

# RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 06/07/2024.



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"

Vinícius Barbosa Henrique

**Extração de edificações por *Deep Learning* e  
combinação de dados LiDAR e imagens ópticas**

Presidente Prudente

2024

Vinícius Barbosa Henrique

**Extração de edificações por *Deep Learning* e combinação de dados LiDAR e imagens ópticas**

Dissertação apresentada ao Programa de Pós Graduação em Ciências Cartográficas (PPGCC) na Faculdade de Ciências e Tecnologia (FCT) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), como parte dos requisitos para a obtenção do título de Mestre em Ciências Cartográficas.

Orientador: Prof. Dr. Maurício Galo

Coorientador: Prof. Dr. Milton Hirokazu Shimabukuro

Presidente Prudente

2024

H519e

Henrique, Vinícius Barbosa

Extração de edificações por Deep Learning e combinação de dados LiDAR e imagens ópticas / Vinícius Barbosa Henrique. -- Presidente Prudente, 2024

95 p. : il., tabs., fotos, mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Maurício Galo

Coorientador: Milton Hirokazu Shimabukuro

1. Aprendizado profundo. 2. Extração de edificações. 3. LiDAR. 4. Modelos de cores. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

## IMPACTO<sup>1</sup> ESPERADO DESTA PESQUISA

O tema central desta Dissertação é a extração de edificações urbanas por meio da combinação de modelos digitais de superfície normalizados e imagens ópticas usando aprendizado profundo. Essa abordagem aprimora a qualidade da identificação de edificações, permitindo a atualização eficiente de informações urbanas, úteis para o planejamento e gestão das cidades, detecção de expansões e uso do solo.

## POTENTIAL IMPACT OF THIS RESEARCH

The central theme of this Dissertation is the extraction of urban buildings through the combination of normalized digital surface models and optical images using deep learning. This approach enhances the quality in building identification, allowing the efficient updating of urban information for city planning and management, including the detection of urban expansion and land use.

---

<sup>1</sup> Informação inserida de acordo com a Portaria Unesp nº 117, de 21 de dezembro de 2022.

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO:** Extração de edificações por *Deep Learning* e combinação de dados LiDAR e imagens ópticas

**AUTOR:** VINÍCIUS BARBOSA HENRIQUE

**ORIENTADOR:** MAURICIO GALO

**COORIENTADOR:** MILTON HIROKAZU SHIMABUKURO

Aprovado como parte das exigências para obtenção do Título de Mestre em Ciências Cartográficas, área: Aquisição, Análise e Representação de Informações Espaciais pela Comissão Examinadora:

  
Prof. Dr. MAURICIO GALO (Participação Presencial)  
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Prof. Dr. HIDEO ARAKI (Participação Virtual)  
Departamento de Geomática / Universidade Federal do Paraná

Prof. Dr. NILTON NOBUHIRO IMAI (Participação Presencial)  
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente



Presidente Prudente, 06 de janeiro de 2023

Dedico este trabalho ao Prof. Dr. Amilton Amorim (*in memoriam*).  
Obrigado por ter me orientado no início desta jornada.

## AGRADECIMENTOS

Agradeço aos meus orientadores, Prof. Dr. Maurício Galo e Prof. Dr. Milton Hirokazu Shimabukuro, pela paciência, sugestões e principalmente por acreditarem e terem me motivado.

Agradeço ao Programa de Pós Graduação em Ciências Cartográficas (PPGCC) e a UNESP pela oportunidade de desenvolver este projeto, e aos meus amigos que fiz durante este período em Presidente Prudente.

Agradeço aos meus familiares pelo apoio e amor, que me ajudaram a finalizar esta pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## RESUMO

A extração de edificações a partir de imagens ópticas constitui um importante passo para o planejamento urbano e desenvolvimento territorial para as cidades, e os processos automáticos, como o *machine learning* e *deep learning*, beneficiam indiretamente os gestores públicos no contexto de processos decisórios. O *deep learning* é uma subárea do aprendizado de máquina, que se concentra no treinamento de redes neurais profundas com enfoque no aprendizado com base em dados com alta variabilidade, como luminosidade, posição, textura e outros, e difere de outras técnicas por utilizar redes neurais na aprendizagem e envolver o uso de múltiplas camadas de neurônios artificiais para realizar tarefas complexas de processamento de dados. Há diversas redes disponíveis na literatura para os mais variados fins, incluído o contexto de segmentação de edificações em área urbana, ou redes inicialmente designadas para uma finalidade que apresenta compatibilidade em aprendizado de edificações (como a *U-Net* desenvolvida para segmentação em imagens da área médica). Independente da rede considerada, é comum variações com implementações de módulos e novos processos, como ocorre na *ResUNet-a* baseada na *U-Net*, com o objetivo de aprimorar sua segmentação. A melhoria nos resultados dessas redes também pode ser produzida pela entrada de informações adicionais, modificando os dados usados no treinamento para melhorar a identificação de um objeto, como o uso de dados LiDAR (*Light Detection And Ranging*) e diferentes modelos de cores, por exemplo. A presente pesquisa se propõe a avaliar o efeito da modificação dos dados de entrada nas redes de *deep learning*, com uso de dados LiDAR e composição de cores, na arquitetura *ResUNet-a*, e testar a hipótese de aprimoramento da segmentação. Para avaliar a hipótese foi realizado um experimento prático visando testar o uso da rede *ResUNet-a*, sendo possível atingir 96.0% de correspondência na segmentação semântica de edificações pela métrica F1-Score com um modelo de parâmetros treinado em 60 épocas durante o período de 4 dias. A partir disso, a metodologia foi proposta para avaliar o desempenho utilizando dois *datasets*: *HInDSM* (composto por informação de *matiz*, intensidade e *MDSn* - modelo digital de superfície normalizado) e imagem RGB, usado para efeito de comparação. O resultado pelo *HInDSM* alcançou 96,601% de F1-Score, correspondendo a um acréscimo de 1,89% se comparado ao uso apenas da imagem RGB. Além disso, o modelo apresenta melhorias qualitativas na segmentação de edificações, e conclui-se que o uso de *MDSn* combinado com os componentes *matiz* (*hue*) e intensidade, do modelo de cor HSI trazem melhorias na acurácia e desempenho na identificação de edificações em meio urbano.

**Palavras-chave:** Aprendizado Profundo. Extração de edificações. LiDAR. Modelo de Cores.

## ABSTRACT

The extraction of buildings from optical images is an important step for urban planning and territorial development for cities, and automated processes, such as machine learning and deep learning, indirectly benefit public decision-makers. Deep learning is a subfield of machine learning that focuses on training deep neural networks with an emphasis on data-driven learning with high variability, such as brightness, position, texture, and others. It differs from other techniques in using neural networks in learning and involving the use of multiple layers of artificial neurons to perform complex data processing tasks. There are several networks available in the literature for various purposes, including the context of building segmentation in urban areas, or networks initially designed for a purpose that is compatible with building learning (such as the *U-Net* developed for segmentation in medical images). Regardless of the network considered, variations with module implementations and new processes are common, as seen in the *ResUNet-a* based on *U-Net*, aiming to improve its segmentation. Improved results in these networks can also be achieved by introducing additional information, modifying the training data to enhance object identification, such as the use of Light Detection and Ranging (LiDAR) data and different color models, for example. This research aims to evaluate the effect of modifying input data in deep learning networks, using LiDAR data and color composition in the *ResUNet-a* architecture, and test the hypothesis of segmentation improvement. To evaluate the hypothesis, a practical experiment was conducted to test the use of the *ResUNet-a* network, achieving a 96.0% correspondence in semantic building segmentation by the *F1-Score* metric with a parameter model trained in 60 epochs over a period of 4 days. Subsequently, the methodology was proposed to evaluate performance using two datasets: *HInDSM* (comprising hue, intensity, and normalized digital surface model (*nDSM*) information) and *RGB* images, used for comparison. The result for *HInDSM* reached a 96.601% *F1-Score*, representing an increase of 1.89% compared to using only *RGB* images. Additionally, the model presents qualitative improvements in building segmentation, and it is concluded that the use of *nDSM* combined with hue and intensity components from the *HSI* color model brings improvements in accuracy and performance in the identification of buildings in urban environments.

**Keywords:** Deep learning. Building Extraction. LiDAR. Color Models.

## LISTA DE ILUSTRAÇÕES

Figura 2.1 – Metodologia de avaliação <i>k-fold</i> . . . . .	21
Figura 2.2 – Representação gráfica de algumas funções de ativação, no qual $x$ é o valor de entrada para o neurônio. . . . .	22
Figura 2.3 – <i>Feedforward</i> e camadas em uma rede neural. . . . .	23
Figura 2.4 – Exemplo de aplicação de camada de <i>max pooling</i> . . . . .	24
Figura 2.5 – Operação de <i>up-convolution</i> . . . . .	27
Figura 2.6 – Arquitetura U-Net. . . . .	28
Figura 2.7 – <i>Atrous Convolutions</i> , em que <i>Rate</i> é a taxa de dilatação da convolução. . . . .	29
Figura 2.8 – Arquitetura ResUNet-a. . . . .	31
Figura 2.9 – Bloco residual da ResUNet-a. . . . .	32
Figura 3.10–Técnicas de varredura LiDAR. . . . .	34
Figura 3.11–Múltiplos ecos em um pulso. . . . .	35
Figura 3.12–Modelos digitais. . . . .	37
Figura 4.13–Representação gráfica e vetorial da mistura de cores, em que as quantidades de energia são representadas pelos vetores $P_1, P_2$ e $P_3$ , em (a) os planos triangulares são definidos pela combinação das energias, e em (b) a projeção do plano triangular define o diagrama de vetores em que pode-se expressar a quantidade de energia por coordenadas. . . . .	39
Figura 4.14–Geometria hexacone para representação do modelo IHS. . . . .	39
Figura 5.15–Identificação e localização do <i>tp</i> , <i>tn</i> , <i>fp</i> e <i>fn</i> ao fazer a comparação entre uma edificação de referência (a), o resultado da edificação gerada por um algoritmo de classificação (b) e na sobreposição da edificação de referência e a gerada por um algoritmo (c). . . . .	43
Figura 6.16–Fluxograma da metodologia proposta. . . . .	45
Figura 6.17–Mosaico do <i>dataset</i> de <i>Potsdam</i> . . . . .	46
Figura 6.18–Exemplo da composição presente no <i>dataset</i> de <i>Potsdam</i> . (a) é a true ortofoto, (b) é o MDSn, e (c) são os rótulos. . . . .	46
Figura 6.19–Composição <i>HInDSM</i> - <i>top_potsdam_3_12</i> . . . . .	48
Figura 6.20–Combinação RGB - <i>top_potsdam_3_12</i> . . . . .	49
Figura 6.21– <i>Tiles</i> de imagem. . . . .	50
Figura 6.22– <i>Tiles</i> de rótulo de referência. . . . .	51
Figura 6.23– <i>Tiles</i> de bordas. . . . .	51
Figura 6.24– <i>Tiles</i> de mapa de distâncias. . . . .	51
Figura 7.25–Acurácia do treinamento em cada modelo. . . . .	58
Figura 7.26–Função custo em cada modelo. . . . .	58
Figura 7.27–Média móvel da função custo. . . . .	59

Figura 7.28–Acurácia da validação. . . . .	63
Figura 7.29–F1-Score da validação. . . . .	63
Figura 7.30–Predição 1 com o modelo <i>HInDSM_205</i> em <i>tiles</i> de validação. . . . .	64
Figura 7.31–Predição 2 com o modelo <i>HInDSM_205</i> em <i>tiles</i> de validação. . . . .	65
Figura 7.32–Predição 3 com o modelo <i>RGB_201</i> em <i>tiles</i> de validação. . . . .	66
Figura 7.33–Predição 4 com o modelo <i>RGB_201</i> em <i>tiles</i> de validação. . . . .	66
Figura 7.34–Dados dos <i>HInDSM</i> e <i>RGB</i> da área 1. . . . .	67
Figura 7.35–Predição da área 2 no <i>RGB_201</i> e <i>HInDSM_205</i> . . . . .	67
Figura 7.36–Dados dos <i>datasets HInDSM</i> e <i>RGB</i> da área 2 . . . . .	68
Figura 7.37–Predição da área 1 no <i>RGB_201</i> e <i>HInDSM_205</i> . . . . .	69
Figura 7.38–Predição da área 1 no <i>HInDSM_203</i> e <i>RGB_200</i> . . . . .	70
Figura 7.39–Predição da área 2 no <i>HInDSM_203</i> e <i>RGB_200</i> . . . . .	70
Figura A.1–Resultado da predição no modelo <i>Potsdam256_4</i> , sendo a primeira linha composta pelo <i>tile</i> , rótulo e predição. Na segunda linha é visto o <i>tile</i> , as bordas e a predição das bordas. Na terceira linha são apresentadas as distâncias das bordas, e a inferência das distâncias das bordas; e na quarta linha são mostrados o primeiro <i>tile</i> , o <i>tile</i> reconstruído, e o <i>tile</i> em HSI. . . . .	87
Figura A.2–Resultado da predição no modelo <i>Potsdam256_1</i> , sendo a primeira linha composta pelo <i>tile</i> , rótulo e predição. Na segunda linha é visto o <i>tile</i> , as bordas e a predição das bordas. Na terceira linha são apresentadas as distâncias das bordas, e a inferência das distâncias das bordas; e na quarta linha são mostrados o primeiro <i>tile</i> , o <i>tile</i> reconstruído, e o <i>tile</i> em HSI. . . . .	88
Figura A.3–Resultado do modelo <i>Potsdam128_4</i> , sendo a primeira linha composta pelo <i>tile</i> , rótulo e predição. Na segunda linha é visto o <i>tile</i> , as bordas e a predição das bordas. Na terceira são apresentadas as distâncias das bordas, e a inferência das distâncias das bordas; na quarta linha são mostrados o primeiro <i>tile</i> , o <i>tile</i> reconstruído, e o <i>tile</i> em HSI. . . . .	89
Figura A.4–Evolução da função custo <i>Tanimoto with dual</i> pelo número de épocas. . . .	90

## LISTA DE TABELAS

Tabela 2.1 – Redes selecionadas para a revisão. . . . .	26
Tabela 6.2 – Parâmetros de recorte. . . . .	50
Tabela 6.3 – Parâmetros de normalização. . . . .	53
Tabela 7.4 – Maiores acurácias do treinamento para cada <i>dataset</i> . . . . .	57
Tabela 7.5 – Matriz de confusão do modelo na época 205 - <i>HInDSM</i> . . . . .	60
Tabela 7.6 – Métricas do modelo na época 205 - <i>HInDSM</i> . . . . .	60
Tabela 7.7 – Matriz de confusão do modelo na época 201 - RGB. . . . .	60
Tabela 7.8 – Métricas do modelo na época 201 - RGB. . . . .	60
Tabela 7.9 – Diferença percentual entre as métricas estimadas para o modelo <i>HInDSM</i> (época 205) e <i>RGB</i> (época 201). . . . .	61
Tabela 7.10–Matriz de confusão do modelo na época 203 - <i>HInDSM</i> . . . . .	61
Tabela 7.11–Matriz de confusão do modelo na época 200 - RGB. . . . .	61
Tabela 7.12–Métricas do modelo <i>HInDSM</i> na época 203. . . . .	62
Tabela 7.13–Métricas do modelo RGB na época 200. . . . .	62
Tabela 7.14–Diferença percentual entre as métricas estimadas para o modelo <i>HInDSM</i> (época 203) e <i>RGB</i> (época 200). . . . .	62
Tabela A.1 – Conjunto de dados preparados para o treinamento. . . . .	81
Tabela A.2 – Matriz de confusão do conjunto <i>Potsdam256_4</i> . . . . .	86
Tabela A.3 – Matriz de confusão do conjunto <i>Potsdam256_1</i> . . . . .	86
Tabela A.4 – Métricas do conjunto <i>Potsdam256_4</i> em porcentagem. . . . .	86
Tabela A.5 – Métricas do conjunto <i>Potsdam256_1</i> em porcentagem. . . . .	86
Tabela A.6 – Matriz de confusão do conjunto <i>Potsdam128_4</i> . . . . .	89
Tabela A.7 – Métricas do conjunto <i>Potsdam128_4</i> em porcentagem. . . . .	90

## LISTA DE ABREVIATURAS E SIGLAS

CLS	- Classificador
CUDA	- <i>Compute Unified Device Architecture</i>
cuDNN	- <i>CUDA Deep Neural Network library</i>
DL	- <i>Deep Learning</i>
DSM	- <i>Digital Surface Model</i>
DTM	- <i>Digital Terrain Model</i>
FN	- <i>False Negative</i>
FP	- <i>False Positive</i>
FPN	- <i>Feature Pyramid Network</i>
GIS	- <i>Geographic Information System</i>
GPU	- <i>Graphics Processing Unit</i>
GSD	- <i>Ground Sample Distance</i>
HInDSM	- Composição hue, intensidade e MDSn
HSV	- <i>Hue, Saturation e Value</i>
IHS	- <i>Intesidade, Hue e Saturação</i>
IoU	- <i>Intersection Over Union</i>
LASER	- <i>Light Amplification by Stimulated Emission of Radiation</i>
LiDAR	- <i>Light Detection And Ranging</i>
MCC	- <i>Matthews Correlation Coefficient</i>
MDS	- Modelo Digital de Superfície
MDSn	- Modelo Digital de Superfície Normalizado
MDT	- Modelo Digital de Terreno
mIoU	- <i>Mean Intersection Over Union</i>
nDSM	- <i>Normalized Digital Surface Model</i>

OA	- <i>Overall accuracy</i>
RBF	- <i>Radial Basis Function</i>
REG	- <i>Regressão</i>
ReLU	- <i>Rectified Linear Unit</i>
RGB	- <i>Red, Green e Blue</i>
RPN	- <i>Region Proposal Networks</i>
TN	- <i>True Negative</i>
TP	- <i>True Positive</i>
VRAM	- <i>Video Random Access Memory</i>
Wcov	- <i>Weighted By Coverage</i>

## CÓDIGOS

B.1	Código destinado à preparação do <i>dataset HInDSM</i> . . . . .	92
B.2	Cálculo dos parâmetros de normalização. . . . .	93
B.3	Aplicação da função de agregação de gradientes. . . . .	94
B.4	Aplicação da função visando inicializar a lista de gradientes com o valor nulo. . . . .	94
B.5	Código utilizado para a execução da agregação manual de gradientes. . . . .	94

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Hipótese</b>	<b>18</b>
<b>1.2</b>	<b>Objetivos</b>	<b>18</b>
1.2.1	Objetivo geral	18
1.2.2	Objetivos específicos	18
<b>2</b>	<b>DEEP LEARNING E REDES NEURAIAS CONVOLUCIONAIS</b>	<b>20</b>
<b>2.1</b>	<b>Deep learning</b>	<b>20</b>
<b>2.2</b>	<b>Redes Neurais Convolucionais</b>	<b>24</b>
<b>2.3</b>	<b>Arquiteturas para segmentação de edificações</b>	<b>25</b>
2.3.1	Arquitetura U-Net	26
2.3.2	Arquitetura <i>ResUNet-a</i>	28
<b>3</b>	<b>LASER E SISTEMAS LIDAR</b>	<b>33</b>
<b>3.1</b>	<b>Princípios</b>	<b>33</b>
3.1.1	Modelos digitais de superfície	36
<b>4</b>	<b>MODELOS DE CORES RGB E IHS</b>	<b>38</b>
<b>5</b>	<b>EXTRAÇÃO DE EDIFICAÇÕES</b>	<b>41</b>
<b>5.1</b>	<b>Métricas para avaliação das extrações</b>	<b>42</b>
<b>6</b>	<b>MATERIAL E METODOLOGIA</b>	<b>44</b>
<b>6.1</b>	<b>Material</b>	<b>44</b>
<b>6.2</b>	<b>Metodologia</b>	<b>44</b>
6.2.1	Área de estudos	45
6.2.2	Preparação dos dados	47
6.2.2.1	<i>Dataset HInDSM</i>	47
6.2.2.2	<i>Dataset RGB</i>	48
6.2.2.3	Rótulos	49
6.2.2.4	Recorte dos <i>datasets</i> - criação dos <i>tiles</i>	49
6.2.3	Treinamento	51
6.2.3.1	Ambiente de treinamento	52
6.2.3.2	Parâmetros de treinamento	52
6.2.3.2.1	Transformações	53
6.2.3.2.2	Agregação manual de gradiente	54
6.2.3.2.3	Critério de parada	55

6.2.3.3	Treinamento no <i>dataset HInDSM e RGB</i> . . . . .	55
6.2.4	Validação dos modelos . . . . .	56
<b>7</b>	<b>RESULTADOS</b> . . . . .	<b>57</b>
<b>7.1</b>	<b>Treinamento do modelo de parâmetros</b> . . . . .	<b>57</b>
<b>7.2</b>	<b>Validação do modelo</b> . . . . .	<b>59</b>
7.2.1	Avaliação Quantitativa . . . . .	60
7.2.2	Avaliação Qualitativa . . . . .	64
<b>8</b>	<b>CONCLUSÕES</b> . . . . .	<b>72</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>74</b>
	 <b>APÊNDICES</b>	 <b>79</b>
	<b>APÊNDICE A – EXPERIMENTO PRELIMINAR</b> . . . . .	<b>80</b>
<b>A.1</b>	<b>Metodologia do experimento preliminar</b> . . . . .	<b>80</b>
A.1.1	Preparação dos dados . . . . .	80
A.1.2	Treinamento do modelo de parâmetros da rede . . . . .	82
A.1.3	Validação dos modelos . . . . .	83
<b>A.2</b>	<b>Resultados e análise do experimento preliminar</b> . . . . .	<b>84</b>
	 <b>APÊNDICE B – ALGORITMOS/CÓDIGOS</b> . . . . .	 <b>92</b>

# 1 INTRODUÇÃO

A extração automática de edificações em meio urbano não é uma necessidade recente, e tem se tornado cada vez mais essencial no cotidiano administrativo das cidades, como no planejamento urbano e desenvolvimento territorial. Para as cidades, a representação detalhada da complexidade territorial é um tópico relevante e de grande importância por fomentar a resolução de questões de direito ao solo, permitir implantação de políticas públicas direcionadas e, dessa forma, melhorar as condições de vida do cidadão. Todavia, segundo Sohn e Dowman (2007) extrações automáticas não atingem 100% de sucesso por algumas razões, como, por exemplo: a complexidade nas cenas, oclusões e dependência dos sensores. Como consequência, pesquisas sobre esse assunto são importantes meios para melhorar essas segmentações e reduzir a necessidade de extrações manuais.

Há uma gama de metodologias para a extração automática de feições a partir de dados obtidos por sensores remotos, com o uso de algoritmos de aprendizagem de máquina ou de classificadores lineares, que podem ser aplicados a dados distintos. Duas informações importantes nesse sentido são imagens de sensores ópticos, por permitirem ampla visão da superfície terrestre, e dados LiDAR advindos de sistemas de varredura *LASER (Light Amplification by Stimulated Emission of Radiation)*, sendo esse amplamente utilizado por ser uma fonte de aquisição de nuvens de pontos com coordenadas tridimensionais em alta densidade. Apesar da alta resolução desses produtos, a segmentação manual em imagens de sensores ópticos para extração de objetos urbanos possui complexidades relacionadas à extração perfeita e exaustiva por um operador humano, e o desenvolvimento computacional em *hardwares* e linguagens de programação envolvidos nesses processos não acompanharam o aumento da oferta de imagens de alta resolução, contexto já observado por (WILKINSON, 2005). O ritmo de mudanças também é um fator de relevância no espaço urbano, devido às cidades estarem em constantes processos de alterações, requerendo procedimentos rápidos de modo a permitir constantes atualizações. Ao se tratar da extração em dados LiDAR, tem-se que essa também pode ser conduzida por algoritmos computacionais ou manualmente, filtrando os objetos de interesse na nuvem de pontos ou aplicando algoritmos que identificam uma geometria específica, e segmenta esses dados.

Yuan (2018) diz que em 2006 o interesse em pesquisas envolvendo redes neurais de aprendizagem profunda, ou *deep learning*, aumenta com investigações voltadas para a criação de algoritmos capazes de aprenderem uma determinada função, como reconhecimento facial ou diagnósticos em saúde, com processos de funcionamento semelhante ao do neurônio humano. Segundo Yuan (2018), o uso de *deep learning* surge, nesse aspecto da extração de informações, como uma alternativa aos classificadores lineares, por permitir a inclusão da aprendizagem profunda em imagens ópticas e de dados LASER, transferindo o foco da extração dos objetos urbanos para a eficácia das redes neurais em aprender, detectar e segmentar, e introduzindo o

conceito de generalização ilimitada como habilidade em distinguir objetos em sua diversidade de cores e formas.

Há uma variedade de redes que podem ser enquadradas na segmentação semântica de edificações, como a *U-Net* e *ResUNet-a* que utilizam poucas imagens para treino, com enfoque na definição das bordas dos objetos. Portanto, a aplicação de algoritmos de *deep learning* para a extração de edificações em áreas urbanas torna-se promissora devido a esse ambiente possuir características com alta variabilidade entre si, como a diversidade de coberturas de edificações, com a presença de materiais distintos, em diferentes formas e condições físicas. Apesar dessa possibilidade, os dados usados para segmentação semântica em ambiente urbano ainda são majoritariamente imagens aéreas com composição de cores em RGB (*Red, Green e Blue*), por ser um dado com maior disponibilidade. Como as redes neurais permitem a inserção de dados distintos, o aprendizado com outras composições ou a adição de outras informações podem impactar nos resultados apresentados pelas redes. Desse modo, este trabalho tem por objetivo avaliar o uso combinado de dados LiDAR e imagens ópticas no modelo de cor HSI (Hue, Saturação e Intensidade) para o aprimoramento dos resultados das redes neurais, com enfoque na extração de edificações em ambiente urbano.

## 1.1 Hipótese

Este trabalho avaliou o uso combinado de imagens ópticas com diferentes modelos de cor e modelos de elevação de superfície normalizados, advindos de dados *LiDAR*, para a extração de edificações com uso da rede neural convolucional *ResUNet-a*. Estima-se que a rede neural usada permita a identificação e segmentação semântica desses objetos urbanos pelo uso combinado desses dados como camadas de entrada na rede. Desse modo, a hipótese colocada é: “A acurácia e desempenho na segmentação semântica de edificações em meio urbano, por técnicas de *deep learning*, são aprimorados pela combinação de imagens ópticas e dados *LiDAR*”.

## 1.2 Objetivos

### 1.2.1 Objetivo geral

Avaliar a extração de edificações por redes neurais convolucionais, inserindo a combinação de modelos digitais de superfície normalizados advindos de dados *LiDAR* e imagens ópticas representadas nos modelos de cores IHS e RGB.

### 1.2.2 Objetivos específicos

Como objetivos específicos tem-se que:

- Construir, treinar e gerar um modelo por rede neural convolucional a partir da combinação de dados *LiDAR* e imagens ópticas em diferentes modelos de cor, e aplicá-lo para segmentação semântica de edificações em área urbanizada;
- Comparar a segmentação semântica obtida, com a resultante de imagens sem transformação (*RGB*) e dados *LiDAR*.

## 8 CONCLUSÕES

Os resultados encontrados no modelo usando a composição de cores *Hue* e Intensidade unido com o modelo digital de superfície normalizado demonstrou ser superior ao uso do modelo tradicional (com a composição de cores *RGB*). Apesar do resultado positivo visto nas métricas, como o ganho de 1,89% em termos de *F1-Score*, outros pontos se destacaram na metodologia proposta, como a estabilidade do treinamento, alta convergência e consequente redução do tempo de treinamento do modelo de parâmetros. Ao comparar o uso do modelo *HInDSM* ao uso das componentes *RGB* foi possível identificar que o uso do primeiro modelo permitiu uma convergência mais rápida da rede, tanto em termos de custo quanto acurácia.

O destaque do *HInDSM* pode ser justificado por permitir a arquitetura reconhecer as edificações em meio urbano devido à presença da informação de altura. O uso desta informação convertida em uma imagem com níveis de cinza facilita ao modelo diferenciá-los de outros objetos urbanos, como vias, carros e vegetação. O emprego do modelo de cor *Hue* e Intensidade como camadas da imagem composta também trazem maior contexto ao treinamento.

A cada canal adicionado no processamento, aumenta-se o custo computacional para o aprendizado. Deste modo, apesar de ser viável adicionar o *MDSn* em níveis de cinza juntamente com a informação de cor, o custo computacional se amplia quando comparado ao utilizado nesta pesquisa devido à presença de mais um canal de processamento pela arquitetura, limitando mais o uso de GPUs nos processos de *deep learning*. Deste modo, alternativas como a proposta nesta pesquisa, que reduzem o número de dados com melhoria dos resultados, trazem avanços aos estudos de segmentação de edificações por *deep learning*.

A principal limitação de *hardware* detectada na pesquisa foi a impossibilidade de se aumentar o *batch* do treinamento devido ao tamanho da memória *VRAM* presente nas máquinas do *Google Colab*. O problema decorre do processamento de imagens ter alto consumo de memória. Tal limitação foi superada ao utilizar a metodologia de agregação manual de gradientes, permitindo o *batch* ser aumentado para 264 *tiles* para cada atualização dos pesos, trazendo mais estabilidade aos processos e maior contexto para a arquitetura durante o treinamento.

Apesar de a segmentação ter atingido 96,6% de *F1-Score*, o tempo de treinamento da arquitetura ainda é um fator a se considerar. O alto custo computacional havia sido detectado no experimento preliminar realizado, antes do treinamento da metodologia proposta, e apesar do número reduzido de canais ter diminuído o tempo de processamento, ainda é um custo computacional considerável da arquitetura utilizada. O tempo de treinamento poderia ser reduzido pela sincronização com outras *GPUs*, dividindo os processos entre cada instância e utilizando a técnica de agregação manual de gradientes para o cálculo dos pesos.

Na análise qualitativa a metodologia proposta também apresentou melhores resultados

no modelo *HInDSM*, com baixo ruído e alta correspondência com a referência. Apesar disto, em ambos os modelos há erros na segmentação de edificações, que corresponde ao visto nas métricas avaliadas. Os principais pontos identificados foram a dificuldade da arquitetura com objetos urbanos sem altura, como canteiros de avenidas e quadras esportivas. Ao comparar os modelos, o modelo *HInDSM* conseguiu reduzir os problemas mencionados, mas não extinguiu-os. Não obstante, o modelo *HInDSM* tem mais regularidade nas edificações.

Os resultados produzidos poderiam ser aplicados na atualização rápida e constante de informações urbanas requeridas pelos gestores públicos. De fato, o tempo de processamento ainda é um fator determinante no treinamento das arquiteturas de *deep learning*, que devem ser reduzidos com o aprimoramento das técnicas e métodos, e a maior disponibilidade de *hardware* eficientes em processos de aprendizagem profunda. Após o modelo treinado, a inferência é rápida e pode ser realizada continuamente em imagens ópticas e dados *LiDAR*, permitindo aplicações como detecção de expansão urbana, uso irregular do solo, atualização de atributos e outras aplicações. Além do meio urbano, trabalhos futuros podem verificar a aplicação da metodologia na melhoria dos modelos de *deep learning* na detecção de desmatamento e pontos de fogo em áreas de proteção ambiental.

Conclui-se que a acurácia e o desempenho na segmentação semântica de edificações em meio urbano, por *deep learning*, foi aprimorado pela combinação de imagens ópticas e dados *LiDAR*, validando a hipótese inicialmente colocada para esta pesquisa.

## REFERÊNCIAS

- AHMED, N.; MAHBUB, R. B.; RAHMAN, R. M. Learning to extract buildings from ultra-high-resolution drone images and noisy labels. *International Journal of Remote Sensing*, Taylor & Francis, v. 41, n. 21, p. 8216–8237, 2020. ISSN 0143-1161. DOI: <https://doi.org/10.1080/01431161.2020.1763496>.
- AMORIM, A.; PELEGRINA, M. A.; JULIÃO, R. P. *Cadastro e gestão territorial: uma visão luso-brasileira para a implementação de sistemas de informação cadastral nos municípios*. São Paulo: Editora Unesp Digital, 2018. ISBN 9788595462823.
- APACHE-MXNET. *MXNet*. 2022. Disponível em: <<https://mxnet.apache.org/versions/1.5.0/>>. Acesso em: 2022-03-08.
- AWRANGJEB, M.; RAVANBAKSH, M.; FRASER, C. S. Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 65, n. 5, p. 457–467, 2010. ISSN 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2010.06.001>.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 12, p. 2481–2495, dez. 2017. ISSN 1939-3539. DOI: <https://doi.org/10.1109/TPAMI.2016.2644615>.
- BERALDIN, J. A.; BLAIS, F.; LOHR, U. Laser scanning technology. In: VOSSELMAN, G.; MAAS, H.-G. (Org.). *Airborne and Terrestrial Laser Scanning*. 1. ed. Dunbeath: CRC Press, 2010. p. 1–39. ISBN 978-1-4398-2798-7.
- BRENNER, C. Building extraction. In: VOSSELMAN, G.; MAAS, H.-G. (Org.). *Airborne and Terrestrial Laser Scanning*. 1. ed. Dunbeath: CRC Press, 2010. p. 169–207. ISBN 978-1-4398-2798-7.
- BRIESE, C. Extraction of digital terrain models. In: VOSSELMAN, G.; MAAS, H.-G. (Org.). *Airborne and Terrestrial Laser Scanning*. 1. ed. Dunbeath: CRC Press, 2010. p. 135–163. ISBN 978-1-4398-2798-7.
- BUJÁN, S. et al. Land use classification from lidar data and ortho-images in a rural area. *The Photogrammetric Record*, v. 27, n. 140, p. 401–422, 2012. DOI: <https://doi.org/10.1111/j.1477-9730.2012.00698.x>.
- CASTREJON, L. et al. Annotating Object Instances with a Polygon-RNN. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, 2017. p. 4485–4493. ISBN 978-1-5386-0457-1. DOI: <https://doi.org/10.1109/CVPR.2017.477>.
- CHEN, L.-C. et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: FERRARI, V. et al. (Ed.). *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018. v. 11211, p. 833–851. ISBN 978-3-030-01233-5. DOI: [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49).

- CHENG, D. et al. DARNet: Deep Active Ray Network for Building Segmentation. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019. p. 7423–7431. DOI: <https://doi.org/10.1109/CVPR.2019.00761>.
- DENG, L.; YU, D. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, v. 7, n. 3–4, p. 197–387, 2014. ISSN 1932-8346. DOI: <https://doi.org/10.1561/20000000039>.
- DIAKOGIANNIS, F. I. et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 162, p. 94–114, abr. 2020. ISSN 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- EL-SHEIMY, N.; VALEO, C.; HABIB, A. *Digital terrain modeling: acquisition, manipulation, and applications*. Boston: Artech House, 2005. ISBN 978-1-58053-921-0.
- ETTEN, A. V. City-scale road extraction from satellite imagery v2: Road speeds and travel times. In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2020. p. 1775–1784. ISSN 2642-9381. DOI: <https://doi.org/10.1109/WACV45572.2020.9093593>.
- FREUDENBERG, M. et al. Large scale palm tree detection in high resolution satellite images using U-Net. *Remote Sensing*, Multidisciplinary Digital Publishing Institute (MDPI), v. 11, n. 3, p. 312, 2019. DOI: <https://doi.org/10.3390/rs11030312>.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. 3. ed. Upper Saddle River, N.J: Prentice Hall, 2008. ISBN 978-0-13-168728-8.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>. Acesso em: 20 abr. 2021.
- HE, K. et al. Mask R-CNN. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 2980–2988. ISSN 2380-7504. DOI: <https://doi.org/10.1109/ICCV.2017.322>.
- HE, K. et al. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016. p. 770–778. ISBN 978-1-4673-8851-1. DOI: <https://doi.org/10.1109/CVPR.2016.90>.
- HUANG, J. et al. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 151, p. 91–105, maio 2019. ISSN 09242716. DOI: <https://doi.org/10.1016/j.isprsjprs.2019.02.019>.
- INTEL. *2D Max Pooling Forward Layer*. 2017. Disponível em: <[https://software.intel.com/sites/products/documentation/doclib/daal/daal-user-and-reference-guides/daal\\_prog\\_guide/GUID-CCB814DD-945A-46DD-989A-8BC39D2D01CA.htm](https://software.intel.com/sites/products/documentation/doclib/daal/daal-user-and-reference-guides/daal_prog_guide/GUID-CCB814DD-945A-46DD-989A-8BC39D2D01CA.htm)>. Acesso em: 15 out. 2021.
- IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015. DOI: <http://doi.org/10.48550/ARXIV.1502.03167>.
- ISPRS-WG-III/4. *2D Semantic Labeling*. 2018. Disponível em: <<https://www2.isprs.org/commissions/comm2/wg4/benchmark/semantic-labeling>>. Acesso em: 20 abr. 2021.
- JAIN, A. K. *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall, 1989. (Prentice Hall information and system sciences series). ISBN 978-0-13-336165-0.

- JIAO, C. et al. Burn image segmentation based on Mask Regions with Convolutional Neural Network deep learning framework: more accurate and more convenient. *Burns and Trauma*, v. 7, 02 2019. ISSN 2321-3876. DOI: <https://doi.org/10.1186/s41038-018-0137-9>.
- JUPYTER. *JupyterNotebook*. 2022. Disponível em: <<https://jupyter.org>>. Acesso em: 2022-03-08.
- KANG, W. et al. EU-Net: An Efficient Fully Convolutional Network for Building Extraction from Optical Remote Sensing Images. *Remote Sensing*, v. 11, n. 23, p. 2813, nov. 2019. ISSN 2072-4292. DOI: <https://doi.org/10.3390/rs11232813>.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. v. 2, p. 1137–1143. ISBN 1558603638. DOI: <https://doi.org/10.5555/1643031.1643047>.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015. ISSN 1476-4687. DOI: <https://doi.org/10.1038/nature14539>.
- LI, Z.; WEGNER, J. D.; LUCCHI, A. Topological Map Extraction From Overhead Images. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, 2019. p. 1715–1724. ISBN 978-1-72814-803-8. DOI: <https://doi.org/10.1109/ICCV.2019.00180>.
- LI, Z.; ZHU, Q.; GOLD, C. *Digital terrain modeling: principles and methodology*. New York: CRC Press, 2005. ISBN 978-0-415-32462-5.
- MENESES, P. R. Modelos de cores aplicados às imagens. In: MENESES, P. R.; ALMEIDA, T. d. (Org.). *Introdução ao processamento de imagens de sensoriamento remoto*. Brasília: Universidade de Brasília e CNPq, 2012. p. 121–134.
- MEYER, L.; LEMARCHAND, F.; SIDIROPOULOS, P. A deep learning architecture for batch-mode fully automated field boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, XLIII-B3-2020, p. 1009–1016, 2020. ISSN 2194-9034. DOI: <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1009-2020>.
- MILLER, C. L.; LAFLAMME, R. A. *The Digital Terrain Model: Theory & Application*. [S.l.]: MIT Photogrammetry Laboratory, 1958.
- MIROŚLAW-ŚWIĄTEK, D. et al. Developing an algorithm for enhancement of a digital terrain model for a densely vegetated floodplain wetland. *Journal of Applied Remote Sensing*, SPIE, v. 10, n. 3, p. 1 – 16, 2016. DOI: <https://doi.org/10.1117/1.JRS.10.036013>.
- MISHRA, D. *Transposed Convolution Demystified*. 2020. Disponível em: <<https://towardsdatascience.com/transposed-convolution-demystified-84ca81b4baba>>. Acesso em: 2021-06-06.
- MOUJAHID, A. *A Practical Introduction to Deep Learning with Caffe and Python // Adil Moujahid // Data Analytics and more*. 2016. Disponível em: <<http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/>>. Acesso em: 2022-03-08.

- NASCIMENTO, G. A. G. d. *Verificação da Aplicabilidade de Dados Obtidos por Sistema LASER Batimétrico Aerotransportado à Cartografia Náutica*. Dissertação (Mestrado em Ciências Cartográficas) — Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente, 2019. Disponível em: <<https://repositorio.unesp.br/handle/11449/181407>>. Acesso em: 20 abr. 2021.
- PEDRINI, H.; SCHWARTZ, W. R. *Análise de imagens digitais princípios, algoritmos e aplicações*. São Paulo: Thomson Learning, 2008. OCLC: 319215118. ISBN 978-85-221-0595-3.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – (MICCAI)*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4. DOI: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- SANTOS, R. C. dos; GALO, M.; CARRILHO, A. C. Building boundary extraction from lidar data using a local estimated parameter for alpha shape algorithm. *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS TC I Mid-term Symposium "Innovative Sensing - From Sensors to Methods and Applications"*, V.XLII-1, p. 127–132, 2018. ISSN 2194-9034. DOI: <https://doi.org/10.5194/isprs-archives-XLII-1-127-2018>.
- SANTOS, R. C. dos; GALO, M.; HABIB, A. F. Regularization of Building Roof Boundaries from Airborne LiDAR Data Using an Iterative CD-Spline. *Remote Sensing*, v. 12, n. 12, p. 1904, jun. 2020. ISSN 2072-4292. DOI: <https://doi.org/10.3390/rs12121904>.
- SCIKIT-LEARN. *scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation*. 2022. Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 2022-03-08.
- SHAN, J.; TOTH, C. K. *Topographic laser ranging and scanning: principles and processing*. 1. ed. Boca Raton: CRC Press, 2008. ISBN 978-1-4200-5142-1.
- SHI, Y.; LI, Q.; ZHU, X. X. Building footprint generation using improved generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, v. 16, n. 4, p. 603–607, 2019. ISSN 1545-598X, 1558-0571. DOI: <https://doi.org/10.1109/LGRS.2018.2878486>.
- 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*, v. 62, n. 1, p. 43–63, 2007. ISSN 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2007.01.001>.
- SOKOLOVA, M.; JAPKOWICZ, N.; SZPAKOWICZ, S. Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. In: *AI 2006: Advances in Artificial Intelligence*. Berlin: Springer, 2006. p. 1015–1021. ISBN 978-3-540-49788-2. DOI: [https://doi.org/10.1007/11941439\\_114](https://doi.org/10.1007/11941439_114).
- TRANSFER, K. *PyTorch K-Fold Cross-Validation using Dataloader and Sklearn*. 2021. Disponível em: <<https://androidkt.com/pytorch-k-fold-cross-validation-using-dataloader-and-sklearn/>>. Acesso em: 01 set. 2021.
- WANG, Y.; ZHANG, D.; DAI, G. Classification of high resolution satellite images using improved U-Net. *International Journal of Applied Mathematics and Computer Science*, v. 30, n. 3, p. 399–413, 2020. DOI: <https://doi.org/10.34768/AMCS-2020-0030>.
- WEHR, A.; LOHR, U. Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 54, n. 2-3, p. 68–82, jul. 1999. ISSN 09242716. DOI: [https://doi.org/10.1016/S0924-2716\(99\)00011-8](https://doi.org/10.1016/S0924-2716(99)00011-8).

- WEI, S.; JI, S.; LU, M. Toward Automatic Building Footprint Delineation From Aerial Images Using CNN and Regularization. *IEEE Transactions on Geoscience and Remote Sensing*, v. 58, n. 3, p. 2178–2189, mar. 2020. ISSN 0196-2892, 1558-0644. DOI: <https://doi.org/10.1109/TGRS.2019.2954461>.
- WILKINSON, G. G. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, n. 3, p. 433–440, 2005. ISSN 1558-0644. DOI: <https://doi.org/10.1109/TGRS.2004.837325>.
- XU, B. et al. Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *International Journal of Remote Sensing*, Taylor & Francis, v. 41, n. 21, p. 8121–8142, 2020. ISSN 0143-1161. DOI: <https://doi.org/10.1080/01431161.2020.1734245>.
- XU, Y. et al. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, Multidisciplinary Digital Publishing Institute (MDPI), v. 10, n. 1, p. 144, 2018. DOI: <https://doi.org/10.3390/rs10010144>.
- YEKEEN, S. T.; BALOGUN, A.; YUSOF, K. B. W. A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 167, p. 190–200, 2020. ISSN 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.07.011>.
- YUAN, J. Learning building extraction in aerial scenes with convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 40, n. 11, p. 2793–2798, 2018. ISSN 1939-3539. DOI: <https://doi.org/10.1109/TPAMI.2017.2750680>.
- ZHANG, Z.; LIU, Q.; WANG, Y. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, v. 15, n. 5, p. 749–753, 2018. ISSN 1558-0571. DOI: <https://doi.org/10.1109/LGRS.2018.2802944>.
- ZHENG, X. et al. Parsing very high resolution urban scene images by learning deep ConvNets with edge-aware loss. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 170, p. 15–28, dez. 2020. ISSN 09242716. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.09.019>.