PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DOUTORADO EM MEIO AMBIENTE E
DESENVOLVIMENTO REGIONAL

MAYARA MAEZANO FAITA PINHEIRO

APRENDIZAGEM PROFUNDA NA SEGMENTAÇÃO SEMÂNTICA DE RIOS EM IMAGENS DE ALTA RESOLUÇÃO ESPACIAL

Presidente Prudente - SP
2023

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DOUTORADO EM MEIO AMBIENTE E
DESENVOLVIMENTO REGIONAL

MAYARA MAEZANO FAITA PINHEIRO

APRENDIZAGEM PROFUNDA NA SEGMENTAÇÃO SEMÂNTICA DE RIOS EM IMAGENS DE ALTA RESOLUÇÃO ESPACIAL

Tese apresentada Pró-Reitoria de Pesquisa e Pós-Graduação, Universidade do Oeste Paulista, como parte dos requisitos para obtenção do título de Doutora em Meio Ambiente e Desenvolvimento Regional – Área de concentração: Ciências Ambientais.

Orientadora: Dra. Ana Paula Marques Ramos
Co-orientador: Dr. Marcelo Rodrigo Alves

Colaborador: Lucas Prado Osco (UNOESTE).
Colaborador externo: Keiller Nogueira (Universidade de Stirling – Escócia, Reino Unido).

Presidente Prudente - SP
2023

526.982
P654a

Pinheiro, Mayara Maezano Faita.
Aprendizagem profunda na segmentação semântica de rios em imagens de alta resolução espacial / Mayara Maezano Faita Pinheiro. – Presidente Prudente, 2023.
102f.: il.

Tese (Doutorado em Meio Ambiente e Desenvolvimento Regional) - Universidade do Oeste Paulista – Unoeste, Presidente Prudente, SP, 2023.
Bibliografia.
Orientadora: Dra. Ana Paula Marques Ramos
Co-orientador: Dr. Marcelo Rodrigo Alves

1. Mapeamento de corpos d'água. 2. Sensoriamento Remoto. 3. Redes de aprendizagem profunda. I. Título.

Catalogação na fonte: Maria Letícia Silva Vila Real – CRB 8/10699

MAYARA MAEZANO FAITA PINHEIRO

APRENDIZAGEM PROFUNDA NA SEGMENTAÇÃO SEMÂNTICA DE RIOS EM IMAGENS DE ALTA RESOLUÇÃO ESPACIAL

Tese apresentada Pró-Reitoria de Pesquisa e Pós-Graduação, Universidade do Oeste Paulista, como parte dos requisitos para obtenção do título de Doutora em Meio Ambiente e Desenvolvimento Regional – Área de concentração: Ciências Ambientais.

Presidente Prudente, 24 de fevereiro de 2023.

BANCA EXAMINADORA

_____
Prof. Dra. Ana Paula Marques Ramos
Universidade do Oeste Paulista – Unoeste
Presidente Prudente - SP

_____
Prof. Dr. Marcelo Rodrigo Alves
Universidade do Oeste Paulista – Unoeste
Presidente Prudente – SP

_____
Prof. Dr. Lucas Prado Osco
Universidade do Oeste Paulista – Unoeste
Presidente Prudente – SP

_____
Prof. Dr. José Marcato Júnior
Universidade Federal de Mato Grosso do Sul
Campo Grande – MS

_____
Prof. Dra. Tatiana Sussel Gonçalves Mendes
Universidade Estadual Paulista - UNESP
São José dos Campos – SP

DEDICATÓRIA

Ao meu marido Rafael e filho Pedro pelo apoio, compreensão e paciência durante minhas ausências nos últimos anos. Aos meus pais, por incentivarem e serem minha base para chegar até aqui.

AGRADECIMENTOS

A Deus e Nossa Mãe Maria, que sempre sustentaram minhas caminhadas, sempre cuidaram e olharam por mim e minha família, por derramarem tantas graças e colocarem pessoas especiais em meu caminho.

Ao meu marido Rafael e filho Pedro, por me manterem firme na caminhada, pela paciência em minhas ausências, pelo apoio, carinho e amor infinito.

Aos meus pais, Carlos e Karen, que me apoiam desde sempre, incentivam e me fazem acreditar que sou capaz, por ensinarem a ser determinada e me levarem ao caminho da fé.

Aos amigos de caminhada, Bruno Magro Rodrigues, Felipe Gomes, Letícia Ap. Costa Magro, Renata Mafra, Fabio Friol, Jacqueline Tamashiro, companheiros de grupo de estudos e parte da vida. Ao amigo Lucas Prado Osco, pela paciência, franqueza e solicitude, sempre com contribuições de alto nível, enriquecendo o trabalho e a vida.

À professora Dra. Ângela Kinoshita, por mesmo que por um curto período, ter compartilhado seu conhecimento, e enriquecido minha experiência acadêmica e pessoal. Aos Prof. José Marcato Júnior, Prof. Keiller Nogueira, Maximilian Melo, que mesmo distante, contribuíram significamente com os avanços do trabalho.

À orientadora Dra. Ana Paula, que considero uma amiga, e tem ensinado o caminho da pesquisa científica de forma tão leve. Com plenitude, consegue extrair o melhor de cada um, nos ensina a caminhar sozinhos, porém sabendo do suporte que estará de prontidão para auxiliar quando preciso. Muito obrigada, por contribuir em minha transformação como pessoa e profissional.

Aos professores que fizeram parte da banca avaliadora com contribuições e críticas construtivas essenciais para enriquecimento do trabalho.

Aos professores e funcionários do PPGMADRE, nos ensinando e dando suporte nessa caminhada. E a todos que de alguma forma me incentivaram, apoiaram e contribuíram para que este trabalho fosse concluído. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – (Brasil) CAPES – Código de Financiamento 001.

“Mas aqueles que contam com o
Senhor renovam suas forças;
Ele dá-lhes assa de águia.
Correm sem se cansar,
Vão pra frente sem se fatigar.”
Isaías, 40:31.

RESUMO

Aprendizagem Profunda na Segmentação Semântica de Rios em Imagens de Alta Resolução Espacial

O mapeamento de rios é fundamental para o diagnóstico e planejamentos ambiental e o sensoriamento remoto tem auxiliado em investigações relacionadas a este recurso hídrico. Um método que tem ganhado atenção e auxiliado na tarefa de mapear rios é o método de aprendizagem profunda. Os métodos de aprendizagem profunda mais utilizados na tarefa de mapeamento (segmentação semântica) são baseados em arquiteturas de rede em convolução e, mais recentemente, Vision Transformers. São métodos robustos que tem apresentado uma boa performance para esse tipo de tarefa, porém, ainda existem desafios a serem superados ao mapear feições de rios estreitos. Nesse sentido, considerando a diversidade de feições de rios e ausência de uma abordagem que se aprofunde em suas características como a largura dos mesmos, o objetivo deste estudo é mapear rios com diferentes tamanhos em imagens aéreas RGB de alta resolução espacial utilizando redes profundas de segmentação semântica. Para isso, o presente trabalho buscou avaliar o desempenho de redes profundas de segmentação semântica ao identificar rios largos (largura superior a 10 metros) em imagens RGB de alta resolução espacial; caracterizar as capacidades e limitações das redes profundas de segmentação semântica em identificar rios estreitos (largura inferior a 10 metros) em imagens RGB de alta resolução espacial; e avaliar o desempenho de redes profundas de segmentação semântica baseadas em Transformers para segmentar rios, comparando-as com as redes neurais profundas tradicionais baseadas em convoluções. As imagens aéreas possuem 1 m de resolução e foram treinadas e testadas em ambiente computacional, utilizando os métodos de segmentação semântica baseados em convolução e Vision Transformer. Em seguida, os resultados foram comparados qualitativamente, por meio visual, e quantitativamente, por meio de métricas de avaliação (Acurácia, F1-Score, Precisão, Recall, IoU). Os resultados mostraram que o desempenho das redes profundas de segmentação semântica varia conforme a largura dos rios. A melhor abordagem para segmentar rios largos (superior a 10 metros) e rios estreitos (inferior a 10 metros) é usar ambos os tipos de larguras de rios no treinamento das redes profundas de segmentação semântica. Outra descoberta é que as redes profundas baseadas em Vision Transformer superaram a performance das redes profundas baseadas em convoluções para a tarefa de segmentação semântica de rios largos e estreitos em imagens RGB de 1 m de resolução. O SegFormer superou as métricas de avaliação para segmentação de rios com um F1-Score cima de 98%. Os trabalhos futuros devem continuar investigando as redes profundas baseadas em Vision Transformer, podendo explorar imagens multi-temporais e multimodais, a fim de melhorar o monitoramento de recursos hídricos.

Palavras-chave: Mapeamento de corpos d’água. Sensoriamento Remoto. Redes de aprendizagem profunda.

ABSTRACT

Deep Learning in Semantic Segmentation of Rivers in High Spatial Resolution Images

Mapping rivers is essential for diagnosis and environmental planning, and remote sensing has helped in research related to this water resource. A method that has gained attention and helped in the task of mapping rivers is the deep learning method. The most used deep learning methods in the mapping task (semantic segmentation) are based on convolutional network architectures (CNN - Convolutional Neural Networks) and, more recently, Vision Transformers. They are robust methods that have shown good performance for this type of task, however, there are still challenges to be overcome when mapping features of narrow rivers. In this sense, considering the diversity of river features and the absence of an approach that goes deeper into their characteristics such as their width, the objective of this study is to map rivers with different sizes in RGB aerial images of high spatial resolution using deep segmentation networks semantics. For this, the present work sought to evaluate the performance of deep semantic segmentation networks when identifying large rivers (width greater than 10 m) in RGB images of high spatial resolution; characterize the capabilities and limits of deep semantic segmentation networks to identify narrow rivers (width less than 10 m) in high spatial resolution RGB images; and to evaluate the performance of deep semantic segmentation networks experimented in Vision Transformer to segment rivers, comparing them with traditional neural networks experimented in convolutions. The aerial images have 1 m of resolution and were trained and tested in a computational environment, using semantic segmentation methods based on convolution and Vision Transformer. Then, the results were compared qualitatively, visually, and quantitatively, using evaluation metrics (Accuracy, F1-Score, Precision, Recall, IoU). The results appreciated that the performance of deep semantic segmentation networks varies according to the width of the rivers. The best approach to segmenting large rivers (greater than 10 m) and narrow rivers (less than 10 m) is to use both river widths to train the deep semantic segmentation networks. Another finding is that the deep networks in Vision Transformers outperformed the deep networks in convolutions for the semantic segmentation task of large and narrow rivers in RGB images of 1 m resolution. SegFormer outperformed as evaluation indicators for river segmentation with an F1-Score above 98%. Future work should continue to investigate the aspiring deep networks in Vision Transformers, being able to explore multitemporal and multimodal images, in order to improve the monitoring of water resources.

Keywords: Mapping of water bodies. Remote sensing. Deep learning networks.

LISTA DE SIGLAS

AC	- Active Contour
ACC	- Accuracy
ANFIS	- Adaptive Neuro-Fuzzy Interference System
ASPP	- Atrous Spatial Pyramid Pooling
BOA	- Boundary Overall Accuracy
CAPES	- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CFOA	- Chaotic Forest Optimization Algorithm
CHEOS	- China High-Resolution Earth Observation System
CNN	- Convolutional Neural Network
CRF	- Conditional Random Field
CSWS	- Cosine Similarity-Based Watershed Segmentation
CUDA	- Compute Unified Device Architecture
CV	- Computer Vision
DA	- Dual Attention
DCNNs	- Deep Convolutional Neural Networks
DFE	- Deep Feature Extraction Module
DL	- Deep Learning
DNN	- Deep Neural Network
DPHE	- Double Plateaus Histogram Equalization
DSFF	- Encoder And Decoder Semantic Feature Fusion
DT	- Decision Tree
ENN	- Elman Neural Network
EOA	- Edge Overall Accuracy
FCM	- Fuzzy C-Means
FCN	- Fully Convolutional Network
FFN	- Feed-Forward Network
FFU	- Feature-Fusion Upsample Module
FN	- False Negative
FP	- False Positive
FWIoU	- Frequency Weighted Intersection over Union
Gi3A	- Geomática e Inteligência Artificial Aplicada a Análise Ambiental
GID	- Gaofen Image Dataset
GSD	- Ground Sample Distance
HRRS	- High-Resolution Remote Sensing
IoU	- Intersection over Union
Kappa	- Kappa coefficient
KM	- K-Means
MBA	- Multi-Branch Aggregation Module
MCC	- Mathews correlation coefficient
MCNN	- Modified Convolution Neural Network
MEC	- Multi-Feature Extraction And Combination
MIoU	- Mean Intersection over Union
MLP	- Multilayer Perceptron
MNDWI	- Modified Normalized Difference Water Index
MPA	- Mean Pixel Accuracy
MPF	- Multi-Scale Prediction Fusion
NDWI	- Normalized Difference Water Index

NEAmb	- Núcleo de Estudos Ambientais
NN	- Neural Network
OA	- Overall Accuracy
P	- Precision
PA	- Pixel Accuracy
PVT	- Pyramid Vision Transformer
R	- Recall
RF	- Random Forest
RG	- Region Growing
RGB	- Red, Green, Blue
RNN	- Recurrent Neural Network
RRF DeconvNet	- Restricted Receptive Field Deconvolution Network
RS	- Remote Sensing
SAR	- Synthetic Aperture Radar
SFO	- Shape Feature Optimization
SSL	- Self-Supervised Learning
SVM	- Support Vector Machine
TN	- True Negative
TP	- True Positive
UAV	- Unmanned Aerial Vehicle
UTFPR	- Universidade Tecnológica Federal do Paraná
VANT	- Veículo Aéreo não Tripulado
ViT	- Vision Transformer

LISTA DE FIGURAS

Figure 1- Word Cloud From Articles With “Deep Learning” And “Remote Sensing” String 25
Figure 2- Filters And Quantity Of Related Articles 28
Figure 3- Temporal Distribution Of Articles From 2017 To June 2022 29
Figure 4- Percent Of Articles Published Per Country 30
Figure 5- Top 5 Influential Journals 31
Figure 6- Top 5 Main Topics Research In Water Body Fields 32
Figure 7- Platforms Sensors For The Article 34
Figure 8- Image Spatial Resolution Of The Studies 35
Figure 9- Distribution Of Image Spatial Resolution Over The Years 35
Figure 10- Data Type Percent 36
Figure 11- Top 5 Deep Learning Methods 37
Figure 12- Most Used Metrics 38
Figure 13- The Main Steps Of The Method Are Summarized In A Flowchart 54
Figure 14- Study Area 55
Figure 15- Examples Of Rivers That Composed The Dataset 56
Figure 16- Flowchart Of The Experiments 57
Figure 17- U-Net Architecture 59
Figure 18- Pspnet Architecture 60
Figure 19- Deeplabv3+ Architecture 61
Figure 20- Qualitative Comparison Of The Classification Of Deep Learning Models In Large Rivers 64
Figure 21- Qualitative Comparison Of The Classification Of Deep Learning Models In Narrow Rivers 66
Figure 22- Inference Of The Best Model In Entire Orthophotos 67
Figure 23- Qualitative Results From Experiment 5 69
Figure 24- Details Of River Segmentation From Orthophoto 5 In Experiments 4 And 5 70
Figure 25- Details Of River Segmentation From Orthophoto 6 In Experiments 4 And 5 71
Figure 26- Main Steps Carried Out In The Approach Of Our Study 86
Figure 27- Study Area And Water Labeling 87
Figure 28- Deeplabv3+ Architecture 89

Figure 29- Segformer Framework 89
Figure 30- Results From Water Surface Semantic Segmentation In The Orthophotos. The First 2 Rows Are Examples Of Narrow Rivers And The 2 Last Rows Represent Large Rivers 94
Figure 31- Examples Of False Positives And False Negatives Water Segmentation 95

LISTA DE TABELAS

Table 1- Distribution of train and test datasets 57
Table 2- Metrics for evaluating the performance of methods 61
Table 3- Quantitative comparison of deep learning models 62
Table 4- Quantitative performance of the DeepLabV3+ model to segment entire orthophotos 67
Table 5- Quantitative performance of the DeepLabV3+ model trained in experiment 5 to segment entire orthophotos.....68
Table 6- Segmentation results of all water surfaces.....92
Table 7- Segmentation results of narrow rivers.....92

SUMÁRIO

1 CONSIDERAÇÕES INICIAIS 16

1.1 Introdução geral 17

1.2 Problemática 19

1.3 Objetivo geral 21

1.4 Objetivos específicos 21

1.5 Contribuições 22

CAPÍTULO 1- DEEP LEARNING APPLIED TO WATER BODY EXTRACTION IN REMOTE SENSING IMAGES - A SCIENTOMETRIC ANALYSIS 23

1. INTRODUCTION 24

2. METHOD 27

3. RESULTS 29

3.1 Temporal analysis and Research Countries 29

3.2 Influential Journals and Topics in the water body subject 31

3.3 Detailed analysis of the extraction of the water body using deep learning methods 33

4. DISCUSSION 39

5. CONCLUSION 42

REFERENCES 44

CAPÍTULO 2 - MAPPING LARGE AND NARROW RIVERS USING SEMANTIC SEGMENTATION IN HIGH SPATIAL RESOLUTION RGB IMAGES 48

1. INTRODUCTION 50

2. MATERIALS AND METHOD 53

2.1 Data labeling and split 54

2.2 Deep Semantic Segmentation networks and Experimental Setup 58

2.2.1 U-Net 58

2.2.2 PSPNet 59

2.2.3 DeepLabV3+ 60

2.3 Evaluation metrics 61

3. RESULTS AND DISCUSSION 62

4. CONCLUSION 71

REFERENCES 73

CAPÍTULO 3– RIVER MAPPING USING TRANSFORMER-BASED ARCHITECTURE IN HIGH SPATIAL RESOLUTION RGB IMAGES 81
1. INTRODUCTION 82
2. MATERIALS AND METHOD 85
2.1 Data acquisition, processing, and labeling 86
2.2 Semantic Segmentation Background 88
2.3 Experimental details 90
3. RESULTS AND DISCUSSION 91
3.1 Evaluation of Quali-Quantitative Performance 91
4. CONCLUSION 95
REFERENCES 96
2 CONSIDERAÇÕES FINAIS 99
REFERÊNCIAS 101

1 CONSIDERAÇÕES INICIAIS

Graduada em Engenharia Ambiental pela Universidade do Oeste Paulista (2010). Especialista em Gestão Ambiental em Municípios pela Universidade Tecnológica Federal do Paraná (UTFPR) (2015). Mestre em Meio Ambiente e Desenvolvimento Regional pelo Programa de Pós Graduação da Universidade do Oeste Paulista (PPGMADRE - UNOESTE) (2019). Durante o mestrado fui bolsista CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e desenvolvi a pesquisas na área de ciências ambientais, a partir do uso de geotecnologias. Na dissertação de mestrado, atuei no desenvolvimento de uma abordagem geoespacial para a definição de área para aterro sanitário em escala intermunicipal (aterro consorciado). Integrante dos grupos de pesquisa: Núcleo de Estudos Ambientais (NEAmb) desde 2015 e Geomática e Inteligência Artificial Aplicada a Análise Ambiental (Gi3A) desde 2019. Entrei no doutorado em abril de 2019, também com bolsa CAPES, e tenho trabalhado com métodos de aprendizado profundo no processamento de dados de sensoriamento remoto para mapeamento de recursos hídricos. Dentro do PPGMADRE, me mantive na linha de pesquisa 2 - Planejamento Ambiental e Desenvolvimento Regional, que envolve 4 eixos temáticos, dos quais, a minha pesquisa se insere no eixo “estratégias de manejo e conservação ambiental e de bacias hidrográficas”. Ainda nesta linha de pesquisa, os estudos procuram analisar o impacto da gestão do conhecimento e da informação no planejamento ambiental e desenvolvimento regional, bem como no desenvolvimento econômico e social e seu impacto ambiental.

Este documento está organizado em: Introdução Geral, Capítulo I, Capítulo II, Capítulo III, Considerações Finais. A *introdução geral*, apresenta a problemática e objetivos da presente pesquisa. O *Capítulo I* é composto por um breve levantamento científico sobre a literatura que compõe a pesquisa em formato de manuscrito. Neste capítulo, o levantamento é feito por uma análise cienciométrica de estudos que usaram aprendizagem profunda para extrair feições de corpo d’água em imagens de alta resolução espacial. Em seguida, no *Capítulo II* é apresentado uma abordagem para mapear rios com diferentes larguras (rios largos e rios estreitos) usando aprendizagem profunda em imagens RGB de alta resolução espacial. O *Capítulo III* apresenta a delineação de um manuscrito que completa o objetivo geral da presente pesquisa. Neste manuscrito é apresentado um novo método de aprendizagem

profunda para mapear rios (largos e estreitos) em imagens RGB de alta resolução espacial. Por fim as *Considerações finais* apresentam as principais conclusões da tese e os direcionamentos para trabalhos futuros.

1.1 Introdução geral

O mapeamento de recursos hídricos é fundamental para o planejamento ambiental, pois ajuda no desenvolvimento de projetos de conservação da água, monitoramento fluvial, prevenção de desastres e uso sustentável dos recursos hídricos (GUI *et al*., 2022). Ademais, esse tipo de mapeamento favorece investigações de detecção de seca, diagnóstico e planejamentos futuros. O mapeamento de rios pode ser uma tarefa extenuante para ser realizada exclusivamente em campo, principalmente quando se trata de grandes áreas geográficas, ou de mapeamento em períodos diferentes do ano. As imagens de sensoriamento remoto têm, por conseguinte auxiliado em tarefas como essas relacionadas aos recursos hídricos, como: mapeamento de águas superficiais (TANG *et al*., 2022), monitoramento de enchentes (SHAHABI *et al*., 2020), monitoramento da qualidade da água (SAGAN *et al*., 2020), análise e predição batimétrica (MA *et al*., 2020). Um estudo (LIU *et al*., 2019) foi capaz de identificar materiais, separadamente, inorgânicos e orgânicos em suspensão em lagos eutrofizados utilizando somente imagens de satélite (OLCI/Sentinel-3A). Outro trabalho (SUN *et al*., 2018), investigou a confiabilidade da calibração de um modelo hidrológico para bacias regionais não calibradas, com base nas larguras da superfície da água do rio derivadas de imagens ópticas de satélite (QuickBird, IKONOS e WorldView-1), e demonstram como os dados de sensoriamento remoto podem ser mais efetivamente integrados à modelagem hidrológica. Esses e outros estudos (ELHAG *et al*., 2019; KUHN *et al*., 2019) demonstram a viabilidade de se mapear a feição de rios em imagens de sensoriamento remoto e, inclusive, de se identificar características relacionada aos aspectos físico-químico, como a poluição, nesses corpos d'água.

As imagens de sensoriamento remoto podem ser adquiridas a partir de diferentes veículos e plataformas (NOVO, 2010), como plataformas orbitais, por meio de satélites, e plataformas aéreas, por meio de aviões ou Veículo Aéreo não Tripulado (VANT). As imagens orbitais têm sido amplamente usadas para o mapeamento de rios. Uma das razões é a facilidade de acesso a imagens gratuitas, como o caso do

satélite Landsat (USGS, 2020), e da resolução espacial média que apresentam, por exemplo, uma resolução de 30 m no caso das imagens Landsat. No entanto, imagens de média, ou mesmo baixa resolução, limitam a escala de mapeamento, por não permitirem a extração de rios menos espessos, como os rios estreitos. Uma necessidade neste caso é a adoção de imagens de maior resolução espacial que poderão não só permitir o mapeamento de rios mais estreitos como também a produção de um mapa em maior nível de detalhe. Nesse caso, espera-se que o uso de padrões espaciais no lugar de informações espectrais possa facilitar a identificação de rios.

Os métodos tradicionais mais usados, até então, para a extração de rios em imagens são baseados em características espectrais dos corpos d'água, definindo limites e usando índice para classificar e extrair informações sobre o corpo d'água (FEYISA *et al*., 2014; LI *et al*., 2016; WANG *et al*., 2018; BAO; LV; YAO, 2021; YULIANTO *et al*., 2022). Além desses, os métodos classificação avançaram no sentido de automatização e melhores desempenhos usando machine learning para extrair feições de rios (ACHARYA; SUBEDI; LEE, 2019; LI *et al*., 2021a; LI; FAN; QIN, 2021). Os métodos propostos têm por objetivo reduzir a quantidade de processos manuais realizados pelo usuário e, em paralelo, facilitarem a identificação de rios por meio de realces ou técnicas de pré-processamento de imagens e extração de características. Embora nota-se progressos nesses métodos, os mesmos ainda apresentam dificuldades em diferenciar feições de água com sombras, áreas edificadas ou nuvens.

Recentemente, uma nova abordagem tem sido explorada em imagens de sensoriamento remoto. Essa abordagem vem do campo da inteligência artificial, a aprendizagem profunda (Deep Learning – DL). Como uma subclasse do aprendizado de máquina, a aprendizagem profunda possui redes neurais mais profundas dos que as de aprendizado de máquina. Além disso, dispõe de uma representação hierárquica dos dados, permitindo maiores recursos de aprendizado, maior desempenho e precisão do que métodos mais comuns (LECUN; BENGIO; HINTON, 2015). As redes neurais profundas necessitam de maior potência computacional e alta demanda de dados rotulados. A segmentação semântica realiza a rotulagem em nível de pixel, ou seja, categoriza cada pixel em um conjunto de objetos (MINAEE *et al*., 2021). Contudo, as redes apresentam desempenhos impressionantes em diversas tarefas, como na segmentação semântica de pomares (OSCO *et al*., 2021), de copas de árvores

(MARTINS *et al*., 2021), construções urbanas (YI *et al*., 2019; SEONG; CHOI, 2021), estimativa de nível de água (MUHADI *et al*., 2021), entre outros.

1.2 Problemática

A segmentação semântica de imagens também tem sido utilizada na tarefa de detecção e extração de rios em imagens de sensoriamento remoto. Um dos primeiros trabalhos que investigaram essa perspectiva (ISIKDOGAN; BOVIK; PASSALACQUA, 2017) usou uma abordagem baseada em aprendizado profundo para o mapeamento de águas superficiais. O estudo propôs uma rede neural totalmente convolucional (FCN - Fully Convolutional Network) chamado Deep-WaterMap, para segmentar água em imagens Landsat. Apesar de apresentarem F1-Score de 90% o modelo proposto foi testado apenas em imagens multiespectrais de média-resolução. Estudos recentes tem usado a arquitetura de rede CNN que tem apresentado resultados importantes (ALAM *et al*., 2021). Mais recentemente, uma nova arquitetura de rede tem sido investigada, a arquitetura baseada em Transforers (VASWANI *et al*., 2017) que, ao obter excelentes resultados na área de programação de linguagem natural, chamou atenção da área de visão computacional. Inicialmente esta arquitetura de rede foi aplicada em conjunto ou substituindo componentes em redes convolucionais, a Vision Transformer (ViT) aplicada para classificação de imagens em visão computacional mostrou o potencial dessa arquitetura ao ser usada puramente (DOSOVITSKIY *et al*., 2020). Esta arquitetura apresentou excelentes resultado e ao mesmo tempo exigindo menor recurso computacional para treinamento. Baseado em ViT, o SegFormer (XIE et al., 2021) estruturado com um codificador de arquitetura hierárquica, menor do que o ViT, porém, capaz de capturar recursos de alta resolução grosseiros e finos de baixa resolução (XIE *et al*., 2021). Outra importante característica da estrutura do SegFormer é o uso de um decodificador leve e compacto baseado em MLP (Multilayer Perceptron) que exige menor custo computacional (XIE et al., 2021). O SegFormer tem sido usado em diversos contextos, como para mapear vegetação urbana (GEORGES GOMES, 2022) e mapeamento de áreas queimadas (GONÇALVES *et al*., 2023). Porém, tem sido pouco explorado no contexto de recursos hídricos. Considerando o potencial do SegFormer em obter ganhos de contexto e em multiescala, investigar o seu potencial para mapear recursos hídricos como os rios poderá trazer novas descobertas científicas.

Uma outra questão a ser considerada é quando se trata de imagens de baixa resolução espectral, como imagens RGB (Red, Green, Blue), o desafio de mapear corpos d'água pode ser maior quando comparadas com imagens multiespectrais, por exemplo. Por não se beneficiarem das faixas do infravermelho próximo e médio, não se beneficiam, portanto, do maior contraste entre a água e demais alvos. Por outro lado, as imagens RGB podem se beneficiar quando trabalhadas em alta resolução, por exemplo. Imagens com maior nível de detalhes apresentam menor mistura de pixels, porém, possuem maior variação de pixels e maior volume dados, precisando, portanto, de modelos mais robustos para capturar as feições do alvo.

Outro problema a ser superado é a rotulagem das feições de interesse (rios) para compor a base de dados. Por exigir muitos recursos humanos e de tempo para rotular um grande volume de amostras, a deficiência de conjuntos de dados rotulados é algo que precisa ser superado. Segundo um trabalho atual de revisão de literatura (LI *et al*., 2022), existem apenas 10 conjuntos de dados disponíveis publicamente que podem ser aplicados para a avaliação e aprendizado de supervisão da classificação de corpos d'água a partir de imagens ópticas de sensoriamento remoto de alta resolução. Ainda, segundo os autores, os conjuntos de dados de imagens de alta resolução existentes não são suficientes para fornecer características globais de corpos d'água tão diversificados. Essa diversidade de características de corpos d'água é outra questão a ser levantada, sendo difícil defini-los com um padrão unificado. Os corpos d'água podem ser divididos, de uma maneira geral, em grandes (rios principais e lagos) e pequenos (afluentes e lagoas), apresentando características multiescalares e, portanto, muitos desafios para a generalização em escala das abordagens existentes (LI *et al*., 2022). Além disso, as diferenças de distribuição devem ser consideradas, por exemplo, imagens obtidas de regiões ou sensores divergentes apresentam grandes diferenças de iluminação, tom de cor, textura e aparência (LI *et al*., 2022). Rios largos, por exemplo, possuem cor, tamanho, forma e características espectrais mais homogêneas. Enquanto rios estreitos apresentam sinuosidade, sedimentação e pixels de mistura, o que costuma tornar essas feições mais difíceis de segmentar. Portanto, fatores como variação de formato, tamanho, distribuição, complexidade de cena, assinatura espectral, bordas complexas e difíceis de serem delimitadas devem ser considerados no mapeamento de rios por meio do sensoriamento remoto e métodos de aprendizado profundo.

Estudos que investiguem a extração de feições de rios em imagens RGB de alta resolução espacial explorando novas redes de segmentação semática são recentes (MIAO *et al*., 2018; WANG *et al*., 2021; LI *et al*., 2021b; GAUTAM; SINGHAI, 2022; HU *et al*., 2022). Contudo, os estudos mensionados apresentam alguns problemas a serem enfrentados em comum, como a dificuldade em se segmentar rios estreitos com alta acurácia. Nenhum trabalho até o momento, avaliou a performance e desempenho das redes de segmentação semântica em diferentes larguras de rios, como os rios estreitos, e a atual proposta visa o mapeamento dessas feições em grande escala cartográfica. Além disso, o mapeamento de rios em imagens de baixa resolução espectral, como as imagens as RGB, usando redes neurais ViT não foram testadas, constituindo assim, outra lacuna científica importante que requer ser suprida. Nesse sentido, como originalidade da presente proposta se tem a aplicação de métodos de aprendizado profundo (redes profundas de segmentação semântica) em imagens aéreas RGB de alta resolução espacial para mapear rios largos e estreitos, considerando os métodos do estado-da-arte em segmentação semântica para essa tarefa.

1.3 Objetivo geral

Mapear rios com diferentes tamanhos em imagens aéreas RGB de alta resolução espacial utilizando redes profundas de segmentação semântica.

1.4 Objetivos específicos

Como objetivos específicos, tem-se: (1) avaliar o desempenho de redes profundas de segmentação semântica ao identificar rios com largura superior a 10 metros em imagens RGB de alta resolução espacial; (2) caracterizar as capacidades e limitações das redes profundas de segmentação semântica em identificar rios de pouca largura (inferior a 10 metros) em imagens RGB de alta resolução espacial; e (3) avaliar o desempenho de redes profundas de segmentação semântica baseadas em ViT, comparando-as com as redes tradicionais baseadas em convoluções.

1.5 Contribuições

As principais contribuições deste trabalho envolvem disponibilizar uma estratégia de mapeamento acurado de rios com diferentes características a partir de imagens de alta resolução espacial, o que é importante para auxiliar tarefas de planejamento e gestão de recursos hídricos. Com esse tipo de produto é possível responder questões do tipo: “Quais tipos e dimensões de rios existem em determinada área geográfica?”, o que é importante para agilizar e priorizar medidas de planejamento, gestão e conservação dos mesmos. Outra contribuição deste estudo é tornar público o conjunto de dados rotulado. A tarefa de anotação de feições em imagens é onerosa em termos de tempo e nem sempre considerada fácil. Assim, disponibilizar esses dados irá permitir que novos modelos de aprendizagem profunda sejam treinados e aprimorados para a segmentação semântica de rios. Ademais, a estrutura deste trabalho poderá ser replicada por outros estudos, envolvendo novos conjuntos de dados e podendo beneficiar outras regiões do Brasil e do mundo.

CAPÍTULO 1 - DEEP LEARNING APPLIED TO WATER BODY EXTRACTION IN REMOTE SENSING IMAGES - A SCIENTOMETRIC ANALYSIS

Abstract: Deep learning methods have been used for many tasks, including water body extraction. Understanding which types of images, sensor platforms, image resolution, deep net, and metrics, most explored in recent years can give us information about trends or gaps in the literature related to the subject of study. So far, there is a lack of study of the quantitative aspects of science and scientific production referring to the information mentioned above. Thus, our work aimed to compose a scientometric analysis for the extraction of water bodies in remote sensing images using deep learning methods and to obtain an overview of current research. We analyzed 366 articles and verified the trends and paths taken so far within the adopted theme. Results indicated an increase over the years and a great interest in water body extraction in remote sensing images using deep learning. Another finding is that the subject of water bodies extraction has been studied mostly from multispectral images, with more than 10 m of resolution images. In this way, a specific approach using RGB images in high-resolution remote sensing (HRRS) images to extract water bodies with deep learning methods has been little explored in the last few years. In addition, there is another lacune to investigate the narrow rivers segmentation from HHRS RGB images exploring deep learning approaches accurately. This study is expected to open paths and contribute to the advancement of research in this field.

Keywords: Deep-learning; Remote sensing; High-Resolution; Water body mapping.

Resumo: Os métodos de aprendizado profundo têm sido usados para muitas tarefas, incluindo a extração de corpos d'água. Entender quais tipos de imagens, plataformas de sensores, resolução de imagens, redes profundas e métricas mais exploradas nos últimos anos pode nos viabilizar informações sobre tendências ou lacunas na literatura relacionada ao tema de estudo. Até o momento, falta o estudo dos aspectos quantitativos da ciência e da produção científica referente às informações acima mencionadas. Assim, nosso trabalho teve como objetivo compor uma análise cienciométrica para extração de corpos d'água em imagens de sensoriamento remoto utilizando métodos de aprendizado profundo e obter um panorama das pesquisas atuais. Analisamos 366 artigos e verificamos as tendências e caminhos percorridos

até o momento dentro do tema adotado. Os resultados indicaram um aumento ao longo dos anos e um grande interesse na extração de corpos d'água em imagens de sensoriamento remoto usando aprendizado profundo. Outra constatação é que o tema da extração de corpos d'água tem sido estudado principalmente a partir de imagens multiespectrais, com mais de 5 metros de resolução de imagens. Desta forma, uma abordagem específica usando imagens RGB de sensoriamento remoto de alta resolução espacial para extração de corpos d'água com métodos de aprendizado profundo tem sido pouco explorada nos últimos anos. Além disso, há outra lacuna para investigar a segmentação de rios estreitos a partir desse tipo de imagem explorando abordagens de aprendizado profundo com precisão. Espera-se que este estudo abra caminhos e contribua para o avanço das pesquisas nesta área.

Palavras-chave: Aprendizagem profunda; Sensoriamento remoto; Alta resolução; Mapeamento de corpos d'água.

1. INTRODUCTION

Deep learning (DL) methods consist of powerful and robust techniques to improve the mapping of the Earth's surface (OSCO *et al*., 2021b). Specifically for DL tasks including images, the most frequently used architecture are based on CNN (Convolutional Neural Network), which can work with object or region detection, segmentation, and recognition (LECUN, 1989; LECUN; BENGIO; HINTON, 2015). The DL method with CNN architecture is widely used for a variety of tasks in remote sensing images, such as for extracting buildings (WANG; MIAO, 2022), counting plants, and detecting plantation rows (OSCO *et al*., 2021a), apple classification (SUN *et al*., 2021), plant leaf disease (LU; TAN; JIANG, 2021), water level prediction (PAN *et al*., 2020), water quality (BARZEGAR; AALAMI; ADAMOWSKI, 2020), and water body segmentation (ZHANG *et al*., 2021). To summarize it, we create a word cloud with a simple search for articles based on the Web of Science database using the words: Deep Learning AND Remote Sensing (Figure 1). Thus, it is clear that words, such as CNN, segmentation, detection, and classification are very frequent, reinforcing the use of CNN for image classification issues. Another observation is the following cloud words: urban, road, water, crop, and tree, which appear in that respective order of frequency. This demonstrates how much the topic related to water has been widely studied.

Figure 1- Word cloud from articles with “deep learning” and “remote sensing” string.

Fonte: Autor (2023).

The importance of water resources for human existence and the management of life cycles in nature, studying their dynamics and comportment, has been essential for our survival and maintenance of life. Based on remote sensing images, water resources like the sea, water bodies, streams, lakes, and ponds have been widely studied. One of the ways to study the characteristics of water has been the extraction or mapping of water resources using DL methods applied in remote sensing images (ZHAO; FENG, 2022; MORADKHANI; FATHI, 2022; LUO; TONG; HU, 2021; YANG; WANG; ZENG, 2021; JAMES; SCHILLACI; LIPANI, 2021). A study (HOESER; BACHOFER; KUENZER, 2020) afford an overview of the CNN for image segmentation and object detection in Earth observation research by analyzing 429 publications. The authors concluded regards the CNN architecture in earth observation is that is an established tool and will further continue to increase its relevance. Another conclusion reveled for the authors is related to the availability of datasets with very high spatial resolution and suggests that studies focused on object dynamics will widely employ

the CNN DL model. A review (SIT *et al*., 2020) focused on DL methods in the water subject reviewed 129 studies. The information evaluated by them consisted of Architecture, Framework/Library/Programming Language, Dataset, Source Code Sharing, Reproducibility, Subfield: Deep Learning, and Subfield: Environment. One of the conclusions pointed out by the authors indicated that the water topic will continue to integrate DL methods and this methodology contributed to a profusion of opportunities in the application and research of hydrological science. However, this review just focuses on the tendency of DL methods. More recently, another study (YANG *et al*., 2022) contributed to a systematic review of DL methods in the water subject. Still, in this case, the authors investigate water quality detection and monitoring in addition to water extraction. In conclusion, the research regarding water body detection and water quality monitoring has been realized separately, and linking the two can allow synoptic water quality monitoring. Moreover, the integration of remote sensing, machine learning, and DL methods has considerable potential to address water resource monitoring and management.

Recently, a study (LI *et al*., 2022a) summarized the current common methods of water extraction based on optical and radar images. The methods such as the threshold method, support vector machine (SVM), decision tree, object-oriented extraction, and deep learning were evaluated in Landsat Collection 2 Level-2 images. The authors pointed a few issues that required further study, such as: (a) fusion of multisource remote sensing data to bring more possibilities for water body extraction, (b) efforts to solve water body mixed pixels issue, (c) complex optical properties and lack universality of the methods for inland water bodies, (d) lack of uniform evaluation standard for the results of water extraction, which is not conducive to a comparison between various methods and (e) a concentration of research in a small or local area may be improved with the increase of the available platforms of cloud computing technologies.

Currently, research (LI *et al*., 2022b) summarized and analyzed the achievements, and perspectives for water body classification from high-resolution optical remote sensing imagery. They considered five challenges of the water bodies' features and pointed five respective opportunities combined with advanced deep learning techniques. Further, they selected two representative benchmarks to employ for evaluating 10 typical approaches, and discussed their performance. The study affirmed that the water bodies from high-resolution images commonly have a variety

of shapes and sizes, wide and complex distribution scenes, boundaries winding, and exists a lack of data sets that can be used for supervised training. Despite of the challenging extracting the water body from high-resolution optical remote sensing images, the authors points that with the rapid development of computer vision techniques and remote sensing technologies the challenges can be overcome in the in the future.

Although the mentioned studies analyze DL methods involving the theme of water, to date, no review study points out how the literature has dealt with the semantic segmentation of water bodies studying the quantitative aspects, answering questions such as: "Which types of images", "Which sensors", and "Which spatial resolution", "Which metrics ", "Which countries", and "Which journals" have been most used to investigate water body mapping. Therefore, our objective in this work is summarize the studies that worked with the extraction of water bodies in remote sensing images using deep learning methods and to answer the mentioned questions to obtain an overview of current research in this regard.

2. METHOD

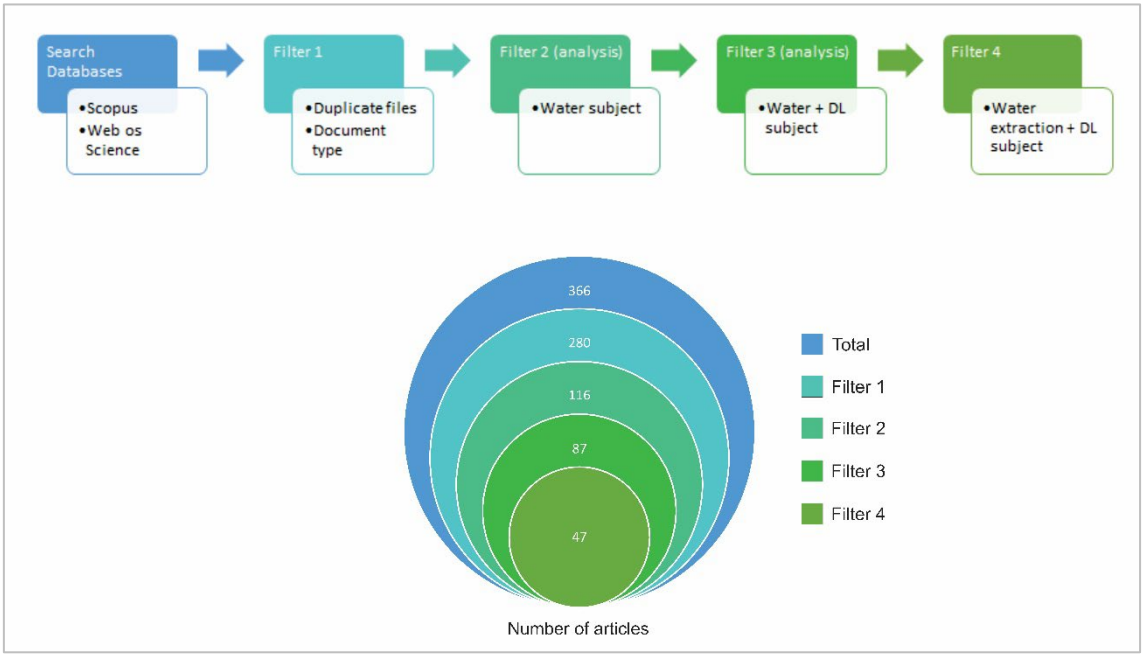
We select Scopus and Web of Science (WOS) databases to realize the scientometric analysis of water body extraction in remote sensing images using deep learning. These databases have a significant number of publications, and journals with high-impact indices, and are very appreciated in the academic field. In both databases we used the advanced search with the query string: ALL = (TS = ((deep learning OR convolutional neural network OR CNN OR Deep convolutional neural networks OR DCNN OR Semantic Segmentation OR transformers) AND (mapping river OR water extraction OR body water mapping OR Water body detection OR river identification) AND (remote sensing OR multispectral imagery OR RGB OR aerial images OR orbital images OR drone OR UAV))). Scopus base search from 1960-2022 and WOS from 1945-2022. Our study was actualized until June 2022.

Although the mentioned studies analyze DL methods involving the theme of water, to date, no review study points out how the literature has dealt with the semantic segmentation of water bodies studying the quantitative aspects, answering questions such as: "Which types of images", "Which sensors", and "Which spatial resolution", "Which metrics ", "Which countries", and "Which journals" have been most used to investigate water body mapping. Therefore, our objective in this work is summarize the

studies that worked with the extraction of water bodies in remote sensing images using deep learning methods and to answer the mentioned questions to obtain an overview of current research in this regard.

We found a total of 366 papers until June 2022, 118 from the Scopus and 248 from WOS databases. Although the search string allows for an extensive range of data, it can increase the possibility of appearing data that is not exactly what we are aimed. An example is articles that used deep learning for agriculture, medicine, construction, and vegetation problems, which are different from the theme we were searching for. Therefore, we filter the result manually, organizing in tables, to better select the data (Figure 2). Filter 1 excludes duplicate articles and document types, which separates conference documents, books, book chapters, reviews, letters, and proceedings papers. Filter 2 selects all the articles with the water theme. From this selection, we manually reviewed the papers one by one to extract specific publication information such as year of publication, co-authorship, journal, and country. Filter 3 selects all the papers with the water and deep learning approach themes and shows the topics on water body subjects. Lastly, filter 4 selects only water body extraction topics with a deep learning approach, and analyzes data type, sensor platform, resolution, network, and metrics. Then, we discuss about the possible tendencies and lacunes from the mentioned analyses from articles of this last filter, obtaining an overview from literature.

Figure 2- Filters and quantity of related articles.



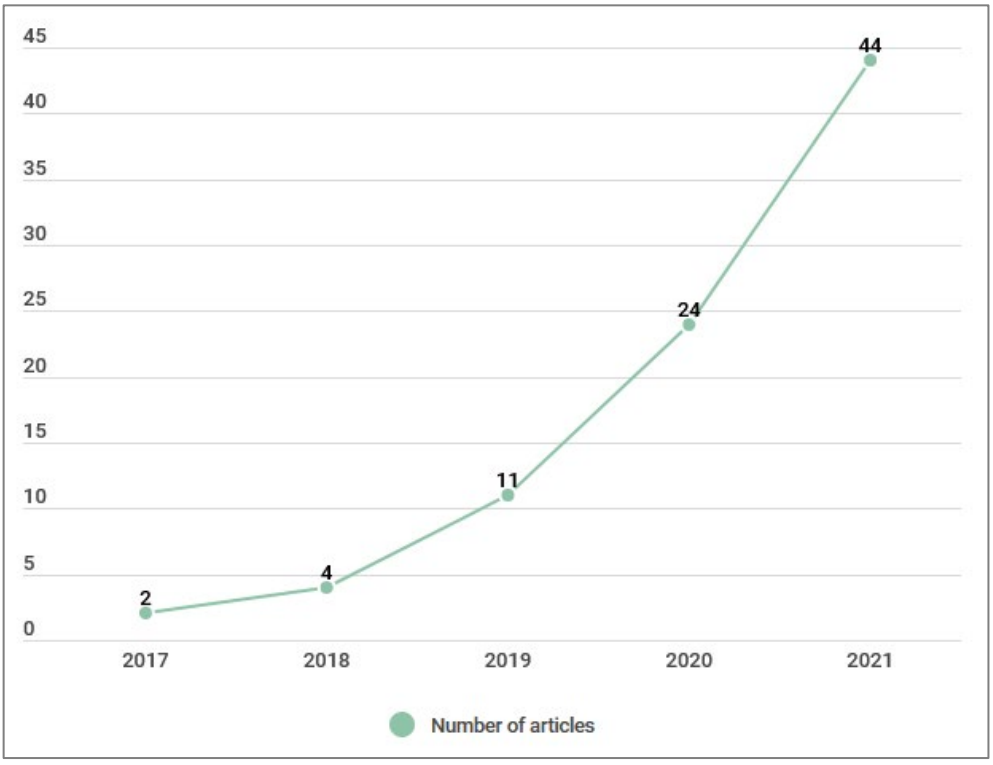
Fonte: Autor (2023).

3. RESULTS

3.1 Temporal analysis and Research Countries

The dataset analysis comes from filter 2, i.e., from the total of 366 documents, we eliminated duplicates, documents that are not published as journal articles, and themes distinct from the water theme, the remaining 116 documents. Figure 3 shows the number of these articles distributed over the years. The first publications found in the year 2017 studied sea-land segmentation and surface water mapping. This year's highlight is the article titled "Surface Water Mapping by Deep Learning" by Isikdogan, Bovik and Passalacqua in the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. The authors proposed a fully convolutional neural network (FCN) called DeepWaterMap to separate water from land, snow, ice, clouds, and shadows in Landsat images. This initiative to use deep learning to map surface water surpasses the previous techniques such as modified normalized difference water index (MNDWI) and the traditional multilayer perceptron (MLP) approach, opening up ways to use the deep learning model to map water bodies.

Figure 3- Temporal distribution of articles from 2017 to June 2022.

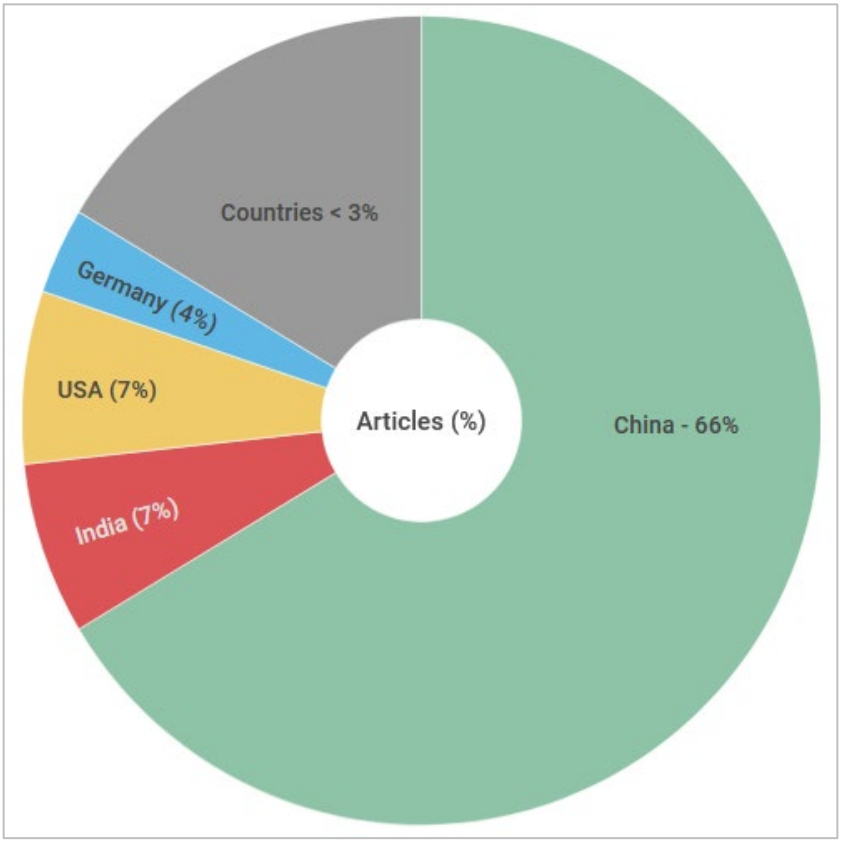


Fonte: Autor (2023).

The temporal line (Figure 3) presents the number of articles climbing significantly over the years. Although the search for this study went up to June 2022, we plotted the results only for full years, therefore, from 2017 to 2021. From 2017 to 2018 the number doubled and continues to rise in the coming years. Each year there is an average increase of more than 50% and from 2017 to 2021 there is an increase of almost 95%. This analysis demonstrates the exponential growth of studies related to water body extraction in remote sensing images using deep learning.

Regarding the country's publications, considering that some articles may have authors from one or more countries, for a better analysis, we separated the countries of the first authors (Figure 4). China is the country with massive publications (66%) compared with other countries. USA and India represent 7% each, and Germany with 4%. The other countries represent less than 3% of papers published. These analysis results demonstrate that China is investing efforts and have a lot of interest in studies related to water and deep learning subjects.

Figure 4- Percent of articles published per Country.

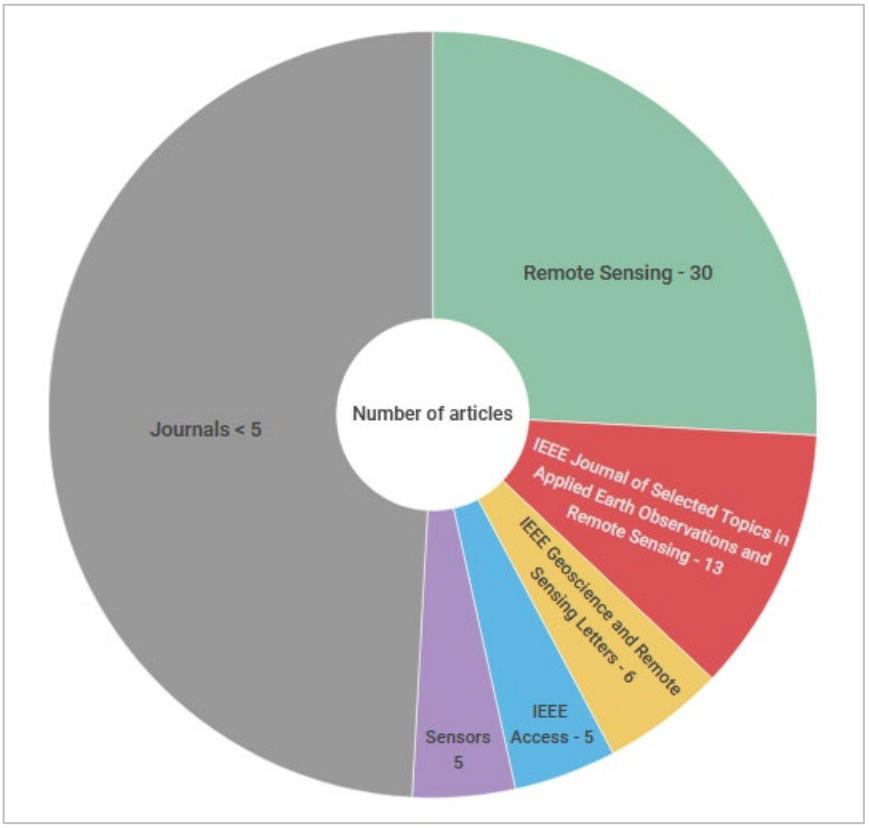


Fonte: Autor (2023).

3.2 Influential Journals and Topics in the water body subject

The 116 articles appeared in 46 different journals. Of all journals, a total of 32 (70%) published only 1 article. For better visualization, we grouped journals with less than 5 publications (89%). Then, we selected the top 5 journals with the highest number of publications, that is, more than 5 publications, to be represented in Figure 5. The journal with the highest number of publications, Remote Sensing, published 30 articles in 4 years. Then the journal IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing with 13 publications, starting with the “Surface Water Mapping by Deep Learning” publication by Isikdogan, Bovik and Passalacqua cited previously. Afterward the IEEE Geoscience and Remote Sensing Letters (6), IEEE Access (5), and MDPI Sensors (5). Comparing the journals selected in the top 5 (Figure 5), the remote sensing journal represents 51% of the total (59) articles published in this selection. Therefore, the journal Remote Sensing has received substantial attention from researchers and could become one of the most powerful drivers of deep learning and water research in the future.

Figure 5- Top 5 influential journals.

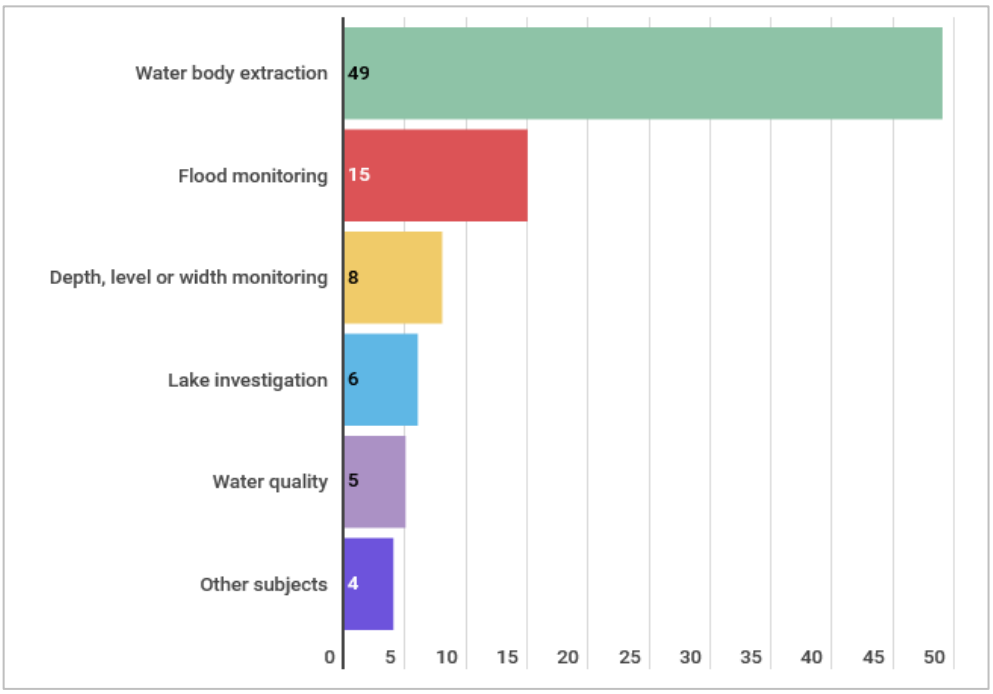


Fonte: Autor (2023).

Afterward, we applied Filter 3 in the previous dataset (116 articles) and selected articles that have studied water bodies using deep learning methods (87 articles). At this moment, we excluded studies with machine learning, wetlands, and object recognition subjects. Still, there are different subtopics in water body subjects, such as flooding, water quality, specific lake investigation, and others. Thus, we separated the subtopics in the water theme and the topic of interest “water body extraction with methods of deep learning” appears in 47 articles (Filter 4) analyzed in this research’s next item.

From the subtopics of the water body theme (Figure 6), the “flood monitoring” topic is present in 15 articles, being the second most studied topic. The topic “depth, level, or width monitoring” may be related to “flood monitoring”, however, we analyzed only the title name, thus totaling 9 articles on this topic. Despite the “Lake investigation”, “Water quality”, and “Other subjects” topics that have appeared in some documents, “Flood monitoring” has been of greater research interest in recent years.

Figure 6- Top 5 main topics research in water body fields.

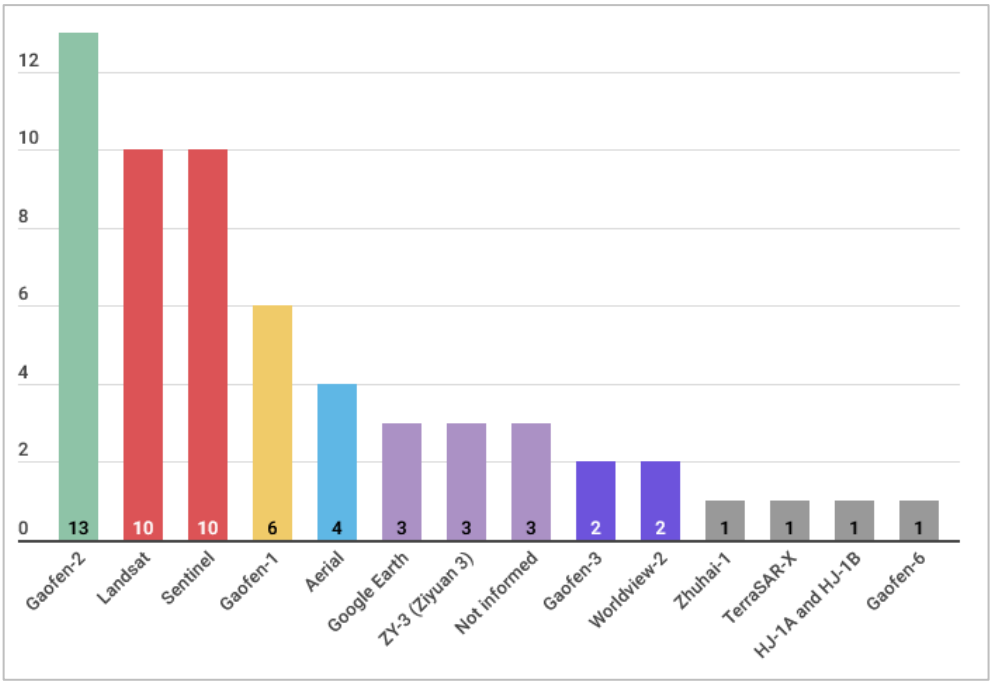


Fonte: Autor (2023).

3.3 Detailed analysis of the extraction of the water body using deep learning methods

Considering the 47 articles filtered (Filter 4) in the previous stage, the filter includes articles related to water body extraction using a deep learning method from 2017 to June 2022, therefore, we analyzed each article seeing the following topics: sensor platform, image resolution, image type, metrics, and network architecture used. For the sensor platform analysis, we separate each platform, even though the article used more than one platform in the study. Although we analyzed 47 articles, the number of platforms found was more, due to some articles using more than one platform in the study. Just four articles adopted aerial platforms, all others used orbital platforms (Figure 7). Gaofen-2 was the most used orbital platform, appearing in 13 articles, followed by Landsat and Sentinel, both, resulting in 10 articles each, and Gaofen-1 in six articles (Figure 7). The Gaofen series of satellites are Chinese high-resolution Earth imaging satellites that are part of the China High-Resolution Earth Observation System (CHEOS) program. Although they are not freely available to the population, their wide use may be due to the vast majority of studies realized in China (item 3.2). Widely known and used in several studies, Landsat and Sentinel appears as the most used satellite on our planet. Both satellites can have the images acquired free of charge, through the US website Geological Survey and Copernicus (Earth Observation Program of the European Union), respectively.

Figure 7- Platforms sensors for the article.



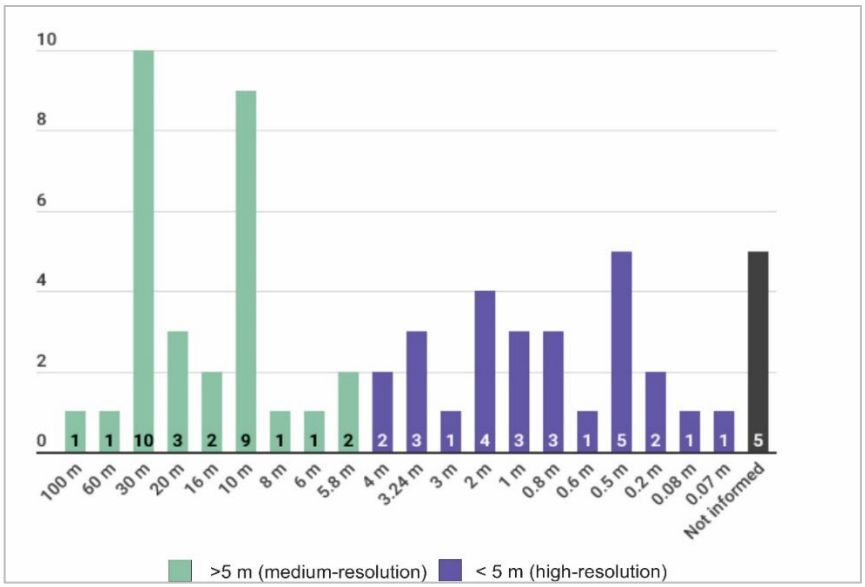
Fonte: Autor (2023).

Regarding the spatial resolution, 30 m was the resolution mostly explored in the databases, followed by 10 m resolution (Figure 8). This result can be explained due to the popularity of Landsat and Sentinel, as mentioned previously (Figure 7). Although Gaofen-2 was the most studied, this satellite has different resolutions, such as 3.24 m (multispectral), and 0.8 m (panchromatic), and was combined resulting in other resolutions, such as 2 m. In Landsat, for example, the studies work with just 30 m resolution.

Resolution values range from 0.2 m to 100 m. We divided the spatial resolutions into: above 5 m, and less than 5 m, which means, medium resolution and high spatial resolution. Although images with more than 5 m were the most explored (54%), compared to all images with less than 5 m, they represent 46% of all articles analyzed, demonstrating that high spatial resolution is almost equivalent to medium-resolution images studied. Among the submeter images (less than 1 m), which represent 23% of the analyzed articles, the resolutions of 0.5 m and 0.8 m were the most investigated. Despite the submeter representing 1/4 of the proportion of articles analyzed, with technological advances and the increase in the availability of high-resolution images, this scenario may change over the years. Another visual representation (Figure 9), demonstrates the temporal distribution of the spatial

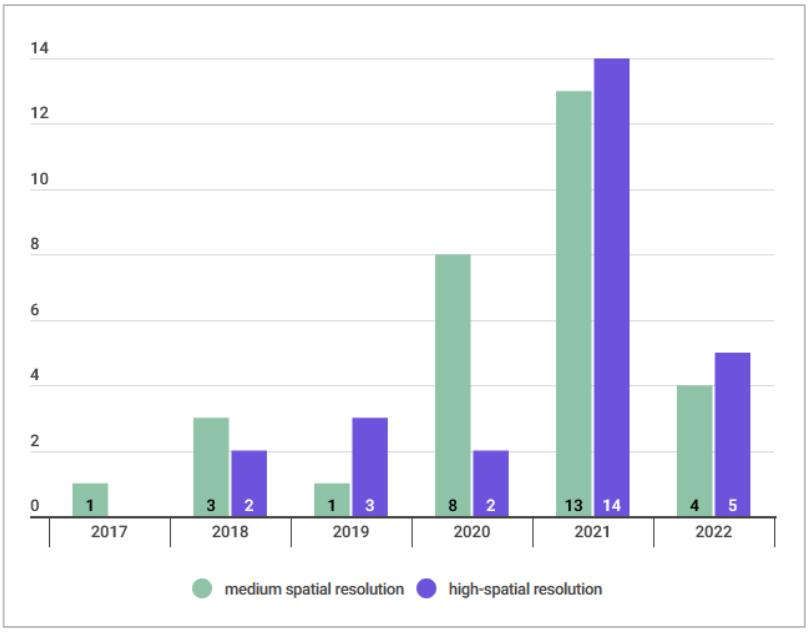
resolutions. We note that in 2022 high spatial resolution overcome the medium spatial resolution in the articles analyzed and may indicate a tendency to explore high-spatial images in the next years.

Figure 8- Image spatial resolution of the studies.



Fonte: Autor (2023).

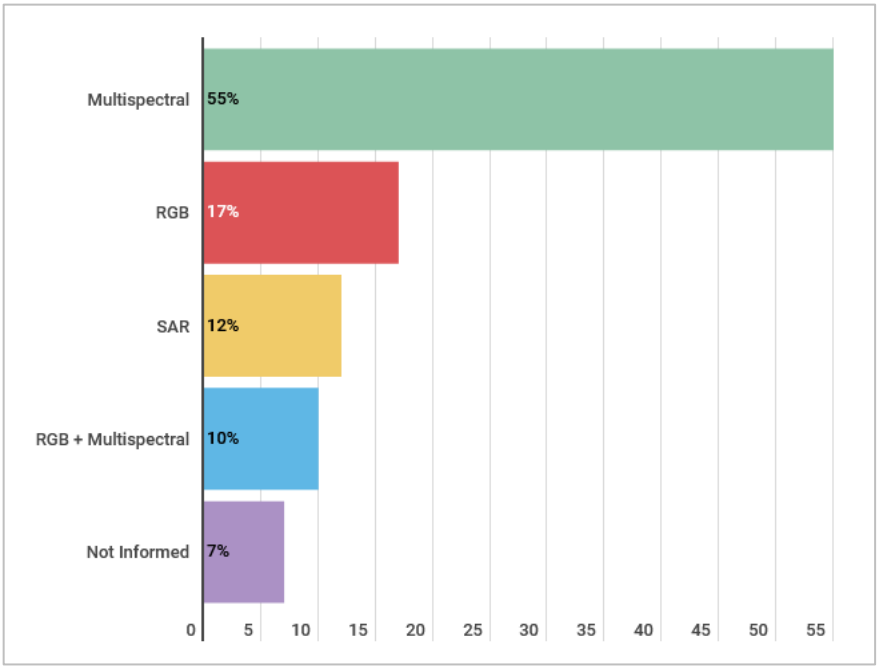
Figure 9- Distribution of image spatial resolution over the years.



Fonte: Autor (2023).

The data type investigated in the studies was multispectral data, RGB data, both multispectral and RGB data, and Synthetic Aperture Radar (SAR) data. Many works benefit from the multispectral data, because there may be greater reflectance in the length of the infrared spectrum, especially if the water quality changes. Although most works use multispectral images (55%), some sensors capture only the RGB bands, being the second most used type of data by scientific studies (Figure 10).

Figure 10- Data type percent.

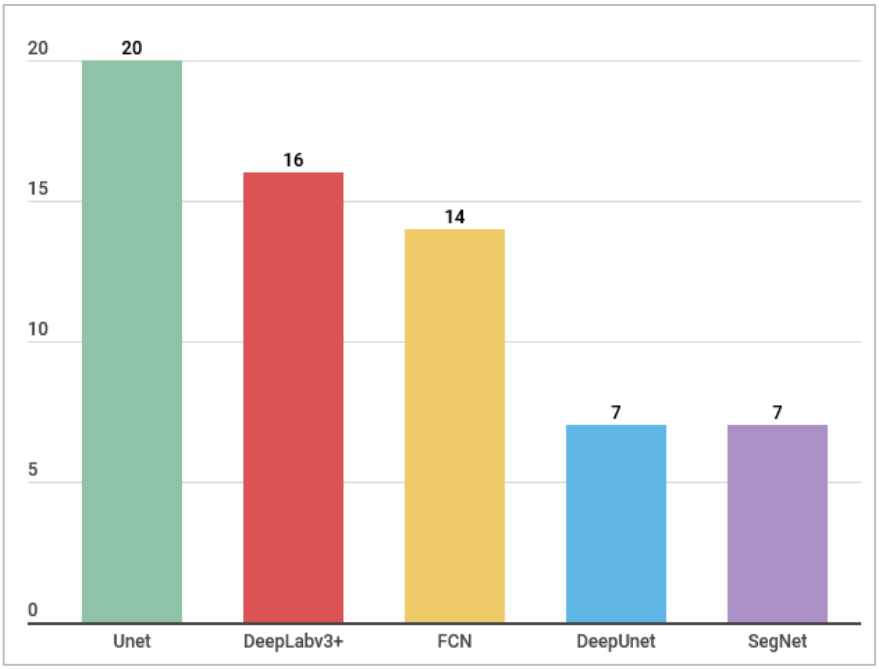


Fonte: Autor (2023).

About 80% of the articles analyzed used new network architectures based on convolutional neural networks (CNN) or modified deep neural networks to extract water bodies in remote sensing images. All articles used different methods and neural networks to compare the proposed method. Most compared it with traditional deep neural networks, machine learning methods (Support Vector Machine - SVM and Random Forest - RF), and threshold-based methods such as NDWI (Normalized Difference Water Index) and MNDWI (Modified Normalized Difference Water Index). To aid the analysis, we selected the 5 main deep neural network architectures adopted in the articles studied (Figure 11). The Unet (20), DeepLabV3+ (16), and FCN (14) architectures were widely used, presenting a slight difference in terms of the number of times used in the articles studied. DeepUnet and SegNet, cited in 7 articles each,

were among the top 5 CNNs used, with just under half usage compared to the other architectures. Therefore, the most used CNN networks to extract bodies of water from SR images in the last 5 years were Unet, DeepLabV3+, and FCN.

Figure 11- Top 5 deep learning methods.



Fonte: Autor (2023).

All the methods are evaluated with metrics that are widely adopted in most articles. Metrics are calculated from the confusion matrix and the classified pixels are compared to the ground truth. From all pixels correctly predicted to the class of interest, it is the true positive (TP); for all correctly predicted to the other class (background), it is the true negative (TN). For all pixels falsely predicted to the class of interest, it is a false positive (FP), and the missing and incoming label is the false negative (FN). All the metrics are calculated as the equations below (DUAN *et al.*, 2021):

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN}$$

(3)

$$Overall\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(4)

$$IoU = \frac{TP}{TP + FN + FP}$$

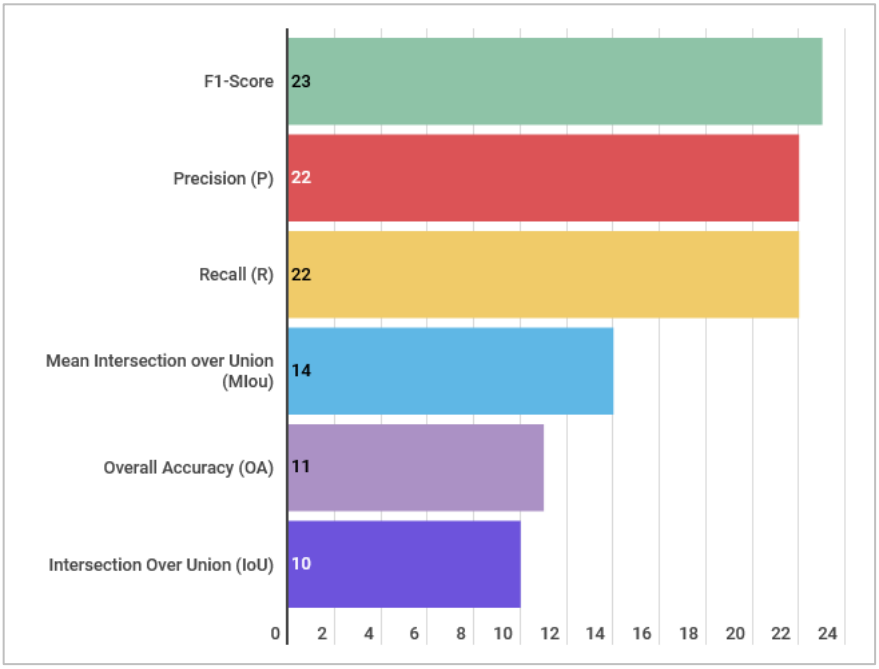
(5)

$$MIoU = \frac{1}{n+1}\sum_{i=0}^{n} \quad \frac{TP}{TP + FP + FN}$$

(6)

The mostly metrics used were: F1-Score, Precision (P), Recall (R), Mean Intersection over Union (MIoU), Overall Accuracy (OA), and Intersection over Union (IoU) respective (Figure 12). Other metrics were used, for example, Pixel Accuracy (PA), Kappa Coefficient (Kappa), Commission Error, Omission Error, Mean Pixel Accuracy (MPA), Dice Coefficient Loss, Producer's Accuracy, User's Accuracy, Jaccard Loss, Boundary Overall Accuracy (BOA), Edge Overall Accuracy (EOA), Specificity and Mathews Correlation Coefficient (MCC). However, were used just under 10 times, thus, we focus on the top six metrics mostly used in the articles.

Figure 12- Most used metrics.



Fonte: Autor (2023).

4. DISCUSSION

In the last 5 years, the subject of water bodies extraction has been studied mostly from multispectral images, with more than 5 m of resolution images. The method most used was CNN architecture, mainly with a proposed or modified architecture, then using Unet, and DeepLabV3+. In this way, a specific approach using RGB images in HRRS (high-resolution remote sensing) images to extract water bodies with deep learning methods has been little explored in the last few years. A study (MIAO *et al*., 2018) proposed a segmentation network named restricted receptive field deconvolution network (RRF DeconvNet) to extract water bodies automatically. To demonstrate their method, they adopted a high-resolution RGB dataset from Google Earth images. The images were RGB pan-sharpened with 0.5 m resolution and showed lakes, reservoirs, rivers, ponds, paddies, and ditches in the water-body class. They achieved 96.5% of accuracy in the referred water segmentation, however, they used just one metric to evaluate the segmentation. Overall Accuracy is widely used for this task, however, when the dataset has imbalance class problems, like segmenting a minority class (water) from an image, this metric may not be the best to evaluate the performance network. Because this metric is greatly affected by the proportion of the majority class, in cases of class imbalance, it is important to employ an evaluation metric that is appropriate to the problem, which, in this case, can favor the minority class (HOSSIN; SULAIMAN, 2015).

Recently, Wang *et al*. (2021), proposed the MobileNetV2 for water body extraction in multisensor high-resolution remote sensing images. The method was performed in RGB images from the sensors GaoFen-2 (GF-2), WorldView-2, and UAV orthoimages. GF-2 and WoldView-2 orbital platforms used pan-sharpened RGB images with 1.0 m and 0.5 m resolution, respectively, and the UAV ortho images with 0.2 m resolution. The experiment was compared with other methods: support vector machine (SVM), random forest (RF), and U-Net. F1-Score for the proposed method achieved the best metrics, obtaining 75% for GF-2, 86% for Worldview-2, and 98% for UAV. Despite the excellent result presented in the image provided by the UAV, this dataset contained large lakes and dams. Unlike the other sensors, in the UAV image, there were no narrow rivers, at least visually, in the images presented in the work. Another conclusion is that training with a combination of lower and higher spatial resolution images can be beneficial, however, using just lower resolution imagery can

be limited. Despite obtaining good results with the increase in spatial resolution, the requirement to improve the proposed network for the extraction of mixed and small-area water bodies was pointed out.

Another work (DANG; LI, 2021) proposed a multiscale residual network (MSResNet) that uses self-supervised learning (SSL) for water-body detection. The method was performed in RGB images from the 2020 Gaofen Challenge water-body segmentation dataset with 0.5 m resolution, and multispectral images from the GID dataset with 4 m resolution. The mean intersection over union (MIoU), and the frequency weighted intersection over union (FWIoU) were respectively 85.82%, and 90.88% for RGB images and 94.94%, and 95.08% for multispectral images. Despite the water body being labeled to have a variety like lakes, ponds, rivers, paddies, and seas, the datasets don't present complex spectral mixtures like sedimentation and sinuosity which may bring more difficulty to the task of segmentation. On the other hand, a study (ZHANG *et al*., 2021) designed a specific network to deal with complex spectral mixtures, a multi-feature extraction, and a combination module (MECNet). The network was assembled to reach a rich feature representation, extracting the complete target information from the local space, larger space, and between-channel relationships. In addition, they adopted a multi-scale prediction fusion module and applied an encoder-decoder semantic feature fusion module to promote fusion effects. The method was evaluated with aerial (0.2 m) images, and satellite (0.5 m) images from GF-2, both with RGB bands. The method proposed was compared with U-Net, RefineNet, DeeplabV3+, DANet, and CascadePS and performed better than these methods. The results were around 90% for IoU, with 90,64% for aerial and 90.80% for satellite images. However, the method was designed to focus on the global information of the feature maps, which turns in less attention to the spatial relationship between feature maps. Despite the study working with high spatial resolution RGB images using deep learning to extract bodies of water with complex spectral variations, challenges such as extracting small rivers from the image, for example, were not addressed.

More recently, an investigation (HU *et al*., 2022) proposed a network with multi-scale feature aggregation for water segmentation. The network used ResNet in downsampling feature extraction to obtain rich context information, aggregate spatial information, and semantic information. Furthermore, the multi-branch aggregation module was used for two-channel information communication to provide rich pixel information for the recovery of up-sampling information. The experiment was

performed in Landsat-8 and Google Earth (QuickBird and Earth Sat) using 30 m, 1m, and 0.6 m image resolution. Additionally, the authors used cloud and cloud shadow datasets from Google Earth and LandCover datasets with 0.25 m and 0.50 m resolution to test whether the algorithm has the same segmentation performance in different tasks and generalized the assignment. In the first test, the authors compared 11 networks and achieved better performance with an MIoU of 95.94 %. For the cloud dataset, the proposed method has better results compared with 5 different networks and achieves 87.28% MIoU. This method accomplished 92.89% MIoU from the LandCover dataset, overcoming all 7 networks compared. The method accomplishes these results even with complex backgrounds and small rivers. Despite efforts to obtain rich context information, according to the authors, there are still points to be improved, for example, improving the number of parameters, reducing the weight of the model, relieving the training pressure, optimizing the backbone network, changing the convolution kernel or the convolution type, and even continuing to select a lighter network.

Another recent research (GAUTAM; SINGHAI, 2022), proposed a modified convolution neural network (MCNN) for water body extraction, with attention in achieve accurate water boundaries. For this proposal, the authors modified the raw image with RGB-HIS conversion and enhancement, utilizing the double plateaus histogram equalization (DPHE) algorithm. Then, morphological operations combining erosion and dilation are used and the necessary features are selected from the extracted features using the chaotic forest optimization algorithm (CFOA). The classification was done using MCNN, and, the segmentation process was done using the cosine similarity-based watershed segmentation algorithm (CSWS). To validate the method (MCNN) in the classification step, they compared with NN (neural network), CNN (convolutional neural network), ANFIS (adaptive neuro-fuzzy interference system), DNN (deep neural network), ENN (Elman neural network), and RNN (recurrent neural network) achieving higher metric values of sensitivity, specificity, accuracy, precision, and recall over the mentioned existing deep neural network classifiers. In the segmentation step, they compared CSWS with active contour (AC), K-means (KM), fuzzy C-means (FCM), and region growing (RG) methods based on the aforementioned evaluation metrics. The results demonstrate that the CSWS method gives accurate water boundary delineation. However, the proposed methodology can be enhanced by considering more advanced algorithms. In addition, despite this

method focusing on water boundaries which can bring better water segmentation, the study doesn't focus on and evaluate small water bodies.

A further recent survey (WANG *et al*., 2022) proposed a water bodies extraction method (SADA-Net) for high-resolution remote sensing images. The authors consider the method's multiscale information, context dependence, and shape features. For this, the network framework integrates three components: shape feature optimization (SFO), atrous spatial pyramid pooling (ASPP), and dual attention (DA) modules. The experiment was conducted in RGB and multispectral images from GID (Gaofen Image Dataset) and compared with SegNet, LinkNet, DeeplabV3+, Attention U-Net, MECNet, and MSResNet methods. The results were better for SADA-Net with F1-Score 88.57% in RGB images and 96.14% in multispectral images. Regardless of the study mentioned processing with small water bodies, the work mainly refers to artificial ponds. Thus, specific small water bodies like narrow rivers were not considered.

The analyzed studies have shown advances in the semantic segmentation of rivers in high spatial resolution images. However, even exploring different modules and modifications in deep neural networks to achieve better results, studies still have difficulty in segmenting small rivers (narrows). This is an important challenge to be overcome, which can be explored through specific studies for narrow rivers, and also through other DL approaches.

5. CONCLUSION

This study aimed to provide a scientometric analysis of water body semantic segmentation in remote sensing images with deep learning methods. Water body extraction with a deep learning approach has been a lot of attention and has been a hot topic in the last 5 years. The primary studies started in 2017 and are increasing widely over the years until June 2022. Multispectral images have greater scope for this task, however, RGB images have also been used and, they can open paths for new analysis. Although most studies investigate images with medium-spatial resolution over the last years, high spatial resolution have gained more attention in the last year (2022). Deep learning methods have been mainly updated or created a new proposal based on CNN architecture for the task of water body segmentation. A lot of studies proposed new improving methods to achieve better accuracies to segment the water

body. However, all of the studies indicate the difficulty to segment water complex features, such as sedimented water, spectral mixtures, and mainly small water bodies. Regardless of whether the studies have better results with new methods, small bodies are still a barrier to being passed. Therefore, remaining the lacune to investigate the water body extraction from HHRS RGB images using deep learning and explore other deep learning approaches, such as ViT-based, to accurately segment narrow rivers.

REFERENCES

BARZEGAR, R.; AALAMI, M. T.; ADAMOWSKI, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. **Stochastic Environmental Research and Risk Assessment**, v. 34, n. 2, p. 415-433, 2020. Disponível em:<https://link.springer.com/article/10.1007/s00477-020-01776-2>. Acesso em: 28 de março de 2021.

DANG, B.; LI, Y. MSResNet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery. **Remote Sensing**, v. 13, n. 16, p. 3122, 2021. Disponível em:<https://www.mdpi.com/2072-4292/13/16/3122>. Acesso em: 09 de abril de 2022.

DUAN, Y. *et al*. A New Lightweight Convolutional Neural Network for Multi-Scale Land Surface Water Extraction from GaoFen-1D Satellite Images. **Remote Sensing**, v. 13, n. 22, p. 4576, 2021. Disponível em:<https://www.mdpi.com/1357606>. Acesso em: 12 de fevereiro de 2022.

GAUTAM, S.; SINGHAI, J. Cosine-similarity watershed algorithm for water-body segmentation applying deep neural network classifier. **Environmental Earth Sciences**, v. 81, n. 9, p. 1-16, 2022. Disponível em:< https://link.springer.com/article/10.1007/s12665-022-10376-y>. Acesso em: 19 de março de 2022.

HOESER, T.; BACHOFER, F.; KUENZER, C. Object detection and image segmentation with deep learning on Earth observation data: A review—Part II: Applications. **Remote Sensing**, v. 12, n. 18, p. 3053, 2020. Disponível em:< https://www.mdpi.com/831252>. Acesso em: 29 de março de 2022.

HOSSIN, M.; SULAIMAN, M. N. A review of evaluation metrics for data classification evaluations. **International journal of data mining & knowledge management process**, v. 5, n. 2, p. 1, 2015. Disponível em:< https://www.academia.edu/download/37219940/5215ijdkp01.pdf>. Acesso em: 28 de outubro de 2021.

HU, K. *et al*. Multi-Scale Feature Aggregation Network for Water Area Segmentation. **Remote Sensing**, v. 14, n. 1, p. 206, 2022. Disponível em:< https://www.mdpi.com/article/10.3390/rs14010206>. Acesso em: 23 de junho de 2022.

ISIKDOGAN, F.; BOVIK, A. C.; PASSALACQUA, P. Surface water mapping by deep learning. **IEEE journal of selected topics in applied earth observations and remote sensing**, v. 10, n. 11, p. 4909-4918, 2017. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8013683/>. Acesso em: 19 de maio de 2021.

JAMES, T.; SCHILLACI, C.; LIPANI, A. Convolutional neural networks for water segmentation using sentinel-2 red, green, blue (RGB) composites and derived spectral indices. **International Journal of Remote Sensing**, v. 42, n. 14, p. 5338-5365, 2021. Disponível em:< https://www.tandfonline.com/doi/abs/10.1080/01431161.2021.1913298>. Acesso em: 26 de agosto de 2020.

LECUN, Y. Generalization and network design strategies. **Connectionism in perspective**, v. 19, n. 143-155, p. 18, 1989. Disponível em:< https://www.academia.edu/download/30766382/lecun.pdf>. Acesso em: 17 de setembro de 2019.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015. Disponível em:< https://www.nature.com/articles/nature14539>. Acesso em: 22 de agosto de 2019.

LI, J. *et al.* Satellite detection of surface water extent: A review of methodology. **Water**, v. 14, n. 7, p. 1148, 2022a. Disponível em:< https://www.mdpi.com/2073-4441/14/7/1148>. Acesso em: 23 de setembro de 2022.

LI, Y. *et al*. Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 187, p. 306-327, 2022b. Disponível em:< https://www.sciencedirect.com/science/article/pii/S0924271622000867>. Acesso em: 27 de agosto de 2022.

LU, J.; TAN, L.; JIANG, H. Review on convolutional neural network (CNN) applied to plant leaf disease classification. **Agriculture**, v. 11, n. 8, p. 707, 2021. Disponível em:< https://www.mdpi.com/1204652>. Acesso em: 16 de agosto de 2022.

LUO, X.; TONG, X.; HU, Z. An applicable and automatic method for earth surface water mapping based on multispectral images. **International Journal of Applied Earth Observation and Geoinformation**, v. 103, p. 102472, 2021. Disponível em:< https://www.sciencedirect.com/science/article/pii/S0303243421001793>. Acesso em: 18 de agosto de 2022.

MIAO, Z. *et al*. Automatic water-body segmentation from high-resolution satellite images via deep networks. **IEEE geoscience and remote sensing letters**, v. 15, n. 4, p. 602-606, 2018. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8286914/>. Acesso em: 25 de setembro de 2022.

MORADKHANI, K.; FATHI, A. Segmentation of waterbodies in remote sensing images using deep stacked ensemble model. **Applied Soft Computing**, p. 109038, 2022. Disponível em:< https://www.sciencedirect.com/science/article/pii/S1568494622003465>. Acesso em: 10 de setembro de 2022.

OSCO, L. P. *et al*. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 174, p. 1-17, 2021a. Disponível em:< https://www.sciencedirect.com/science/article/pii/S0924271621000307>. Acesso em: 21 de fevereiro de 2022.

OSCO, L. P. *et al*. A review on deep learning in UAV remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, v. 102, p. 102456, 2021b. Disponível em:< https://www.sciencedirect.com/science/article/pii/S030324342100163X>. Acesso em: 23 de março de 2022.

PAN, M. et al. Water level prediction model based on GRU and CNN. **Ieee Access**, v. 8, p. 60090-60100, 2020. Disponível em:< https://ieeexplore.ieee.org/abstract/document/9044367/>. Acesso em: 27 de outubro de 2021.

SIT, M. et al. A comprehensive review of deep learning applications in hydrology and water resources. **Water Science and Technology**, v. 82, n. 12, p. 2635-2670, 2020. Disponível em:< https://iwaponline.com/wst/article-abstract/82/12/2635/75838>. Acesso em: 15 de abril de 2021.

SUN, L. *et al*. An Improved CNN-Based Apple Appearance Quality Classification Method With Small Samples. **IEEE Access**, v. 9, p. 68054-68065, 2021. Disponível em:< https://ieeexplore.ieee.org/abstract/document/9422830/>. Acesso em: 22 de março de 2022.

WANG, B. *et al*. SADA-Net: A Shape Feature Optimization and Multiscale Context Information-Based Water Body Extraction Method for High-Resolution Remote Sensing Images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 1744-1759, 2022. Disponível em:< https://ieeexplore.ieee.org/abstract/document/9695297/>. Acesso em: 26 de maio de 2022.

WANG, H.; MIAO, F. Building extraction from remote sensing images using deep residual U-Net. **European Journal of Remote Sensing**, v. 55, n. 1, p. 71-85, 2022. Disponível em< https://www.tandfonline.com/doi/abs/10.1080/22797254.2021.2018944>. Acesso em: 11 de maio de 2022.

WANG, Y. *et al*. Lightweight Deep Neural Network Method for Water Body Extraction from High-Resolution Remote Sensing Images with Multisensors. **Sensors**, v. 21, n. 21, p. 7397, 2021. Disponível em:< https://www.mdpi.com/1347686>. Acesso em: 24 de fevereiro de 2022.

YANG, Q.; WANG, C.; ZENG, T. A method of water change monitoring in remote image time series based on long short time memory. **Remote Sensing Letters**, v. 12, n. 1, p. 30-39, 2021. Disponível em:< https://www.tandfonline.com/doi/abs/10.1080/2150704X.2020.1868602>. Acesso em: 16 de novembro de 2021.

YANG, L. *et al*. Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing. **Sensors**, v. 22, n. 6, p. 2416, 2022. Disponível em< https://www.mdpi.com/1424-8220/22/6/2416>. Acesso em: 17 de Agosto de 2022.

ZHANG, Z. *et al*. Rich CNN Features for water-body segmentation from very high resolution aerial and satellite imagery. **Remote Sensing**, v. 13, n. 10, p. 1912, 2021. Disponível em:< https://www.mdpi.com/2072-4292/13/10/1912>. Acesso em: 18 de março de 2022.

ZHAO, Y.; FENG, C. Remote Sensing Image Water Body Recognition Algorithm Based on Deep Convolution Generating Network and Combined Features. **Wireless Communications and Mobile Computing**, v. 2022, 2022. Disponível em:< https://www.hindawi.com/journals/wcmc/2022/9932251/>. Acesso em: 26 de setembro de 2022.

CAPÍTULO 2 - MAPPING LARGE AND NARROW RIVERS USING SEMANTIC SEGMENTATION IN HIGH SPATIAL RESOLUTION RGB IMAGES

Abstract: Accurate mapping of water resources is essential for planning the sustainable use of this important resource. Deep learning (DL) methods for semantic segmentation tasks based on the convolutional neural network (CNN) architectures, to deal with remote sensing images have been explored to map rivers. However, mapping rivers with high accuracy is a difficult task, mainly due to the complex scenarios of a water body, such as size variations and pixel mixtures, usually detected in high-resolution RGB images. Although there are studies with new DL methods to map water bodies, there are still difficulties to segment narrow rivers. In this way, there is still no in-depth analysis of the performance of semantic segmentation networks in mapping large (over a width of 10 m) and narrow rivers (a width less than 10 m). An approach that points out the differences between the performance of CNNs for semantic segmentation of large and narrow rivers can contribute to advances in river mapping. In this context, we propose an approach based on DL methods for semantic segmentation of large and narrow rivers in high spatial resolution RGB images. We conducted experiments applying different CNNs (DeepLabV3+, Unet, PSPNet) for mapping large rivers (width over 10 m) and narrow rivers (width of fewer than 10 m). The results showed that the CNN performance is different depending on the characteristics of this river, since narrower rivers tend to have different characteristics from larger water bodies. Training with the combination of large rivers and narrow rivers outcomes in better effects for both river widths. All the CNNs performed better in large river segmentation, with the best F1-Score 97.43% presented by UNet. However, there is a decrease in all CNNs performances for narrow rivers segmentation. This finding reinforces the difficulty in mapping rivers smaller than 10 m in RGB high-resolution images and opens perspectives for defining strategies to deal with this specific type of river. A CNN that can map narrow rivers accurately will have a good chance of mapping large rivers accurately. Future studies may explore the CNNs in multisensor or multitemporal datasets, and explore other new architectures, such as Transformer-based architectures.

Keywords: Deep Learning mapping; Water resources management; Environmental planning.

Resumo: O mapeamento preciso dos recursos hídricos é essencial para o planejamento do uso sustentável desse importante recurso. Métodos de aprendizado profundo (DL) para tarefas de segmentação semântica baseados em arquiteturas de redes neurais de convolução (CNN) para lidar com imagens de sensoriamento remoto foram explorados para mapear rios. No entanto, mapear rios com alta precisão é uma tarefa difícil, principalmente devido aos cenários complexos de um corpo d'água, como variações de tamanho e misturas de pixels, geralmente detectados em imagens RGB de alta resolução. Embora existam estudos com novos métodos de DL para mapear corpos d'água, ainda existem dificuldades para segmentar rios estreitos. Dessa forma, ainda não há uma análise aprofundada do desempenho das redes de segmentação semântica no mapeamento de rios largos (largura superior a 10 metros) e estreitos (largura inferior a 10 metros). Uma abordagem que aponte as diferenças entre o desempenho das CNNs para segmentação semântica de rios grandes e estreitos pode contribuir para possíveis avanços no mapeamento de rios. Neste contexto, propomos uma abordagem baseada em métodos DL para segmentação semântica de rios grandes e estreitos em imagens RGB de alta resolução espacial. Conduzimos experimentos aplicando diferentes CNNs (DeepLabV3+, Unet, PSPNet) para mapear rios grandes (largura superior a 10 metros) e rios estreitos (largura inferior a 10 metros). Os resultados mostraram que o desempenho da CNN é modificado de acordo com a largura do rio. O treinamento com a combinação de rios largos e rios estreitos resulta em melhores efeitos para ambas as larguras dos rios. Todas as CNNs tiveram melhor desempenho na segmentação de rios largos, com o melhor F1-Score 97,43% apresentado pela UNet. No entanto, há uma queda no desempenho de todas as CNNs para a segmentação de rios estreitos. Essa constatação reforça a dificuldade de mapear rios menores que 10 metros em imagens RGB de alta resolução e abre perspectivas para a definição de estratégias para lidar com esse tipo específico de rio. Uma CNN que pode mapear rios estreitos com precisão terá uma boa chance de mapear rios largos com precisão. Estudos futuros podem explorar as CNNs em conjuntos de dados multisensor ou multitemporais e explorar outras novas arquiteturas, como arquiteturas baseadas em Transformer.

Palavras-chave: Mapeamento de Deep Learning; Gestão de recursos hídricos; Planejamento Ambiental.

1. INTRODUCTION

Mapping water resources is essential for environmental planning. Obtaining accurate information about this resource can assist in several issues, such as biodiversity conservation, management of natural disasters, and strategies to minimize climate change (JIANG *et al*., 2018). Additionally, mapping water resources favor river monitoring as it can point out areas to be preserved and the spatial distribution of rivers, which can assist water management and water supply. However, mapping rivers is challenging, especially when dealing with large geographic areas that may be composed of rivers with different characteristics such as shape, composition, spectral signature, and size.

Remote sensing images have supported several environmental tasks development, such as those related to monitoring river channel dynamics (LANGAT; KUMAR; KOECH, 2019; DRUCE *et al*., 2021; CHEN *et al*., 2022), river flow management (SAMBOKO *et al*., 2020), river discharge estimation (KEBEDE *et al*., 2020), river level monitoring (MISHRA; PANT, 2020; KIM *et al*., 2022), flood hazard (NOGUEIRA *et al*., 2018; JANIZADEH *et al*., 2021; KALANTAR *et al*., 2021; CHENG *et al*., 2022), water quality estimation (YOTOVA *et al*., 2021; SUN *et al*., 2022), and anthropogenic impacts in ecological status (VIGIAK *et al*., 2021). In recent years, many of these tasks have been performed using high spatial resolution images because they allow identifying river features in large-scale detail. The literature has presented many attempts using different methods to map rivers using images of high spatial resolution (PEKEL *et al*., 2016; JIANG *et al*., 2018; CHEN *et al*., 2018; JIN *et al*., 2021; ZHENG; CHENG, 2021; MORADKHANI; FATHI, 2022). One of the primary methods to map water bodies is based on spectral characteristics, which consists of the single-band thresholds (SHIH,1985; ZHANG *et al*., 2017) and the multi-band thresholds. The multi-band thresholds use the spectral relationship method (DU; ZHOU, 1998), and the water index method (MCFEETERS, 1996; ROGERS; KEARNEY, 2004; XU, 2006; WANG *et al*., 2018) adopts band operation to extract information about water, benefiting from the water spectral signature.

The most common algorithm used to predict water bodies' class in remote sensing images are supervised classifiers, which have to analyze training samples to complete the task. For example, commonly applied algorithms include support vector machine (SVM), decision tree (DT), random forest (RF), and neural networks (NNs)

(HUANG *et al*., 2015; BANGIRA *et al*., 2019; ACHARYA; SUBEDI; LEE, 2019, LU *et al*., 2021; OLIVER *et al*., 2022). Methods that integrate the threshold approach and machine learning were used for mapping water bodies, but they have some gaps and problems that the algorithms cannot solve. In the threshold-based method the performance decline substantially when conducted with narrow and sedimented water bodies. These features are complex because of the spectral signature and it's difficult to define thresholds and derive from the spectral differences. Although machine learning-based methods have come to improve water classification, these methods have difficulty obtaining full context information from the image. For this reason, the algorithm loses a lot of useful image information and is easily affected by noise, declining the performance of the mapping task (KANG *et al*., 2021). Therefore, to fill these gaps, deep learning methods are explored. Representing a subclass of machine learning, deep learning algorithms refer to the state-of-the-art approach for extracting information in remote sensing images. Deep learning makes a hierarchical representation of the data, allowing greater learning resources, greater performance, and more accuracy than more common methods (LECUN; BENGIO; HINTON, 2015; GHAMISI *et al*. 2017; BADRINARAYANAN; KENDALL; CIPOLLA, 2017).

Although deep neural networks require greater computational power and high demand for labeled data, they can achieve impressive performances in several tasks, such as image classification (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; NOGUEIRA; PENATTI; DOS SANTOS, 2017; JAMALI *et al*., 2021), object detection (NOGUEIRA *et al*. 2019a; OSCO *et al*. 2020), semantic segmentation (BADRINARAYANAN; KENDALL; CIPOLLA, 2017; NOGUEIRA *et al*. 2019b; HAO; ZHOU; GUO, 2020), and instance segmentation (BAI; URTASUN, 2017; HAFIZ; BHAT, 2020). Different from image classification which provides a single label for the class(s) of the entire image, semantic segmentation performs pixel-level labeling, categorizing each pixel into a class of information. An instance segmentation task besides categorizing each pixel also detects and delineates each object of interest in the image (MINAEE *et al*. 2021).

The automatic river mapping from high-resolution spatial remotely sensed images has gained attention in both remote sensing and computer vision fields, and deep learning-based semantic segmentation methods, such as convolutional neural networks (CNN), have been applied to extracting rivers in remote sensing images in recent years (ISIKDOGAN; BOVIK; PASSALACQUA, 2017; CHEN *et al*. 2018; WEI *et*

al. 2020; GUO *et al*., 2020; TAMBE; TALBAR; CHAVAN, 2021; VERMA *et al*., 2021; THAYAMMAL *et al*., 2022; GAO *et al*., 2022). Recently, studies that explore water body extraction in high-resolution RGB images have gained attention (MIAO *et al*., 2018; WANG *et al*., 2021; DANG; LI, 2021; GAUTAM; SINGHAI, 2022; HU *et al*., 2022; ZHANG; LI; HUA, 2022).

A recent investigation (KANG *et al*., 2021) used high-resolution optical remote sensing images to propose a multiscale context extractor network (MSCENet) to delineate water bodies. The proposed method demonstrates a performance of 94.2% (F1-Score) for aerial RGB images and 95.35% for multispectral satellite images. However, the mentioned method has problems with the misclassification of shadows and trees, and the incomplete extraction of narrow rivers. Another research (ZHANG *et al*., 2021) proposed a multi-feature extraction and a combination module (MECNet) to extract water bodies with complex spectral mixtures. The combination module involves the multi-scale prediction fusion module and an encoder-decoder semantic feature fusion module (DSFF). The proposed method was evaluated with RGB aerial images (0.2 m) and RGB satellite images (0.5 m) and achieved around 90% of IoU. However, the (DSFF) misses the influence of the spatial relationship between feature maps, and the method still has difficulty in segmenting very narrow rivers. More recently, a study (WANG *et al*., 2022), proposed a method (SADA-Net) that combines components to extract multiscale information, context dependence, and shape features of water bodies. The method focuses on water body extraction from high-resolution remote sensing images, including both multispectral and RGB high-resolution images in the experiment. The proposed method achieves an of F1-Score 88.57% for RGB images and an F1-Score of 96.14% for multispectral images. However, only artificial ponds were considered small water bodies, which means that specific narrow rivers were not mentioned in the dataset.

As aforementioned, there are several attempts to create efficient CNNs for segmenting rivers in remote sensing images. However, there is no study that analyses deeply and verified qualitatively and quantitatively the performance of CNNs in specific river widths, such as large and narrow rivers. In addition, in mentioned studies, the dimension of narrow rivers is not known, for example, if the rivers have a width of fewer than 10 m, even in a high-resolution image, they can be considered narrow and more challenging for segmentation. Remarkably, dealing with narrow rivers is very difficult and demanding. Therefore, there is a gap in the literature regarding the comparison of

the performance of CNNs in the segmentation of large (over 10 m width) and narrow rivers (fewer 10 m width). A comparative approach like this can make it possible to verify more precisely which river size the network presents ease or difficulty in river mapping. In this way, it is easier to specifically address the gaps that the network presents.

The challenge is intensified when we have only the RGB bands to map rivers, because there is lower contrast between the target "water" and the background, different of when the infrared bands (near and medium) are use. Another challenge is the variety of river features in sizes different. Large rivers usually have more homogeneous color, size, shape, and spectral features. On the other hand, narrow rivers have sinuosity, sedimentation, and mixture pixels, which usually turn these features more challenging to segment. In this regard, an approach that investigates the performance of CNNs to map rivers of different sizes in RGB images of high spatial resolution is not evaluated up to the writing moment. This approach may consist of advancement in mapping rivers using a low-cost strategy that provides information at a high level of detail for this environmental feature.

We tested different CNNs and proposed an approach for mapping different sizes of rivers (large rivers and narrow rivers) in high spatial resolution RGB aerial images. The main contribution of this work is to present the performance of CNNs to map large and narrow rivers in RGB images, and make available a low-cost and accurate strategy to extract river features in RGB images. This consists of valuable information for the planning and management of water resources strategies. In addition, due to feature annotations in images being time consuming, troublesome, and hard-task, as another contribution of this study, we make available our dataset publicly[1]. Thus, deep-learning approaches can be evaluated in the future using the same dataset adopted here.

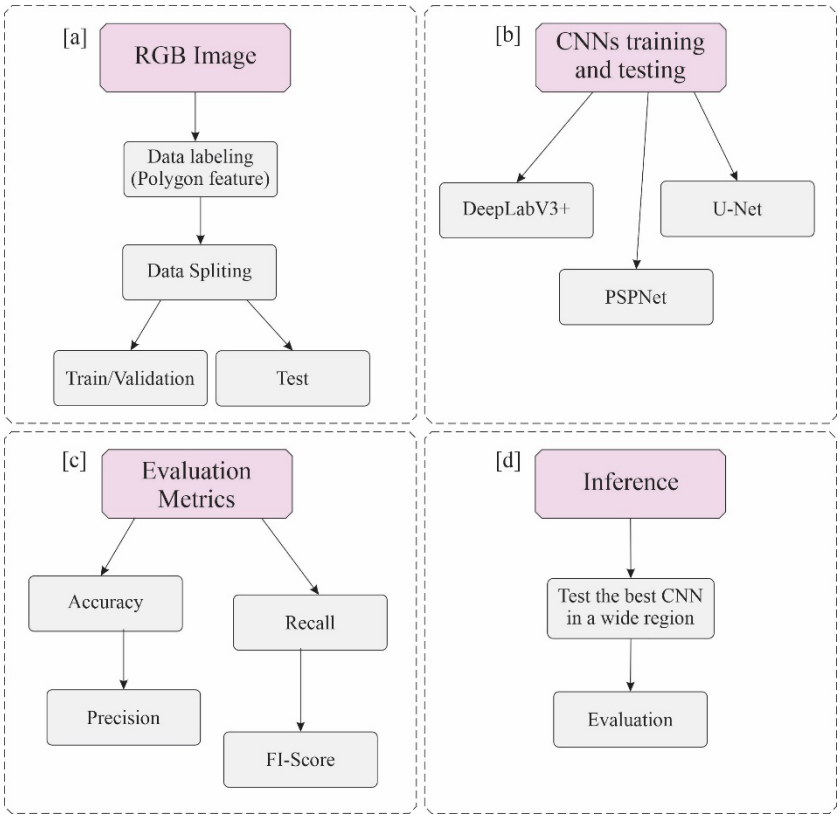
2. MATERIALS AND METHOD

We divided the method into four main stages (Figure 13). Initially, we labeled the rivers (a) in the RGB aerial images conceded by the Mapeia-SP project. Following this, we split the labeled data into training, validation, and testing datasets. In a

[1] https://github.com/mayarafaita/Large-Narrow_River.git

computational environment, we applied CNNs to train and test the images of large and narrow rivers. Then, we performed the evaluation (c) of the CNNs and compared them against each other. Lastly, we apply the best-trained CNN in a larger region (d), evaluate, and confirm if the best CNN could infer in a larger area. Details regarding these stages are described in the following subsections.

Figure 13- The main steps of the method are summarized in a flowchart.

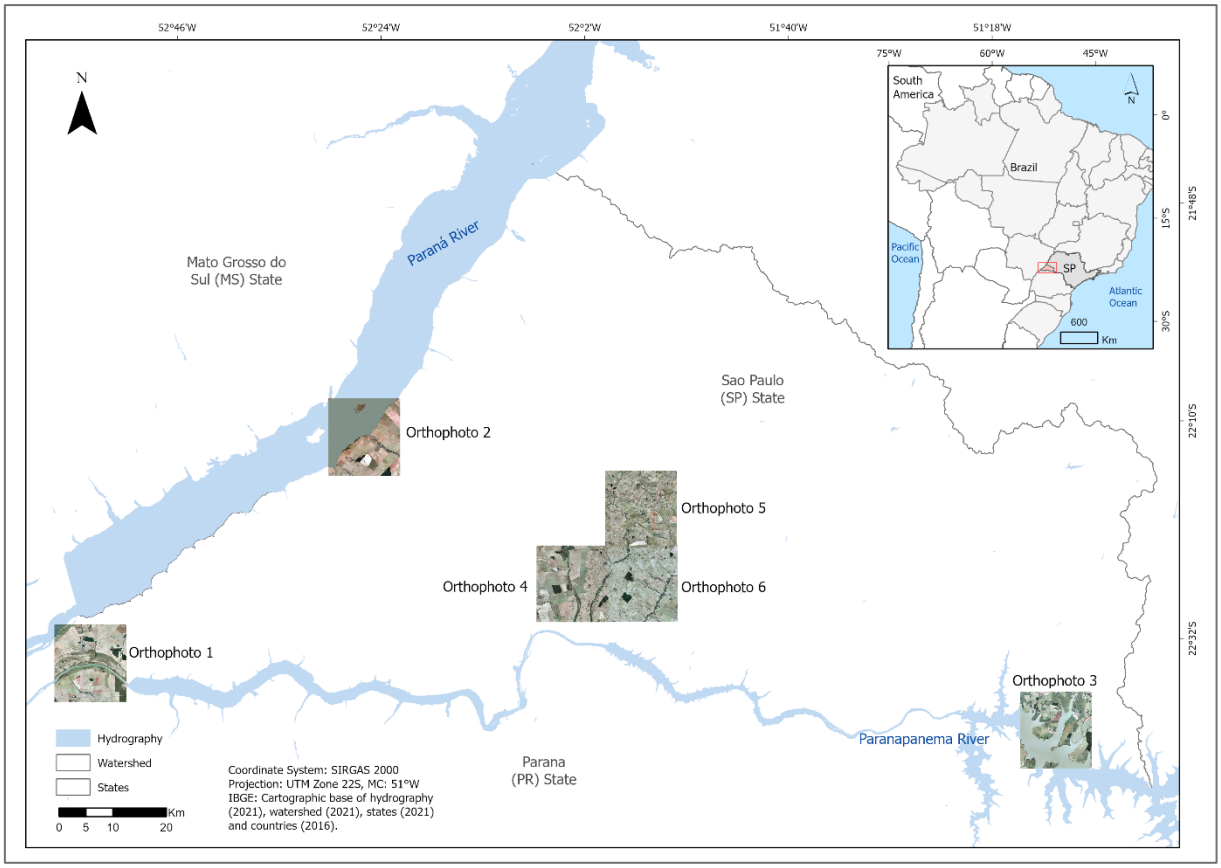


Fonte: Autor (2023).

2.1 Data labeling and split

We use aerial images with high spatial resolution from the Mapeia-SP project in Brazil. This project collected RGB aerial photographs of the entire state of São Paulo with a Ground Sample Distance (GSD) of 0.45 m. These images were orthorectified, i.e., acquired the characteristics of a map (orthophotos), and resembled 1 m of resolution. We selected as the study area a watershed with approximately 12,000 km² in the west of São Paulo state. Inside this area, we selected 6 orthophotos distributed throughout this region as experimental areas (Figure 14).

Figure 14- Study area.



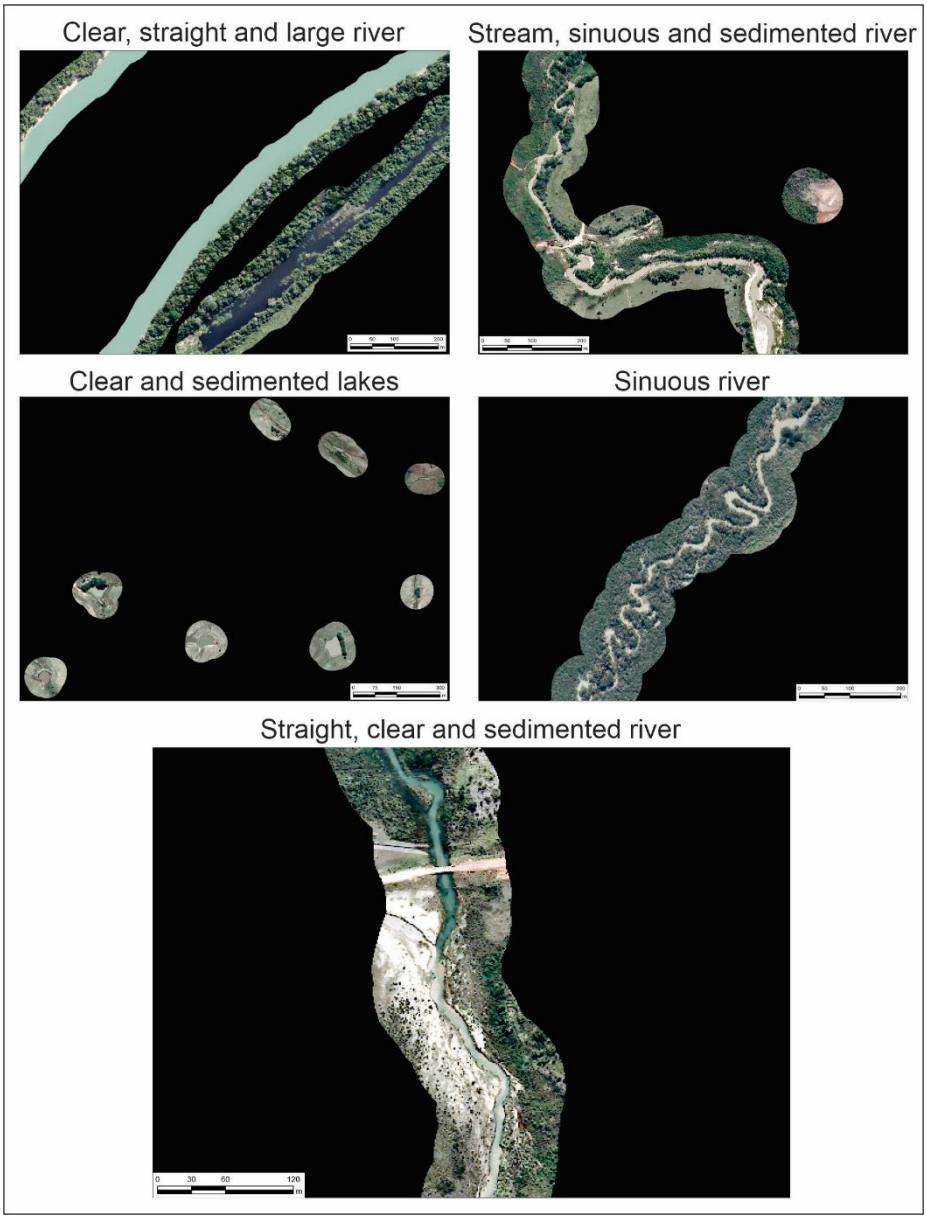
Fonte: Autor (2023).

To compose the training, validation, and testing data, we labeled the river's features in the RGB images. We performed this process by visual interpretation of the images using GIS and classified the labels as 1 for "river" and 2 for "no-river". The dataset is composed of different river characteristics including shapes, sizes, and spectral signatures. In the study area, we can detect large rivers, narrow rivers and lakes, sinuous and straight rivers, and limpid and sedimented rivers (Figure 15). These different characteristics can be confusing due to the wide variety of river samples, composing a complex and challenging dataset.

We conduct the experiment separating rivers with a width greater than 10 m (large rivers) and a width less than 10 m (narrow rivers) and evaluate the performance of deep semantic segmentation networks in the function of river size variations. The large river dataset has a much larger number of pixels than the small river (narrow rivers) dataset, which causes misleading class balancing with the background. The impact of class imbalance is detrimental to the performance of CNNs and this influence increases according to the scale of a task (BUDA; MAKI; MAZUROWSKI, 2018). To

deal with the class imbalance problem and save processing time, we conduct 3 initial experiments of undersampling adopting images that cover up to 50 m beyond the distance from the river's margin (Figure 15).

Figure 15- Examples of rivers that composed the dataset.



Fonte: Autor (2023).

From a total of 6 orthophotos with 13,436 x 14,409 pixels, we defined 5 experiments using different approaches (Table 1 and Figure 16). In experiments 1 and 2 we used 1 orthophoto for training and the other 5 orthophotos for testing the model (Table 1). In 3rd experiment, we combined large rivers and narrow rivers in a mosaic that was composed of 2 orthophotos for training and validation and the other 4

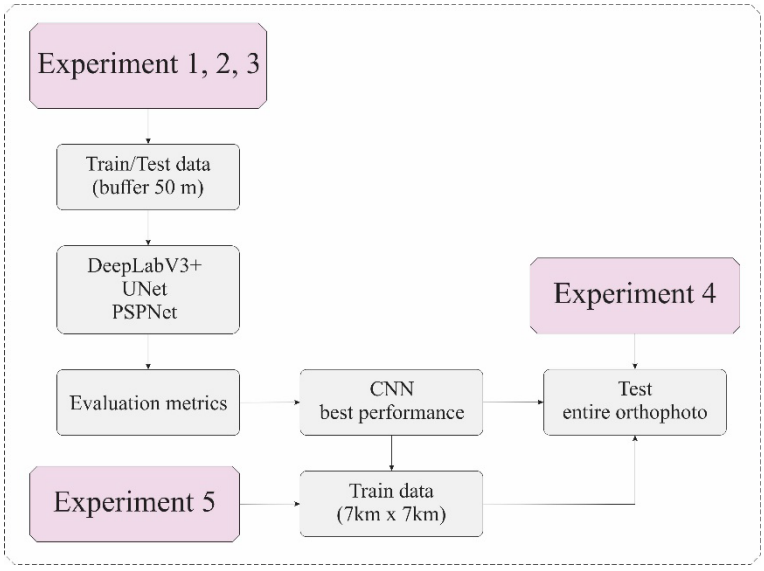
orthophotos to test the model's performance (Table 1). The training and validation stage was performed using non-overlapping patches of 256 x 256 pixels split into 70% for training and 30% for validation.

Table 1- Distribution of train and test datasets.

	Experiment 1	Experiment 2	Experiment 3
Train	Orthophoto 1 (large)	Orthophoto 4 (narrow)	Mosaic: Orthophoto 1 (large) Orthophoto 4 (narrow)
Number of patches (256x256)	381 (70% training; 30% validation)	434 (70% training; 30% validation)	826 (70% training; 30% validation)
Test	Orthophoto 2 (large) Orthophoto 3 (large) Orthophoto 4 (narrow) Orthophoto 5 (narrow) Orthophoto 6 (narrow)	Orthophoto 1 (large) Orthophoto 2 (large) Orthophoto 3 (large) Orthophoto 5 (narrow) Orthophoto 6 (narrow)	Orthophoto 2 (large) Orthophoto 3 (large) Orthophoto 5 (narrow) Orthophoto 6 (narrow)

Fonte: Autor (2023).

Figure 16- Flowchart of the experiments.



Fonte: Autor (2023).

We conduct experiments 4 and 5 (Figure 16), to verify the model's performance and its capability to map rivers in entire orthophotos. In experiment 4, we intend to assess the learning ability of the previously trained models to map rivers in the entire orthophoto. We proposed to test whether the model trained in a small area (orthophoto with buffer 50m) can segment rivers in a larger size area (entire orthophoto). Thereunto, we selected the best model for experiments 1, 2, and 3. Then,

we applied that model to a larger region, the entire orthophotos 2, 3, 5, and 6. Then we organize another training and testing set (experiment 5) using a different orthophoto size. Considering the complex geographical environment of the background of the dataset, we intend to test whether adding counterexamples, such as more image context (background), to the model can improve the performance. Thus, we considered an area of 7 km x 7 km (1/4 of the entire image) in orthophotos 1 and 4 and used this dataset for training. We applied the model with the best performance found in previous experiments (1, 2, and 3) for training this new dataset. Following this, we tested this newly trained model (experiment 5) in the entire orthophotos (ortho 2, 3, 5, and 6).

2.2 Deep Semantic Segmentation networks and Experimental Setup

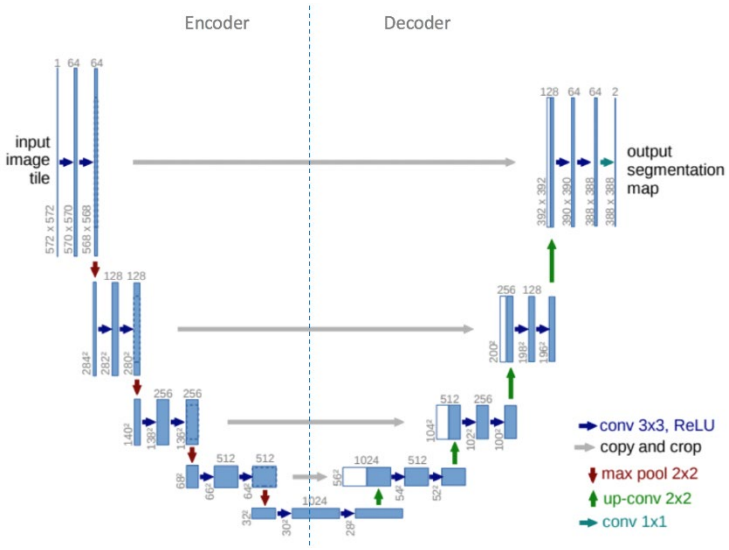
In the task of mapping water body features in RGB images, we worked with deep semantic segmentation networks architectures based on CNN, such as U-Net (RONNEBERGER et al., 2015), Pyramid Scene Parsing Network (PSPNet) (ZHAO et al., 2017), and DeepLabV3 + (CHEN *et al*., 2017). These models are implemented on the ESRI ArcGIS Pro 2.8. The workstation for carrying out the experiments consists of a 64-bit Intel i7-6500U@2.50GHz CPU, with 8 GB memory.

2.2.1 U-Net

U-Net (RONNEBERGER et al., 2015) was one of the pioneer networks to propose encoder-decoder architectures to accomplish semantic segmentation tasks. In the U-Net structure (Figure 17), the encoder has a typical architecture of a convolutional neural network that extracts the features and with the max-pooling operation generates an initial coarse prediction map. Whereas, the decoder is usually composed of convolution, deconvolution (ZEILER; FERGUS, 2014), and/or unpooling layers (GOODFELLOW et al., 2016). The decoder is responsible for further processing the initial prediction map, increasing its spatial resolution gradually, and generating the final prediction. Normally, the decoder has the same number of layers as the encode, but it replaces some of the operations with their counterparts (i.e., convolution with deconvolution, pooling with unpooling, etc). In summary, the decoder can be perceived as a mirrored and symmetrical version of the encoder (Figure 17). In this work, the U-Net architecture from ArcGIS Pro constructs a dynamic U-Net from a backbone pre-

trained on ImageNet and infers the intermediate sizes automatically. For U-Net, we settled the maximum of epochs of 20, batch size of 8, and class balancing with “true”, in which the cross-entropy loss inverse is balanced to the frequency of pixels per class. The learning rate (default) is the optimal rate extracted from the learning curve during the training. The backbone is resnet-34 (pre-trained in Imagenet Dataset) and the early-stopping mechanism is checked to stop the training model when the model is no longer improving.

Figure 17- U-net architecture.



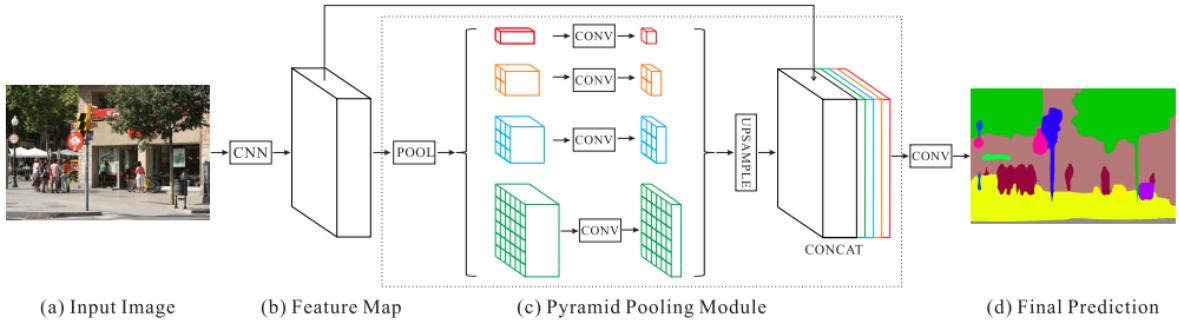
Fonte: Ronneberger et al. (2015).

2.2.2 PSPNet

The winner of the ImageNet Scene Parsing Challenge 2016, the PSPNet (ZHAO et al., 2017) is an improvement from the Fully Convolutional Network (FCN) (LONG et al., 2015). PSPNet differential in that the network aimed to improve the performance of complex-scene parsing, fixing issues like mismatched relationships, confusion categories, and inconspicuous classes (ZHAO et al., 2017). To realize this, they built the pyramid pooling module (Figure 18c). Firstly, the input image has the features extracted through a CNN model. Then, the features were fused under four different pyramid scales, in the pyramid pooling module. In this module, global and local features are extracted through four parallel pooling features and the output is a feature map with variate sizes. Subsequently, a layer with 1 x 1 convolution is used to

maintain the weight of the global feature, reducing the dimension of context representation to 1/4 of the original. Following, the size is restorage through bilinear interpolation and connected with the feature map before polling. Lastly, there is a convolutional layer and the final prediction is generated. For the PSPNEt model, we settled as U-Net (max. of epochs of 20, batch size of 8, class balancing with “true”, learning rate default, backbone resnet-34, and early-stopping mechanism checked). In addition, we used the default pyramid sizes configure from the software, which settles the number and size of convolutions layers as 1, 2, 3, and 6 to the different subregions.

Figure 18- PSPNet architecture.

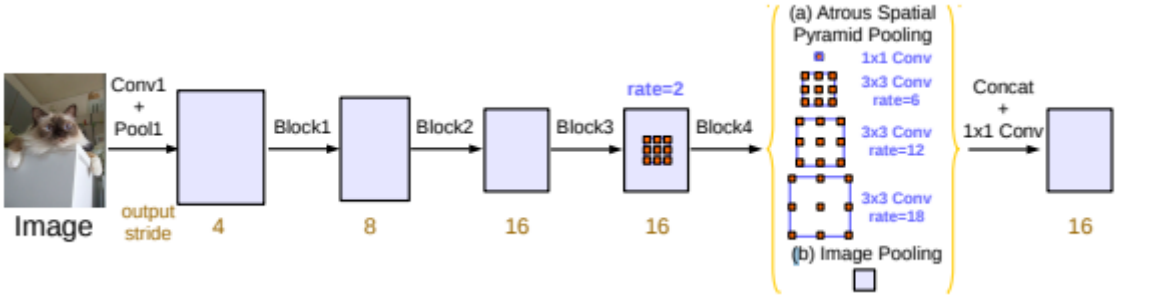


Fonte: Zhao et al. (2017).

2.2.3 DeepLabV3+

The DeepLabV3+ network (Figure 19) is proposed by Chen *et al.* (2017). According to the authors, the network uses a combination of deep convolutional neural networks and fully-connected conditional random fields to achieve semantically accurate predictions and detailed segmentation maps. The differential in this network is first used of dilated convolutional layers. To solve the task of dense feature extraction, they applied atrous convolution with upsampled filters. In addition, they complement the network with Atrous Spatial Pyramid Pooling (ASPP). This special module allows the model to absorb context images in multi-scale dimensions. Capable to establish the receptive field without downsampling the input. In our work, we settled the maximum of epochs of 30, batch size of 8, class balancing with “true”, learning rate default, and backbone resnet-101.

Figure 19- DeepLabV3+ architecture.



Fonte: Chen *et al*. (2017).

2.3 Evaluation metrics

The deep semantic segmentation networks were evaluated using five different metrics, including Accuracy, F1-Score, R-Precision, and Recall (Table 2) (DUAN *et al*., 2021). In addition, due to the test being conducted in more than one orthophoto, we consider the average of the results of the tests in each type of river (large and narrow rivers) in each experiment. For example, in experiment 1, we tested 5 orthophotos (2 in large rivers and 3 in narrow rivers), and we show the mean of the results of these 2 orthophotos in the category “large river” and the mean of these 3 orthophotos in the category “narrow rivers”.

Table 2- Metrics for evaluating the performance of methods.

Evaluation metric	Definition	Formula
Accuracy	The ratio of the correctly predicted pixel numbers to the total pixel numbers	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
F1-Score	Harmonic means for precision and recall	$F1 - Score = \frac{2TP}{2TP + FP + FN}$
R-Precision	The ratio of the correctly predicted river pixel numbers to the number of the labeled river pixels	$Precision = \frac{TP}{TP + FP}$
Recall	The proportion of the number of correctly predicted river pixels and the number of the actual target feature pixels	$Recall = \frac{TP}{TP + FN}$

Fonte: DUAN *et al*. (2021).
Where, True Positive (TP) is the correctly predicting a label (we predicted “water”, and its “water”), True Negative (TN) is the correctly predicting the other label (we predicted “background”, and its “background”), False Positive

(FP) is the falsely predicting a label (we predicted “water”, but it's “background”), False Negative (FN) is the missing and incoming label (we predicted “background”, but it’s “water”).

3. RESULTS AND DISCUSSION

In general, the U-Net model performs better on the semantic segmentation task of rivers in high-resolution RGB images (Table 3), achieving an F1-Score of 97,43% and an Accuracy of 96,82% when the model was tested considering large rivers (experiment 3). However, DeepLabV3+ has better results in narrow rivers, with an F1-Score of 67,51% and an Accuracy of 95,65%, and the second better result for large rivers (F1-Score 95,84% and Accuracy of 95,12%). Both performances occurred when the model was trained with large and thin rivers together (experiment 3). The PSPNet model surpassed the other models mentioned only for the precision metric in river class (R-Precision) with 98.55% for large rivers and 90.98% for narrow rivers. These values were also achieved in experiment 3. Therefore, the best approach to segmenting large and narrow rivers is to make sure that both characteristics are present and balanced in the training model.

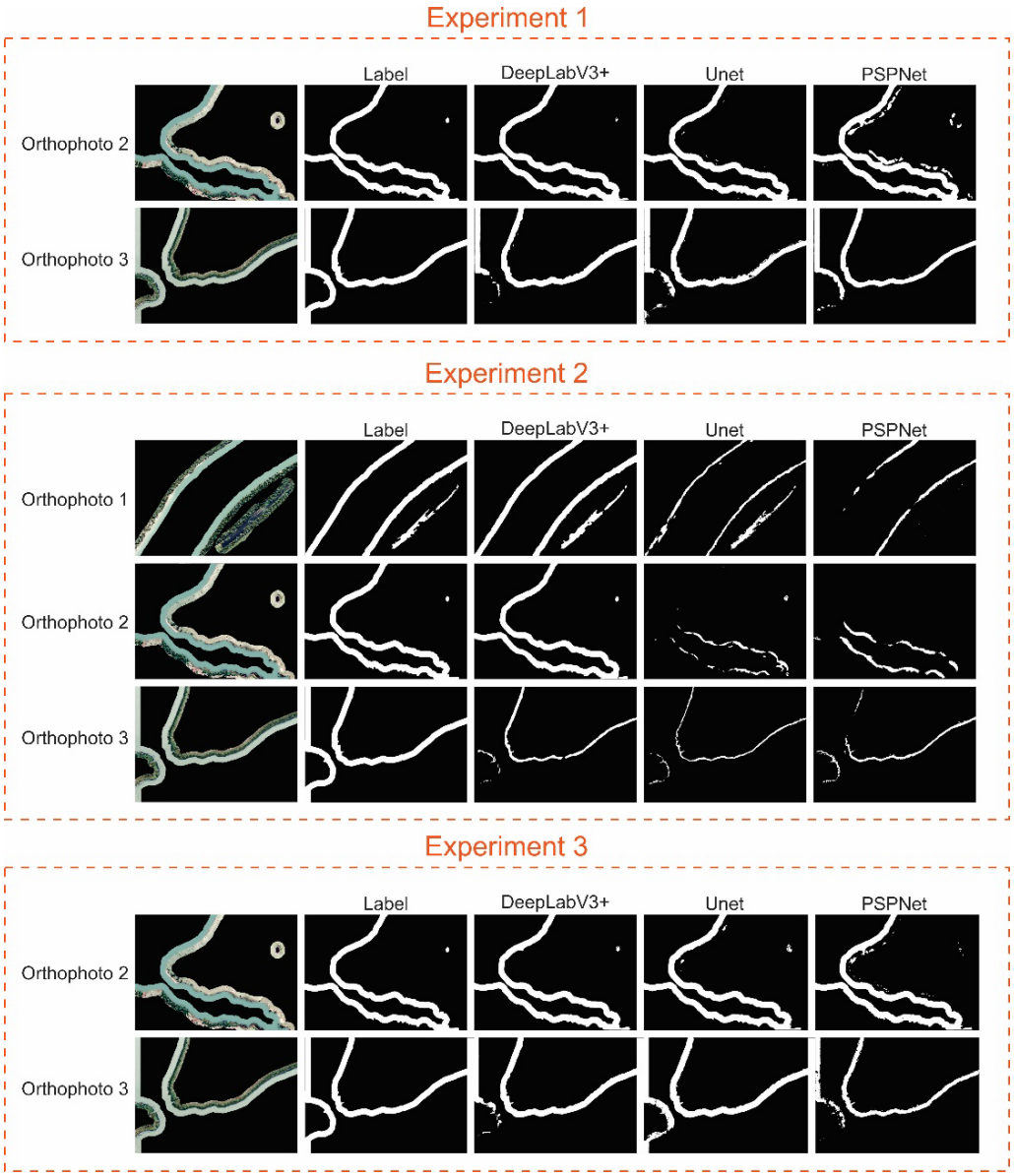
Table 3- Quantitative comparison of deep learning models.

Method	Experiment	River width	Acc	F1-Score	R-Precision	Recall
DeepLabV3+	**1** (Train: large river)	Large river	94.79%	95.58%	96.75%	94.48%
		Narrow river	94.03%	33.55%	79.88%	22.15%
	2 (Train: narrow river)	Large river	82.86%	80.73%	94.89%	74.00%
		Narrow river	95.57%	66.52%	88.86%	53.18%
	3 (Train: both large and narrow rivers)	Large river	95.12%	95.84%	97.89%	93.96%
		Narrow river	**95.65%**	**67.51%**	88.47%	**54.59%**
Unet	**1** (Train: large river)	Large river	94.77%	95.53%	97.62%	93.61%
		Narrow river	93.92%	51.26%	86.24%	43.71%
	2 (Train: narrow river)	Large river	49.68%	25.03%	87.08%	15.34%
		Narrow river	95.30%	64.00%	86.28%	50.88%
	3 (Train: both large and narrow rivers)	Large river	**96.82%**	**97.43%**	96.71%	**98.17%**
		Narrow river	95.47%	66.36%	85.92%	54.06%
PSPNet	**1** (Train: large river)	Large river	93.53%	94.46%	97.35%	92.07%
		Narrow river	93.56%	24.06%	88.06%	15.03%
	2 (Train: narrow river)	Large river	54.69%	36.82%	92.94%	23.70%
		Narrow river	93.53%	44.02%	74.62%	31.41%
	3 (Train: both large and narrow rivers)	Large river	90.99%	91.57%	**98.55%**	86.25%
		Narrow river	94.29%	50.33%	**90.98%**	34.88%

Fonte: Autor (2023).

The performance of the three CNNs exploited to segment large rivers was better in experiments 1 and 3. DeepLabV3+ and UNet had a slight improvement in performance in experiment 3 compared to experiment 1. Only PSPNet had better results in experiment 1 than in experiment 3. This means that for DeepLabV3+ and UNet the information on narrow rivers added in the dataset of experiment 3 improved the performance, even if subtly, except for PSPNet. Another perception of the results for large rivers was that when the model is trained with features of narrow rivers (experiment 2), there is a decrease in the results in all CNNs tested for large rivers. However, DeepLabV3+ showed a smaller performance drop when compared to the other two networks. Although the test orthophoto has large, well-defined rivers, coloration, and a homogeneous spectral signature, the results deteriorate because the training sample is very different. Although all tested networks maintained a precision above 87%, all results slumped in the recall. UNet and PSPNet have an abrupt decrease in this metric, presenting 15,34% and 23,70% respectively, i.e., failing to segment a good part of the river. This finding is noticeable in the qualitative results (Figure 20), in which the segmentation result appears to be a narrower river. This may imply that the trained models learned the narrow-width characteristic of the river and applied it to the segmentation. However, it does not match reality, highlighting the need to also give examples of large rivers, so that the models learn to differentiate these types of rivers.

Figure 20- Qualitative comparison of the classification of deep learning models in large rivers.

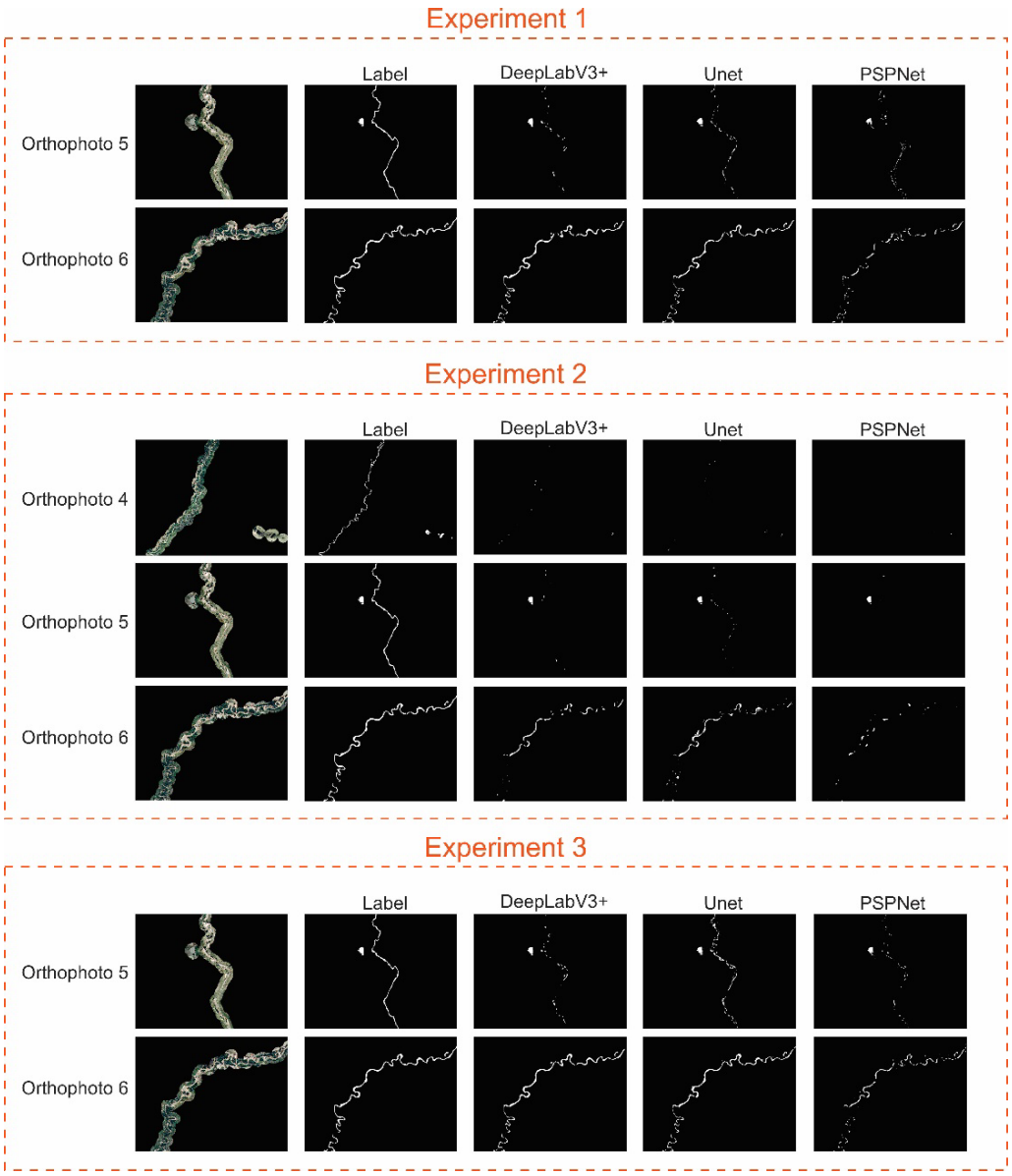


Fonte: Autor (2023).

Regarding the performance in narrow rivers, we observed that both qualitatively and quantitatively, the performance of all CNNs in the experiments gradually improved from experiment 1 to experiment 2 and experiment 3 (Table 3 and Figure 21). Although the DeepLabV3+ and UNet networks obtained better results in experiment 3, both maintained similar values between experiments 2 and 3. This finding reinforces the need to have several samples along the widths of the rivers so that the networks learn the types of rivers studied. For example, in experiment 1 the training was carried out with large rivers, which have well-defined characteristics, with

clean water and few meanders. These characteristics are quite different from the characteristics of narrow rivers, which are narrower, have water with frequent sediments, change color, and spectral signature, and are more sinuous. Therefore, the result in experiment 1 was slightly worse than in the other experiments considering all the evaluated CNNs. Although experiment 2 was trained with narrow rivers and improved the results, the results were still not good, in all networks the recall, and F1-Score were low. Experiment 3, which was trained with both river characteristics, achieved a light increase and improvement in results on all CNNs tested to target narrow rivers. It is also noted that the DeepLabV3+ and U-Net models stand out qualitatively and quantitatively in the precise segmentation of rivers. All CNNs achieved good accuracy (above 93%) and precision (above 74%) results in all experiments. However, for all networks, the recall results were very low, which influenced the F1-Score values to be low as well. The highest recall value was from DeepLabV3+ (54.59%) in experiment 3 and the lowest value was from PSPNet (15.03%) in experiment 1. This means that CNNS had great difficulty in segmenting most of the narrow rivers. Training and testing all networks on orthophotos with different sizes of bodies of water, such as a large river and a thin river, returned different learning from CNNs. This approach demonstrates that different sizes of river datasets affect the performance of networks, due to the different characteristics of the river sizes. In addition, in narrow rivers, there is great difficulty in CNNs learning due to the complexity of the dataset. In addition to having more complex characteristics when compared to large rivers, such as sinuosity, diffuse edges and water with the presence of sediments, the data have a complex background. For example, wetlands, rural roads and exposed soil, which can cause further confusion in river delineation.

Figure 21- Qualitative comparison of the classification of deep learning models in narrow rivers.



Fonte: Autor (2023).

We found that, in general, training with both large and narrow rivers (experiment 3), the results were better both qualitatively and quantitatively on all CNNs. Furthermore, DeepLabV3+ and U-Net performed better than PSPNet in all experiments. Although the CNNs models show little variation between the quantitative results, we adopted DeepLabV3+ as the best CNN for the river segmentation task, as it indicated better results for both large rivers and narrow rivers. Thus, to map larger regions and verify that the results are corroborated, we used the DeepLabV3+ model

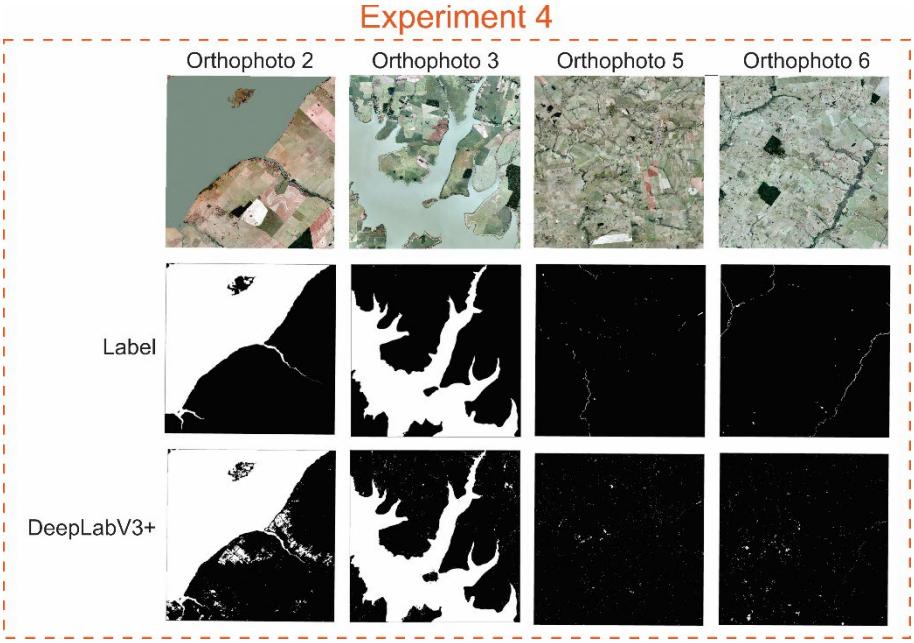
trained in experiment 3 to classify the 4 entire orthophotos in experiment 4, that is, without clipping the 50-meter buffer from the river banks. In this experiment, the result for large rivers, the quantitative performance remained with 97.55% accuracy and 97.24% F1-Score (Table 4). However, qualitatively, the result of orthophoto 3 seems to be better than that of orthophoto 2 (Figure 22). The results were better and very similar to the experiments carried out in the cropped orthophotos. This indicates that for large rivers the approach presented in this study is supported and can be used with strong confidence to segment this type of river.

Table 4- Quantitative performance of the DeepLabV3+ model to segment entire orthophotos.

Method	Experiment	Test	Acc	F1-Score	R-Precision	Recall
DeepLabV3+	4	Large River	97.55%	97.24%	94.98%	99.65%
		Narrow river	98.88%	15.30%	8.70%	64.24%

Fonte: Autor (2023).

Figure 22- Inference of the best model in entire orthophotos.



Fonte: Autor (2023).

For rivers with smaller widths (narrow rivers), DeepLabV3+ did not perform well in segmentation. Despite having obtained 98.88% of correct answers (accuracy), the F1-Score was 15.30% (Table 4). Accuracy shows how well the model correctly classified the two existing classes of all pixels in the image. However, if we analyze

only the class of interest, we find that the precision was very low (8.70%), which means that the model had difficulty in correctly classifying the “river” class. This finding can be noted visually (Figure 22) with a lot of false positives (FP) mainly in the orthophoto 6. The F1-Score value (15.30%) reflects the low recall values (64.24%) and precision, which was less than 10%. In this case, the model had a hard time classifying the river class correctly and failed to classify a good part of the river class. Furthermore, the qualitative perspective (Figure 22) highlights the quantitative findings, in which we can observe a large number of regions wrongly classified as river class. This case highlights how the class imbalance present in the dataset can be confusing if we only look at the high accuracy results (98.8%). It is necessary to reflect on the low values of precision, recall, and F1-Score and, therefore, the need to work with more than one evaluation metric and a qualitative analysis whenever possible.

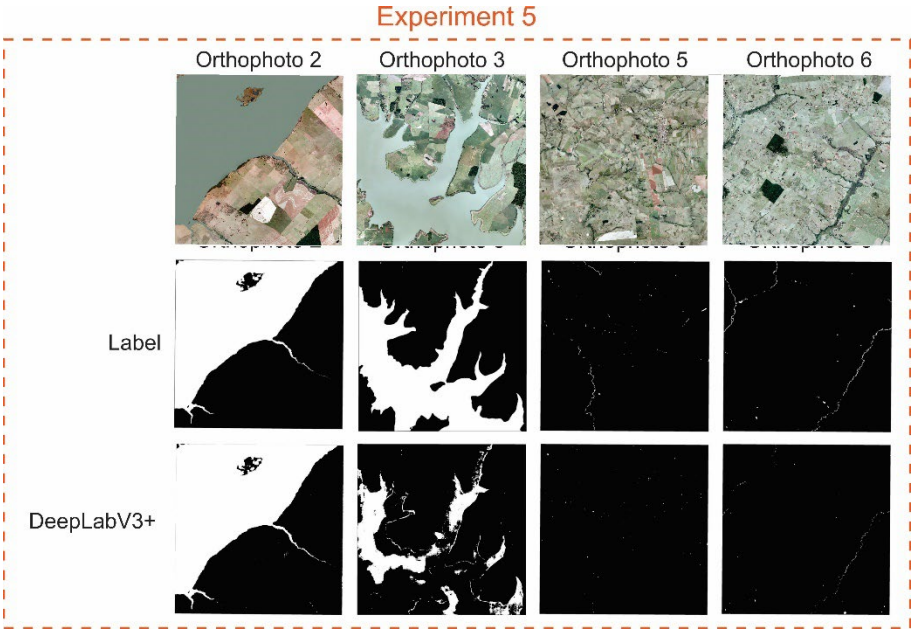
Afterward, we conduct the 5th experiment giving more examples of background and verifying if the DeepLabV3+ can segment better the river class. We apply this approach to certify that we are not missing any feature of the background and to compare if the augmentation of these examples can improve the results. Quantitative results show that this approach has improved 22.9% in F1-Score (38.22%) for segment narrow rivers (Table 5). Despite this rise, the result is still unsatisfactory, staying under 40%. Furthermore, for narrow rivers, there was a large increase in R-Precision values (60.36%), however, recall remained low (27.97%), which means that there are still many regions where the river class was lost and not segmented correctly (Figure 23). On the other hand, the results are still better for large rivers, with an F1-Score of 76.70% and an R-Precision of 99.59%.

Table 5- Quantitative performance of the DeepLabV3+ model trained in experiment 5 to segment entire orthophotos.

Method	Experiment	Test	Acc	F1-Score	R-Precision	Recall
DeepLabV3+	5	Large River	88.13%	76.70%	99.59%	68.30%
		Narrow river	99.86%	38.22%	60.36%	27.97%

Fonte: Autor (2023).

Figure 23- Qualitative results from experiment 5.

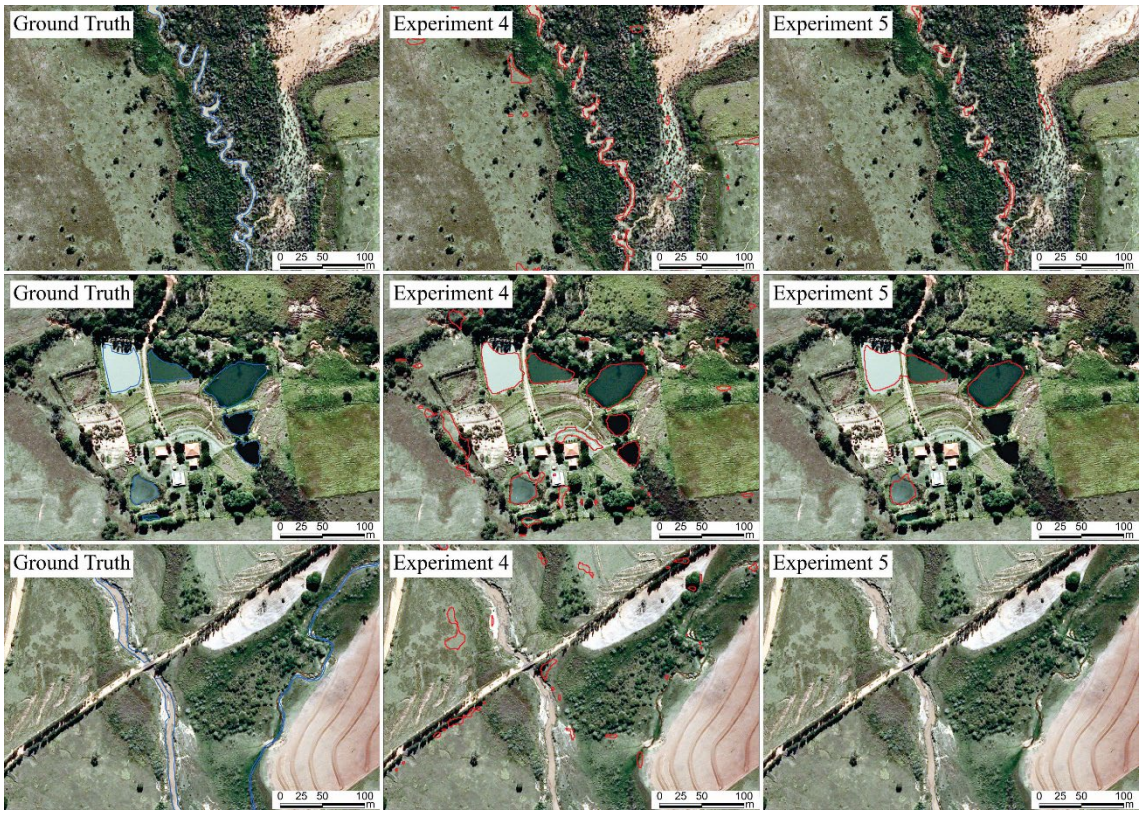


Fonte: Autor (2023).

From a qualitative viewpoint (Figure 23) the results of the large rivers confirm that DeepLabV3+ can segment the rivers, especially if we focus on orthophoto 2. Although the qualitative results of the narrow rivers (ortho 5 and 6) did not present many false positives (FP), as in experiment 4, this approach still did not solve the narrow river segmentation problem. In this case, it is still possible to visually perceive that there are many false negatives (FN), that is, there are still many regions where the river class was not segmented. This result can also be noticed visually, in regions where the rivers were not classified as true rivers (third line of Figures 24 and 25). In these figures' details, it is possible to visualize an example of when the DeepLabV3+ model could segment a good region of the river in both experiments (first line in Figures 24 and 25). Our dataset is also composed of lakes, which, despite not being the objective of the work, were qualitatively evaluated. In the second line (Figures 24 and 25), there is an example of the segmentation of lakes, which has similar results to those of rivers, showing a greater number of false positives in experiment 4 and failure to segment some regions in experiment 5. The third line (Figures 24 and 25), it's a challenging region, where both experiments 4 and 5 had difficulty segmenting the river, and there were even regions where the river was not segmented. All examples showed that experiment 4 has issues with many false positives (FP) when the model classified the river to a lot of pixels of background, which reflects the plummet of the R-Precision.

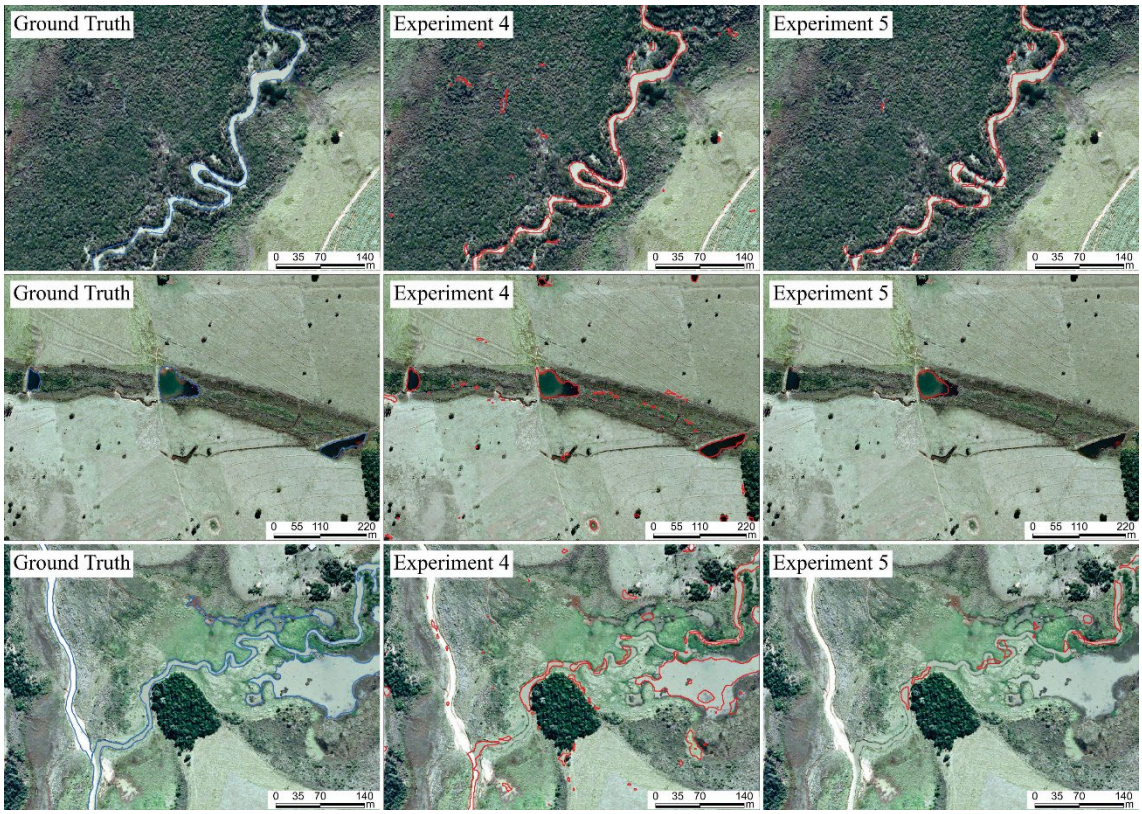
In addition, experiment 5 has issues with many false negatives (FN), failing to truly segment a good part of the river class, and presenting a very low recall. Therefore, even giving more background examples for the CNN (experiment 5) there are still difficulties in segmenting narrow rivers. In this way, there is a need to work and study more in-depth this type of complex data, which are narrow rivers.

Figure 24- Details of river segmentation from orthophoto 5 in experiments 4 and 5.



Fonte: Autor (2023).

Figure 25- Details of river segmentation from orthophoto 6 in experiments 4 and 5.



Fonte: Autor (2023).

4. CONCLUSION

This paper presented an approach to accurately mapping large and narrow rivers in high spatial resolution RGB images using deep semantic segmentation networks. This approach was applied using a challenging dataset including rivers of different sizes, widths such as large rivers (more than 10 m large) and narrow rivers (less than 10 m large), and river shapes. We conclude that even with different approaches, there is still great difficulty in segmenting narrow rivers. his conclusion may be due to the characteristics of narrow rivers, which are usually sinuous, sedimented, have spectral comportment complex, and a small number of pixels that are densely distributed, making this task even more challenging.

In our study, the best results for all the CNNs studied were when the networks were trained with both large and narrow river types combined. The deep learning model (U-Net) achieved the best quantitative result for segmenting large rivers (F1-Score of 97.43%). However, DeepLabV3+ was produced the best model, as the results for large

rivers were very similar to the Unet, and, in addition, it obtained better results for narrow rivers. Furthermore, when CNNs were trained with only large rivers, the performance to target narrow rivers deteriorated and when trained with only narrow rivers, the performance was worse for both large and narrow rivers. In addition, when tested over a larger region, the large rivers maintained good accuracy, but the narrow rivers greatly worsened the performance. Even offering more counterexamples in DeepLabV3+ training (better performance network), the results for large rivers were better and for narrow rivers worse.

We conclude that this high precision is because large rivers present smaller variations in terms of spectral signature and pixel mixing, which may favor the learning of the model. Our experiments show that all tested models have difficulty segmenting narrow rivers, even though the high-resolution RGB images show them in great detail. Although convolutional neural networks have the characteristic of learning characteristics hierarchically, the evaluated models still present difficulties to segment flows and flows. The approach used in this study demonstrated that large areas of water bodies can be successfully segmented, even using RGB images, which makes our approach much broader, without the need for an infrared spectrum. However, narrow rivers have trouble getting accurate metrics in segmentation.

We recommend that future studies explore new alternatives based on deep learning methods as a vision transformer to perform the task of semantic segmentation of smaller rivers. We demonstrate the ability of methods based on deep learning to map water resources in RGB images, contributing to decision-making in water management and environmental studies. As there is still a gap in the segmentation of narrow rivers with accurate precision in high-resolution RGB images, future studies can try strategies to fill this question. One approach would be to work with the architecture of the networks, looking for implementations that gain more image context at this level of scale, such as Transformer-based architectures. Another bias would be to work with the image, so that the features are sharper, which can improve the performance of deep learning models.

REFERENCES

ACHARYA, T. D.; SUBEDI, A.; LEE, D. H. Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 scene of Nepal. **Sensors**, v.19, n.12, p. 2769, 2019. Disponível em:< https://www.mdpi.com/482956>. Acesso em: 23 de abril de 2020.

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, 2017. Disponível em:< https://ieeexplore.ieee.org/abstract/document/7803544/>. Acesso em: 16 de março de 2020.

BAI, M.; URTASUN, R. Deep watershed transform for instance segmentation. *In:* **Proceedings of the IEEE conference on computer vision and pattern recognition,** 2017. p. 5221-5229. Disponível em: < http://openaccess.thecvf.com/content_cvpr_2017/html/Bai_Deep_Watershed_Transform_CVPR_2017_paper.html>. Acesso em: 19 de março de 2020.

BANGIRA, T. *et al*. Comparing thresholding with machine learning classifiers for mapping complex water. **Remote Sensing**, v. 11, n.11, p. 1351, 2019. Disponível em: < https://www.mdpi.com/474688>. Acesso em: 05 de setembro de 2020.

BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. **Neural networks**, v. 106, p. 249-259, 2018. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0893608018302107>. Acesso em: 07 de fevereiro de 2021.

CHEN, J. *et al*. Remote sensing big data for water environment monitoring: current status, challenges, and future prospects. **Earth's Future**, v. 10, n. 2, p. e2021EF002289, 2022. Disponível em: < https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002289>. Acesso em: 14 de abril de 2022.

CHEN, Y. *et al*. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. **Water**, v. 10, n. 5, p.585, 2018. Disponível em: < https://www.mdpi.com/287878>. Acesso em: 17 de maio de 2021.

CHENG, J. *et al*. Satellite and UAV-based remote sensing for assessing the flooding risk from Tibetan Lake expansion and optimizing the village relocation site. **Science of The Total Environment**, v. 802, p. 149928, 2022. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0048969721050038>. Acesso em: 22 de junho de 2022.

DANG, B.; LI, Y. MSResNet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery. **Remote Sensing**, v. 13, n. 16, p. 3122, 2021. Disponível em:<https://www.mdpi.com/2072-4292/13/16/3122>. Acesso em: 09 de abril de 2022.

DU, Y.; ZHOU, C. Automatically extracting remote sensing information for water bodies. **JOURNAL OF REMOTE SENSING-BEIJING-**, v. 2, p. 269-274, 1998. Disponível em:< https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Automatically+extracting+remote+sensing+information+for+water+bodies&btnG=> Acesso em: 19 de junho de 2021.

DUAN, Y. *et al*. A New Lightweight Convolutional Neural Network for Multi-Scale Land Surface Water Extraction from GaoFen-1D Satellite Images. **Remote Sensing**, v. 13, n. 22, p. 4576, 2021. Disponível em:<https://www.mdpi.com/1357606>. Acesso em: 12 de fevereiro de 2022.

DRUCE, D. *et al*. An optical and SAR based fusion approach for mapping surface water dynamics over mainland China. **Remote Sensing**, v. 13, n. 9, p. 1663, 2021. Disponível em:< https://www.mdpi.com/2072-4292/13/9/1663>. Acesso em: 25 de fevereiro de 2022.

GAO, J. *et al*. Camera-LiDAR Cross-Modality Fusion Water Segmentation for Unmanned Surface Vehicles. **Journal of Marine Science and Engineering**, v.10, n. 6, p. 744, 2022. Disponível em: < https://www.mdpi.com/1653078>. Acesso em: 23 de agosto de 2022.

GAUTAM, S.; SINGHAI, J. Cosine-similarity watershed algorithm for water-body segmentation applying deep neural network classifier. **Environmental Earth Sciences**, v. 81, n. 9, p. 1-16, 2022. Disponível em:< https://link.springer.com/article/10.1007/s12665-022-10376-y>. Acesso em: 19 de março de 2022.

GHAMISI, P. *et al*. Advanced spectral classifiers for hyperspectral images: A review. **IEEE Geoscience and Remote Sensing**, v. 5, n.1, p. 8–32., 2017. Disponível em: < https://ieeexplore.ieee.org/abstract/document/7882742/>. Acesso em: 28 de abril de 2021.

GUO, H. *et al*. A Multi-Scale Water Extraction Convolutional Neural Network (MWEN) Method for GaoFen-1 Remote Sensing Images. **ISPRS International Journal of Geo-Information**, v. 9, n. 4, p. 189, 2020. Disponível em: < https://www.mdpi.com/673286>. Acesso em: 11 de junho de 2021.

HAFIZ, A. M.; BHAT, G. M. A survey on instance segmentation: state of the art. **International journal of multimedia information retrieval**, v. 9, n. 3, p. 171-189, 2020. Disponível em: < https://link.springer.com/article/10.1007/s13735-020-00195-x>. Acesso em: 8 de agosto de 2021.

HAO, S.; ZHOU, Y.; GUO, Y. A brief survey on semantic segmentation with deep learning. **Neurocomputing**, v. 406, p. 302-321, 2020. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0925231220305476>. Acesso em: 12 de maio de 2021.

HU, K. *et al*. Multi-Scale Feature Aggregation Network for Water Area Segmentation. **Remote Sensing**, v. 14, n. 1, p. 206, 2022. Disponível em:< https://www.mdpi.com/article/10.3390/rs14010206>. Acesso em: 23 de junho de 2022.

HUANG, X. *et al*. Combining pixel-and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery. **IEEE Journal of Selected topics in applied earth observations and remote sensing**, v. 8, n. 5, p. 2097-2110, 2015. Disponível em: < https://ieeexplore.ieee.org/abstract/document/7094223/>. Acesso em: 27 de abril de 2020.

ISIKDOGAN, F.; BOVIK, A. C.; PASSALACQUA, P. Surface water mapping by deep learning. **IEEE journal of selected topics in applied earth observations and remote sensing**, v. 10, n. 11, p. 4909-4918, 2017. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8013683/>. Acesso em: 19 de maio de 2021.

JAMALI, A. Improving land use land cover mapping of a neural network with three optimizers of multi-verse optimizer, genetic algorithm, and derivative-free function. **The Egyptian Journal of Remote Sensing and Space Science**, v. 24, n. 3, p. 373-390, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1110982320300697>. Acesso em: 15 de março de 2022.

JANIZADEH, S. *et al*. Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. **Journal of Environmental Management**, v. 298, p. 113551, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0301479721016133>. Acesso em: 28 de abril de 2022.

JIN, S. *et al*. River body extraction from sentinel-2A/B MSI images based on an adaptive multi-scale region growth method. **Remote Sensing of Environment**, v. 255, p. 112297, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0034425721000158>. Acesso em: 30 de março de 2022.

JIANG, W. *et al*. Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. **Remote Sensing**, v. 10, n. 5, p. 755, 2018. Disponível em: < https://www.mdpi.com/292956>. Acesso em: 19 de setembro de 2020.

KALANTAR, B. *et al*. Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane, Australia. **Remote Sensing**, v. 13, n. 13, p. 2638, 2021. Disponível em: < https://www.mdpi.com/1177534>. Acesso em: 2 de novembro de 2021.

KANG, J. *et al*. Multi-scale context extractor network for water-body extraction from high-resolution optical remotely sensed images. **International Journal of Applied Earth Observation and Geoinformation**, v. 103, p. 102499, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0303243421002063>. Acesso em: 10 de maio de 2022.

KEBEDE, M. G. *et al*. Remote sensing-based river discharge estimation for a small river flowing over the high mountain regions of the Tibetan Plateau. **International Journal of Remote Sensing**, v. 41, n. 9, p. 3322-3345, 2020. Disponível em: < https://www.tandfonline.com/doi/abs/10.1080/01431161.2019.1701213>. Acesso em: 14 de agosto de 2021.

KIM, K. H. *et al*. Application of CCTV Image and Semantic Segmentation Model for Water Level Estimation of Irrigation Channel. **Journal of The Korean Society of Agricultural Engineers**, v. 64, n. 3, p. 63-73, 2022. Disponível em: < https://www.koreascience.or.kr/article/JAKO202215961131892.page>. Acesso em 23 de setembro de 2022.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, v. 60, n. 6, p. 84-90, 2017. Disponível em: < https://dl.acm.org/doi/abs/10.1145/3065386>. Acesso em: 17 de agosto de 2021.

LANGAT, P. K.; KUMAR, L.; KOECH, R. Monitoring river channel dynamics using remote sensing and GIS techniques. **Geomorphology**, v. 325, p. 92-102, 2019. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0169555X18304070>. Acesso em: 26 de setembro de 2021.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015. Disponível em:< https://www.nature.com/articles/nature14539>. Acesso em: 22 de agosto de 2019.

LU, L. *et al*. Water Body Extraction from High-resolution Remote Sensing images Based on Scaling EfficientNets. *In*: **Journal of Physics: Conference Series**. IOP Publishing, 2021. p. 012100. Disponível em:< https://iopscience.iop.org/article/10.1088/1742-6596/1894/1/012100/meta>. Acesso em: 22 de agosto de 2022.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International journal of remote sensing**, v. 17, n. 7, p. 1425-1432, 1996. Disponível em: < https://www.tandfonline.com/doi/abs/10.1080/01431169608948714>. Acesso em: 13 de setembro de 2021.

MIAO, Z. *et al*. Automatic water-body segmentation from high-resolution satellite images via deep networks. **IEEE geoscience and remote sensing letters**, v. 15, n. 4, p. 602-606, 2018. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8286914/>. Acesso em: 25 de setembro de 2022.

MINAEE, S. *et al*. Image segmentation using deep learning: A survey. **IEEE transactions on pattern analysis and machine intelligence**, v. 44, n. 7, p. 3523-3542, 2021. Disponível em: < https://ieeexplore.ieee.org/abstract/document/9356353/>. Acesso em: 26 de outubro de 2021.

MISHRA, V. K.; PANT, T. Water level monitoring using classification techniques on Landsat-8 data at Sangam region, Prayagraj, India. **IET Image Processing**, v. 14, n. 15, p. 3733-3741, 2020. Disponível em: < https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2020.1078>. Acesso em: 14 de março de 2021.

MORADKHANI, K.; FATHI, A. Segmentation of waterbodies in remote sensing images using deep stacked ensemble model. **Applied Soft Computing**, p. 109038, 2022. Disponível em:< https://www.sciencedirect.com/science/article/pii/S1568494622003465>. Acesso em: 10 de setembro de 2022.

NOGUEIRA, K. *et al*. A tool for bridge detection in major infrastructure works using satellite images. *In*: **2019 XV Workshop de Visão Computacional (WVC)**. IEEE, 2019a. p. 72-77. Disponível em: < https://ieeexplore.ieee.org/abstract/document/8876942/>. Acesso em: 20 de novembro de 2020.

NOGUEIRA, K. *et al*. Dynamic multicontext segmentation of remote sensing images based on convolutional networks. **IEEE Transactions on Geoscience and Remote Sensing**, v. 57, n. 10, p. 7503-7520, 2019b. Disponível em: < https://ieeexplore.ieee.org/abstract/document/8727958/>. Acesso em: 18 de setembro de 2020.

NOGUEIRA, K. *et al*. Exploiting convnet diversity for flooding identification. **IEEE Geoscience and Remote Sensing Letters**, v. 15, n. 9, p. 1446-1450, 2018. Disponível em: < https://ieeexplore.ieee.org/abstract/document/8398414/>. Acesso em: 29 de setembro de 2020.

NOGUEIRA, K.; PENATTI, O. A.; DOS SANTOS, J. A. Towards better exploiting convolutional neural networks for remote sensing scene classification. **Pattern Recognition**, v. 61, p. 539-556, 2017. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0031320316301509>. Acesso em: 11 de outubro de 2020.

OLIVER, J. A. *et al*. A Machine Learning Approach to Waterbody Segmentation in Thermal Infrared Imagery in Support of Tactical Wildfire Mapping. **Remote Sensing**, v. 14, n. 9, p. 2262, 2022. Disponível em: < https://www.mdpi.com/1621920> Acesso em: 15 de outubro de 2022.

OSCO, L. P. *et al.* A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 160, p. 97-106, 2020. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0924271619302989>. Acesso em: 12 de março de 2021.

PEKEL, J. F. *et al*. High-resolution mapping of global surface water and its long-term changes. **Nature**, v., 540, n. 7633, p. 418-422, 2016. Disponível em: < https://www.nature.com/articles/nature20584>. Acesso em: 24 de março de 2020.

ROGERS, A. S.; KEARNEY, M. S. Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices. **International Journal of Remote Sensing**, v. 25, n. 12, p. 2317-2335, 2004. Disponível em: < https://www.tandfonline.com/doi/abs/10.1080/01431160310001618103>. Acesso em: 23 de novembro de 2019.

SAMBOKO, H. T. *et al*. Evaluation and improvement of remote sensing-based methods for river flow management. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 117, p. 102839, 2020. Disponível em:< https://www.sciencedirect.com/science/article/pii/S1474706519301500>. Acesso em: 26 de maio de 2021.

SHIH, S. F. Comparison of ELAS classifications and density slicing Landsat data for water surface area assessment. **Hydrologic applications of space technology**, v. 160, p. 91-97, 1985.

SUN, X. *et al*. Monitoring water quality using proximal remote sensing technology. **Science of The Total Environment**, v. 803, p. 149805, 2022. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0048969721048804>. Acesso em: 17 de maio de 2022.

TAMBE, R. G.; TALBAR, S. N.; CHAVAN, S. S. Deep multi-feature learning architecture for water body segmentation from satellite images. **Journal of Visual Communication and Image Representation**, v. 77, p. 103141, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1047320321000870>. Acesso em: 27 de agosto de 2022.

THAYAMMAL, S. *et al*. Analysis of Water Body Segmentation from Landsat Imagery using Deep Neural Network. **Wireless Personal Communications**, v. 123, n. 2, p. 1265-1282, 2022. Disponível em: < https://link.springer.com/article/10.1007/s11277-021-09178-5>. Acesso em: 8 de junho de 2022.

VERMA, U. *et al*. DeepRivWidth: Deep learning based semantic segmentation approach for river identification and width measurement in SAR images of Coastal Karnataka. **Computers & Geosciences**, v. 154, p. 104805, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0098300421001059>. Acesso em: 18 de março de 2022.

VIGIAK, O. *et al*. Probability maps of anthropogenic impacts affecting ecological status in European rivers. **Ecological indicators**, v. 126, p. 107684, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1470160X21003496>. Acesso em: 9 de abril de 2022.

WANG, B. *et al*. SADA-Net: A Shape Feature Optimization and Multiscale Context Information-Based Water Body Extraction Method for High-Resolution Remote Sensing Images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 1744-1759, 2022. Disponível em:< https://ieeexplore.ieee.org/abstract/document/9695297/>. Acesso em: 26 de maio de 2022.

WANG, X. *et al*. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 68, p. 73-91, 2018. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0303243418300990>. Acesso em: 25 de fevereiro de 2020.

WANG, Y. *et al*. Lightweight Deep Neural Network Method for Water Body Extraction from High-Resolution Remote Sensing Images with Multisensors. **Sensors**, v. 21, n. 21, p. 7397, 2021. Disponível em:< https://www.mdpi.com/1347686>. Acesso em: 24 de fevereiro de 2022.

WEI, Z. *et al*. Global river monitoring using semantic fusion networks. **Water**, v. 12, n. 8, p. 2258, 2020. Disponível em: < https://www.mdpi.com/795410>. Acesso em: 23 de novembro de 2021.

XU, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. **International journal of remote sensing**, v. 27, n. 14, p. 3025-3033, 2006. Disponível em: < https://www.tandfonline.com/doi/abs/10.1080/01431160600589179>. Acesso em: 2 de novembro de 2021.

YOTOVA, G. *et al*. Water quality assessment of a river catchment by the composite water quality index and self-organizing maps**. Ecological indicators**, v. 120, p. 106872, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1470160X20308104>. Acesso em: 12 de março de 2022.

ZHANG, G. *et al*. Automated water classification in the Tibetan plateau using Chinese GF-1 WFV data. **Photogrammetric Engineering & Remote Sensing**, v. 83, n. 7, p. 509-519, 2017. Disponível em: < https://www.researchgate.net/profile/Junli-Li/publication/318400239_Automated_Water_Classification_in_the_Tibetan_Plateau_Using_Chinese_GF-1_WFV_Data/links/62551144ef013420666bd72d/Automated-Water-Classification-in-the-Tibetan-Plateau-Using-Chinese-GF-1-WFV-Data.pdf>. Acesso em: 29 de novembro de 2021.

ZHANG, Z. *et al*. Rich CNN Features for water-body segmentation from very high resolution aerial and satellite imagery. **Remote Sensing**, v. 13, n. 10, p. 1912, 2021. Disponível em:< https://www.mdpi.com/2072-4292/13/10/1912>. Acesso em: 18 de março de 2022.

ZHANG, X.; LI, J.; HUA, Z. MRSE-Net: Multi-Scale Residuals and SE-Attention Network for Water Body Segmentation from Satellite Images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 5049-5064, 2022. Disponível em: < https://ieeexplore.ieee.org/abstract/document/9803863/>. Acesso em: 11 de outubro de 2022.

ZHENG, X.; CHEN, T. High spatial resolution remote sensing image segmentation based on the multiclassification model and the binary classification model. **Neural Computing and Applications**, p. 1-8, 2021. Disponível em: < https://link.springer.com/article/10.1007/s00521-020-05561-8>. Acesso em: 25 de maio nde 2022.

CAPÍTULO 3– RIVER MAPPING USING TRANSFORMER-BASED ARCHITECTURE IN HIGH SPATIAL RESOLUTION RGB IMAGES

Abstract: Mapping water resources such as rivers is important to the planning and management of this vital resource. With advances and technological studies of remote sensing and computer vision, the mapping of these resources has become increasingly accurate. The deep learning semantic segmentation methods has been widely explored and ViT-based networks are the state-of-the-art to extract information from remote sensing images. We evaluated the performance of the SegFormer (transformer-based) neural network to map rivers in high spatial resolution RGB images. To validate the information, we compared the performance of the SegFormer with the DeepLabV3+ neural network (CNN-based). In addition, we separately compare the performance of networks in narrow rivers, as they have more complex characteristics, in addition to being inserted in a complex context. For this, we tested the networks in aerial RGB orthophotos with 1 meter of spatial resolution. Although both evaluated networks were able to segment rivers under the mentioned conditions, SegFormer outperformed DeepLabV3+ in all evaluated metrics (Accuracy: 98.97%, F1-Score: 98.96%, and IoU: 97.96%). Despite the expected performance decline for narrow rivers, SegFormer maintained its excellent performance (Accuracy: 81.76%, F1-Score: 85.99%, and IoU: 75.42%. Thus, we conclude that SegFormer is suitable for segmenting rivers in high-resolution RGB images. spatial resolution, even narrow rivers. Studies using Transformer-based architectures should continue to be investigated in other contexts, due to their ability to gain context, be a lightweight network, and present excellent performance.

Keywords: Water resource mapping; Deep Learning method; High-resolution remote sensing image; Low-spectral resolution.

Resumo: O mapeamento de recursos hídricos, como rios, é importante para o planejamento e gestão desse recurso vital. Com os avanços e estudos tecnológicos de sensoriamento remoto e visão computacional, o mapeamento desses recursos tem se tornado cada vez mais preciso. O método de segmentação semântica de aprendizado profundo tem sido amplamente explorado e as redes baseadas em ViT têm sido o estado da arte para extrair informações de imagens de sensoriamento

remoto. Avaliamos o desempenho da rede neural SegFormer (baseada em transformador) para mapear rios em imagens RGB de alta resolução espacial. Para validar as informações, comparamos o desempenho do SegFormer com a rede neural DeepLabV3+ (baseada em Convolutional Neural Networks). Além disso, comparamos separadamente o desempenho das redes em rios estreitos, pois possuem características mais complexas, além de estarem inseridas em um contexto complexo. Para isso, testamos as redes em ortofotos RGB aéreas com 1 metro de resolução espacial. Embora ambas as redes avaliadas tenham conseguido segmentar rios nas condições mencionadas, o SegFormer superou o DeepLabV3+ em todas as métricas avaliadas (Precisão: 98,97%, F1-Score: 98,96% e IoU: 97,96%). Apesar da esperada queda de desempenho para rios estreitos, o SegFormer manteve seu excelente desempenho (Precisão: 81,76%, F1-Score: 85,99% e IoU: 75,42%. Assim, concluímos que o SegFormer é adequado para segmentar rios em imagens RGB de alta resolução espacial, mesmo em rios estreitos. Estudos com arquitetura baseada em Transformer tem potencial para ser investigados em outros contextos, devido à sua capacidade de ganhar contexto, ser uma rede leve e apresentar excelente desempenho.

Palavras-chave: Mapeamento de recursos hídricos; Método Deep Learning; Imagem de sensoriamento remoto de alta resolução; Baixa resolução espectral.

1. INTRODUCTION

Water resources are essential for maintaining life, and knowing their location and features is necessary for managing this vital resource. The mapping of this resource provides essential subsidies for water management, contributing directly to public health, environment preservation, and economic development. Rivers, streams, lakes, ponds, wetlands, and seas can be categorized as surface water, i.e., it accumulates on the Earth's surface. Rivers have an important role in the water cycle and water management, thus being a topic of analysis for several years. They can be observed and studied through remote sensing (RS) data, i.e., images obtained from orbital, aerial, or terrestrial platforms. Through RS images it is possible to extract the features of rivers and characterize them, however, rivers have complex characteristics and scenes. They can vary in shape (large, narrow, straight, sinuous), texture (polluted or clean rivers), and spectral signature, and are difficult to present well-defined edges.

Furthermore, when using images with lower spectral resolution (e.g., RGB images) to reduce mapping costs, the challenge is greater, because this bands don't have a lot of spectral contrast between the target (water) and background. However, the association with images of higher spatial resolution can help in the mapping, as these present a better level of detail in the image. In addition, with the progress of high-resolution RS imaging technology, image texture, and object geometry are increasingly clear and fine (LI *et al*., 2022).

There are several methods for mapping surface waters. The initial methods are based on spectral characteristics, such as the single-band threshold method, multiband spectral relationship method, and water index method (WANG *et al*., 2021). However, these methods focus only on spectral characteristics and have small automation. Spatial information such as shape, size, texture, edge, shadow, and context semantics are essential to achieve better accuracies (ZHANG *et al*., 2021). There are methods based on machine learning algorithms, such as random forest (KO; KIM; NAM, 2015; TIAN *et al*., 2016) and support vector machine (SUN *et al*., 2014; SARP; OZCELIK, 2017), however, such algorithms process a limited number of samples and cannot extract deep information from the dataset, so the generalization ability is insufficient (LI *et al*., 2021). Deep learning (DL) based methods have been explored in the remote sensing field and have demonstrated superior performance in applications such as object classification, detection, and semantic segmentation (YUAN *et al*., 2021).

Semantic segmentation is a task that aims to infer the knowledge of a scene by establishing a known class for each pixel, a.k.a. pixel-level classification (BRESSAN *et al*., 2022). Firstly, Convolutional Neural Networks (CNN) architecture demonstrates good performances and has been explored intensively for semantic segmentation tasks (ALAM *et al*., 2021). The CNNs explored for this task in remote sensing usually are deeper and have several hidden layers, being called Deep Convolutional Neural Networks (DCNNs) (KOTARIDIS; LAZARIDOU, 2021). Despite the existing traditional DCNNs (SegNet, Unet, DeepLabV3+), in the last years, several studies have designed, modified, and incorporate multiple models to increase the performance in semantic segmentation. One study (ZHANG *et al*., 2021) proposed a rich feature extraction network (MECNet) composed of three modules: Multi-feature Extraction and Combination (MEC), encoder and Decoder Semantic Feature Fusion (DSFF), and Multi-scale Prediction Fusion (MPF). The proposed network was valid for water-body

segmentation in very high-resolution RGB remote sensing images (IoU around 90%). However, the method doesn't pay much attention to the spatial relationship between feature maps, influenced by the design that focuses on the global information of feature maps. Despite the network being designed specifically to deal with complex spectral mixtures, challenges such as extracting small rivers were not addressed. Another investigation (HU *et al*., 2022), proposed a Multi-scale Feature Aggregation network to extract rich context information. Composed of a deep feature extraction module (DFE), to obtain multi-scale features and pay attention to global and local edge information; a multi-branch aggregation module (MBA) to enhance the interconnection and integrate the two types of feature representation, and the feature-fusion upsample module (FFU) to complete feature fusion and location recovery. However, the authors point out some shortcomings, for example, reducing the weight of the model, changing the convolution kernel or the convolution type, and even selecting a lighter network. DCNN has an acknowledged success, however, there are still issues to be solved, for example, the loss of localization accuracy (blurred boundaries) and spatial details (omission of small objects) in the pooling operations (KOTARIDIS; LAZARIDOU, 2021). Thus, aggregating contextual information is essential, especially for semantic segmentation of remote sensing images (KOTARIDIS; LAZARIDOU, 2021).

A recent deep learning architecture has gained attention in the field of computer vision (CV), the Vision Transformer (ViT) (DOSOVITSKIY *et al*., 2020) which is based on the Transformer architecture (VASWANI *et al*., 2017), a state-of-art for translation in the NLP field (Natural Language Processing). Transformers use a self-attention mechanism, without using sequence-aligned RNNs or convolutions in translation tasks (VASWANI *et al*., 2017). The assertive response and lightweight Transformer architecture inspired the uses in other fields of NLP, like in CV. In this context, Transformed-based networks are one of the best-emerging approaches to extracting information, being the state-of-art DL semantic segmentation method. SegFormer is a ViT-based architecture that uses an encoder with a hierarchical structure, maintaining multiscale resource outputs, which significantly contributes to tasks with objects that vary in size, such as rivers. In addition, this architecture is built with lightweight but robust decoders, representing an advantage over other ViT-based networks (XIE *et al*., 2021). This network architecture can maintain a combination of local and global attention with its encoder, aggregating information from different layers of the network to render more powerful representations (GONÇALVES *et al*., 2023).

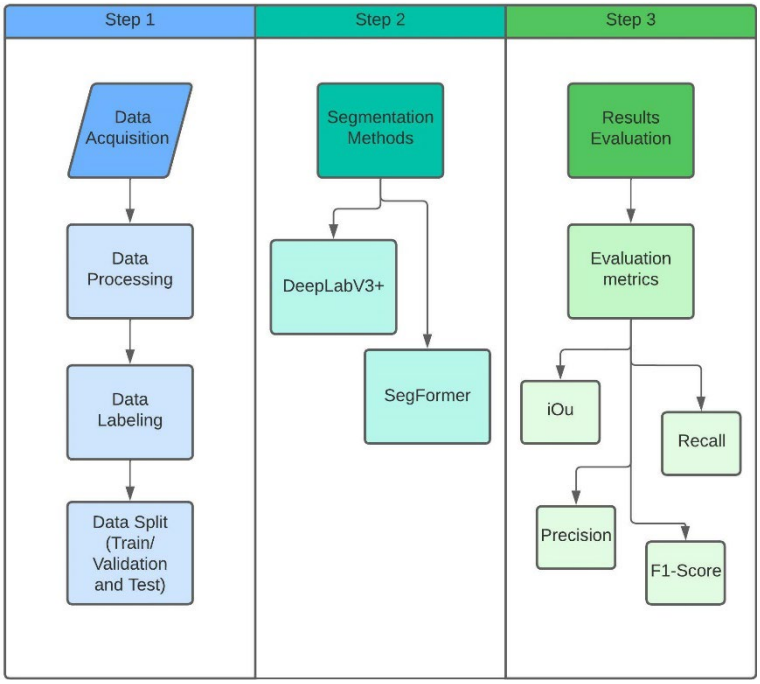
SegFormer (XIE *et al*., 2021) has shown promising results in terms of accuracy and precision in vegetation context (GEORGES GOMES *et al*., 2022), and burned areas (GONÇALVES *et al*., 2023). However, there is a lack of studies on mapping rivers using SegFormer. Only one study (YANG *et al*., 2022) modified the version Segformer-b0 with a module (Deepmask) to pay more attention to the details in the image and use Lovász loss to improve IoU. They used DeepLabv3+model as the teacher model to improve the segmentation, achieving 95.06% mIoU in Sentinel-2A multi-spectral images (10 m resolution). However, there is no information about the performance of this new semantic segmentation method to map rivers using high spatial resolution RGB images, or for narrow rivers, which have complex features, such as, sinuosity, higher variety of pixels and complex spectral information due to, vegetation and solid suspended, presence on the water. Thus, this study aims to evaluate the performance of the transformer-based SegFormer in the semantic segmentation of surface water (large and narrow rivers) in high spatial resolution RGB images. SegFormer will be compared with a recent CNN DeepLabV3+ which has been applied in several contexts (LI *et al*., 2019; MARTINS *et al*., 2021; MUHADI *et al*., 2021; HARIKA *et al*., 2021). The main contribution of this study is to indicate whether Transformer-based neural networks outperform CNN-based neural networks for segmenting rivers in high-resolution RGB images. Another contribution is to make available our dataset publicly[2].

2. MATERIALS AND METHOD

We divide the method into 3 steps and organize the workflow in Figure 26. Initially (step 1) we processed and labeled the RGB aerial images. Then we split the dataset into train/validation and test images. In a computer environment, we conducted semantic segmentation using deep learning methods (step 2). Finally, we evaluated and compared the performance of the methods (step 3).

[2] https://github.com/mayarafaita/River_SemanticSegmentation.git

Figure 26- Main steps carried out in the approach of our study.



Fonte: Autor (2023).

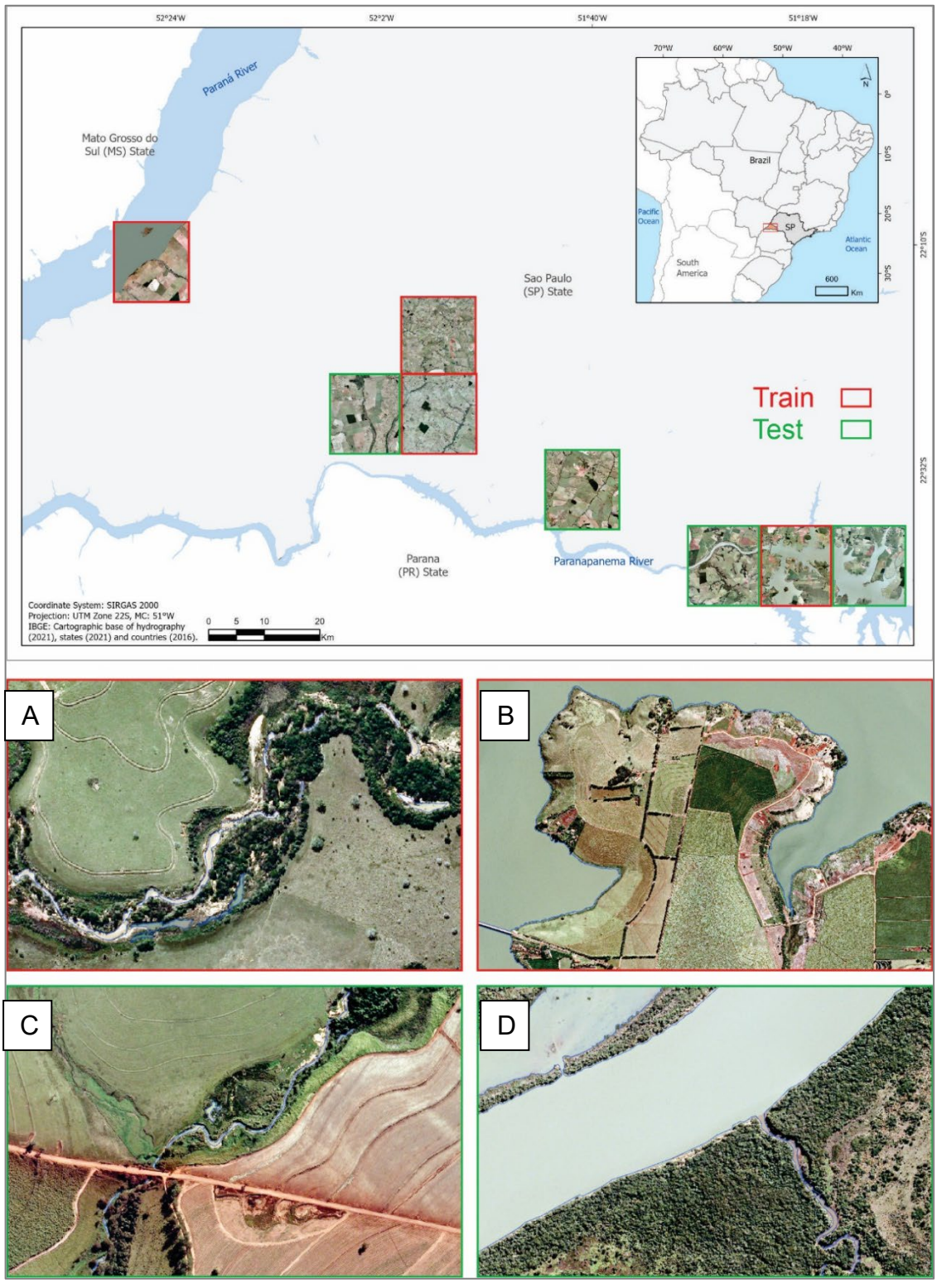
2.1 Data acquisition, processing, and labeling

We used aerial photographs from the Mapeia-SP project conducted in Brazil (2011). They are composed of RGB bands with 13,436 x 14,409 pixels and, a Ground Sample Distance (GSD) of 0.45 m. In imagery, the photographs present geometry and relief distortions, therefore, they undergo an orthorectification process, being called orthophotos. This process corrects distortions, effects of perspectives, and the relief's influence on the image's geometry. This gives more excellent cartographic product quality and results in a 1-meter resolution for our orthophotos.

The study area is a western region of São Paulo state (Figure 27). The region comprises rural roads, crops, pastures, buildings, vegetation, exposed soil, and rivers and lakes with different characteristics, such as large, narrow, sinuous, straight, silted, and clear, with riparian forests and shadows. This diversity of background elements and the diversity of rivers' water features makes the dataset even more challenging with a complex background. In terms of definition, we refer to the "water class" for large and narrow river labels. Furthermore, there is a class imbalance issue, where in some cases the water class presents more pixels than the background, for example, in large

rivers. Whereas, in most orthophotos, the water class has much fewer pixels (narrow rivers) than the background class.

Figure 27- Study area and water labeling.



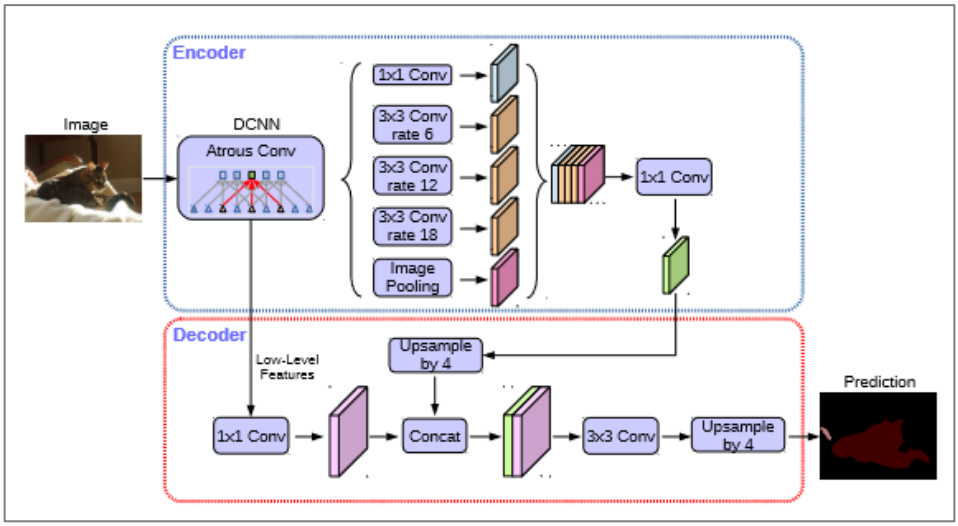
Fonte: Autor (2023).

We selected 8 orthophotos distributed throughout the study region as experimental areas. We manually labeled the river features in all orthophotos by visual interpretation using GIS. From this selection, 4 orthophotos are for training and validation and 4 are for testing (Figure 27). There are different features of rivers in the dataset, such as, narrow rivers (Figure 27 – A and C) and large rivers (Figure 27 – B and D). Narrow rivers are sinuous, have more sediments on the water, and spectral comportment different from clean river, for example. Large rivers don't have such sinuosity and are clearly and homogeneous water. Due to the variety of water and background features, and the class imbalance, we are careful to present samples representing these variations in training and test samples.

2.2 Semantic Segmentation Background

One of the structures of deep neural networks used in semantic segmentation tasks is the encoder-decoder. We are using two different encoder-decoder architectures: CNN-based DeepLabV3+ (CHEN *et al*., 2018) and Transformer-based SegFormer (XIE *et al*., 2021). Initially, DeepLab (CHEN *et al*., 2017a) was proposed with Atrous Convolution, Atrous Spatial Pyramid Pooling (ASPP) module, and a fully connected Conditional Random Field (CRF). This architecture allows an increase in the receptive field of network, and context information at multiple scales, and improves localization performance. However, the model misses some objects' boundaries. Therefore, the DeepLabV3 (CHEN *et al*., 2017b) makes some adjustments, such as Atrous convolution in cascade and batch normalization within ASPP, overcoming the DeepLab. Furthermore, to achieve better results, the CNN was improved with an encoder-decoder model, the DeepLabV3+ (CHEN *et al*., 2018). In this architecture, the encoder module is the structure of DeepLabV3, and the decoder module was added to improve the performance of the object boundaries (Figure 28). In addition, they applied the Xception model to the encoder-decoder for faster and more robust performance.

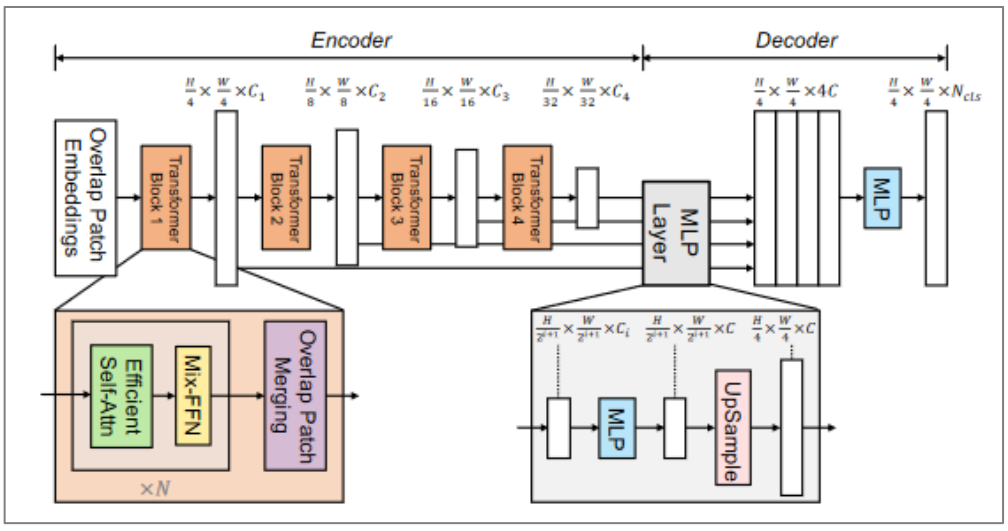
Figure 28- DeepLabV3+ architecture.



Fonte: Chen *et al*. (2018).

SegFormer (XIE *et al*., 2021) is an encoder-decoder based on Transformer-backbone for semantic segmentation tasks. The SegFormer framework consists of a hierarchical Transformer encoder and a lightweight All-MLP decoder (Figure 29). The hierarchically constructed encoder (Mix Transformer) qualifies the model to learn high-resolution coarse and low-resolution fine features and gain processing time and it is smaller than ViT. A further component that produces less computational demand and simplifies the model is the lightweight MLP decoder to predict the semantic segmentation mask.

Figure 29- SegFormer framework.



Fonte: Xie *et al*. (2021).

The SegFormer outsets with an input image split in a 4x4 patch size that goes through 4 transformer blocks and earns multi-level and multi-level features. Each transformer block has a self-attention module in Pyramid Vision Transformer (PVT), which reduces the sequence's length and the self-attention mechanism's complexity. In sequence, a Mix-FFN (feed-forward network) layer is introduced using a 3 × 3 Conv in the feed-forward network (FFN) to reduce the number of parameters and improve efficiency. The last layer in the transformer block, overlapping patch merging, assembles features of the same process size without overlapping.

2.3 Experimental details

For the input of the deep neural networks used in the experiments (SegFormer and DeepLabv3+), we split the images into non-overlapping patches with 512 x 512 pixels. Then we divide it into 2,569 for training, 454 for validation, and 2,968 for testing image patches. We created a second test set of 1,512 images, limited to only those with narrow rivers. The previously-mentioned semantic segmentation methods were implemented using the MMSegmentation toolbox (https://github.com/open-mmlab/mmsegmentation) on the Ubuntu 18.04 operating system. For DeepLabv3+ (CHEN *et al*., 2018) and SegFormer (Xia et al., 2021) we settle the parameters as AdamW optimizer (LOSHCHILOV; HUTTER, 2017) with an initial learning rate of 0.00006 for 80K iterations using a batch size of 2 and were updated by a Poly LR schedule with a factor of 1 by default. Deeplabv3+ was trained with ResNet-101, while SegFormer was trained with the MiT-B5 backbone, as determined to be the best option available in MMSegmentation. All the experiments were performed on a workstation with Intel (R) Core™ i7-5820K CPU @ 3.30GHz × 12 CPU, 32 GB memory, and an NVIDIA Titan V graphics card with 12 GB of graphics memory and 5120 CUDA (Compute Unified Device Architecture) cores.

Transformer-based and CNN-based models were evaluated and compared for performance assessment. We evaluate and compared SegFormer and DeepLabV3+ for all tested orthophotos and also evaluated the performance in the second test set, with only narrow rivers. We use the most consolidated metrics for the semantic segmentation task: Accuracy (Acc), F1-Score (F1), and Intersection over Union (IoU). Acc is the ratio of the correctly predicted pixel numbers to the total pixel number (Equation 1). F1-Score is the harmonic means for water precision and recall (Equation

2). IoU is the ratio of the intersection to the union of the ground truth and the predicted area.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (2)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (3)$$

Where, True Positive (TP) is the correctly predicting a label (we predicted “water”, and its “water”), True Negative (TN) is the correctly predicting the other label (we predicted “background”, and its “background”), False Positive (FP) is the falsely predicting a label (we predicted “water”, but it's “background”), False Negative (FN) is the missing and incoming label (we predicted “background”, but it’s “water”).

3. RESULTS AND DISCUSSION

We present the results and, discuss the related implications in 2 sections. In the first one (Section 3.1), we show the quantitative results and discuss the performance of Transformer-based architecture (SegFormer), and compared it with a CNN-based architecture (DeepLabV3+) to map water surface in high-resolution RGB images. Then (Section 3.2) we analyze the qualitative results comparing the DeepLabV3+ and SegFormer performance segmentation.

3.1 Evaluation of Quali-Quantitative Performance

The rivers’ semantic segmentation results are presented in Table 6. The evaluation metrics are described separately for the background pixels and segmented water pixels to obtain a complete analysis of the results. There is an imbalance of classes, and therefore, there may be inconsistency in the results evaluated on averages since there are more background pixels than water pixels on most images. Both methods demonstrated competence to map water surfaces, presenting accuracies of up to 96% (DeeplabV3+) and 98% (SegFormer) for water class (Table 6). However, SegFormer performed better than DeepLabV3+ in all evaluated metrics

for the two classes (background and water surface). This finding demonstrates the superiority of the Transformer-based over the CNN-based methods. Regarding the background class, both architectures demonstrated excellent performance, exceeding 99%. On the other hand, SegFormer achieved results up to 3.5% higher than DeepLabV3+, in the case of the water class IoU, in which DeepLabV3+ obtained 94.42% and SegFormer 97.96%.

Table 6- Segmentation results of all water surfaces.

All surface water	**Acc (%)**		**F1-Score (%)**		**IoU (%)**	
	BG	Water	BG	Water	BG	Water
SegFormer	**99.88**	**98.97**	**99.88**	**98.96**	**99.76**	**97.96**
DeepLabV3+	99.72	96.63	99.66	97.13	99.33	94.42

Fonte: Autor (2023).

In addition to this analysis, we also investigated the performance of SegFormer and DeepLabV3+ for narrow rivers (Table 7). Due to complex characteristics such as spectral comportment, shape, and more blurred edges, mapping narrow rivers has been a challenge in the scientific community (LI *et al*., 2022). We evaluate two orthophotos that represent narrow rivers, mentioned previously in section 2.3. Both methods maintained a great performance regarding the background classification performance, raising 99%. As predicted, both methods showed a decrease in all evaluated metrics, due to the complex resources of narrow rivers. However, the performance of the SegFormer concerning the segmentation of the water class remained superior for both scenarios. This finding further emphasizes the evidence that the Transformer-based method has been remarkable to the CNN method. Furthermore, for the water class, SegFormer presented a result up to 6.8% higher than DeepLabV3+ about IoU and 4.6% higher about F1-Score.

Table 7- Segmentation results of narrow rivers.

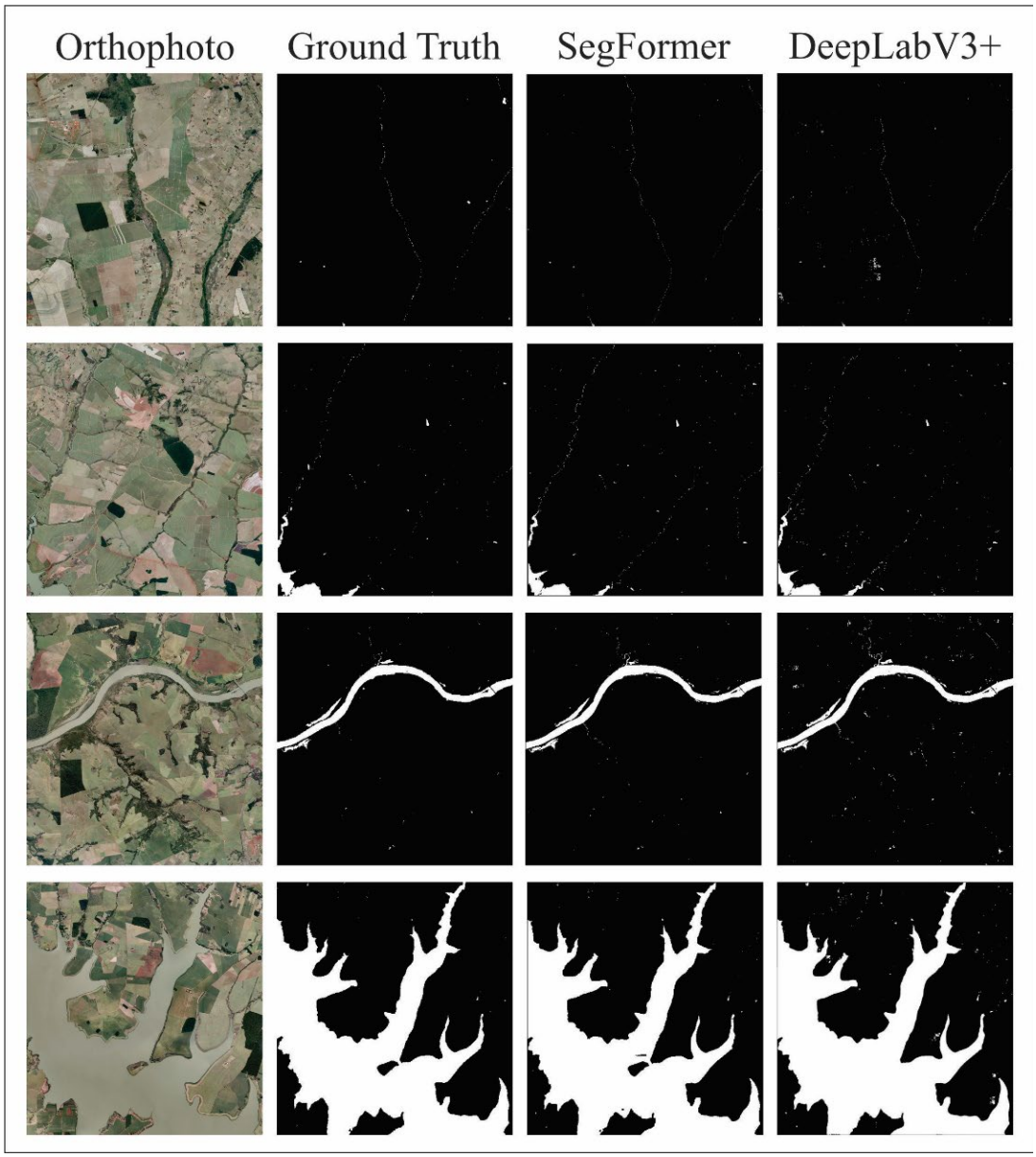
Narrow Rivers	**Acc (%)**		**F1-Score (%)**		**IoU (%)**	
	BG	Water	BG	Water	BG	Water
SegFormer	**99.92**	**81.76**	**99.87**	**85.99**	**99.75**	**75.42**
DeepLabV3+	99.83	80.81	99.83	81.35	99.65	68.56

Fonte: Autor (2023).

In all tested orthophotos, SegFormer presents better results, with segmentation more similar to ground truth (Figure 30). Both segmentation methods

can identify the water surface on the orthophotos. However, visually, DeepLabV3+ performed poorly, producing more false positive pixels of water. In the first two lines, the features of the river are narrow, with sediments in the water, and more complex to delimit. On the other hand, the last two lines are an example of rivers with characteristics of rivers with greater width, clear waters, and better-defined borders. If we compare the mentioned types of rivers, we noted that the large rivers were almost perfectly delimited, that is, there were almost no false negatives, and the two networks were able to identify the rivers. However, both networks still showed some false negatives, even for the large rivers. The biggest difficulty of the networks, mainly for DeepLabV3+, is concerning the examples of what is not a river.

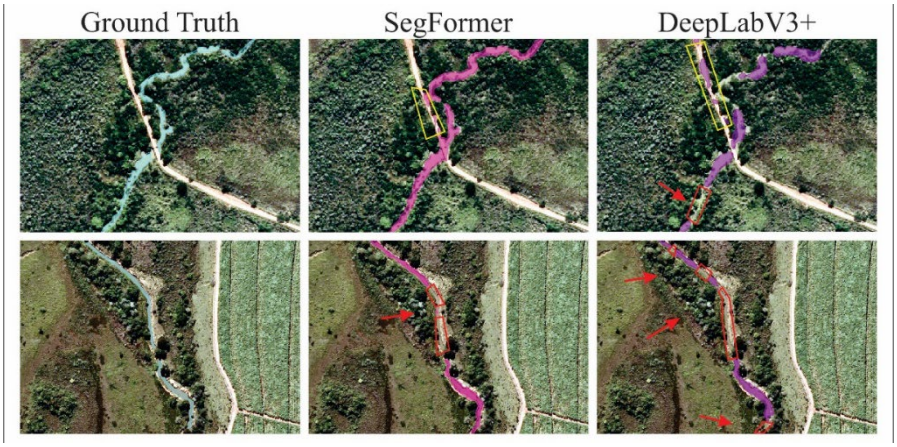
Figure 30- Results from water surface semantic segmentation in the orthophotos. The first 2 rows are examples of narrow rivers and the 2 last rows represent large rivers.



Fonte: Autor (2023).

SegFormer and DeepLabV3+ presented FP in a rural road region (Figure 31 - first line, yellow highlight), however, DeepLabV3+ presented more difficulty, segmenting larger areas of water class erroneously. This example also shows the ability of the networks to segment rivers with complex characteristics, both networks were able to delimit the narrow river. However, SegFormer presents a classification of the river continuously, that is, the river path is segmented continuously, without interruptions in the path. Otherwise, DeepLabV3+ presents the river segmentation in a disconnected way, as shown in Figure 31 (red highlight).

Figure 31- Examples of False Positives and False Negatives water segmentation.



Fonte: Autor (2023).

4. CONCLUSION

We investigated the capacity of a deep neural network based on ViT (SegFormer) in the task of semantic segmentation of rivers in RGB images with high spatial resolution. We compare the performance of the mentioned network with a CNN-based network (DeepLabV3+). The results demonstrated that vision transformer-based networks outperformed traditional CNN architecture (DeepLabV3+) in terms of quantitative and qualitative evaluation. The SegFormer architecture presented superior results in all evaluated metrics, reaching an excellent performance with an F1-Score of 98.96%. Both neural networks were able to segment rivers with great accuracy, above 96%, however, DeepLabV3+ segmentation presented more false positives and false negatives, making its performance worse. We found that the performance of deep neural networks degrades when we evaluate images of narrow rivers separately. As they contain narrow rivers, the features are more complex in spectral aspects, shape, and edge definition, making the segmentation task more challenging. Even so, the tests carried out with SegFormer are remarkable and maintained the best metrics, with an F1-Score of 85.99%. We conclude that the Transformer-based network, SegFormer, outperforms the CNN-based network (DeepLabV3+) and is capable of segment large rivers and narrow rivers in high spatial resolution RGB images. Future studies should continue to explore vision transformer architectures to segment rivers with different features, explore different platforms and sensors, and test the generalization capability of the SegFormer network in segmenting rivers.

REFERENCES

ALAM, M. *et al*. Convolutional neural network for the semantic segmentation of remote sensing images. **Mobile Networks and Applications**, v. 26, p. 200-215, 2021. Disponível em: < https://link.springer.com/article/10.1007/s11036-020-01703-3>. Acesso em: 24 de abril de 2022.

BRESSAN, P. O. *et al*. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. **International Journal of Applied Earth Observation and Geoinformation**, v. 108, p. 102690, 2022. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0303243422000162>. Acesso em: 23 de Agosto de 2022.

CHEN, L. C. *et al*. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. **IEEE transactions on pattern analysis and machine intelligence**, v. 40, n. 4, p. 834-848, 2017a. Disponível em: < https://ieeexplore.ieee.org/abstract/document/7913730/>. Acesso em: 16 de setembro de 2021.

CHEN, L. C. *et al*. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation. *In*: **Proceedings of the European conference on computer vision (ECCV),** p. 801-818, 2018. Disponível em: < http://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html>. Acesso em: 29 de maio de 2022.

CHEN, L. C. *et al*. Rethinking Atrous Convolution for Semantic Image Segmentation. **arXiv preprint arXiv:1706.05587**, 2017b. Disponível em: < https://arxiv.org/abs/1706.05587>. Acesso em : 16 de agosto de 2020.

DOSOVITSKIY, A. *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. **arXiv preprint arXiv:2010**, p. 11929, 2020. Disponível em: < https://arxiv.org/abs/2010.11929>. Acesso em: 28 de março de 2022.

GEORGES GOMES, F. D. *et al*. Urban Trees Mapping Using Multi-Scale Rgb Image and Deep Learning Vision Transformer-Based. **Available at SSRN 4167085**. Disponível em: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4167085>. Acesso em: 12 de setembro de 2022.

GONÇALVES, D. N. *et al*. Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 116, p. 103151, 2023. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1569843222003399>. Acesso em: fevereiro de 2023.

HARIKA, A. *et al*. Extracting Water Bodies in RGB Images Using DEEPLABV3+ Algorithm. **The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 46, p. 97-101, 2022. Disponível em : < https://isprs-archives.copernicus.org/articles/XLVI-M-2-2022/97/2022/isprs-archives-XLVI-M-2-2022-97-2022.html>. Acesso em: 27 de outubro de 2022.

KO, B. C.; KIM, H. H.; NAM, J. Y. Classification of potential water bodies using Landsat 8 OLI and a combination of two boosted random forest classifiers. **Sensors**, v. 15, n. 6, p. 13763-13777, 2015. Disponível em: < https://www.mdpi.com/102012>. Acesso em: 20 de março de 2020.

KOTARIDIS, I.; LAZARIDOU, M. Remote sensing image segmentation advances: A meta-analysis. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 173, p. 309-322, 2021. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0924271621000265>. Acesso em: 14 de abril de 2022.

LI, W. *et al*. Urban water extraction with UAV high-resolution remote sensing data based on an improved U-Net model. **Remote Sensing**, v. 13, n. 16, p. 3165, 2021. Disponível em: < https://www.mdpi.com/2072-4292/13/16/3165>. Acesso em: 13 de março de 2022.

LI, Y. *et al*. Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 187, p. 306-327, 2022b. Disponível em:< https://www.sciencedirect.com/science/article/pii/S0924271622000867>. Acesso em: 27 de agosto de 2022.

LI, Z. *et al*. Multiscale features supported the deeplabv3+ optimization scheme for accurate water semantic segmentation. **IEEE Access**, v. 7, p. 155787-155804, 2019. Disponível em: < https://ieeexplore.ieee.org/abstract/document/8883176/>. Acesso em: 19 de setembro de 2022.

LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. **arXiv preprint arXiv:1711.05101**, 2017. Disponível em: < https://arxiv.org/abs/1711.05101>. Acesso em: 23 de março de 2022.

MARTINS, J. A. C. *et al*. Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. **Remote Sensing**, v. 13, n. 16, p. 3054, 2021. Disponível em: < https://www.mdpi.com/2072-4292/13/16/3054>. Acesso em: 27 de março de 2022.

MUHADI, N. A. *et al*. Deep learning semantic segmentation for water level estimation using surveillance camera. **Applied Sciences**, v. 11, n. 20, p. 9691, 2021. Disponível em: < https://www.mdpi.com/2076-3417/11/20/9691>. Acesso em: 5 de abril de 2022.

SARP, G.; OZCELIK, M. Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey. **Journal of Taibah University for Science**, v. 11, n. 3, p. 381-391, 2017. Disponível em: < https://www.tandfonline.com/doi/abs/10.1016/j.jtusci.2016.04.005>. Acesso em: 14 de maio de 2022.

SUN, F. *et al*. Monitoring dynamic changes of global land cover types: Fluctuations of major lakes in China every 8 days during 2000–2010. **Chinese Science Bulletin**, v. 59, p. 171-189, 2014. Disponível em: < https://link.springer.com/article/10.1007/s11434-013-0045-0>. Acesso em: 12 de novembro de 2019.

TIAN, S. *et al*. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. **Remote Sensing**, v. 8, n. 11, p. 954, 2016. Disponível em: < https://www.mdpi.com/165886>. Acesso em: 15 de maio de 2020.

VASWANI, A. *et al*. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017. Disponível em: < https://proceedings.neurips.cc/paper/7181-attention-is-all>. Acesso em: 26 de outubro de 2022.

WANG, Y. *et al*. Lightweight Deep Neural Network Method for Water Body Extraction from High-Resolution Remote Sensing Images with Multisensors. **Sensors**, v. 21, n. 21, p. 7397, 2021. Disponível em:< https://www.mdpi.com/1347686>. Acesso em: 24 de fevereiro de 2022.

XIE, E. *et al*. SegFormer: Simple and efficient design for semantic segmentation with transformers. **Advances in Neural Information Processing Systems**, v. 34, p. 12077-12090, 2021. Disponível em: < https://proceedings.neurips.cc/paper/2021/hash/64f1f27bf1b4ec22924fd0acb550c235-Abstract.html>. Acesso em: 23 de outubro de 2022.

YANG, X. *et al*. WaterSegformer: A lightweight model for water body information extraction from remote sensing images. **IET Image Processing**, v. 17, n. 3, p. 862-871, 2023. Disponível em: < https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.12678>. Acesso em: 4 de fevereiro de 2023.

YUAN, X.; SHI, J.; GU, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. **Expert Systems with Applications**, v. 169, p. 114417, 2021. DIsponível em: < https://www.sciencedirect.com/science/article/pii/S0957417420310836>. Acesso em: 15 de setembro de 2022.

ZHANG, Z. *et al*. Rich CNN Features for water-body segmentation from very high resolution aerial and satellite imagery. **Remote Sensing**, v. 13, n. 10, p. 1912, 2021. Disponível em:< https://www.mdpi.com/2072-4292/13/10/1912>. Acesso em: 18 de março de 2022.

2 CONSIDERAÇÕES FINAIS

A presente pesquisa avaliou o desempenho de redes de aprendizagem profunda para mapear rios em imagens RGB de alta resolução espacial. No capítulo I foi realizada uma síntese das pesquisas mais recentes que trabalharam com sensoriamento remoto e aprendizagem profunda para segmentar recursos hídricos. Com esse estudo, notamos que existem poucos estudos que trabalham com imagens de baixa resolução espectral, como RGB, em conjunto com alta resolução espacial, para mapeamento de rios, apresentando ainda, problemas para mapear rios estreitos.

Em seguida, Capítulo II, avaliamos redes neurais de aprendizagem profunda baseadas em CNN para segmentar rios largos e rios estreitos. As arquiteturas avaliadas nesta etapa da pesquisa (DeepLabV3+, Unet e PSPNet) apresentaram baixa performance, com um F1-Score abaixo de 70% para segmentar rios de largura inferior a 10 metros (rios estreitos) em imagens RGB mesmo que em alta resolução espacial. Porém, estas CNNs conseguem segmentar rios com largura superior a 10 metros com uma alta performance, com um F1-Score acima de 95% (DeepLabV3+ e Unet). Ao usar a CNN de melhor performance (DeepLabV3+) em uma área de maior abrangência para inferir as regiões de rios, a métrica de se mantém alta (F1-Score 97%) para rios largos. Entretanto, para rios estreitos a acurácia diminui ainda mais (F1-Score 15%). Em razão de trabalharmos um conjunto de dados com características espectrais muito complexas e diversas, houve maior dificuldade em segmentar rios estreitos. Estes apresentam comportamento espectral facilmente confundível com vegetações, áreas úmidas, sombras e até trechos de solo exposto, porque apresentam vegetação e sedimentos em suspensão na água. Portanto, mesmo usando amostras com rios de largura inferior a 10 metros e amostras diversificadas da categoria “não rio”, as redes neurais ainda apresentaram dificuldades para segmentar esse tipo de rio.

Com o objetivo de melhorar a acurácia da segmentação semântica dos rios (largos e estreitos), no Capítulo III, exploramos uma arquitetura de rede de aprendizagem profunda SegFormer, que tem sido o estado-da-arte em segmentação semântica, que são baseadas em visão-Transformer (ViT). Comparada com a rede neural DeepLabV3+, o SegFormer apresentou melhor performance em todas as métricas avaliadas, inclusive para rios estreitos, sendo capaz de segmentar os rios com maior precisão e menor confusão com os elementos do fundo. Concluímos que

o SegFormer foi capaz de mapear rios largos e estreitos com acurácia de 98.97%, sendo adequado para essa tarefa em imagens RGB de 1 metro de resolução.

Recomendamos que trabalhos futuros explorem as redes baseadas em ViT para segmentação semântica em outros contextos. Outra recomendação está na investigação da adaptação de domínio da rede treinada, assim, será possível verificar a generalização da rede, se ela é capaz de segmentar outros tipos de rios em outras regiões além da que foi estudada. Outra análise para trabalhos futuros está relacionada com o mapeamento de recursos hídricos utilizando ViT em imagens de sensoriamento remoto multitemporais, que podem auxiliar no monitoramento, em tempo real, dos recursos hídricos. Uma análise multimodal também poderá contribuir para o monitoramento dos recursos hídricos, considerando que cada sensor tem suas próprias limitações, portanto, modelar uma estrutura de rede robusta com base em imagens multimodais é significativo para um maior desenvolvimento no futuro.

REFERÊNCIAS

ACHARYA, T. D.; SUBEDI, A.; LEE, D. H. Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 scene of Nepal. **Sensors**, v.19, n.12, p. 2769, 2019. Disponível em:< https://www.mdpi.com/482956>. Acesso em: 23 de abril de 2020.

ALAM, M. *et al*. Convolutional neural network for the semantic segmentation of remote sensing images. **Mobile Networks and Applications**, v. 26, p. 200-215, 2021. Disponível em: < https://link.springer.com/article/10.1007/s11036-020-01703-3>. Acesso em: 24 de abril de 2022.

BAO, L.; LV, X.; YAO, J. Water extraction in SAR Images using features analysis and dual-threshold graph cut model. **Remote Sensing**, v. 13, n. 17, p. 3465, 2021. Disponível em: < https://www.mdpi.com/1254334>. Acesso em: 21 de fevereiro de 2022.

DOSOVITSKIY, A. *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. **arXiv preprint arXiv:2010**, p. 11929, 2020. Disponível em: < https://arxiv.org/abs/2010.11929>. Acesso em: 28 de março de 2022.

ELHAG, M. *et al*. Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia. **Water**, v. 11, n. 3, p. 556, 2019. Disponível em:< https://www.mdpi.com/429356>. Acesso em: 22 de março de 2022.

FEYISA, G. L. *et al*. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. **Remote Sensing of Environment**, v. 140, p. 23-35, 2014. Disponível em :< https://www.sciencedirect.com/science/article/pii/S0034425713002873>. Acesso em: 29 de novembro de 2019.

GAUTAM, S.; SINGHAI, J. Cosine-similarity watershed algorithm for water-body segmentation applying deep neural network classifier. **Environmental Earth Sciences**, v. 81, n. 9, p. 1-16, 2022. Disponível em:< https://link.springer.com/article/10.1007/s12665-022-10376-y>. Acesso em: 19 de março de 2022.

GEORGES GOMES, F. D. *et al*. Urban Trees Mapping Using Multi-Scale Rgb Image and Deep Learning Vision Transformer-Based. **Available at SSRN 4167085**. Disponível em: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4167085>. Acesso em: 12 de setembro de 2022.

GONÇALVES, D. N. *et al*. Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 116, p. 103151, 2023. Disponível em: < https://www.sciencedirect.com/science/article/pii/S1569843222003399>. Acesso em: fevereiro de 2023.

GUI, R. *et al*. A River Channel Extraction Method Based on a Digital Elevation Model Retrieved from Satellite Imagery. **Water**, v. 14, n. 15, p. 2387, 2022. Disponível em: < https://www.mdpi.com/1755924>. Acesso em: 3 de setembro de 2022.

HU, K. *et al*. Multi-Scale Feature Aggregation Network for Water Area Segmentation. **Remote Sensing**, v. 14, n. 1, p. 206, 2022. Disponível em:< https://www.mdpi.com/article/10.3390/rs14010206>. Acesso em: 23 de junho de 2022.

ISIKDOGAN, F.; BOVIK, A. C.; PASSALACQUA, P. Surface water mapping by deep learning. **IEEE journal of selected topics in applied earth observations and remote sensing**, v. 10, n. 11, p. 4909-4918, 2017. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8013683/>. Acesso em: 19 de maio de 2021.

KUHN, C. *et al*. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. **Remote Sensing of Environment**, v. 224, p. 104-118, 2019. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0034425719300288>. Acesso em: 19 de abril de 2020.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015. Disponível em:< https://www.nature.com/articles/nature14539>. Acesso em: 22 de agosto de 2019.

LI, A. *et al*. Comparative analysis of machine learning algorithms in automatic identification and extraction of water boundaries. **Applied Sciences**, v. 11, n. 21, p. 10062, 2021a. Disponível em: < https://www.mdpi.com/1332132>. Acesso em: 28 de setembro de 2022.

LI, A.; FAN, M.; QIN, G. Comparative analysis of machine learning algorithms in water extraction. *In*: **Journal of Physics: Conference Series**. IOP Publishing, 2021. p. 012045. Disponível em: < https://iopscience.iop.org/article/10.1088/1742-6596/2076/1/012045/meta>. Acesso em: 17 de março de 2022.

LI, W. *et al*. Urban water extraction with UAV high-resolution remote sensing data based on an improved U-Net model. **Remote Sensing**, v. 13, n. 16, p. 3165, 2021b. Disponível em: < https://www.mdpi.com/2072-4292/13/16/3165>. Acesso em: 13 de março de 2022.

LI, Y. *et al*. An index and approach for water extraction using Landsat–OLI data. **International Journal of Remote Sensing**, v. 37, n. 16, p. 3611-3635, 2016. Disponível em: < https://www.tandfonline.com/doi/abs/10.1080/01431161.2016.1201228>. Acesso em: 22 de agosto de 2022.

LI, Y. *et al*. Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 187, p. 306-327, 2022b. Disponível em:< https://www.sciencedirect.com/science/article/pii/S0924271622000867>. Acesso em: 27 de agosto de 2022.

LIU, D. *et al*. Human-induced eutrophication dominates the bio-optical compositions of suspended particles in shallow lakes: Implications for remote sensing. **Science of the Total Environment**, v. 667, p. 112-123, 2019. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0048969719308733>. Acesso em: 22 de novembro de 2019.

MA, Y. *et al*. Satellite-derived bathymetry using the ICESat-2 lidar and Sentinel-2 imagery datasets. **Remote Sensing of Environment**, v. 250, p. 112047, 2020. Disponível em: < https://www.sciencedirect.com/science/article/pii/S003442572030417X>. Acesso em: 6 de dezembro de 2021.

MARTINS, J. A. C. *et al*. Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. **Remote Sensing**, v. 13, n. 16, p. 3054, 2021. Disponível em: < https://www.mdpi.com/2072-4292/13/16/3054>. Acesso em: 27 de março de 2022.

MIAO, Z. *et al*. Automatic water-body segmentation from high-resolution satellite images via deep networks. **IEEE geoscience and remote sensing letters**, v. 15, n. 4, p. 602-606, 2018. Disponível em:< https://ieeexplore.ieee.org/abstract/document/8286914/>. Acesso em: 25 de setembro de 2022.

MINAEE, S. *et al*. Image segmentation using deep learning: A survey. **IEEE transactions on pattern analysis and machine intelligence**, v. 44, n. 7, p. 3523-3542, 2021. Disponível em: < https://ieeexplore.ieee.org/abstract/document/9356353/>. Acesso em: 26 de outubro de 2021.

MUHADI, N. A. *et al*. Deep learning semantic segmentation for water level estimation using surveillance camera. **Applied Sciences**, v. 11, n. 20, p. 9691, 2021. Disponível em: < https://www.mdpi.com/2076-3417/11/20/9691>. Acesso em: 5 de abril de 2022.

NOVO, E. M.L.M. **Sensoriamento Remoto: princípios e aplicações**. São Paulo: Editora Blucher, 2010.

OSCO, L. P. *et al*. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. **Precision Agriculture**, v. 22, n. 4, p. 1171-1188, 2021. Disponível em: < https://link.springer.com/article/10.1007/s11119-020-09777-5>. Acesso em: 16 de abril de 2022.

SAGAN, V. *et al*. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. **Earth-Science Reviews**, v. 205, p. 103187, 2020. DIsponível em: < https://www.sciencedirect.com/science/article/pii/S0012825220302336>. Acesso em: 10 de abril de 2022.

SEONG, S.; CHOI, J. Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates. **Remote Sensing**, v. 13, n. 16, p. 3087, 2021. Disponível em: < https://www.mdpi.com/2072-4292/13/16/3087>. Acesso em: 27 de maio de 2022.

SHAHABI, H. *et al*. Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. **Remote Sensing**, v. 12, n. 2, p. 266, 2020. Disponível em: < https://www.mdpi.com/616574>. Acesso em: 23 de fevereiro de 2022.

SUN, W. *et al*. Calibrating a hydrological model in a regional river of the Qinghai–Tibet plateau using river water width determined from high spatial resolution satellite images. **Remote Sensing of Environment**, v. 214, p. 100-114, 2018. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425718302414. Acesso em: 24 de maio de 2020.

TANG, H. *et al*. Large-Scale Surface Water Mapping Based on Landsat and Sentinel-1 Images. **Water**, v. 14, n. 9, p. 1454, 2022. Disponível em: < https://www.mdpi.com/2073-4441/14/9/1454>. Acesso em: 28 de agosto de 2022.

USGS. United States Geological Survey. **Earth Explorer**. 2020. Acesso em: 28 jan. 2020. Disponível em: < https://earthexplorer.usgs.gov/>. Acesso em: 18 de abril de 2020.

VASWANI, A. *et al*. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017. Disponível em: < https://proceedings.neurips.cc/paper/7181-attention-is-all>. Acesso em: 26 de outubro de 2022.

WANG, X. *et al*. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 68, p. 73-91, 2018. Disponível em: < https://www.sciencedirect.com/science/article/pii/S0303243418300990>. Acesso em: 25 de fevereiro de 2020.

WANG, Y. *et al*. Lightweight Deep Neural Network Method for Water Body Extraction from High-Resolution Remote Sensing Images with Multisensors. **Sensors**, v. 21, n. 21, p. 7397, 2021. Disponível em:< https://www.mdpi.com/1347686>. Acesso em: 24 de fevereiro de 2022.

XIE, E. *et al*. SegFormer: Simple and efficient design for semantic segmentation with transformers. **Advances in Neural Information Processing Systems**, v. 34, p. 12077-12090, 2021. Disponível em: < https://proceedings.neurips.cc/paper/2021/hash/64f1f27bf1b4ec22924fd0acb550c235-Abstract.html>. Acesso em: 23 de outubro de 2022.

YI, Y. *et al*. Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network. **Remote sensing**, v. 11, n. 15, p. 1774, 2019. Disponível em: < https://www.mdpi.com/504832>. Acesso em: 22 de novembro de 2019.

YULIANTO, F. *et al*. Evaluation of the Threshold for an Improved Surface Water Extraction Index Using Optical Remote Sensing Data. **The Scientific World Journal**, v. 2022, 2022. Disponível em: < https://www.hindawi.com/journals/tswj/2022/4894929/>. Acesso em: 14 de maio de 2022.