, p. 1–9. IEEE, 2016.
- [86] ZHANG, L.; SHI, Z. ; WU, J.. **A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery.** IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(10):4895–4909, 2015.

- [87] SAITO, S. .; YAMASHITA, T. . ; AOKI, Y.. Multiple object extraction from aerial imagery with convolutional neural networks multiple object extraction from aerial imagery with convolutional neural networks. *Journal of Imaging Science and Technology*, 1(60):1–9, 2016.
- [88] CASTELLUCCIO, M.; POGGI, G.; SANSONE, C. ; VERDOLIVA, L.. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015.
- [89] ŠEVO, I.; AVRAMOVIĆ, A.. Convolutional neural network based automatic object detection on aerial images. *IEEE Geoscience and Remote Sensing Letters*, 13(5):740–744, 2016.
- [90] YANG, Y.; NEWSAM, S.. Bag-of-visual-words and spatial extensions for land-use classification. In: *PROCEEDINGS OF THE 18TH SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS*, p. 270–279. ACM, 2010.
- [91] PENATTI, O. A. B.; NOGUEIRA, K. ; DOS SANTOS, J. A.. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2015-Octob:44–51, 2015.
- [92] ZHANG, F.; DU, B. ; ZHANG, L.. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1793–1802, 2016.
- [93] MNIH, V.. Machine learning for aerial image labeling. PhD thesis, University of Toronto (Canada), 2013.
- [94] LONG, J.; SHELHAMER, E. ; DARRELL, T.. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015:3431–3440, 2015.
- [95] GERKE, M.. Use of the stair vision library within the isprs 2d semantic labeling benchmark (vaihingen). Technical report, University of Twente, 2015.
- [96] CHOLLET, F.; OTHERS. Keras. <https://github.com/fchollet/keras>, 2015.

- [97] DENG, C.; WU, C. ; WANG, L.. Improving the housing-unit method for small-area population estimation using remote-sensing and gis information. *International Journal of Remote Sensing*, 31(21):5673–5688, 2010.
- [98] LINARD, C.; GILBERT, M. ; TATEM, A. J.. Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*, 76(5):525–538, 2011.
- [99] DOBSON, J. E.; BRIGHT, E. A.; COLEMAN, P. R.; DURFEE, R. C. ; WORLEY, B. A.. Landscan: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing*, 66(7):849–857, 2000.
- [100] MAANTAY, J.; MAROKO, A.. Mapping urban risk: Flood hazards, race, & environmental justice in new york. *Applied Geography*, 29(1):111–124, 2009.
- [101] TENERELLI, P.; GALLEG0, J. F. ; EHRLICH, D.. Population density modelling in support of disaster risk assessment. *International journal of disaster risk reduction*, 13:334–341, 2015.
- [102] MARTIN, D.; WILLIAMS, H. C.. Market-area analysis and accessibility to primary health-care centres. *Environment and Planning A*, 24(7):1009–1019, 1992.
- [103] THIEKEN, A. H.; MÜLLER, M.; KLEIST, L.; SEIFERT, I.; BORST, D. ; WERNER, U.. Regionalisation of asset values for risk analyses. *Natural Hazards and Earth System Science*, 6(2):167–178, 2006.
- [104] WEBER, N.; CHRISTOPHERSEN, T.. The influence of non-governmental organisations on the creation of natura 2000 during the european policy process. *Forest policy and economics*, 4(1):1–12, 2002.
- [105] VINKX, K.; VISÉE, T.. Usefulness of population files for estimation of noise hindrance effects. icao committee on aviation environmental protection. caep/8 modelling and database task force (modtf). In: 4TH MEETING. SUNNYVALE, USA, p. 20–22, 2008.
- [106] ROY, D.; BLASCHKE, T.. A grid-based approach for refining population data in rural areas. *Journal of Geography and Regional ...*, 7(3):47–57, 2014.

- [107] GRASLAND, C.; MADELIN, M.. **The modifiable areas unit problem**. Final report. ESPON–European Spatial Planning Observation Network, 2006.
- [108] HOGAN, D. J.; D'ANTONA, A. D. O. ; CARMO, R. L.. **Dinâmica demográfica recente da amazônia**. *Amazônia: natureza e sociedade em transformação*. São Paulo: EDUSP, p. 71–116, 2008.
- [109] MARTINE, G.; SCHENSUL, D.. **The demography of adaptation to climate change**. UNFPA, IIED, and El Colegio de México, 2013.
- [110] HOLT, J. B.; LO, C. ; HODLER, T. W.. **Dasymetric estimation of population density and areal interpolation of census data**. *Cartography and Geographic Information Science*, 31(2):103–121, 2004.
- [111] RUTHER, M.; LEYK, S. ; BUTTENFIELD, B. P.. **Comparing the effects of an nlcd-derived dasymetric refinement on estimation accuracies for multiple areal interpolation methods**. *GIScience & Remote Sensing*, 52(2):158–178, 2015.
- [112] BALK, D.; YETMAN, G.. **The global distribution of population: evaluating the gains in resolution refinement**. New York: Center for International Earth Science Information Network (CIESIN), Columbia University, 2004.
- [113] GOODCHILD, M. F.; ANSELIN, L. ; DEICHMANN, U.. **A framework for the areal interpolation of socioeconomic data**. *Environment and Planning A*, 25(3):383–397, 1993.
- [114] GOODCHILD, M.; LAM, N.. **Areal interpolation: a variant of the traditional spatial problem**. *Geo-Processing*, 1:297–312, 1980.
- [115] OPENSHAW, S.. **Developing gis-relevant zone-based spatial analysis methods**. *Spatial analysis: modelling in a GIS environment*, p. 55–73, 1996.
- [116] PLUMEJEAUD, C.; PRUD'HOMME, J.; DAVOINE, P.-A. ; GENSEL, J.. **Transferring indicators into different partitions of geographic space**. In: *INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ITS APPLICATIONS*, p. 445–460. Springer, 2010.
- [117] BUENO, M. D. C.. **Grade estatística: uma abordagem para ampliar o potencial analítico de dados censitários**. PhD thesis, UNCAMP, 2014.

- [118] WRIGHT, J. K.. **A method of mapping densities of population: With cape cod as an example.** *Geographical Review*, 26(1):103–110, 1936.
- [119] LANGFORD, M.; UNWIN, D. J.. **Generating and mapping population density surfaces within a geographical information system.** *The Cartographic Journal*, 2013.
- [120] EICHER, C. L.; BREWER, C. A.. **Dasymetric mapping and areal interpolation: Implementation and evaluation.** *Cartography and Geographic Information Science*, 28(2):125–138, 2001.
- [121] LAM, N. S.-N.. **Spatial interpolation methods: a review.** *The American Cartographer*, 10(2):129–150, 1983.
- [122] PETROV, A.. **One Hundred Years of Dasymetric Mapping: Back to the Origin.** *Cartographic Journal*, The, 49(3):256–264, 2012.
- [123] MARTIN, D.; LLOYD, C. ; SHUTTLEWORTH, I.. **Evaluation of gridded population models using 2001 northern ireland census data.** *Environment and Planning A*, 43(8):1965–1980, 2011.
- [124] MENNIS, J.. **Dasymetric mapping for estimating population in small areas.** *Geography Compass*, 3(2):727–745, 2009.
- [125] MENNIS, J.. **Generating surface models of population using dasymetric mapping*.** *The Professional Geographer*, 55(1):31–42, 2003.
- [126] URAL, S.; HUSSAIN, E. ; SHAN, J.. **Building population mapping with aerial imagery and gis data.** *International Journal of Applied Earth Observation and Geoinformation*, 13(6):841–852, 2011.
- [127] LUNG, T.; LÜBKER, T.; NGOCHOCH, J. K. ; SCHAAB, G.. **Human population distribution modelling at regional level using very high resolution satellite imagery.** *Applied Geography*, 41:36–45, 2013.
- [128] REIBEL, M.; BUFALINO, M.. **A test of street weighted areal interpolation using geographic information systems.** *Environment and Planning A*, 37(1):127–139, 2005.
- [129] BAKILLAH, M.; LIANG, S.; MOBASHERI, A.; JOKAR ARSANJANI, J. ; ZIPF, A.. **Fine-resolution population mapping using openstreetmap points-of-interest.** *International Journal of Geographical Information Science*, 28(9):1940–1963, 2014.

- [130] MAANTAY, J. A.; MAROKO, A. R. ; HERRMANN, C.. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (ceds). *Cartography and Geographic Information Science*, 34(2):77–102, 2007.
- [131] KAR, B.; HODGSON, M. E.. A process oriented areal interpolation technique: a coastal county example. *Cartography and Geographic Information Science*, 39(1):3–16, 2012.
- [132] MITSOVA, D.; ESNARD, A.-M. ; LI, Y.. Using enhanced dasymetric mapping techniques to improve the spatial accuracy of sea level rise vulnerability assessments. *Journal of Coastal Conservation*, 16(3):355–372, 2012.
- [133] JIA, P.; QIU, Y. ; GAUGHAN, A. E.. A fine-scale spatial population distribution on the high-resolution gridded population surface and application in alachua county, florida. *Applied Geography*, 50:99–107, 2014.
- [134] JIA, P.; GAUGHAN, A. E.. Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in alachua county, florida. *Applied Geography*, 66:100–108, 2016.
- [135] ZANDBERGEN, P. A.. Dasymetric mapping using high resolution address point datasets. *Transactions in GIS*, 15(s1):5–27, 2011.
- [136] BRIGGS, D. J.; GULLIVER, J.; FECHT, D. ; VIENNEAU, D. M.. Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote sensing of Environment*, 108(4):451–466, 2007.
- [137] SRIDHARAN, H.; QIU, F.. A spatially disaggregated areal interpolation model using light detection and ranging-derived building volumes. *Geographical Analysis*, 45(3):238–258, 2013.
- [138] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Grade estatística*. Technical report, 2016.