

UNIVERSIDADE FEDERAL DO PARANÁ

JORGIANA KAMILA TEIXEIRA DO CRATO

DETECÇÃO DE ESTRADAS RURAIS EM IMAGENS PLANET USANDO REDE
CONVOLUCIONAL U-NET

CURITIBA

2023

JORGIANA KAMILA TEIXEIRA DO CRATO

DETECÇÃO DE ESTRADAS RURAIS EM IMAGENS PLANET USANDO REDE
CONVOLUCIONAL U-NET

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em ciências geodésicas.

Orientador: Prof. Dr. Jorge Antônio Silva Centeno

Coorientador: Dr. Mario Ernesto Jijón Palma

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Crato, Jorgiana Kamila Teixeira do

Detecção de estradas rurais em imagens Planet usando rede
convolucional U-Net / Jorgiana Kamila Teixeira do Crato. – Curitiba, 2023.
1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de
Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas.

Orientador: Jorge Antônio Silva Centeno

Coorientador: Mario Ernesto Jijón Palma

1. Estradas rurais – Detecção. 2. Redes neurais. 3. Sensoriamento
Remoto. 4. Imagens multiespectrais. 5. Google Colab. I. Universidade
Federal do Paraná. II. Programa de Pós-Graduação em Ciências
Geodésicas. III. Centeno, Jorge Antônio Silva. IV. Palma, Mario Ernesto Jijón.
V. Título.

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **JORGIANA KAMILA TEIXEIRA DO CRATO** intitulada: **DETECÇÃO DE ESTRADAS RURAIS EM IMAGENS PLANET USANDO REDE CONVOLUCIONAL U-NET**, sob orientação do Prof. Dr. JORGE ANTONIO SILVA CENTENO, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Agosto de 2023.

Assinatura Eletrônica

26/08/2023 16:48:54.0

JORGE ANTONIO SILVA CENTENO

Presidente da Banca Examinadora

Assinatura Eletrônica

25/08/2023 16:38:07.0

HIDEO ARAKI

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

28/08/2023 14:35:43.0

VERALDO LIESENBERG

Avaliador Externo (UNIVERSIDADE DO ESTADO DE SANTA CATARINA)

Centro Politécnico - Caixa Postal 19001 - CURITIBA - Paraná - Brasil

CEP 81531-980 - Tel: (41) 3361-3153 - E-mail: cpqcg@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 309733

Para autenticar este documento/assinatura, acesse <https://siga.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 309733

Dedico este trabalho aos meus pais que, mesmo diante das dificuldades, sempre me incentivaram a estudar, apoiando-me e prezando pelo meu melhor.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder as forças necessárias para enfrentar os desafios da vida. Sua orientação e apoio constante têm sido fundamentais para meu crescimento e superação.

Também não posso deixar de agradecer à minha família, que sempre esteve ao meu lado, oferecendo amor incondicional, compreensão e encorajamento. Suas presenças em minha jornada tornam cada passo mais significativo e especial.

Aos meus amigos da pós, expresso minha sincera gratidão por compartilharem risos, lágrimas e momentos inesquecíveis. Sua amizade é um tesouro que valorizo profundamente, e sou grato por poder contar com vocês em todos os momentos.

Ao meu orientador Professor Dr. Jorge Antônio Silva Centeno e Coorientador Dr. Mario Ernesto Jijón Palma, pelas orientações e paciência nesse caminho.

Ao programa NCFI que disponibilizou o acesso as imagens da constelação Planet possibilitando o desenvolvimento desta pesquisa.

Enfim agradeço a todas as pessoas que fizeram parte da minha jornada nos dois últimos anos.

RESUMO

O Brasil, um dos países mais extensos do mundo, possui uma significativa parcela de vias de rodagem situada em ambiente rural, sem a devida manutenção, o que dificulta a extensão de serviços à população rural. Muitas dessas estradas desempenham um papel fundamental na gestão territorial, uma vez que são responsáveis pelo escoamento da produção agrícola do interior do país e pela conectividade das comunidades rurais. A manutenção destas estradas e sua exploração para a extensão de serviços básicos, como energia e água, é somente possível com uma adequada atualização cartográfica da rede viária. Mais recentemente, o uso de métodos de aprendizado profundo para a análise de imagens orbitais tem crescido significativamente. Dentro desta nova realidade, esta pesquisa propõe uma abordagem baseada em técnicas de sensoriamento remoto aliadas a ferramentas de inteligência artificial, com o intuito de contribuir para solucionar o problema do mapeamento de estradas em áreas rurais. Para isto, se propõe o uso das redes convolucionais. Utilizando a arquitetura U-Net, foi possível identificar um potencial promissor na detecção de estradas rurais em imagens da constelação Planet. A taxa de detecção alcançada foi notável, atingindo uma acurácia de 92%. Contudo, é importante ressaltar a necessidade de aprimoramentos, visto que outras métricas de avaliação, como a precisão (76,66%) e o f1-score (69,48%), indicam margem para otimização dos parâmetros utilizados. No estudo também é feita uma análise comparativa entre o uso dos interpretadores na nuvem, do Google Colab (em ambiente virtual) e Pyzo (em ambiente local, utilizando o computador desktop/workstation fornecido pela UFPR). Verificou-se que o Colab apresenta vantagens em termos de custo e acesso a recursos de processamento. Entretanto, é relevante destacar que o uso do Colab também traz consigo algumas limitações, as quais requerem uma abordagem cuidadosa ao ajustar a complexidade do modelo e o tamanho do conjunto de dados.

Palavras-chave: U-Net; Planet; Google Colab; Detecção de estradas em ambiente rural.

ABSTRACT

Brazil, one of the largest countries in the world, has a significant number of roads in rural areas that are not properly maintained, making it difficult to extend services to the rural population. Many of these roads play a fundamental role in land management, as they are responsible for transporting agricultural produce from the interior of the country and for connecting rural communities. Maintaining these roads and exploiting them to extend basic services, such as energy and water, is only possible with a proper cartographic update of the road network. More recently, the use of deep learning methods to analyse orbital images has grown significantly. Within this new reality, this research proposes an approach based on remote sensing techniques combined with artificial intelligence tools, with the aim of helping to solve the problem of mapping roads in rural areas. To this end, the use of convolutional networks is proposed. Using the U-Net architecture, it was possible to identify promising potential for detecting rural roads in images from the Planet constellation. The detection rate achieved was remarkable, reaching an accuracy of 92 per cent. However, it is important to highlight the need for improvement, since other evaluation metrics, such as accuracy (76.66%) and f1-score (69.48%), indicate room for optimization of the parameters used. The study also makes a comparative analysis between the use of interpreters in the cloud, Google Colab (in a virtual environment) and Pyzo (in a local environment, using the desktop/workstation computer provided by UFPR). Colab was found to have advantages in terms of cost and access to processing resources. However, it is important to emphasize that the use of Colab also brings with it some limitations, which require a careful approach when adjusting the complexity of the model and the size of the data set.

Keywords: U-Net; Planet; Google Colab; Rural road detection.

LISTA DE FIGURAS

FIGURA 1 –	Esquema de correlação entre inteligência artificial (IA), <i>machine learning</i> (ML) e <i>Deep learning</i> (DL)	30
FIGURA 2 –	Esquematização do neurônio artificial	32
FIGURA 3 –	Exemplos de funções de ativação: A) limiar (degrau); B) linear; C) sigmoide logística, D) sigmoide tangente hiperbólica	34
FIGURA 4 –	Modelo esquemático de uma rede neural de três camadas, com camada de entrada, saída e camada oculta	36
FIGURA 5 –	Aplicação do conceito de neurônio a imagens digitais (A) processamento de um pixel com várias bandas. (B) processamento de uma região em uma banda	40
FIGURA 6 –	Etapas de processamento da metodologia	47
FIGURA 7 –	Exemplo de recortes da imagem PLANET e identificação visual de trechos de estradas. A) Imagem original; B) Imagens com pixels marcados e C) Máscara de estradas	48
FIGURA 8 –	Esquema de uma <i>autoencoder</i>	50
FIGURA 9 –	Exemplo da Rede U-Net aplicada a imagens de 32x32 pixels com baixa resolução	51
FIGURA 10 –	Exemplo de recorte contendo padrão espacial que caracteriza as estradas, A) Recorte contendo exemplo de estradas, B) Máscara com pixels marcados pela previsão da rede e C) Interpretação do trecho pela rede	54
FIGURA 11 –	Exemplo de recorte contendo padrão espacial que caracteriza as estradas em ambiente com elementos antrópicos que podem causar confusão da rede. A) Recorte contendo exemplo de estradas, B) Máscara com pixels marcados pela previsão da rede e C) Interpretação do trecho pela rede	54
FIGURA 12 –	Mapa de localização das 12 cenas Planet sobre a extensão territorial do estado do Paraná em relação a região urbana de Curitiba – Paraná	61
FIGURA 13 –	A) Recorte 40 x 40pixel no RGB e B) imagem rotulada 40 x 40 pixels correspondente ao trecho de estrada	62
FIGURA 14 –	Primeira série de experimentos: A) Imagem original sobreposta pelo resultado da rede; B) Mascara Binária resultante do primeiro teste	66
FIGURA 15 –	Segunda série de experimentos: A) Imagem original sobreposta pelo resultado da rede; B) Mascara Binária resultante do primeiro teste	70
FIGURA 16 –	A) Recorte de entrada; B) Mascara binária baseada na verdade de campo; C) resposta resultante da detecção usando a rede U-Net para 100 épocas com 16 neurônios; e D) resposta resultante da detecção usando a rede U-Net para 100 épocas 32 neurônios	72
FIGURA 17 –	Primeiro Recorte de imagem Planet para teste	82
FIGURA 18 –	Comparação do primeiro recorte: vermelho previsão; azul digitalização; preto coincidência	82
FIGURA 19 –	Segundo recorte de imagem Planet para teste	84

FIGURA 20 –	Comparação do segundo recorte: vermelho predição; azul digitalização; preto coincidência	84
--------------------	--	----

LISTA DE QUADROS

QUADRO 1 –	Trabalhos usando redes neurais convolucionais aplicadas diretamente ao sensoriamento remoto	42
QUADRO 2 –	Trabalhos usando U-Net aplicada diretamente ao sensoriamento remoto	43
QUADRO 3 –	Exemplo de possíveis saídas da rede	56
QUADRO 4 –	Correlação entre resposta TP, FP e FN consideradas para os cálculos das métricas de avaliação	57
QUADRO 5 –	Resultado bem sucedido para primeira série de experimentos: A) exemplo de entrada; B) estrada; C) detecção resultante	67
QUADRO 6 –	Resultado bem sucedido para segunda série de experimentos: A) exemplo de entrada; B) estrada; C) Detecção resultante	71
QUADRO 7 –	Resultado bem sucedido para terceira série de experimentos: A) exemplo de entrada; B) estrada; C) detecção resultante	75
QUADRO 8 –	Primeira comparação visual entre as series de experimentos: exemplos de entrada na rede e marcaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos	77
QUADRO 9 –	Segunda comparação visual entre as series de experimentos: exemplos de entrada na rede e marcaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos	78
QUADRO 10 –	Terceira comparação visual entre as series de experimentos: exemplos de entrada na rede e marcaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos	79

LISTA DE TABELAS

TABELA 1 –	Resultantes para primeira série de testes usando plataforma Colab, 9760 amostras de treinamento e 1783 amostras de teste, 50 épocas de treinamento	65
TABELA 2 –	Resultantes para segundo conjunto de testes usando plataforma Colab, 9760 amostras de treinamento e 1783 amostras de teste, 100 épocas de treinamento	69
TABELA 3 –	Resultantes para terceiro conjunto de testes usando workstation, 9760 amostras de treinamento e 1783 amostras de teste, 300 épocas de treinamento	73
TABELA 4 –	Comparação das métricas resultantes para os experimentos...	76
TABELA 5 –	Matriz de confusão da região L: 1501 - 1900, C: 1701 – 2400. AU. Acurácia do usuário. AP. Acurácia do produtor	83
TABELA 6 –	Matriz de confusão da região L: 2301 - 3000, C: 501 - 900. AU. Acurácia do usuário. AP. Acurácia do produtor	84

LISTA DE ABREVIACÕES E SIGLAS

CNT	Confederação Nacional de Transporte
GNSS	Sistema Global de Navegação por Satélite
BRDF	<i>Bidirectional Reflectance Distribution Function</i>
CNN	<i>Convolutional Neural Network</i>
Colab	<i>Google Colaboratory</i>
CRF	<i>Conditional Random Field</i>
DL	<i>Deep Learning</i>
FA	<i>Função de Ativação</i>
FCN	<i>Fully Convolutional Network</i>
FCNN	<i>Fully Convolutional Neural Network</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
GAN	<i>Generative Adversarial Networks</i>
GPU	Unidades de Processamento Gráfico
IA	Inteligência Artificial
LIDAR	<i>Light Detection and Ranging</i>
MDS	Modelo Digital de Superfície
MD-UNET	<i>Multi-Scale Dense Dilated Convolution U-Net</i>
ML	<i>Machine Learning</i>
NICFI	<i>Norway's International Climate & Forests Initiative</i>
OBIA	<i>Object Based Image Analysos</i>
RCNN-UNET	<i>Recurrent Convolutional Neural Network U-Net</i>
RELU	<i>Rectified Linear Units</i>
STDCNN	<i>Semi-Transfer Deep Convolutional Neural Networks</i>
SVM	<i>Support Vector Machine</i>
TP	<i>True Positive</i>
VGG	<i>Visual Geometry Group</i>

SUMÁRIO

1	INTRODUÇÃO	16
1.1	JUSTIFICATIVA.....	18
1.2	HIPÓTESE	20
1.3	OBJETIVOS	20
1.3.1	Objetivo geral	20
1.3.2	Objetivos específicos.....	20
2	REVISÃO DE LITERATURA.....	22
2.1	DETECÇÃO E EXTRAÇÃO DE ESTRADAS	22
2.1.1	Semiautomáticos	23
2.1.2	Automáticos	24
2.2	MÉTODOS BASEADOS EM INTELIGÊNCIA ARTIFICIAL	27
2.2.1	Considerações.....	28
2.3	INTELIGÊNCIA ARTIFICIAL (IA).....	29
2.4	<i>MACHINE LEARNING</i> (ML)	30
2.5	<i>DEEP LEARNING</i> (DL).....	30
2.6	REDES NEURAIS ARTIFICIAIS.....	32
2.6.1	Redes Neurais Convolucionais.....	38
2.6.2	Redes neurais convolucionais (CNN) em Sensoriamento Remoto	41
3	MATERIAL E MÉTODOS.....	45
3.1	MATERIAL.....	45
3.1.1	Imagens PLANET	45
3.1.2	<i>Google Colab</i>	46
3.1.3	<i>Hardware</i>	46
3.1.3.1	<i>Pyzo</i>	47
3.2	MÉTODOS	47
3.2.1	Preparação dos recortes e rotulação.....	47
3.2.2	Conjuntos para Teste, Treinamento e Validação.....	48
3.2.3	<i>O conceito de Autoencoder</i>	49
3.2.4	U-Net	50
3.2.5	Parâmetros	52
3.2.6	Treinamento	53
3.2.7	Teste.....	55

3.2.8	Validação.....	55
3.2.9	Avaliação de resultados.....	55
4	EXPERIMENTOS	60
4.1	ÁREA DE ESTUDO	60
4.1	AQUISIÇÃO DE DADOS	61
4.2	SERIE DE EXPERIMENTO 1	63
4.3	SERIE DE EXPERIMENTO 2.....	63
4.4	SERIE DE EXPERIMENTO 3.....	63
5	RESULTADOS E DISCUSSÕES	65
5.1	SERIE DE EXPERIMENTO 1	65
5.2	SERIE DE EXPERIMENTO 2.....	69
5.3	SERIE DE EXPERIMENTO 3.....	73
5.4	COMPARAÇÃO DOS RESULTADOS DOS EXPERIMENTOS.....	76
5.5	AVALIAÇÃO DO MODELO	81
6	CONCLUSÃO.....	86
6.1	RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	87
	REFERÊNCIAS	88

1 INTRODUÇÃO

O acesso a serviços públicos é um direito da população que garante melhor qualidade de vida, saúde e trabalho. Em países continentais, como o Brasil, a expansão de serviços públicos encontra dificuldades, o que posterga o acesso aos mesmos por parte da população de regiões distantes aos centros urbanos, até mesmo quando se dispõe do serviço.

Pode-se tomar como exemplo o caso do acesso à energia elétrica. O fornecimento de energia elétrica é uma maneira do estado melhorar as condições de saúde, educação e acesso à informação da população em geral. Embora o Brasil seja um grande produtor de energia elétrica com base em aproveitamentos de seus recursos hídricos, o estado ainda encontra dificuldades para distribuir esta energia a todos os cantos do seu território, especialmente nas regiões rurais, em função das grandes distâncias e a baixa densidade de concentração populacional nessas áreas.

As empresas do setor elétrico aproveitam a infraestrutura preexistente para planejar e construir extensões das redes de distribuição. A rede elétrica, por exemplo, é ampliada acompanhando a rede viária, pois as estradas e ruas permitem o acesso à fiação e postes para fins de manutenção dentro de um espaço público. A falta de mapeamento de estradas vicinais impossibilita que concessionárias de energia elétrica planejem e implementem extensões da rede de transmissão de energia elétrica de forma adequada. Assim, pode-se afirmar que a existência o mapeamento atualizado da rede de estradas e rodovias influencia diretamente na qualidade de vida da população principalmente os habitantes de comunidades rurais.

Embora se disponha de mapas atualizados das principais estradas de rodagem, menor esforço é dedicado à atualização do mapeamento de estradas menores, o que dificulta sua manutenção, comprometendo o escoamento da produção e acesso a serviços públicos – saúde, água e esgoto tratados, escola, entre outros. De acordo com os dados oficiais da Confederação Nacional de Transporte (CNT, 2021) apenas 12,4% das estradas de rodagem nacional são pavimentadas, sendo cerca de 213.500,0 km, sua grande maioria, presente na região sul e sudeste do país, em comparação com 1.350.100,0 km de vias não pavimentadas no restante do território nacional.

O problema da manutenção e atualização de uma base de dados da malha viária nas áreas rurais é sério no Brasil, afetando diretamente várias áreas, como o

planejamento e distribuição de insumos médicos, o escoamento da produção agrícola, indo muito fundo na questão da qualidade de vida da população rural. Existem diferentes maneiras de resolver este problema, que vão desde levantamentos diretos e expeditos, como o uso de veículos dotados de antenas Sistema Global de Navegação por Satélite (GNSS) percorrendo as estradas, até o uso de ferramentas de sensoriamento remoto. Em termos de cobertura global, o avança em tecnologias de imageamento por sensoriamento remoto, principalmente as imagens obtidas por satélite, somadas a melhoria na qualidade de resolução espacial dessas imagens, vem propiciando o acesso a áreas mais remotas no globo, permitindo o mapeamento por exemplo de estradas em meio rural (CAZES et al., 2005; CHAMBON, 2011; CHENG et al., 2014).

Usar imagens de satélite para a extração de informações sobre o traçado de estradas em meio rural, é uma possibilidade viável, visto que atualmente há disponibilidade de imagens de alta resolução e de forma gratuita, o que reduz grande parte dos custos de levantamento de dados. Por exemplo, Anil e Natarajan (2010) utilizam da abordagem computacional para detectar estradas por meio da união entre pontos, a partir de imagem de alta resolução.

A utilização de imagens digitais permite o uso de algoritmos computacionais para a identificação de trechos de estradas, aplicando métodos de processamento digital de imagens e extração automática de informação. Por exemplo, tem-se o método proposto por Mendes, Fazan e Dal Poz (2004) para delinear estradas vicinais baseado na combinação entre uma técnica de extrapolação e de correlação. Outro exemplo foi proposto por Silva e Centeno (2010A) que usa métodos semiautomáticos usando algoritmos genéticos para a extração de estradas vicinais, a partir de imagens digitais.

O uso de métodos de aprendizado profundo vem ganhando espaço dentro da automação de extração de feições em imagens. Estes métodos se baseiam no uso de redes profundas capazes de “aprender por exemplos” para identificar objetos nas imagens. Isto pode facilitar o processo de extração de trechos de estradas em imagens, mas é evidente que, a variação de aparência de estradas em imagens é um problema a ser resolvido quando se deseja implementar métodos automáticos.

Neste projeto, pretende-se explorar o potencial das redes U-Net, um modelo de rede convolucional, para extração de estradas em regiões rurais a partir de imagens multiespectrais de média resolução, imagens da constelação Planet com 5m de

resolução espacial. Assim, pretende-se contribuir com o desenvolvimento de ferramentas para a atualização de mapas da malha viária como suporte para a expansão do suprimento de serviços público de forma geral para as áreas rurais. Os resultados desta pesquisa podem trazer diversos benefícios à população residente nas zonas rurais, pois podem contribuir para reduzir a falta de informações referentes sobre essas regiões e facilitar a manutenção e expansão de outros serviços públicos com o intuito de apoio à população rural.

1.1 JUSTIFICATIVA

Tradicionalmente, métodos de fotointerpretação restituição aerofotogramétrica foram utilizados para extração de informações a partir de imagens aéreas para a produção de bases cartográficas. Estes métodos, apesar de serem excelentes em termo de acurácia, são caros, o que dificulta ações de atualização massiva de cartas. A despesa com os levantamentos e o trabalho de interpretação visual é um dos principais motivadores para a falta de atualização cartográfica no Brasil. Como consequência, grande parte dos mapas topográficos que fazem parte do mapeamento sistemático são o resultado de levantamentos aéreos realizados entre as décadas dos anos 1970 a 1980.

No início da década de 1970, os primeiros sistemas de sensores instalados em satélites foram lançados, motivando assim o uso de imagens orbitais para mapear a superfície da Terra. Os primeiros produtos obtidos por imagens de satélite orbitais não eram apropriados ao mapeamento devido à sua baixa resolução espacial. Esta limitação foi progressivamente superada, pois foram desenvolvidos sensores capazes de obter imagens com cada vez maior resolução espacial.

Na atualidade, sistemas comerciais oferecem imagens de alta resolução espacial, alcançando resoluções abaixo de um metro. Com isto, tornou-se possível realizar o mapeamento mais detalhado da superfície do planeta, com riqueza de detalhes que possibilitam a identificação de objetos com dimensões cada vez menores. Isto melhorou a perspectiva da identificação de trechos de estradas em imagens, tanto em áreas rurais quanto urbanas.

A principal vantagem de se obter informações a partir de imagens é a redução de levantamentos de campo, o que conseqüentemente reduz os custos para a obtenção do produto final. Porém, processar dados de satélite necessita de mão de

obra especializada e de computadores mais potentes (SILVA e CENTENO, 2010a; STATELLA e SILVA, 2002).

A disponibilidade de imagens digitais abriu espaço para o desenvolvimento de métodos automáticos para detecção de estradas com ferramentas de processamento de imagens. Na última década, o potencial da aplicação de métodos de processamento de imagens foi aumentado com o uso de inteligência artificial, o que permitiu automatizar muitas partes do processo de mapeamento, diminuindo ainda mais os custos (DAL POZ, ZANIN e VALE, 2007; MENDES, FAZAN e DAL POZ, 2004). Metodologias baseadas em redes neurais permitem que os programas reconheçam padrões e características das estradas rurais e sejam capazes de extrair traçados de forma mais precisa. Com esta premissa, é esperado que o presente projeto de pesquisa contribua para a obtenção de mapas da malha viária de forma mais eficiente e rápida, que possibilite fornecer informações pertinentes a gestão pública de municípios.

Alguns algoritmos computacionais usam modelos de aprendizado supervisionado, como as redes neurais artificiais, para simular o funcionamento do cérebro humano e a interconexão entre neurônios do sistema nervoso. (MA et al., 2019). Redes Neurais podem ser combinadas para resolver problemas mais complexos. Por exemplo, as Redes Neurais Convolucionais são basicamente a combinação de redes neurais que permitem detectar automaticamente padrões nas imagens. De acordo com Ma et al. (2019), as redes convolucionais são soluções promissoras para o problema de detecção de objetos em imagens, pois usam recursos de aprendizagem complexos para otimizar a função discriminante do modelo de classificação. No entanto, esta é uma abordagem extremamente dependente da fase de aprendizagem (Liu et al., 2019).

Uma arquitetura de Redes Convolucionais que vem apresentando resultados promissores no campo da detecção e extração de estradas é a U-Net, como é descrito em Seenadera (2021). Outro exemplo do uso desta arquitetura para a extração de estradas é apresentado em Hou et al. (2021), os que usam uma estrutura complementar, a C-U-Net, que possui métodos adaptados a partir estrutura U-Net.

1.2 HIPÓTESE

A pesquisa se baseia nas seguintes hipóteses para sustentar a proposta do objetivo:

- Se na atualidade existem imagens de satélite disponibilizadas de forma gratuita e de acesso público, com resolução espacial suficiente para detectar e extrair informações sobre a malha viária em meio rural;
- E dado que para a análise de imagens foram recentemente criados métodos de extração e identificação de objetos com base na inteligência artificial, como o caso das redes convolucionais, que permitem simular o processo de identificação visual de objetos;
- Então, é possível desenvolver métodos de mapeamento da rede viária de estradas rurais a partir de imagens de satélite de alta resolução espacial aplicando algoritmos de inteligência artificial, como as redes convolucionais, com acurácia compatível com projetos de implantação de rede de distribuição elétrica.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Verificar a viabilidade de mapear estradas em ambiente rural em imagens da constelação Planet usando redes neurais convolucionais.

1.3.2 Objetivos específicos

- Avaliar o desempenho da rede U-Net para a detecção de estradas em imagens da constelação PLANET;
- Avaliar o desempenho da rede em relação aos elementos presentes na cena e as vantagens do uso de ambiente Google Colab como opção ao uso de desktop local;

- Avaliar a arquitetura da rede em relação aos parâmetros, batch size, quantidade de camadas, convolução e neurônios por camada, e a quantidade épocas de treinamento. Correlacionando-os com a detecção de estradas em imagens da constelação PLANET.

2 REVISÃO DE LITERATURA

Neste capítulo será apresentada uma revisão de trabalhos encontrados na literatura que servem como base para mostrar a viabilidade da detecção de estradas em imagens. Depois, se dará ênfase à solução deste problema com o uso de ferramentas de inteligência artificial. Também são apresentados exemplos de uso de redes neurais artificiais, redes neurais convolucionais e outras ferramentas de aprendizado profundo (*Deep Learning*) como soluções para detecção e extração de estradas.

2.1 DETECÇÃO E EXTRAÇÃO DE ESTRADAS

É possível detectar e extrair informações usando diferentes produtos de sensoriamento remoto com diferentes resoluções espaciais e espectrais ou até mesmo em dados LiDAR (*Light Detection And Ranging*). Cada fonte de dados demanda um tratamento diferente e pode gerar produtos com diferentes graus de acurácia. Na literatura pode ser encontrada uma variedade de métodos de detecção e extração de estradas a partir da análise de dados de sensoriamento remoto (CHAMBON, 2011; CHENG et al., 2014).

Tradicionalmente, o delineamento da rede de estradas era feito por análise visual de imagens, traçando o eixo da estrada na tela do computador. Este método, embora laborioso, permite obter uma boa aproximação da malha de rodagem, pois o ser humano é capaz de contornar vários problemas, como a oclusão da estrada por superposição de vegetação ou sombras e eventuais variações de cor decorrentes do uso de diferentes materiais (SILVA e CENTENO, 2010B). Além disso, o cérebro baseia sua análise não apenas na informação espectral, mas explora outros atributos da imagem, como a forma, tamanho e o contexto.

Os métodos clássicos de automação da detecção de estradas se baseiam na hipótese de que elas contrastam com o fundo e que, por esse motivo, podem ser identificadas ao detectar bordas nas imagens. Existem várias formas de realizar a detecção de bordas nas imagens digitais, sendo as mais comuns aquelas baseadas na filtragem espacial, pela convolução de uma janela móvel e a imagem. A detecção de bordas por filtragem pode ser obtida com operadores clássicos como Sobel, Prewitt e Canny, que foram desenvolvidos para detecção de descontinuidades na imagem. A

detecção de bordas por convolução de janelas móveis, na literatura, é apresentada pelos operadores de Gradiente e Laplaciano (AHMED, 2018).

Extraír estradas de imagens é um processo complexo e pode ser um desafio principalmente devido ao grande volume de informações contidas nas imagens. Por isso é preciso realizar o preparo das imagens a ser processadas de forma a minimizar a interferência de bandas ruidosas e separar informações de acordo com suas características espectrais e espaciais.

Em sua revisão a respeito de métodos de extração de estradas os autores Lian et al. (2020) apontam que grande parte dos métodos heurísticos baseados em ferramentas de processamento de imagens na literatura podem ser agrupados em dois grandes grupos, segundo o grau de interação necessário: semiautomáticos e automáticos.

2.1.1 Semiautomáticos

Os métodos semiautomáticos demandam a interação humana para definir parâmetros ou ajustar o processo. Por exemplo, em alguns deles é necessário apontar informação inicial, como por exemplo o início e fim de um trecho de estrada, e os algoritmos se encarregam de detectar a melhor aproximação de uma estrada que une estes dois pontos (ANIL e NATARAJAN, 2010; GALLIS et al., 2005; LIAN e HUANG, 2020). Após a extração, o usuário deve avaliar os resultados e muitas vezes editar o produto. No caso dos métodos totalmente automáticos, apenas alguns parâmetros globais são definidos em função da escala e todo fluxo de trabalho é executado sem supervisão humana (DAL POZ, ZANIN e VALE, 2007). Entre os métodos semiautomáticos apresentados na literatura encontram-se: modelos de contornos ativos (*Active Contour Model*), caminho geodésico (*Geodesic Path*), programação dinâmica (*Dynamic Programming*), e morfologia matemática.

O modelo *Active Contour* foi originalmente proposto por (KASS, WITKIN e TERZOPOULOS, 1988) e é comumente chamado de *snake*. Este método se baseia na definição de uma função de energia em um trecho de contorno desenhado pelo usuário. Esta informação é usada para aproximar o contorno da estrada procurando o valor mínimo da função de energia. Este método é utilizado na detecção de bordas, por isso pode ser também aplicado para a detecção de trechos de estradas (ANIL, NATARAJAN, 2010; MEDHI, SAHA, 2019).

O conceito de *Geodesic Path* é usado nos métodos nos que o usuário especifica dois pontos na imagem, definindo o início e fim do trecho de estrada, e o algoritmos aproximam a melhor curva que liga estes dois pontos, usando uma curva suave e procurando minimizar a variação dos valores digitais ao longo deste percurso. Um exemplo de aplicação é encontrado em Chambon, (2011) que aplica este método para detectar fissuras em estradas. Outro exemplo por Miao et al. (2014) onde o método é utilizado para extrair os segmentos rodoviários e ligar as sementes caracterizadas pelo usuário.

Métodos baseados nos princípios da programação dinâmica também podem ser aplicados para aproximar estradas em imagens, como é mostrado por Gallis et al. (2005). Neste caso, o usuário identifica na imagem pontos sementes que pertencem às estradas e o algoritmo procura aproximar a geometria da estrada usando técnicas de otimização que garantem o menor, ou menos custoso, percurso que liga estas sementes (GRUEN e LI, 1995)

Outro método semiautomático baseado em ferramentas de processamento de imagens é o uso de morfologia matemática (SILVA et al., 2009; STATELLA, SILVA, 2002). A imagem é inicialmente binarizada e as bordas correspondentes aos objetos são segmentadas. Logo, operadores matemáticos são usados para eliminar segmentos espúrios e preservar os maiores, aqueles que conformam a rede viária. Pelo apontado pelos autores Statella e Silva (2002) o uso da morfologia matemática depende da escolha adequada de dois principais fatores: o elemento estruturante e alguns limiares. Isto faz com que o método seja altamente dependente da decisão do usuário, quem deve adaptar estes parâmetros para cada caso de imagem, tornando esta opção mais difícil de ser usada em larga escala. A maior problemática para esse método está na interação entre os elementos das estradas e os elementos presentes na cena, não sendo possível separar a estrada de outros elementos com base na imagem binária sem necessitar informações adicionais (MENDES, FAZAN e POZ, 2004; STATELLA e SILVA, 2002).

2.1.2 Automáticos

Os métodos semiautomáticos de extração de estradas dependem muito da intervenção humana, o que influencia diretamente na sua eficiência. Por outro lado, os métodos automáticos são executados sem intervenção humana, produzindo

resultados similares para todos os tipos de imagens de sensoriamento remoto. Na literatura alguns exemplos de métodos automáticos como os métodos baseados em análise da vizinhança, segmentação (*Methods Based on Segmentation*), análise de borda (*Edge Analysis Methods*), métodos envolvendo inteligência artificial e *machine learning* (ML) como métodos baseado em inteligência de enxame (*Swarm Intelligence-Based Methods*), métodos baseados em objetos (*Object-Based Methods*), métodos de segmentação de imagens multiespectral (*Multispectral Segmentation Methods*).

O método baseado em janelas móveis podem ser aqueles baseados em análise de correspondência (template matching) ou derivação de parâmetros espaciais da região em torno do pixel. A análise de correspondência (*template matching*) parte da hipótese que se a forma dos trechos de uma estrada segue padrões definidos, esses padrões podem ser modelados com o uso de janelas móveis contendo modelos de trechos de estradas. Assumindo que a geometria das estradas é relativamente simples, dependendo da resolução da imagem, é considerado que o número de janelas para sua identificação é baixo (SILVA e CENTENO, 2009).

Existem outros métodos, também baseados na análise da vizinhança de um pixel usando janelas móveis, como por exemplo o uso da transformada de Radon. Neste caso, a imagem é varrida usando uma janela de tamanho ajustado à dimensão da estrada em função da resolução espacial da imagem e nesta janela é analisada a variação dos valores digitais e sua orientação, como maneira de detectar padrões lineares, no entanto, é um método sensível à presença de obstruções nas estradas avaliadas (SILVA e CENTENO, 2010C). Outro exemplo da aplicação de janelas moveis é apresentado pelos autores Lian e Huang, (2020) que aplica o método para a extração da linha central da estrada baseados na segmentação semântica, empregando uma rede neural profunda, no entanto observa-se que o método não consegue extrair os segmentos de estradas quando a presença de ruídos nas imagens.

A detecção de estradas com o método baseado em inteligência de enxame é baseada na resolução de problemas utilizando algoritmos baseados no comportamento coletivo auto-organizado dos insetos sociais, como é o caso das abelhas (ZEDADRA et al., 2018). No entanto, o próprio algoritmo tem algumas falhas, sendo difícil encontrar uma constante adequada como peso, onde o peso inadequado pode fazer com que o algoritmo tenha oscilações na proximidade da solução ótima e

podem causar uma partícula a alocar-se no extremo local, assim o algoritmo terá problemas para convergir ou ir convergir para os mínimos locais (LI, HE, WEN, 2015).

Quando se trata de imagens multiespectrais os métodos de segmentação tendem a fundir mais de uma técnica durante seu processo, como é o caso de (WAN et al., 2007) que usa a lógica Fuzzy, e o algoritmo baseado na extração do esqueleto morfológico para extração de estradas em imagens multiespectrais. Outra opção é segmentar a imagem em regiões potencialmente contendo estradas e depois classificar estas regiões considerando atributos espaciais e espectrais, por exemplo, usando a abordagem OBIA (Object Based Image Analysis). Um exemplo é encontrado em Hossain, Chen, (2019). A abordagem OBIA inicia tentando detectar regiões uniformes nas imagens, supostamente correspondentes aos objetos nela presentes, ou pelo menos partes deles. A partir disto, a análise não se foca no pixel, mas sim na região ou segmento, o que muda a maneira de analisar a imagem pois podem incluir, além dos dados espectrais, parâmetros espaciais dos segmentos como a forma, tamanho ou textura, entre muitos outros. Estas informações são usadas na fase de classificação. No entanto, a aplicação de OBIA não é imediata para detectar estradas, pois alguns segmentos correspondentes a estradas encontram-se desconectados, além disso, há a necessidade de treinar caso a caso, o que limita a sua aplicabilidade prática (MIAO et al., 2015; MEDHI, SAHA, 2019).

Outro método aplicado a extração de estradas é o método de grafos, que consiste em uma etapa de pós-processamento que vetorizar a segmentação de estradas prevista pela CNN. Para os autores Tan et al., (2020) o método de pós-processamento consiste em binarizar a imagem a partir de um limiar ideal e com as características morfológica obter o esqueleto rodoviário contido em cada pixel. Em geral, o algoritmo de grafos utiliza um limiar para toda a imagem, no entanto, a utilização de um único limiar causa a sub-segmentação das regiões mesmo que o valor estabelecido seja o ideal, também encontra problemas sob oclusões severas (CHENG et al., 2014; YUAN et al., 2014; CHENG, HAN, 2016).

Uma vez detectados os pontos de uma estrada, eles devem ser interligados para compor a malha viária. Para descrever o trecho de estrada podem ser usadas diferentes informações, como cor, intensidade ou textura. Uma variação deste método é o uso da transformada Randon, como é descrito em (SILVA e CENTENO, 2010A) que usou esta abordagem para detectar estradas a partir de dados lidar.

2.2 MÉTODOS BASEADOS EM INTELIGÊNCIA ARTIFICIAL

Mais recentemente, os avanços de informática permitiram o desenvolvimento de métodos que simulam parte do raciocínio humano com computadores. Entre eles estão as redes neurais, suporte vector machine, Random forest e os métodos de aprendizado profundo.

Os métodos de SVM (*Support Vector Machine*) podem ser usados para detectar segmentos lineares de potenciais trajetórias de estradas. Para isto, são analisados o contraste espectral e a trajetória linear local, o que permite aproximar a estrutura das estradas. No entanto o resultado é viável apenas quando existem poucos obstáculos, quando obstáculos for como árvores grandes, veículos grandes e edifícios altos, não apresenta bons resultados (DAS, MIRNALINEE, VARGHESE, 2011). Os autores ainda observam que a confusão entre as respostas das classes de Edifícios, estacionamento e entroncamentos de estradas, devido às características semelhantes de textura, é necessário a combinação da segmentação com características de textura e características topológicas (WAN et al., 2007).

Outras metodologias como a modelagem estatística CRF (*Conditional Random Field*), obtém resultados expressivos na acurácia da segmentação e correção topológica, porém este método necessita da entrada manual de parâmetros de amostragem não sendo capaz de apreender automaticamente (WEGNER, MONTOYA-ZEGARRA, SCHINDLER, 2013).

A extração de estradas, usando uma metodologia automática pode ser obtida pelo método do *Random Forest*, método esse que consiste em um conjunto de árvores de decisão binárias. Com base no *random forest* uma imagem de entrada é dividida em pequenas regiões do mesmo tamanho, e cada região é representada por um rótulo de entrada no classificador. Contudo a maior parte do erro de classificação é causado pela suposição incorreta de diferença de cor entre regiões da estrada e o ambiente de fundo (CHOI, SONG, LEE, 2012).

Uma categoria de métodos é classificada como Data-Driven, pois neste caso a solução é obtida analisando amostras dos próprios dados para deduzir as regras de decisão. Exemplos de métodos automáticos Data-Driven são aqueles que se baseiam nas redes neurais, com a aprendizagem profunda de máquinas, sendo divididos de acordo com a arquitetura da rede neural. Apesar de obter resultados promissores, estes métodos ainda apresentam problemas de incompatibilidade entre o tamanho do

grupo de dados e o tamanho da rede neural, ou seja, quando há um grande volume de dados e uma rede neural pequena o processo tende a falhar (LIAN et al., 2020; REICHSTEIN et al., 2019). Outro exemplo é apresentado por Lian, Huang, (2020) que aplicam a segmentação semântica para a extração da linha central da estrada, empregando uma rede neural profunda. No entanto observa-se que o método não consegue extrair os segmentos de estradas quando a presença de ruídos nas imagens é alta.

Redes convolucionais também podem ser usadas para mapear estradas em imagens. De acordo com Zhang, Zhou e Lin (2018), o método FCN (Fully Convolutional Network) foi inicialmente desenvolvido para a segmentação semântica de imagens, baseado no modelo VGG (*Visual Geometry Group*) devido ao seu bom desempenho e elevada eficiência. De acordo com os autores Zhang, Zhang, Du, (2016) a cada camada de convolução consegue prever não apenas o objeto de interesse do processo como também a localização espacial e os detalhes do contorno. O método emprega a aprendizagem profunda para o treinamento de ponta a ponta, onde é combinada com campos aleatórios condicionais e pós-processado para melhorar o desempenho da segmentação (DUNG, ANH, 2019).

Para a detecção de estradas em imagens, é necessário realizar a segmentação da imagem como parte do processo, o que pode ser atingido usando uma rede GAN. O modelo GAN (*Generative Adversarial Networks*) é uma CNN (*Convolutional Neural Network*) que usa a segmentação como parte de seu processo. Um exemplo de aplicação do modelo GAN para gerar e discriminar os dados referentes a detecção de estradas pode ser visto no trabalho de Abdollahi et al. (2020).

2.2.1 Considerações

Nota-se um avanço na proposta de soluções conforme novas ferramentas de processamento e análise de imagens são desenvolvidas. Inicialmente, um modelo era proposto pelo analista, o analista também era responsável por definir os parâmetros ou descritores espaciais e espectrais que deveriam ser derivados nas imagens, como por exemplo variações de tonalidade ou bordas dos objetos.

Na atualidade, a tendência é o uso de ferramentas de inteligência artificial, capazes de detectar automaticamente padrões através de exemplos e estes padrões são usados para desenvolver um modelo implícito da estrada, que é usado no

processo de detecção. Por este motivo, neste trabalho será usado um método baseado em redes convolucionais para detectar estradas em imagens de alta resolução espacial.

A seguir é apresentada uma breve revisão dos conceitos fundamentais de redes convolucionais para depois apresentar a proposta da metodologia.

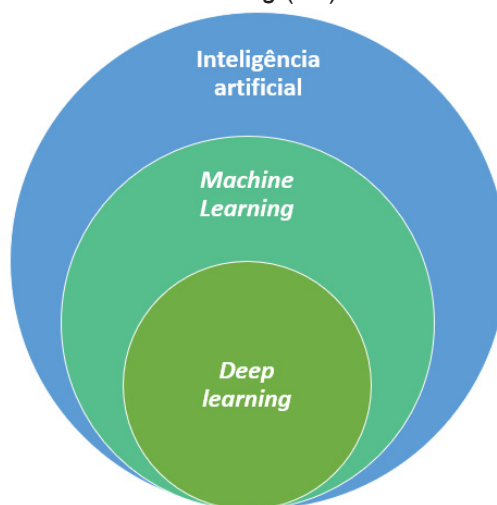
2.3 INTELIGÊNCIA ARTIFICIAL (IA)

O desenvolvimento de tecnologias e o avanço nas áreas de pesquisas computacionais, tem impulsionado o uso de sistemas totalmente autônomos. Considerar que uma máquina possa obedecer a um comando com poucas informações é tentador. Para criar máquinas independentes ou que necessitem de poucas interações humanas, foi necessário desenvolver algoritmos que pudessem replicar o sistema nervoso humano e sua forma de aprendizado. A inteligência artificial IA, é um campo de pesquisa estabelecido com base na máquina ser capaz de aprender espelhando-se no comportamento humano.

A forma mais básica da IA são estruturas programadas para “imitar” o comportamento humano usando uma ampla gama de informações. Eles podem variar desde espelhar o comportamento de colônias de abelhas, como o caso de algoritmos de detecção *bee colony algorithm* (OZKAN et al., 2011) a detecção de estruturas complexas como estradas a partir de imagens de satélite (MOKHTARZADE, ZOEJ, 2007).

O desenvolvimento de métodos de IA tem possibilitado novas áreas de conhecimento como a área do *machine learning* (ML), termo traduzido para o português como Aprendizado de Maquinas e com a subárea de conhecimento do ML, o *Deep Learning* (DL), traduzido diretamente para o português como aprendizado profundo. Na Figura 1, compreende-se melhor a relação entre as áreas da inteligência artificial.

FIGURA 1 – Esquema de correlação entre inteligência artificial (IA), *machine learning* (ML) e *Deep learning* (DL).



Fonte: Adaptado de Sharma, Sharma e Jindal (2021).

2.4 MACHINE LEARNING (ML)

A aprendizagem de máquinas é um campo dentro da ciência da computação para permitir que os computadores aprendam a tomar decisões sem a necessidade de programar explicitamente as regras, ou seja, o sistema deve aprender os parâmetros básicos, com base em informações mais amplas sobre determinado problema (MEI, DOOMERNIK, 2018). Métodos tradicionais de *machine learning* levam em consideração em suas aplicações poucas interferências humanas e considerável complexidade dos dados analisados (SHEN et al., 2020). São exemplos o algoritmo de *Random Forest* e SVM (CHOI, SONG, LEE, 2012; FAKHRI, SHAH-HOSSEINI, 2022).

2.5 DEEP LEARNING (DL)

Deep learning é a área de machine learning desenvolvida a partir da combinação de redes neurais para formar redes mais complexas, empilhadas, que permitem uma representação mais profunda do problema. Os métodos de deep learning são aplicados no reconhecimento de objetos em imagens e na classificação (CAZES et al., 2005), mas eles têm aplicação em diferentes campos da ciência (RONNEBERGER, FISCHER e BROX, 2015 e SILVEIRA, SHIGUEMORI, GOMES, 2016).

Métodos de *Deep learning* podem ser aplicados dentro de esquemas supervisionados ou não supervisionados, dependendo da finalidade do processo. A maior vantagem no uso *Deep learning* é a extração de feições pela máquina, não sendo necessário que o usuário proponha o cálculo de determinadas feições ou selecione as mais relevantes. Isto permite solucionar problemas complexos com menor participação humana, com aumento da precisão de classificação e redução do erro em problemas de regressão, desde que se conte com um conjunto de dados é volumoso, suficiente para identificar os descritores mais relevantes (KAMILARIS, PRENAFETA-BOLDÚ, 2018).

A estrutura do modelo que funciona como uma rede e a capacidade de aprendizagem estabelecido por modelos de *Deep learning* é adaptável e flexível por isso pode ser aplicado a diversas áreas: sistemas autônomos para reposicionamento de plataformas/robôs (LUNA et al., 2018), agricultura (KAMILARIS, PRENAFETA-BOLDÚ, 2018), e determinação da qualidade de papel reciclado (FLAMIA et al., 2021) são alguns exemplos de áreas onde o *Deep learning* já aplicados.

O aprendizado profundo é um modelo funcional que percorre os dados de entrada camada a camada estruturando uma rede de operações sequenciais, onde cada camada executa uma operação particular na matriz de dados. Esse processo ocorre em cadeia, sendo a saída de dados de uma camada, a entrada da camada subsequente, quando o processamento passa pela camada de saída, resulta em um recurso ou uma classificação de saída (CHEN, RAN, 2019).

Em problemas de classificação da cobertura do solo com imagens de satélite, muitas vezes a complexidade das imagens impede a aplicabilidade de métodos clássicos de classificação estatística, como os classificadores pixel a pixel ou de segmentos (VENTURIERI, ROBERTO, 2005; SILVEIRA, SHIGUEMORI, GOMES, 2016). A descrição do elemento de classificação, seja pixel ou região, pode ser uma tarefa complexa, especialmente quando se lida com imagens de maior resolução espacial. Nestes casos, as ferramentas do *Deep learning* para classificação têm vantagens significativas em termo de acurácia de classificação especialmente em áreas urbanas onde a complexidade de informações sobre o solo aumenta (YUAN et al., 2020). Assim como Guidici, Clark, (2017) esclarece que o *Deep learning* proporciona técnicas de classificação com precisão e acurácia, sendo capaz de classificar a cobertura do solo em imagens hiperespectrais.

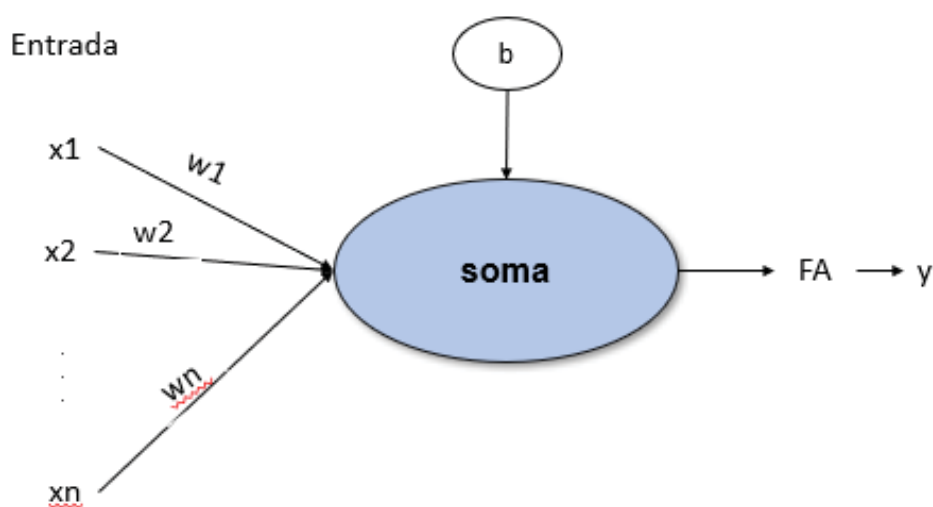
Uma rede profunda funciona seguindo os mesmos princípios de uma rede neural artificial, por isso, esse conceito é explicado a seguir.

2.6 REDES NEURAIS ARTIFICIAIS

Uma das formas iniciais de simular o funcionamento do cérebro com computadores foi a proposta do neurônio matemático e sua combinação para formar redes neurais artificiais. Redes neurais artificiais, termo advindo do inglês neural network, são algoritmos de inteligência artificial que objetivam emular o funcionamento do sistema nervoso humano e principalmente o fluxo de informações no cérebro. Para isto, eles possuem uma estrutura sequencial de neurônios artificiais, esses são estrutura matemáticas composta por funções que artificialmente simula o comportamento do neural biológico.

Quando se pensa no neurônio humano, como os autores Mei, Doomernik, (2018) descrevem, sua função mais básica é receber uma ou mais entradas, combinar estas entradas para produzir um pulso de saída excitatório ou inibidor que pode ser repassado a outro, ou outros, neurônios. O agrupamento de neurônios matemático, segue o mesmo princípio, ocorrendo a entrada de dados, e dentro do neurônio realiza-se cálculos matemáticos para obter uma saída, como apresentado na Figura 2.

FIGURA 2 – Esquematização do neurônio artificial.



Fonte: Adaptado de Silveira, Shiguemori e Gomes (2016).

Na prática, o modelo do neurônio matemático, com “n” entradas, pode ser formalizado segundo as Equações 1 e 2, onde “x” são as entradas, “y”, a saída, “p” os pesos de cada entrada e “b” um bias.

$$y = FA(S) \quad (1)$$

O modelo mostra que as entradas são combinadas linearmente, sob forma de um polinômio de primeiro grau, para produzir uma saída que será modulada pela função de ativação “FA”.

$$S = b + \sum_{i=1}^n p_i \cdot x_i \quad (2)$$

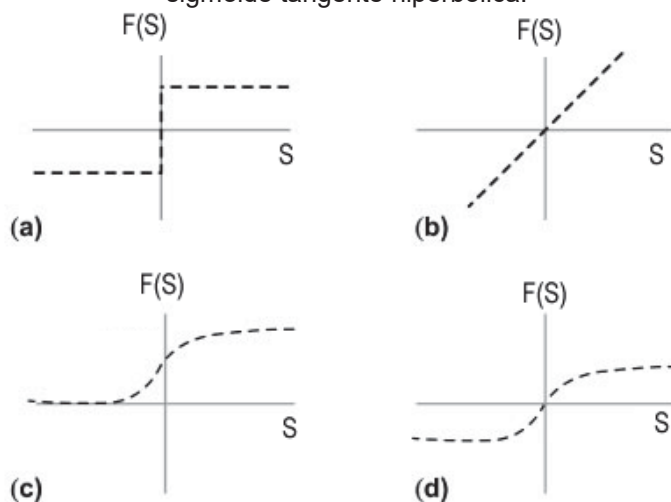
A expressão da Equação 2 pode ser reformulada incluindo as bias no somatório, sendo ele dependente de um novo peso que multiplica um valor 1. Neste caso, $b = p_{i+1} * 1$. Com isto, o somatório pode ser modificado para a Equação 3.

$$S = \sum_{i=1}^n p_i \cdot x_i \quad (3)$$

Nota-se aqui que os pesos servem para ponderar a influência de cada entrada de forma a se obter a saída desejada.

A função de modulação, também chamada de função de ativação, é usada para formatar a saída e alterar sua escala para valores que podem ser entendidos pelos seguintes neurônios artificiais da rede ou para adequá-la à saída desejada. De acordo com Fiorin et al. (2011) funções de ativação mais comumente utilizadas são a função limiar, linear, sigmoide logística e tangente hiperbólica.

FIGURA 3 – Exemplos de funções de ativação: A) limiar (degrau); B) linear; C) sigmoide logística, D) sigmoide tangente hiperbólica.



Fonte: adaptado de Fiorin et al. (2011).

As funções de ativação são basicamente as responsáveis pela funcionalidade do neurônio artificial, ou seja, as funções de ativação são as responsáveis pela ativação ou não do neurônio.

A função limiar, definida na Equação 4, tem variação de valores para a saída do neurônio (γ_j) no intervalo entre -1 e 1. O valor de γ é ajustado pelo bias para adequar-se ao melhor valor de nível de atividade interna do neurônio (v_j) com isso obtém-se a saída do neurônio.

$$FA = \begin{cases} +\gamma & \text{se } v_j \geq 0 \\ -\gamma & \text{se } v_j < 0 \end{cases} \quad (4)$$

A Função linear, apresentada na Equação 5 segundo Fiorin et al. (2011) é definida pelo parâmetro α e modifica a saída de forma linear (“rampa”) usando uma linha que passa pela origem.

$$FA = \alpha * V_j \quad (5)$$

De acordo com Fiorin et al. (2011) as funções sigmóides são as funções mais empregadas nas camadas internas de uma rede neural de múltiplas camadas, por serem contínuas, crescentes, diferenciáveis e não-lineares. A função sigmoide logística, apresentada na Equação 6, tem intervalo geralmente variando no intervalo entre 0 e 1.

$$FA = \frac{1}{1 + \exp(-\alpha * v_j)} \quad (6)$$

A função sigmoide tangente hiperbólica (tgh) apresentada na Equação 7, tem o mesmo intervalo que a função limiar, variando no intervalo de -1 a 1.

$$FA = tgh \frac{v_j}{2} = \frac{1 - \exp(-v_j)}{1 + \exp(-v_j)} \quad (7)$$

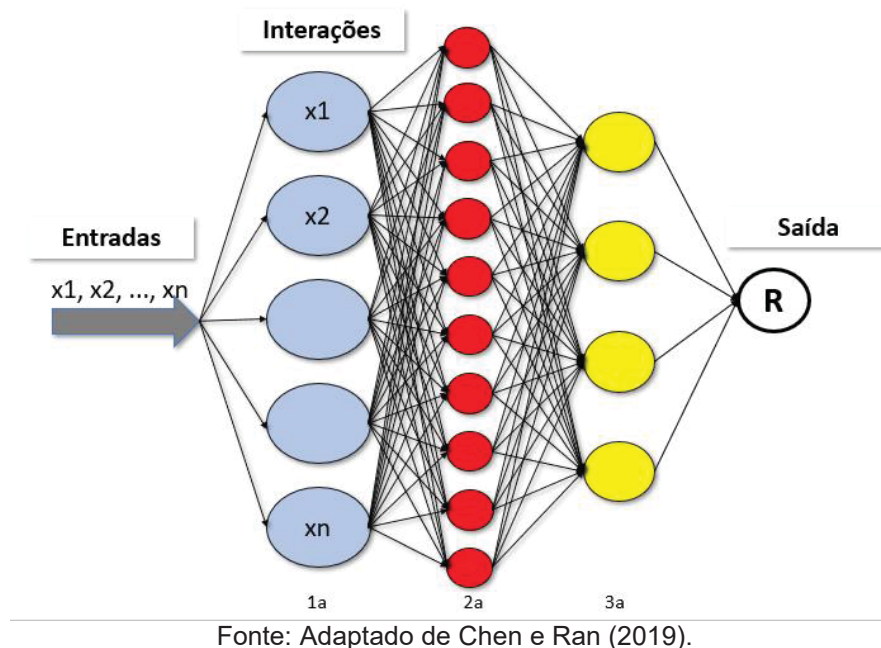
A variação de estas funções pode ser vista na Figura 3. Uma função de ativação com base em um limiar se assemelha a um degrau e produzira apenas valores binários. Variações contínuas são obtidas com as outras funções

Um neurônio isolado não oferece muita capacidade de processamento, por isso, vários neurônios são usados para resolver problemas complexos. Tradicionalmente, os neurônios são organizados em camadas. Cada camada vários neurônios que funcionam em paralelo.

A rede de neurônios, organizada em camadas sequenciais, permite realizar conexões entre neurônios e propagar os efeitos das entradas ao longo da rede, produzindo novas formas de representar as entradas na medida em que se aprofunda na rede, até produzir uma saída. Através do ajuste do processamento efetuado em cada neurônio é possível adaptar a rede de forma que a saída seja compatível com um objetivo desejado. Na Figura 4 a seguir ilustra a estrutura de uma rede neural composta de três camadas, sendo estas a camada de entrada, saída e camada oculta.

Na Figura 4 é mostrado um exemplo de uma rede neural onde se reconhecem três camadas. A primeira camada, serve apenas para receber os dados de entrada, geralmente normalizados previamente. Estes valores são repassados à segunda camada, dita oculta, que no exemplo é composta por 10 neurônios.

FIGURA 4 – Modelo esquemático de uma rede neural de três camadas, com camada de entrada, saída e camada oculta.



Em cada neurônio, as entradas são ponderadas e somadas gerando-se uma saída que é repassada a todos os neurônios da terceira camada. A terceira camada usa como entradas os resultados da segunda camada e produz um novo sinal que é apresentado como resultado. Na camada de saída, cada neurônio pode descrever um aspecto do problema.

Por exemplo, no caso de uma classificação, a camada de entrada receberá os valores das variáveis de entrada, as quais serão combinadas nas camadas escondidas e será gerada uma saída apresentada na camada de saída. Cada neurônio desta última camada pode representar a probabilidade de uma determinada classe calculada a partir das entradas. A segunda camada é chamada de oculta porque ela é imperceptível ao usuário, quem visualiza a camada de entrada e a de saída apenas. A camada oculta pode ser substituída por uma série de várias camadas, onde cada camada tem uma função específica e com quantidade diferente de neurônios, sendo cada neurônio responsável por cada função de ativação diferentes.

A principal problemática em uma rede neural consiste em determinar os pesos em cada neurônio. Uma grande quantidade de pesos pode estar sendo usada em uma rede com várias camadas. A determinação do melhor conjunto de pesos é basicamente um problema de otimização e é solucionada dentro de um processo iterativo. Para isto, dados de entrada com saída conhecida são usados em uma fase de treinamento da rede, quando se ajustam os pesos, desconhecidos no início.

Mesmo usando um conjunto de pesos aleatórios, para cada entrada é possível calcular uma saída, mesmo que seja errada. Com base na saída produzida é possível determinar o erro cometido com a rede comparando a saída com o valor correto. Este erro é usado para corrigir os pesos dentro da rede, repassado aos neurônios parte do erro levando em consideração A) o tamanho da entrada correspondente e B) o erro cometido.

Para isto, parte-se da última camada (saída) e se propaga o erro ao longo da rede de trás para frente até se chegar na primeira camada. Este processo é conhecido como a retro propagação do erro (*backpropagation*) e é repetido para todas as entradas, dentro de um processo iterativo, até se atingir um erro mínimo tolerável para todas as entradas. Teoricamente, o erro tolerado poderia ser fixado em zero, porém na prática adota-se valores baixos próximos de zero.

O processo de alimentar a rede com amostras de entrada e saída conhecidas para ajustar os pesos da rede é conhecido como a fase de treinamento. Em uma etapa posterior, uma vez encontrado o melhor conjunto de pesos, a rede treinada pode ser usada para predizer novas saídas a partir de um novo conjunto de entrada, mesmo que não tenha sido usado na fase de treinamento.

Mei e Doornik (2018) comentam que um grande volume de neurônios é capaz de, a partir de determinada quantidade de amostras de treinamento, detectar e reconhecer padrões, classificar imagens, jogar, criar sugestões de busca entre outros. Sua aplicabilidade em diversos campos se deve à sua funcionalidade operacional. Silveira, Shiguemori e Gomes, (2016) expõem que “o processo de funcionamento de uma rede neural contém duas etapas, o treinamento e a operação”. O treinamento de uma rede tem o intuito de ensinar ela, através de amostras de treinamento, a detectar certo padrão.

As redes neurais artificiais são usadas, geralmente, para resolver problemas de regressão e classificação. Porém, elas oferecem muitas variações que abrem um grande leque de possibilidades, como o desenvolvimento de redes neurais convolucionais e autoencoders.

2.6.1 Redes Neurais Convolucionais

As redes neurais convolucionais termo traduzido do termo em inglês *Convolutional Neural network* (CNN). Consiste em uma série de operações convolucionais, operações pooling e operações não lineares. De acordo com Cheng et al. (2017) e Sharma et al. (2017) as características dos modelos básicos da CNN necessárias para a execução da arquitetura, são:

- **Operações de convolução:** conecta um neurônio há uma pequena região (*Receptive field*), fazendo a ligação entre múltiplas operações de ativação para a camada anterior, obtendo um único neurônio de saída, e computando o produto resultante entre os pesos e o campo receptivo da camada anterior.
- **Conexão local entre camadas:** A conexão da rede CNN ocorre de forma local, Composta por neurónios nas camadas mais ocultas conectado unicamente a um campo receptivo de um neurônio da segunda camada cobrindo uma área maior em termos de entrada, do que um neurónio relacionado na primeira camada, a CNN tem mais camada do que outras estrutura de redes neurais convencionais, podendo modelar dependências de longo alcance de forma mais eficiente e consequentemente sendo necessários menos parâmetros e menos cálculos.
- **Pesos de ligação entre neurónios:** neurónios pertencentes a um determinado mapa de características compartilham os mesmos pesos de ligação que ocupam todo o campo receptivo dos neurónios, assim como as operações e filtros pertencente a camada. Os pesos de ligação compartilhado entre os neurónios pertencem ao mapa de características, reduz significativamente o número de parâmetros, e o mapa de características usa um único filtro para extrair a mesma característica em todas as respectivas camadas.
- **Operações de normalização de lotes:** operações de normalização, são executadas para manter as características de entrada e realiza a distribuição em mini lotes pela camada, aplicando a distribuição Gaussiana com media zero

e variâncias unitária, acarretando em aceleração no processo de aprendizado maior aproveitamento da etapa de aprendizado a CNN.

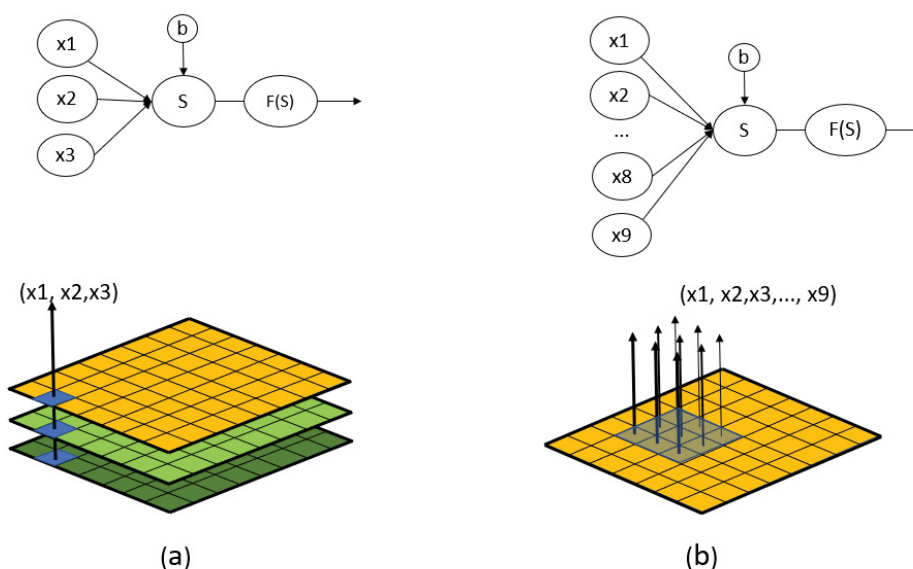
- **Operações/funções de ativação:** além das funções de ativação apresentadas no capítulo 2.7 – Redes Neurais Artificiais – uma outra operação comumente usada para arquiteturas CNN trata-se da *Rectified Linear Units* (Relu, traduzido do inglês como Unidade Linear Retificada), por usar um elemento de ativação em conjunto há um limiar não positivo assumindo o valor de zero, consequentemente fazendo com que os valores positivos permaneçam inalterados.
- **Operações de pooling:** operações de pooling são executadas em sub amostragem espacial em conjuntos não sobrepostos no mapa de características, em que um conjunto de neurônios vizinhos são substituídos por um único neurônio. Os neurônios gerados pelas operações de pooling substituem o conjunto de neurônios vizinhos na camada oculta seguinte, o que reduz o número de neurônios.
- **Operações de unpooling:** como o próprio nome sugere, são operações reversas as operações de pooling, nessas operações as conexões da rede são utilizadas para reconstruir as características de acordo com a sua localização original no mapa de características.
- **Operação de convolução:** as operações de convolução são executadas após realizar as operações de unpooling, os filtros armazenados durante o processo de aprendizado da rede nas operações anteriores são carregados e usados para conectar uma única operação de ativação de entrada da camada anterior há múltiplos neurônios de saída, computando o produto entre os pesos e o campo receptivo pertencente a camada anterior, retornando sua localização original no mapa de características.
- **Operação de limiar (*Threshold*):** estabelece o limiar para separação dos elementos do mapa de características, considerando que tudo que for

detectado acima do limiar será interpretado com o valor de um e o restante recebe o valor de zero.

- **Quantidade de camadas:** as camadas na estrutura de uma CNN têm grande importância para resolução de problemas simples ou de grande complexidade. Redes mais simples/rasas não conseguem resolver problema de alta complexidade. Observado por Cheng et al. (2017) e Rodrigues et al. (2022) uma maior a quantidade de camadas mais profundas possibilitam resolver problemas de alta complexidade devido a sua estrutura hierárquica.

Ao aplicar uma CNN a uma imagem, as entradas podem ser os valores digitais de um pixel em diferentes bandas espectrais, como é mostrado na Figura 5a. Neste caso, os pesos são aplicados a valores digitais medidos em diferentes regiões espectrais, o que pode ser uma solução para o problema de classificação multiespectral.

FIGURA 5 – Aplicação do conceito de neurônio a imagens digitais (A) processamento de um pixel com várias bandas. (B) processamento de uma região em uma banda.



Fonte: A autora (2023).

Outra opção consiste em usar como entradas os valores digitais de uma vizinhança, em uma mesma banda, para detectar padrões espaciais decorrentes da variação dos valores digitais (Figura 5B). Por exemplo, a ocorrência de um certo

arranjo de valores altos e baixos pode caracterizar a ocorrência de uma borda e assim a rede ser treinada para detectar bordas.

Esta capacidade foi notada e aplicada no processamento de imagens para, a partir de um conjunto de amostras, treinar uma rede para detectar padrões espaciais. Podem ser usados vários neurônios, cada um deles treinado para detectar um certo padrão. Neste caso, fala-se em uma camada convolucional, pois se trata de um conjunto de neurônios que detectam padrões analisando a vizinhança do pixel.

O resultado da aplicação de uma camada convolucional é um novo conjunto de imagens, cada uma delas contendo a informação relacionada a um determinado padrão básico. Aplicando-se uma camada convolucional podem ser detectados padrões simples, mas o potencial de uma rede convolucional aumenta quando estes resultados são usados como entradas para novas camadas convolucionais que podem derivar padrões mais complexos a partir deles.

O processo de varrer uma imagem, se assemelha à filtragem no domínio espacial, motivo pelo qual este processo também é descrito como a aplicação de filtros nas camadas, operações convolucionais. Cada filtro é uma pequena matriz de pesos que percorre a imagem de entrada efetuando a convolução do filtro e a imagem, resultando em uma nova imagem de saída (LIMA, MENEZES, 2019). Como cada camada convolucional produz uma nova série de imagens resultado, a quantidade de dados cresce significativamente, motivo pelo que se costuma aplicar uma redução espacial dos resultados, através de uma camada de pooling.

A CNN pode ser aplicada em diversos campos da ciência como estudo de eficiência de dispositivos moveis (ZHANG, ZHOU, LIN, 2018), estudos de falhas de maquinas (JIA et al., 2018), identificação de câncer de mama por sequenciamento genético (ELBASHIR et al., 2019), estudos de turbulência (BORDE, SONDAK, PROTOPAPAS, 2021).

2.6.2 Redes neurais convolucionais (CNN) em Sensoriamento Remoto

As CNNs têm grande aplicabilidade na área do sensoriamento remoto devido à característica hierárquica do modelo. Segundo Cheng e Han (2016) podem ser usados quatro tipos de modelos: métodos baseados em correspondência, baseado em conhecimento, baseados em orientação para objetos (OBIA) e métodos baseados em aprendizado de máquina. Hossain e Chen, (2019) afirmam que o modelo clássico de

CNN trabalha com a extração de objetos, classificando pixel a pixel da imagem, com a problemática de que esse modelo tem um gasto de memória da máquina muito alto.

Os modelos de CNNs podem variar significativamente de acordo com seu principal objetivo de aplicação e os dados disponíveis. No caso das imagens (2-D), Guidici, Clark, (2017) caracterizam o modelo CNN como um classificador líder, onde o objetivo é reconhecer objetos. Quando aplicado aos dados para a classificação de cobertura do solo as Redes Convolucionais te apresentando bons resultados. No Quadro 1, pode-se outros exemplos do uso de arquiteturas CNN no sensoriamento remoto.

QUADRO 1 – Trabalhos usando redes neurais convolucionais aplicadas diretamente ao sensoriamento remoto.

AUTOR	TITULO	RESUMO
(SHARMA et al., 2017)	<i>A patch - based Convolutional neural network for remote sensing image classification.</i>	Propõem classificar imagens de media resolução uma CNN baseada nas características pixel a pixel.
(CHENG et al., 2017)	<i>Automatic Road Detection and Centerline Extraction via Cascaded End – to - End Convolutional Neural Network.</i>	Propõem usar uma rede neural convolucional em cascata (CasNet). Composta de duas redes, onde uma é capaz de delimitar bem oclusões e a outra para a extração da linha central de estradas.
(HUANG, ZHAO, SONG, 2018)	<i>Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery.</i>	Propõem usar <i>Semi-Transfer Deep Convolutional Neural Networks</i> (STDCNN), para classificar uso de e cobertura de solos em áreas urbanas em imagem de multiespectrais.
(DUNG, ANH, 2019)	<i>Autonomous concrete crack detection using fully convolutional neural network.</i>	Propõem a detectar fissuras em estruturas, usando uma <i>Fully Convolutional Neural Network</i> (FCNN) a partir de imagens de frame de vídeo com de 227x227 pixels.
(LIU et al., 2019)	<i>Relation-Shape Convolutional Neural Network for Point Cloud Analysis.</i>	Propõem usar a combinação de grid regular da CNN para uma configuração irregular para analisar nuvem de pontos.
(YANG, ROTTENSTEINE R e HEIPKE, 2019)	<i>Towards better classification of land cover and land use based on convolutional neural networks.</i>	Propõem usar arquitetura CNN em imagens aéreas de alta resolução combinadas com MDS (modelo digital de superfície) para a detecção e classificação do uso e ocupação dos solos.
(DANG et al., 2020)	<i>A Convolutional Neural Network for Coastal Classification Based on ALOS and NOAA Satellite Data.</i>	Propõem a utilização de uma CNN para classificar e detectar diferentes características em regiões costeiras, tanto para terra quanto para o mar.
(REN et al., 2020)	<i>Full Convolutional Neural Network Based on Multi-Scale Feature Fusion for the Class Imbalance Remote Sensing Image Classification.</i>	Propõem usar imagens Sentinel-2, de regiões na China, combinadas ao modelo DeepLabV3, para obter alta acurácia global para a classificação.

AUTOR	TITULO	RESUMO
(YANG, ROTTENSTEINER e HEIPKE, 2020)	<i>Investigations on Skip-Connections With an Additional Cosine Similarity Loss for Land cover Classification.</i>	Propõem uma nova função de perda de similaridade do cosseno ao investigar o <i>Skip-connections</i> (conexões por salto) na arquitetura CNN em imagens aéreas de alta resolução para uso e ocupação dos solos.
(FAKHRI, SHAH-HOSSEINI, 2022)	<i>Improved Road Detection Algorithm Based on Fusion of Deep Convolutional Neural Networks and Random Forest Classifier on VHR Remotely Sensed Images.</i>	Propõem extrair estradas rodoviárias usando arquitetura U-Net, para um conjunto de dados da região de Massachusetts, EUA, imagens RGB.

Fonte: a autora (2023).

Do Quadro 1 pode-se ver que as CNNs têm grande aplicação para diferentes problemas de sensoriamento remoto. Também é constatado que para cada problema é desenvolvida uma arquitetura particular. Um exemplo foi citado anteriormente, um *autoencoder*, é uma rede que pode ser utilizada para extrair informações de um grande conjunto de dados e detectar redundâncias.

Um caso particular de rede é a U-Net, uma forma particular de *autoencoder*, usada nesta pesquisa e detalhada na seção de metodologia. A estrutura U-Net teve suas primeiras aplicações no campo da biomedicina, como visto no trabalho de Ronneberger, Fischer e Brox, (2015) que usam esta rede para identificar células em imagens de baixa resolução. Essa arquitetura tem sido aplicada também aos campos do sensoriamento remoto, como os casos apresentados no Quadro 2. Observa-se várias aplicabilidades da arquitetura U-Net e de algumas da arquitetura modificada a partir da U-Net como *RCNN-UNet*, *RAO-UNet*, *I-UNet* e *MD-UNet* combinadas a metodologias de sensoriamento remoto para detecção e extração de objetos em imagens.

QUADRO 2 – Trabalhos usando U-Net aplicada diretamente ao sensoriamento remoto.

AUTOR	TITULO	RESUMO
(CONSTANTIN, DING, LEE, 2018)	<i>Accurate road detection from satellite images using modified U-Net</i>	Propõem modificar a função de perda da arquitetura U-Net com funções de entropia cruzada e a função da distância Jaccard afim de evitar falsos positivos, em comparação com a arquitetura convolucional atrous.
(WULFF, et al., 2018)	<i>Early fusion of camera and Lidar for robust road detection based on U-Net FCN</i>	Propõem combinar imagens de uma câmera com dados Lidar para identificação de estradas por segmentação semântica baseado na arquitetura modificada da U-Net, usando a arquitetura FCN.

AUTOR	TITULO	RESUMO
(ZHANG, LIU, WANG, 2018)	<i>Road extraction by deep residual U-Net</i>	Propõem extrair estradas por segmentação semântica de redes neurais pela aprendizagem residual combinada a arquitetura U-Net
(LIU et al., 2019)	<i>Computer vision-based concrete crack detection using U-Net fully convolutional networks.</i>	Adota o uso da arquitetura U-Net para detectar fissuras em rodovias pavimentadas.
(YANG et al., 2019)	<i>Road detection and centerline extraction via deep recurrent convolutional neural network U-Net.</i>	Propõem o uso do RCNN-UNet (<i>Recurrent Convolutional Neural Network U-Net</i>) para a detecção de estradas e a extração de linhas centrais a partir de imagens aéreas.
(WANG, MA, YE, 2020)	<i>Computer vision-based road crack detection using an improved I-UNet convolutional network.</i>	Propõem o algoritmo melhorado da U-Net (I-UNet) para detecção de fissuras em estradas pavimentadas.
(FAN et al., 2021)	<i>RAO-UNet: a residual attention and octave U-Net for road crack detection via balance loss.</i>	Adota a arquitetura RAO-UNet para detecção de fissuras em pistas pavimentadas, baseado na frequência de imagens.
(HOU et al., 2021)	<i>C-UNet: Complement UNet for Remote Sensing Road Extraction.</i>	Propõem o método complementar a UNet, ao segmentar imagens de alta resolução usando a U-Net e combinar seu produto com a MD-UNet (<i>Multi-scale dense dilated convolution UNet</i>) obtendo uma segmentação secundária.
(SENADEERA, 2021)	<i>Automation of road feature extraction from high resolution images</i>	Propõem a utilização de uma rede neural semântica combinada há aprendizagem por transferência presente na arquitetura U-Net para extração com alta precisão de rodovias a partir de imagens de satélite.
(GIURGI et al., 2022)	<i>Real-time road detection implementation of U-Net architecture for autonomous driving.</i>	Propõem o uso da estrutura U-Net para detecção de veículos automotivos em tempo real, levando em consideração variáveis climática e de iluminação natural e/ou artificial.

Fonte: a autora (2023).

Essa pesquisa propõe aplicar a arquitetura *U-Net* para detecção e extração de estradas rurais em imagens media resolução, como apresentado a seguir.

3 MATERIAL E MÉTODOS

Neste capítulo apresenta-se, brevemente, as imagens que foram usadas, o ambiente interativo do *Google Colab*, usado para o desenvolvimento das rotinas em *Python*, assim como a contextualização do processo de execução da metodologia da rede *U-Net* para a detecção de estradas em ambiente rural.

3.1 MATERIAL

3.1.1 Imagens PLANET

O conjunto de dados consiste em imagens da constelação *PLANET* disponibilizadas pelo governo da Noruega pelo programa de Iniciativa Internacional do Clima e Florestas (*Norway's International Climate & Forests Initiative - NICFI Program Resource Center*). Esse programa disponibiliza mosaicos mensais atualizados desde setembro de 2020. Ideais para o monitoramento visual, sendo disponibilizados dados em três níveis de acesso:

- **Level 0 - Público e Aberto:** Acesso público para apenas visualização aos Mosaicos Visuais (mosaicos históricos e futuros) através dos sites de *Purpose Ally*.
- **Level 1 - Aberto o quanto Possível:** Acesso para download de mosaicos de Reflexo de Superfície da Plataforma Planet. Grupos de usuários incluem Participantes Terceiros, por exemplo, as pessoas e organizações que utilizam os dados para alcançar o Propósito do NICFI. Disponibiliza-se imagens com correção atmosférica e geométricas, com resolução espacial de 3 e 5 metros.
- **Level 2 - Parceiros seletos:** Parceiros estratégicos definidos pelo Ministério do Clima e Meio Ambiente da Noruega têm acesso às camadas de dados do Nível 1, bem como às cenas subjacentes da Planet e aos dados históricos selecionados do arquivo da Airbus desde 2002. As imagens disponíveis têm resolução espacial de 0,50 m e com correções atmosférica e geométricas

embutidas, no entanto, o acesso é exclusivo a parceiros definidos pelo governo da Noruega.

Para esse trabalho foram utilizadas as cenas disponibilizadas no *Level 1*, onde foi obtido 12 cenas atrás de download, essas imagens disponibilizadas com correções e com resolução espacial de até cinco metros.

3.1.2 Google Colab

O *Google Colab* ou *Colaboratory*, trata-se de uma plataforma online que permite escrever e executar códigos em *Python* no próprio navegador Web, com acesso gratuito a Unidades de Processamento Gráfico do inglês *Graphics Processing Units* (GPUs) virtuais, de fácil compartilhamento e sem obrigatoriedade de configurações de máquina necessária para sua execução.

No ambiente Colab há disponibilidade de uso de três tipos de GPU (T100, A100 e T4), no entanto apenas o GPU básico (T4) está acessível ao uso gratuito sendo liberados 0,18 células de processamento por hora para realizar processamentos, porem ao ocupar todo espaço de memória do GPU Gratuito o ambiente interrompe o uso impedindo de realizar novos processamentos por 24 horas.

Apesar das limitações de uso do GPU no ambiente colab, este possui grande aplicabilidade nas áreas de *machine learning* pois inclui muitas bibliotecas de *machine learning* em *python*. Por isso foi usado para a implementação das rotinas de programação deste projeto, aproveitando as principais bibliotecas de manipulação de dados numéricos e imagem NumPy, *MatPlotLib*, *PiL*, juntamente com as bibliotecas de aprendizado de maquinas *TensorFlow* e *Keras*, e a biblioteca de modelos de segmentação pré-treinados *Segmentation Model*.

3.1.3 Hardware

O interpretador pyzo foi usado no workstation disponível para processamento de imagens disponibilizado no laboratório de Fotogrametria e Sensoriamento Remoto pertencente a Universidade Federal do Paraná (UFPR).

As configurações internas: memoria RAN 8 gb, processador AMD Ryzen 5 5600X 6-Core Processor 3.70 GHz, Sistema operacional de 64 bits, processador

baseado em x64 e placa de vídeo do modelo NVIDIA T600 com memória de dedicada de 8gb.

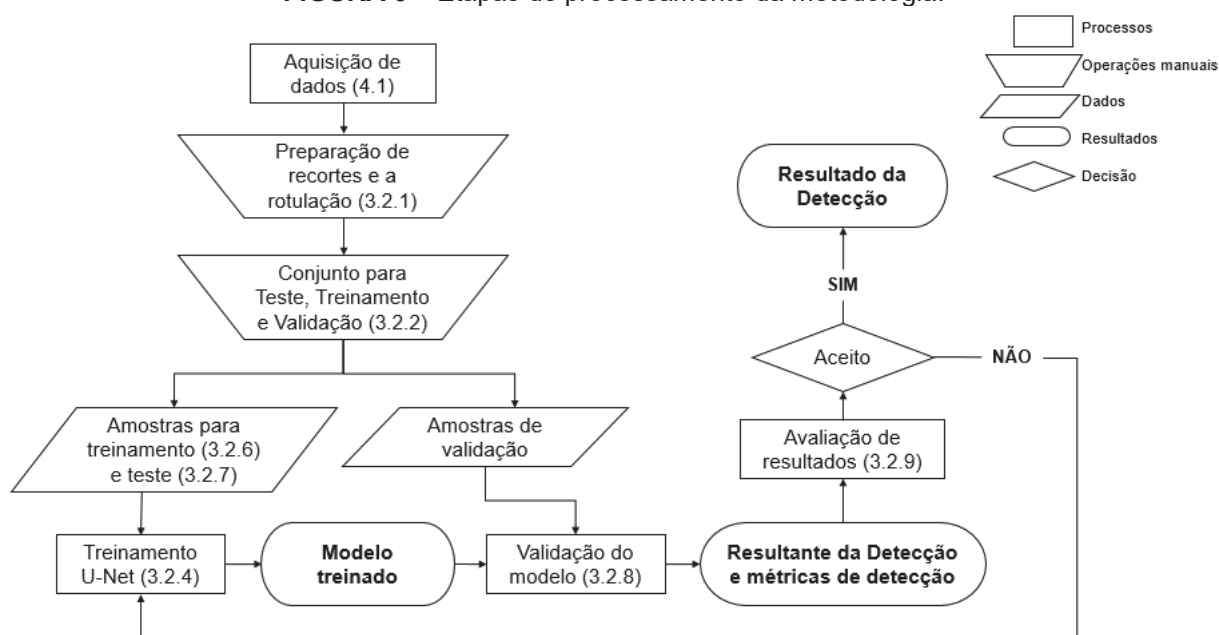
3.1.3.1 Pyzo

Pyzo é um ambiente de desenvolvimento integrado (IDE) Python que funciona com qualquer interpretador Python instalado no sistema interno, incluindo ambientes Anaconda e Conda. O IDE tem como objetivo a interatividade e a simplicidade, e consiste em um editor, um shell e um conjunto de ferramentas.

3.2 MÉTODOS

A metodologia proposta foi dividida em obtenção de imagens, preparo das imagens, o treinamento e validação do uso da rede U-Net (Figura 6).

FIGURA 6 – Etapas de processamento da metodologia.



Fonte: a autora (2023).

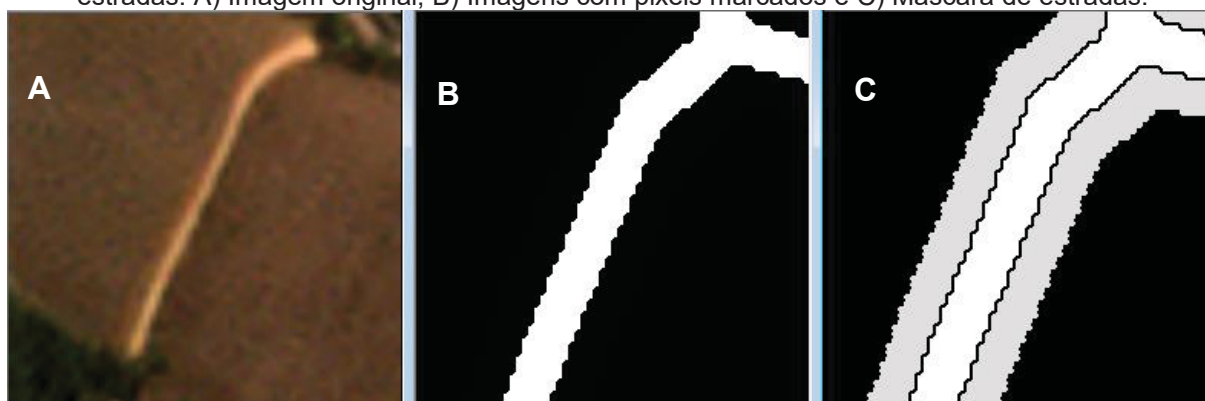
Primeiro realiza-se o preparo das imagens de entradas da rede. Nesta etapa, amostras de treinamento são coletadas na imagem, compondo um conjunto de recortes das bandas de entrada e a saída esperada para cada recorte. O conjunto de amostras é dividido em três grupos: treinamento, teste e validação.

3.2.1 Preparação dos recortes e rotulação

As imagens PLANET abrangem extensas áreas, porém, especial ênfase deve ser dada às regiões onde estradas são visíveis. Nessa fase, foi realizada a preparação das imagens para a construção de conjuntos de amostras contendo informações representativas de trechos de estradas, com o objetivo de treinar a rede. Inicialmente, a intenção era comparar o resultado obtido utilizando imagens de satélite com o resultado obtido por câmeras RGB a bordo de drones. Por isso, optou-se por não incluir a banda do infravermelho próximo nas imagens utilizadas.

Para rotular as amostras de estradas foi usado o software *Qgis*, onde foi realizada a vetorização manual das estradas sobre as imagens. A vetorização foi armazenada como imagens, como apresentado na Figura 7.

FIGURA 7 – Exemplo de recortes da imagem PLANET e identificação visual de trechos de estradas. A) Imagem original; B) Imagens com pixels marcados e C) Máscara de estradas.



Fonte: a autora (2023).

Na Figura 7a é mostrado um exemplo de recorte da imagem RGB contendo um cruzamento de estradas rurais. Na Figura 7b é mostrado o resultado da digitalização. Para levar em consideração eventuais erros de digitalização, foi criado um buffer de 2 pixels em torno das estradas. Assim, no processo de rotulação (*Labeling*) cada pixel do recorte recebeu um valor dentro da escala, a estrutura da estrada (corpo do objeto) recebeu o valor 1, 2 para o Fundo (*background*) e o buffer valor 3. Essa escala foi baseada na rede preexistente usada de base para o desenvolvimento desta pesquisa.

A rotulagem tem a função de informar a saída desejada no processo de aprendizado. Deve-se evitar rótulos ruidosos e a presença de estruturas repetidas, de modo simples agrupando sequencias de imagens com objetos semelhantes (HE, LU, SCLAROFF, 2018).

3.2.2 Conjuntos para Teste, Treinamento e Validação

Os recortes foram agrupados em três grupos: treino, teste e validação.

As amostras dos conjuntos de treino e teste foram obtidas em 10 imagens originais de satélite, com 4096*4096 pixels. Um total de 9760 recortes de 256*256 pixels foram selecionados. Destas amostras, foram separadas, de forma aleatória, dois conjuntos: 70% das amostras para treinamento da rede e 30% para o teste.

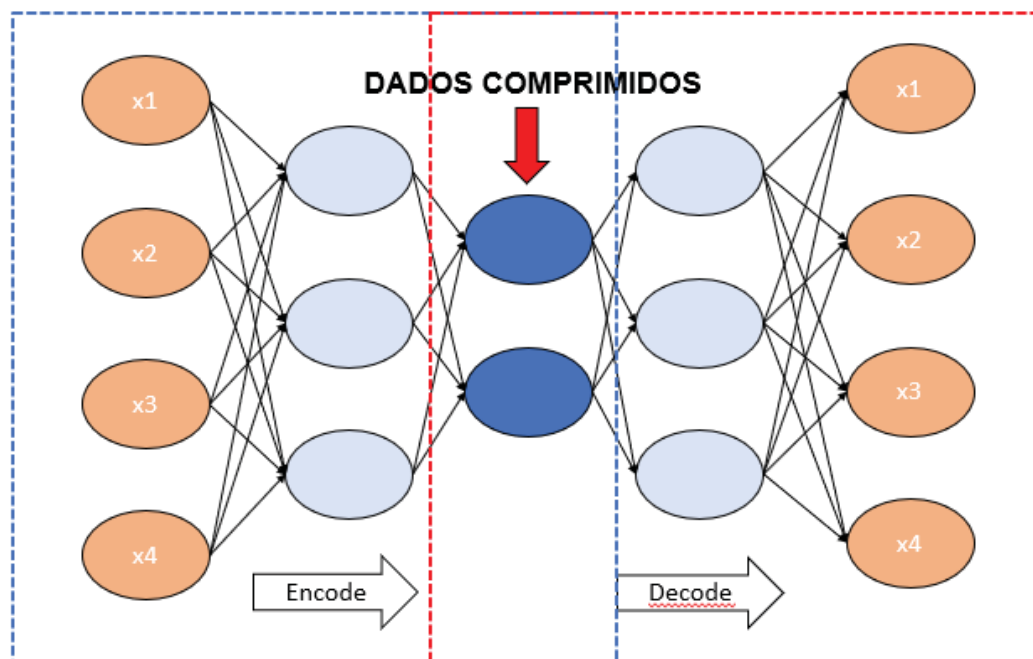
Para a validação da rede, foram selecionadas duas imagens originais de satélite. Dessas imagens, foram extraídos um total de 1703 recortes de 256*256 pixels. Esses recortes foram exclusivamente destinados para a verificação da capacidade da rede em identificar estradas rurais em novas imagens, sendo que tais imagens não foram incluídas na seleção de amostras de treino e teste.

3.2.3 O conceito de Autoencoder

Uma rede neural específica é o *autoencoder*, caracterizado por sua estrutura simétrica que se afunila no centro. Ele é dividido em duas partes principais: a codificação e a decodificação (BENGIO, GOODFELLOW, COURVILLE, 2015). Na fase de codificação, a quantidade de neurônios é sistematicamente reduzida até atingir um mínimo e, em seguida, o número de neurônios é aumentado novamente para corresponder à quantidade presente na camada de entrada. Uma característica distintiva do *autoencoder* é que suas saídas são idênticas às entradas.

Durante a fase de codificação, operações matemáticas de contração são aplicadas aos dados de entrada (*encoder*), e após essa etapa, os dados são expandidos (*decoder*) para reconstruir as entradas na saída, preservando as mesmas características dos dados de entrada (WANG et al., 2019). A Figura 8 ilustra esse processo, mostrando como os dados de entrada são combinados no lado esquerdo para produzir um número reduzido de combinações. Posteriormente, na fase de decodificação, essas combinações são expandidas para reconstruir as entradas originais (BENGIO, GOODFELLOW, COURVILLE, 2015; HOU et al., 2021).

FIGURA 8 – Esquema de uma *autoencoder*.



Fonte: Adaptado de Bengio, Goodfellow e Courville, (2015).

Ao reduzir progressivamente o tamanho das camadas está se compactando os dados, obtendo-se assim uma representação reduzida (indicadas em azul escuro na Figura 8) das entradas que ainda permite representar a informação contida nas variáveis de entrada. Isto propicia a redução da dimensionalidade. No centro do *autoencoder* a quantidade de neurônios (ou variáveis) é mínima e ela é usada para reconstruir as variáveis de entrada no braço direito do *autoencoder*.

Se é possível reconstruir as variáveis de entrada a partir da representação armazenada na camada central, então pode-se dizer que esta camada armazena uma representação compactada (codificada) das variáveis de entrada. Por isso, se diz que a metade esquerda da rede é a fase de codificação e a direita de decodificação.

3.2.4 U-Net

A arquitetura da rede U-Net, como visto na Figura 9, se assemelha muito ao princípio de um *autoencoder*, com algumas diferenças. Os dois braços do autoencoder são compostos de camadas convolucionais de diferentes tamanhos. As camadas convolucionais do lado esquerdo da arquitetura realiza a codificação dos dados de entradas e em seguida no lado direito está a fase de decodificação.

FIGURA 9 – Exemplo da Rede U-Net aplicada a imagens de 32x32 pixels com baixa resolução.

baixos para níveis elevados, possibilitando obter resultados de segmentação mais precisos com menos camadas (SENADEERA, 2021).

A principal diferença da U-Net, nesse sentido é que ao fazer a decodificação e expandir os mapas, realiza simultaneamente a concatenação de mapas de feições de maior resolução da fase de codificação para produzir uma melhor representação para as demais camadas convolucionais.

A U-Net tem aplicação de aprendizado ponta a ponta para todos os parâmetros, incluindo os da parte de decodificação, onde as informações espaciais e as características são combinadas através de funções de convolução e concatenação. As operações de *pooling* não apenas tornam os recursos aprendidos invariantes às transformações da imagem, mas também ampliam o campo receptivo para incorporar mais informações de contexto de maneira implícita (YANG, ROTTENSTEINER e HEIPKE, 2019)

Os recortes das amostras de treinamento são apresentados à rede U-Net, como entrada. As amostras são processadas na fase do *encoder*, com a execução das funções de convolução organizada em sequência. Aqui, os recortes passaram por filtros bidimensionais sendo calculados os descritores na imagem original e assim em sequência passa pelas camadas de convolução, obtendo os mapas de características das imagens, no entanto, com dimensões reduzidas e informações filtradas.

3.2.5 Parâmetros

Para realizar o treinamento da rede e obter os melhores pesos deve ser definidos os parâmetros ideais, o *batch size* variável que define o tamanho dos lotes de imagens, quantidade de camadas, quantidade de convolução e neurônios por camada, e a quantidade de épocas a ser treinado.

O *batch size* é um parâmetro que determina como as amostras do conjunto inicial são divididas em lotes menores, os quais são utilizados durante a etapa de treinamento. Para essa pesquisa foram utilizados o conjunto de amostras inicial composto por 9760 imagens com *batch size* ideal de 32 obtendo 305 lotes. Já com um *batch size* de 64, obteve-se 153 lotes. Essa divisão em conjuntos menores facilita o processamento dos dados durante o treinamento. Vale ressaltar que ao separar as amostras em lotes menores, foi possível alcançar resultados melhores no processo de treinamento.

Cada camada desempenha um papel importante no processamento dos dados e no aprendizado das características relevantes para a tarefa em questão. A quantidade de camadas da rede determina a profundidade da rede e quais as operações a serem executadas, quanto mais camadas mais operações são necessárias, acarretando no aumento do esforço computacional. Nesse caso sendo usadas um total de 13 camadas, onde sete camadas de operações de convolução e seis de deconvolução.

Dentro de cada camada ocorre as operações de convolução, essas operações funcionam como um conjunto de filtros aplicados as amostras de entradas para extração de características relevante. Cada filtro funciona como uma matriz de pesos que varre o conjunto de amostras, calculando respostas entre as entradas e os pesos do filtro, esses pesos são armazenados como memória do treinamento. As operações usadas foram o filtro de convolução 3X3 com a função de ativação *Relu*, *função max pool 2X2* e operações de *deconvolução 2X2*.

A operação executada nas camadas são conectadas pelos neurônios artificiais que funcionam como as sinapses neurais humanas, interligando as operações camada a camada, por tanto a quantidade de neurônios por camada determina a complexidade da rede, ou seja, com o aumento do número de neurônios acarreta o aumento da profundidade da rede e conseqüentemente aumentando a complexidade de solução de problemas, com isso para o caso das estradas rurais 32 neurônios foi suficiente para armazenar as características necessárias há detecção.

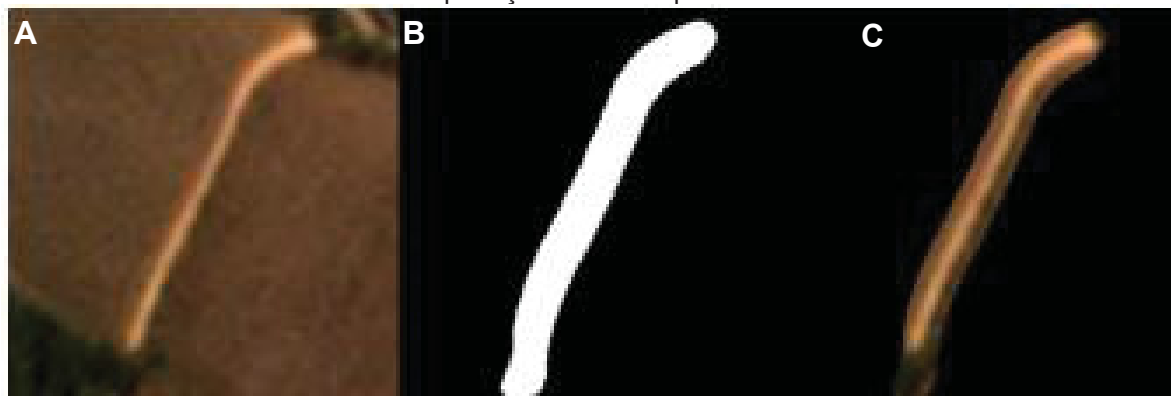
Determinar quantas vezes a rede será executada por completo a fim de aprender as características do problema é o que define as épocas de treinamento. Em cada execução do treinamento, a rede terá um valor de aprendizado armazenado, permitindo que, na execução seguinte, o objeto buscado seja melhor identificado. Usando 100 épocas de treinamento, os conjuntos de imagens percorrem as camadas e executam as operações para apreender as características do objeto procurado, neste caso, a geometria da estrada rural.

3.2.6 Treinamento

Para o treinamento da rede, foram usadas amostras contendo o recorte da imagem em torno de estradas e sua verdadeira classificação (imagem rotulada), de

forma a permitir que a rede aprenda com as amostras o padrão espacial que caracteriza as estradas.

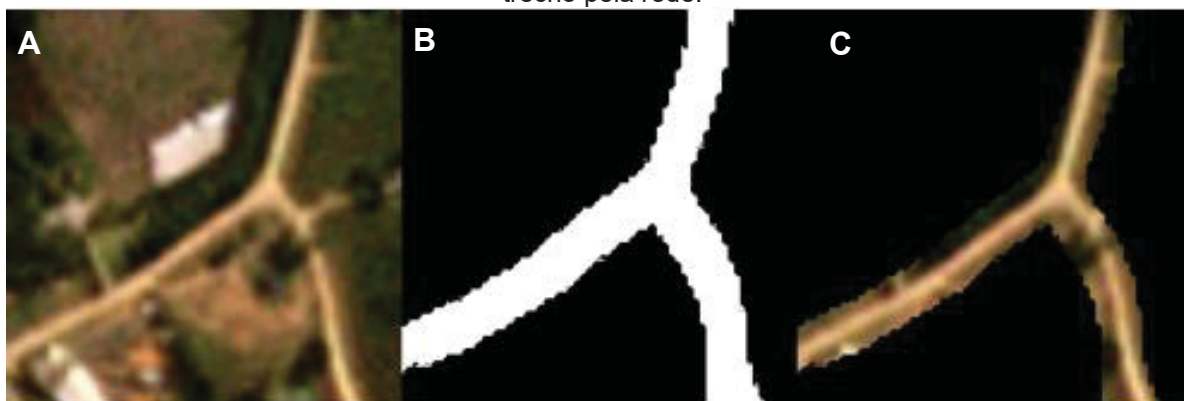
FIGURA 10 – Exemplo de recorte contendo padrão espacial que caracteriza as estradas, A) Recorte contendo exemplo de estradas, B) Máscara com pixels marcados pela previsão da rede e C) Interpretação do trecho pela rede.



Fonte: a autora (2023).

As Figuras 10 e 11 apresenta um trecho de estradas bem delimitado, seguido da sua máscara contendo os pixels marcados e a interpretação que a rede tem durante a etapa de treinamento.

FIGURA 11 – Exemplo de recorte contendo padrão espacial que caracteriza as estradas em ambiente com elementos antrópicos que podem causar confusão da rede. A) Recorte contendo exemplo de estradas, B) Mascara com pixels marcados pela previsão da rede e C) Interpretação do trecho pela rede.



Fonte: a autora (2023).

O processo de treinamento da rede consiste em modificar os pesos da rede de forma iterativa usando o algoritmo de retro propagação. Este ajuste é realizado pela minimização dos erros dentro de um processo iterativo. O algoritmo de retro propagação é encarregado de ajustar os pesos dos neurônios partindo da saída ao longo de toda a rede (SENADEERA, 2021). Cada treinamento gera pesos por época,

esse foram armazenados no arquivo do tipo “.h5”, esse arquivo referente a arquitetura e pesos da rede treinada, que podem ser usados posteriormente na fase de validação.

3.2.7 Teste

O processo de teste da rede ocorreu em conjunto com o processo de treinamento. Inicialmente um conjunto de amostra foi separado de forma randômica, onde 70% para treino e 30% para teste, esse processo de separação ocorre dentro da execução de cada época de treinamento do modelo, e estabelecendo o tamanho dos lotes de imagens.

3.2.8 Validação

Para essa etapa, usou-se duas imagens Planet, recortadas de 40 x 40 pixels, que foram redimensionadas dentro da rede para 256*256 pixels. Essas imagens foram separadas do restante das imagens de satélites obtidas inicialmente, para evitar a redundância de informações durante o processamento, o conjunto de imagens usados para validação não contem imagens usadas na etapa de treinamento.

Os recortes das imagens de validação também se baseiam na posição das estradas rurais em suas máscaras correspondentes (verdade de campo). Os recortes de validação totalizam 1783 recortes. Apesar das máscaras referentes as imagens de validação serem usadas para executar o recorte das imagens de satélite, as mesmas não foram usadas para executar as máscaras de previsão.

A comparação o resultado do processamento da imagem RGB na U-Net e a rotulação manual permite avaliar e medir o sucesso da detecção de trechos de estradas.

3.2.9 Avaliação de resultados

No Quadro 3 são mostradas possíveis situações que podem ser encontradas ao comparar a saída da rede com a saída esperada. A rede pode receber como entrada uma região com ou sem estradas; No primeiro caso (A) o sistema funcionou corretamente e não detectou a estrada; No segundo (B), o sistema prevê a existência de estrada em locais onde ela não existe, um erro. No caso (C), o resultado é correto,

pois a saída coincide com a entrada; No último caso (D), o sistema não identifica a estrada.

QUADRO 3 – Exemplo de possíveis saídas da rede.

CASO	ENTRADA	SAÍDA ESPERADA	RESULTADO OBTIDO
A] (Verdadeiro Negativo)			
B] Falso positivo			
C] Verdadeiro Positivo			
D] Falso Negativo			

Fonte: a autora (2023).

Para a avaliação dos resultados obtidos foram usadas as métricas de avaliação sugeridas por Xu et al. (2019), que são recall (R), precisão (P), F_1 score, Acurácia e IoU. Para o cálculo destas métricas é necessário contabilizar a quantidade de verdadeiros positivos (TP), falsos positivos (FP) e falsos negativos (FN).

O processo constituiu-se na comparação entre a máscara referente à verdade de campo e a máscara prevista na etapa de validação, a partir do total pixels bem classificados, onde foram avaliadas as ocorrências de respostas verdadeiro positivo, falso negativo e falso positivo, em relação ao total de respostas classificadas. Como

a rede não é alimentada com exemplos de verdadeiros negativos, eles não são considerados nas métricas de avaliação de qualidade.

- **Verdadeiro Positivo:** identificado com TP (*True Positive*), todas as repostas resultantes da máscara prevista como “estradas” correlacionada ao pixel marcado pertencente a máscara *ground truth* para “verdade estrada”, ou seja, os pixels de “estradas” marcados corretamente como “estrada” (Quadro 4).
- **Falso Negativo:** indicados como FN (*False Negative*), todas as repostas resultantes da máscara prevista como “não estradas” correlacionada ao pixel marcado pertencente a máscara verdade de campo para “verdade estrada”, ou seja, os pixels de estradas que foram erroneamente marcados como *background* (Quadro 4).
- **Falso Positivo:** Descreve a situação quando uma região da imagem é classificada como estrada sendo que na verdade a região não contém uma estrada. Esses casos são indicados como FP (*False Positive*).

QUADRO 4 - Correlação entre resposta TP, FP e FN consideradas para os cálculos das métricas de avaliação.

	Resultado “estrada”	Resultado “Não estrada”	Soma
Verdade “estrada”	TP	FN	TP + FN
Verdade “Não estrada”	FP	---	FP
Soma	TP + FP	FN	

Fonte: A autora (2023).

A taxa de recall (R), descrita na Equação (10), calculando a razão entre a quantidade de amostras corretamente classificadas em relação ao número de amostras da classe “estrada” disponíveis. Ela equivale a acurácia do ponto de vista do usuário.

$$R = \frac{TP}{TP + FN} \quad (8)$$

A Precisão (P), Equação (11), calculando a razão entre a quantidade de amostras corretamente classificadas em relação ao número total de amostras classificadas como “estrada”. Esta taxa equivale à acurácia do ponto de vista do

usuário e descreve a percentagem de acertos verdadeiros em relação a todos os elementos identificados como estradas.

$$P = \frac{TP}{TP + FP} \quad (9)$$

O parâmetro F_1 score, visto na Equação (12), calcula a média harmônica entre os parâmetros recall e precisão. Usado para assegurar a acurácia de um *data set* desbalanceado, em que o número de pixels marcados e os pixels de fundo correspondem a quantidades diferentes.

$$F_1score = 2 * \frac{R * P}{R + P} \quad (10)$$

Para o caso desta pesquisa, que avalia a detecção de um objeto (estradas rurais) e o restante é interpretado com *background*, o outro parâmetro usado para a avaliação de qualidade a exatidão das respostas obtidas foi a acurácia (AC) apresentado na Equação (13), na qual considera a relação dos pixels detectados corretamente (TP) em relação os pixels detectados no processo como um todo (n). A acurácia indique o quão exato o método de detecção esta, assim o quão próximo a máscara prevista pelo modelo está próximo da máscara vetorizada (verdade de campo).

$$Ac = \frac{TP}{n} \quad (11)$$

Por fim a métrica IoU, ou índice de Jaccard, que é a relação entre a quantidade de amostras corretamente classificadas e o total de amostras classificadas como estrada mais o total de falsos negativos, conforme a Equação (14). Essa métrica avalia os elementos comuns entre as imagens, ou seja, a semelhança entre a imagem testada e a resposta detectada, avaliando a semelhança pixel a pixel entre as imagens.

$$IoU = \frac{TP}{TP + FP + FN} \quad (12)$$

4 EXPERIMENTOS

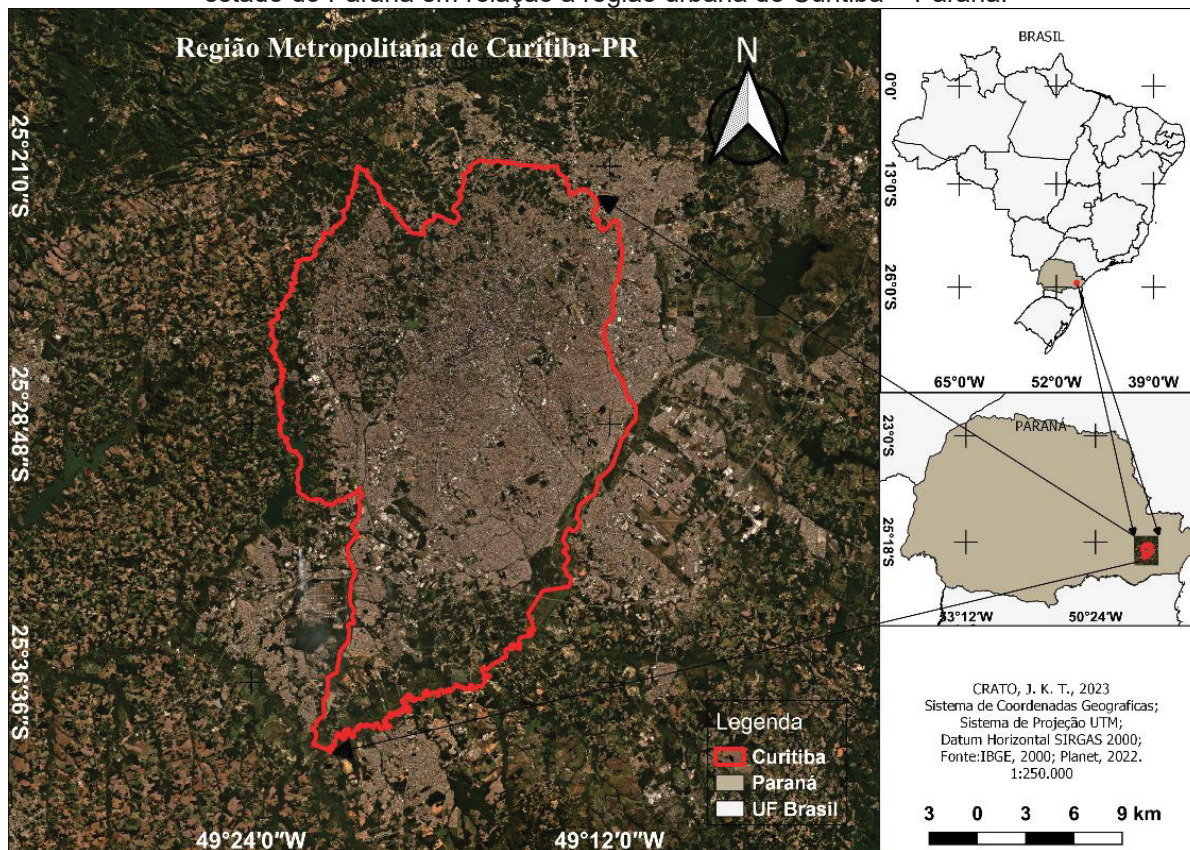
Para o desenvolvimento desta pesquisa, foram realizados três conjuntos de experimentos usando um total de 12 imagens do Planeta para treinamento, teste e validação da rede. Cada uma dessas imagens do Planeta é bastante extensa, com dimensões de 4096x4096 pixels, uma resolução espacial de 5 metros e resolução espectral composta por quatro bandas. Devido ao tamanho original das imagens, foi necessário realizar recortes, com foco especial nas estradas, a fim de melhorar o treinamento da rede.

Portanto, para cada um dos experimentos, utilizou-se um total de 9760 recortes, cada um medindo 40 x 40 pixels. Isso equivale a 10 imagens do Planeta usadas para treinamento e teste da rede em cada conjunto. Além disso, foram reservados 1783 recortes adicionais de 40 x 40 pixels, equivalentes a duas imagens adicionais do Planeta, exclusivamente para a validação da rede.

4.1 ÁREA DE ESTUDO

Para o desenvolvimento desta pesquisa foi selecionada uma região nas proximidades do município de Curitiba, no estado do Paraná, com coordenada central 25°25'48" S e 49°16'15" O. Essa região possui uma área equivalente a 5.000 km². Esta área está situada na região metropolitana de Curitiba, englobando partes da Serra do Mar e áreas rurais circundantes à região metropolitana. Podendo ser visualizada no mapa de localização na Figura 12.

FIGURA 12 – Mapa de localização área de estudo, imagens Planet, sobre a extensão territorial do estado do Paraná em relação a região urbana de Curitiba – Paraná.



Fonte: A autora (2023).

4.1 AQUISIÇÃO DE DADOS

As imagens que foram utilizadas, foram adquiridas no portal <https://www.planet.com/explorer/>. A plataforma *explorer*, foi desenvolvida pelo governo da Noruega para disponibilizar imagens diárias e mosaicos mensais das regiões tropicais do Brasil.

Foram utilizadas as imagens de acesso *Level 1*, disponibilizadas para desenvolvimento acadêmico, com acesso gratuito, as cenas foram selecionadas para o mês de maio do ano de 2022.

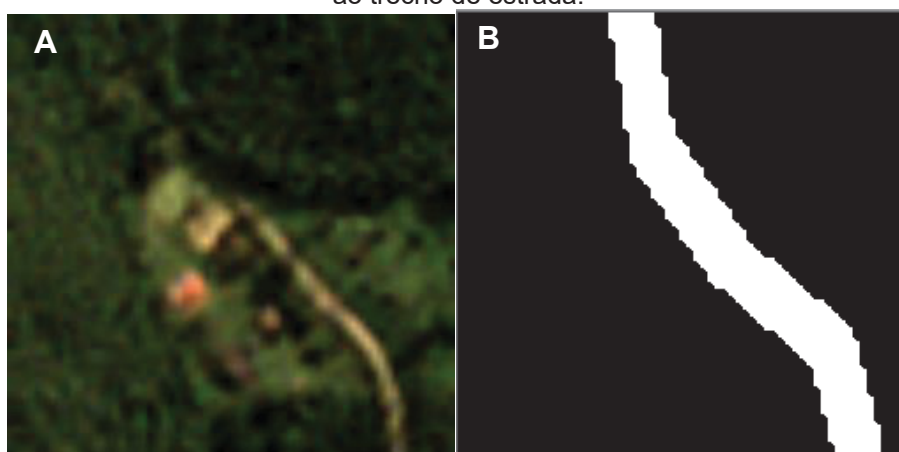
O critério de escolha se deu pelos fatores: as cenas com ausência de nuvens, e disponibilizadas com correções atmosféricas. Essas cenas também passam pelo processo de normalização, assim como o processo de normalização bidirecional da função de distribuição da refletância (BRDF - *Bidirectional Reflectance Distribution Function*).

As características das imagens obtidas são:

- Resolução espacial: 5m por pixel;
- Resolução espectral: NIR, Vermelho, Verde, Azul (4 bandas);
- Resolução temporal:
 - Arquivo de dezembro de 2015 - agosto de 2020 em cadência bianual;
 - Monitoramento a partir de setembro de 2020 - a partir da cadência mensal.

Inicialmente, a imagem de satélite possui dimensões de 4.096 x 4.096 pixels, o que equivale a uma área de cerca de 20,5 x 20,5 km (aproximadamente 420 km² em cada imagem). A partir dela, foram obtidos recortes que correspondem a regiões da imagem contendo trechos de estradas, cada um com o tamanho de 40 x 40 pixels. Isso foi feito considerando a proporção do tamanho de uma estrada em relação à resolução da imagem, como ilustrado na Figura 13. Posteriormente, esses recortes foram reamostrados para o tamanho de 256 x 256 pixels. A Figura 13.A apresenta um exemplo de um desses recortes, que retrata uma estrada rural na cena Planet, enquanto a Figura 13.B mostra o trecho rotulado correspondente. Este último foi usado como base para a janela de recorte nas imagens RGB e serviu como conjunto de rótulos adicionados junto com as imagens de entrada para a rede.

FIGURA 13 – A) Recorte 40 x 40pixel no RGB e B) imagem rotulada 40 x 40 pixels correspondente ao trecho de estrada.



Fonte: A autora (2023).

A escolha dos parâmetros, como *batch size*, épocas e quantidade de neurônios, é essencial em um experimento de treinamento de rede neural e pode influenciar significativamente os resultados obtidos. No geral, é importante realizar experimentos

com diferentes configurações de parâmetros para encontrar a combinação que melhor se adapta ao conjunto de dados específico e ao problema. Neste contexto, foram realizadas três séries de experimentos, cujos detalhes estão descritos a seguir.

4.2 SERIE DE EXPERIMENTO 1

Nesse experimento foram realizados dois testes no ambiente *colab*, com os parâmetros usados, batch size de 32 imagens por lote, obtendo 305 lotes de recortes; desses, 213 lotes são de treinamento (correspondente a 70% das amostras) e 92 lotes de teste (referentes a 30% das amostras), separadas de forma aleatória, com 13 camadas convolucionais, das quais a parte de codificação possui sete camadas e a decodificação com seis camadas, usando 50 épocas de treinamento. Foram realizados os Teste 1 com 16 neurônios e teste 2 com 32 neurônios.

4.3 SERIE DE EXPERIMENTO 2

Nesse experimento foram realizados dois teste no ambiente *colab*, com os parâmetros usados, batch size de 64 imagens por lote, considerando aumentar o número de imagens por lote, diminui o esforço computacional, obtendo 152 lotes de recortes; Desses, 106 lotes são de treinamento (correspondente a 70% das amostras) e 46 lotes de teste (referentes a 30% das amostras), separadas de forma aleatória, sendo mantido 13 camadas convolucionais, com sete camadas para codificação e a decodificação com seis camadas.

4.4 SERIE DE EXPERIMENTO 3

A fim de verificar o desempenho da rede implementada em outro interpretador da linguagem Python, foram conduzidos os experimentos da terceira série no ambiente do Pyzo, executado em um computador workstation com configurações específicas no tópico 3.1.3 Hardware. Mantiveram-se os parâmetros estabelecidos nos experimentos anteriores, com um *batch size* de 32 imagens por lote e o mesmo número de camadas (13 camadas), além de ter sido escolhido o menor número de neurônios testado (16 neurônios).

Com o objetivo de aprimorar o desempenho do modelo, com base nos experimentos anteriores e optou-se por aumentar o número de épocas no treinamento para 300. A expectativa era que esse aumento resultasse em um resultado refinado. Entretanto, constatou-se que essa quantidade de épocas tornou o tempo de treinamento excessivamente lento, tornando-o inviável.

5 RESULTADOS E DISCUSSÕES

O conjunto de amostras para o treinamento da rede totaliza 9760 recortes de 40 x 40 pixels referentes a 10 imagens originais de satélite, posteriormente expandidas dentro da rede para 256*256 pixels. E com mais duas imagens originais de satélite para avaliação da detecção e validação da rede.

5.1 SERIE DE EXPERIMENTO 1

Os testes obtiveram resultados estatísticos/métricas de avaliação (Tabela 1) e a série de mascaras contendo as previsões da rede (Figura 14).

TABELA 1 – Resultantes para primeira série de testes usando plataforma Colab, 9760 amostras de treinamento e 1783 amostras de teste, 50 épocas de treinamento.

GOOGLE COLAB			
Amostras de Treinamento = 9760	Amostra validação – 1783		
	Épocas	50	
	Batch Size	32	
	Neurônios	16	32
	Acurácia	91.85%	91.95%
	Precisão	75.80%	75.83%
	Recall	63.71%	65.92%
	F1-Score	68.99%	69.07%
	Mean IoU	53.94%	55.01%
	Tempo	04h:53min	06h:08min

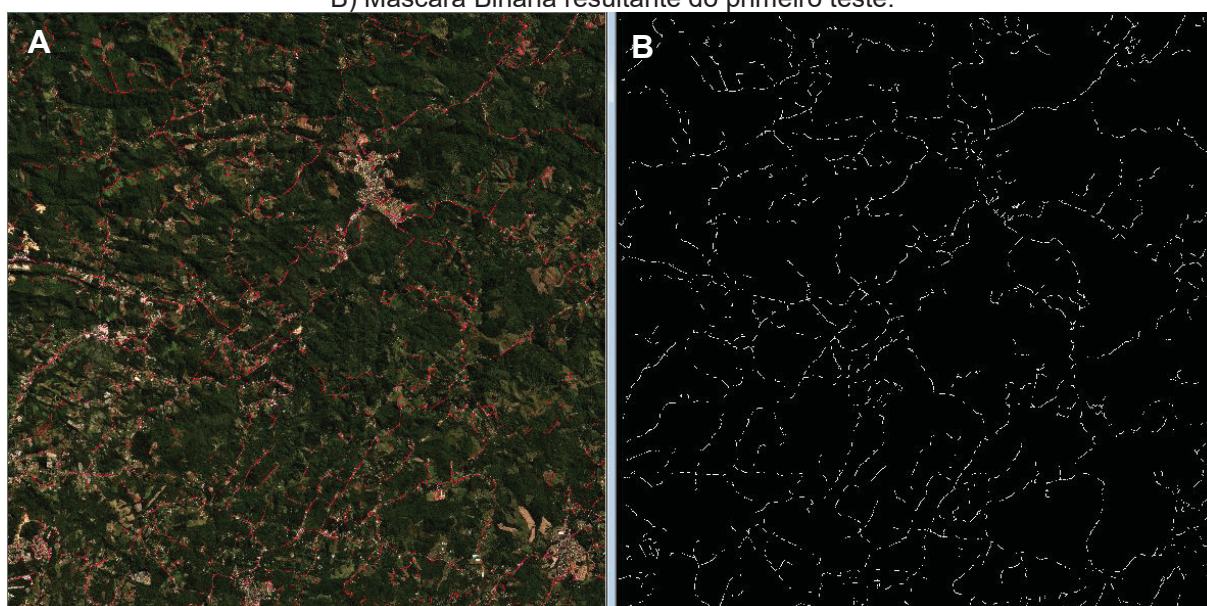
Fonte: A autora (2023).

Ao comparar as métricas obtidas para os teste 1 com 16 neurônios e teste 2 com 32 neurônios, observa-se que com o aumento do número de neurônios, resultou no aumento da demanda de tempo de processamento, esse inicialmente em torno de quatro horas passando no segundo teste para seis horas, manter o mesmo número de camadas e tamanho da camadas e a mesma quantidade de épocas de treinamento para ambos os teste, apenas alterando o neurônios aumentou o esforço computacional com baixo ganho nas estatísticas. O ganho significativamente baixo para a acurácia com melhoria de 0,10%, a precisão com 0,03% e F1-Score com 0,08% mesmo para métricas onde a melhoria foi um pouco mais altas como *Recall* que entre os testes tem melhora de 2,21% e a *Mean IoU* com 1,07%. Com o aumento do tempo

de processamento e sem ganho nas estatísticas a rede pode ser dita como acurada com tudo ainda não sendo a melhor solução para o problema.

Observando a Figura 14 com os resultados da primeira série de experimentos, A) a imagem original usada na validação com o resultado sobre posto e B) resultado obtido pela rede para o primeiro teste é possível ver a silhueta principal da estrada, no entanto por se tratar de uma imagem de dimensões compatíveis com as imagens originais usadas na rede, nota-se as silhuetas das estradas.

FIGURA 14 – Primeira serie de experimentos: A) Imagem original sobre posta pelo resultado da rede; B) Mascara Binaria resultante do primeiro teste.



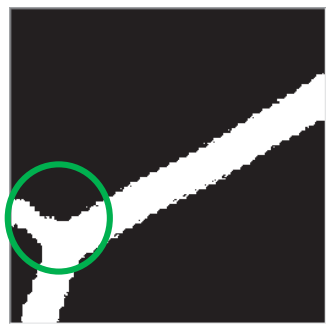
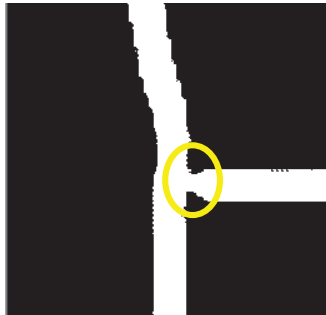
Fonte: A autora (2023).

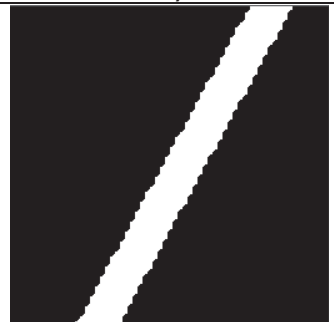
No Quadro 5, são apresentados exemplos da saída produzida pela rede neural para esta série de testes. A Figura (A) mostra a região de entrada, (B) a máscara compatível a verdade de campo e (C) o resultado produzido, que obteve sucesso utilizando 16 neurônios.

É perceptível que a rede foi capaz de detectar o trecho digitalizado e exibiu resultados além da detecção marcada como "verdade de campo". Além disso, ela também conseguiu detectar um trecho que não estava presente na máscara, como mostrado nos exemplos 1 (pertencente à região urbana) e 2 (pertencente à região urbana com vegetação densa) e destacados em verde. Esses trechos não foram demarcados devido à ferramenta utilizada para a vetorização da rede, que possuía um traço mais grosso do que a própria estrada, cobrindo também a vegetação ao redor da estrada.

Entretanto, observa-se que existem alguns gargalos e falta de junção nos entroncamentos dos trechos, representados pelos exemplos 3 e 4, marcados em amarelo. Além disso, nos exemplos 5 e 6, a detecção foi idêntica à máscara de campo, ou seja, houve uma coincidência perfeita com a verdade de campo.

QUADRO 5 – Resultado bem sucedido para primeira série de experimentos: A) exemplo de entrada; B) estrada; C) detecção resultante.

EX	A)	B)	C)
1			
2			
3			
4			

EX	A)	B)	C)
5			
6			

Fonte: A autora (2023).

A hipótese considerada foi que, ao aumentar o número de neurônios, aumenta-se a complexidade da rede, uma vez que os neurônios representam as ligações entre as operações realizadas dentro de cada camada. Com base nessa hipótese, foi conduzido um terceiro teste para verificar a possibilidade de alcançar uma melhor detecção.

No terceiro teste, foram utilizados os seguintes parâmetros: batch size de 32 imagens por lote, 13 camadas convolucionais e 50 épocas de treinamento. Além disso, o número de neurônios foi aumentado para 64. No entanto, esse teste foi bloqueado pelo ambiente *Colab*, resultando na ausência de resultados de detecção. Esse bloqueio pode ter ocorrido devido ao aumento da complexidade da rede com o aumento do número de neurônios.

O ambiente *Colab* é uma plataforma comercial pertencente ao Google e possui limitações em sua versão gratuita. Informa-se que a execução de processos com alta demanda de GPU virtual requer o uso do '*Colab Pro*', uma conta paga que permite utilizar GPU virtual com melhores configurações gráficas. Portanto, ao aumentar o número de neurônios, aumenta-se a complexidade da rede, e o ambiente *Colab* impede a execução do processo devido a essas limitações.

5.2 SERIE DE EXPERIMENTO 2

Com os lotes de treinamento maiores há hipótese de o processo ter menor gasto de tempo, considerando a possibilidade do processo mais rápido e que um maior volume de treinamentos resultaria no produto refinado, por tanto foram usadas 100 épocas de treinamento para 16 e 32 neurônios por camada.

TABELA 2 – Resultantes para segundo conjunto de testes usando plataforma Colab, 9760 amostras de treinamento e 1783 amostras de teste, 100 épocas de treinamento.

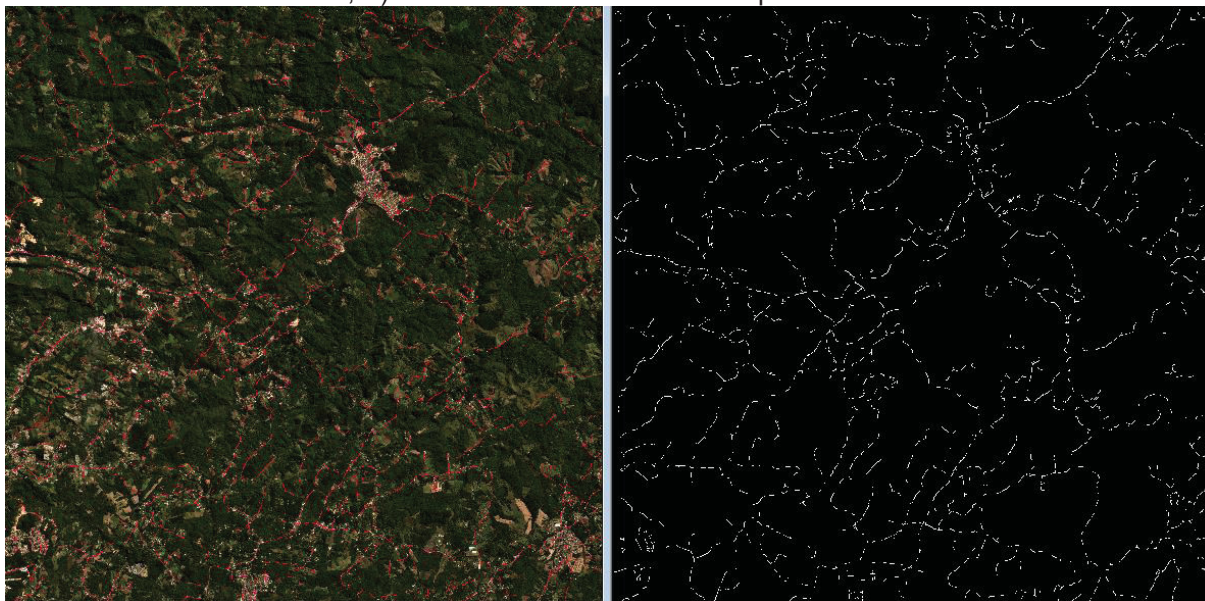
GOOGLE COLAB			
Amostras de Treinamento = 9760	Teste – 1783 Recortes		
	Épocas	100	
	Batch Size	64	
	Neurônios	16	32
	Acurácia	91.89%	92.05%
	Precisão	76.22%	76.66%
	Recall	63.97%	66.30%
	F1-Score	67.93%	69.48%
	Mean IoU	53.92%	55.56%
	Tempo	08h:32min	10h:21min

Fonte: A autora (2023).

Aumentando o número de neurônios com 100 épocas produziu resultados sem ganho significativo para as métricas (Tabela 2) com melhora para acurácia em 0,16%, precisão 0,44%, *Recall* 2,33%, F1-score 1,55% e *Mean IoU* 1,62%, porém aumenta significativamente a demanda de tempo de processamento saltando de oito horas para 10 horas de treinamento.

Nota-se que as estradas ficaram mais bem delimitadas na segunda série de experimentos (Figura 15), sendo possível notar que a mais estradas delimitadas, nos outros exemplos a seguir (Quadro 6) percebe-se os resultados refinado produzidos pela rede, principalmente para o teste 2.

FIGURA 15 – Segunda série de experimentos: A) Imagem original sobre posta pelo resultado da rede; B) Mascara Binaria resultante do primeiro teste.



Fonte: A autora (2023).

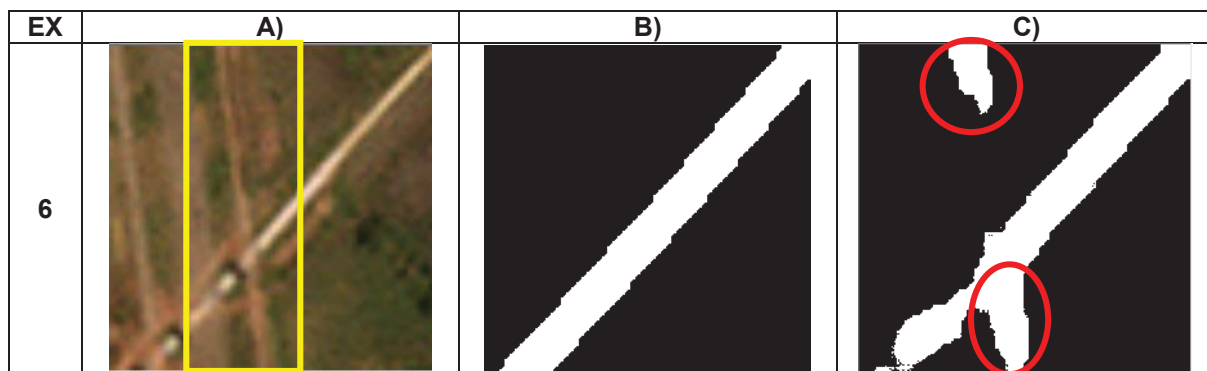
No Quadro 6, são apresentados exemplos da saída produzida pela rede neural, composta por: (A) Um exemplo da região de entrada, (B) A máscara correspondente à verdade de campo para a região de entrada, e (C) O resultado produzido pela rede para esta série de experimentos, obtendo sucesso com o uso de 16 neurônios.

O Quadro 6 exibe alguns exemplos bem-sucedidos da segunda série de experimentos, como observado nos exemplos 1 e 2, nos quais as respostas coincidem com a verdade de campo. Além disso, é notável que a rede foi capaz de ir além do esperado em alguns trechos, como exemplificado pelos exemplos 3 e 4, detectando até mesmo trechos que não estavam incluídos na máscara verdade de campo (demarcados em verde no Quadro 6).

No entanto, é possível observar que a rede identificou uma coluna de vegetação plantada (exemplo 5, Figura (A), demarcada em amarelo) como estrada (exemplo 5, Figura (C), demarcada em vermelho), resultando em uma resposta de "falso positivo". Além disso, no exemplo 6 (Figura (C), demarcado em vermelho), é visível a detecção de trechos de um caminho mais estreito (exemplo 6, Figura (A), demarcado em amarelo), provavelmente delineado pelo pisoteio constante de animais e/ou pessoas a pé.

QUADRO 6 – Resultado bem sucedido para segunda série de experimentos: A) exemplo de entrada; B) estrada; C) detecção resultante.

EX	A)	B)	C)
1			
2			
3			
4			
5			

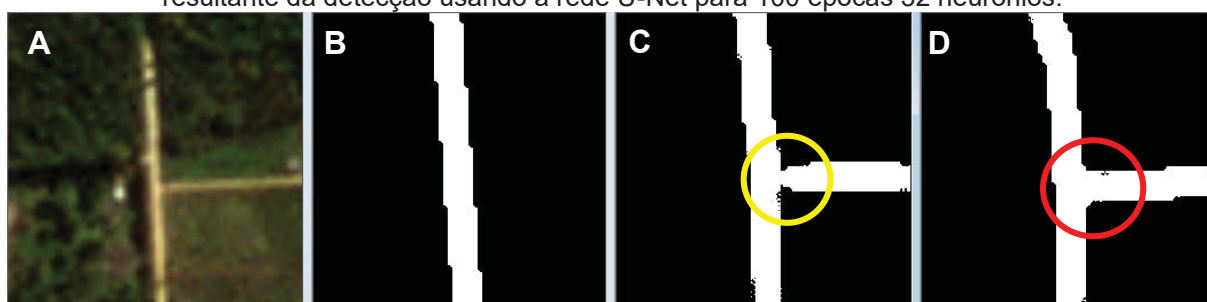


Fonte: A autora (2023).

Na Figura 16 observa-se os resultados para a série de experimentos 1 e 2, A) recorte de entrada, B) máscara da verdade de campo, C) resposta detectada para a primeira série de experimentos 1 e, d) resposta detectada para a primeira série de experimentos 2.

Ao comparar as Figuras 16 C) e D), fica evidente que os resultados obtidos no experimento 2 foram mais bem-sucedidos. Notavelmente, no teste 1 com 16 neurônios e 100 épocas, o entroncamento entre as duas estradas (delimitado em amarelo na Figura 16 C) não foi bem definido, enquanto no segundo teste com 32 neurônios e 100 épocas, esse mesmo entroncamento não apresentou mais falhas (delimitado em vermelho na Figura 16 D).

FIGURA 16 – A) Recorte de entrada; B) Mascara binaria baseada na verdade de campo; C) resposta resultante da detecção usando a rede U-Net para 100 épocas com 16 neurônios; e D) resposta resultante da detecção usando a rede U-Net para 100 épocas 32 neurônios.



Fonte: A autora (2023).

O ambiente *Colab*, sendo um implementador com fins comerciais, possui limitações em relação ao tempo de execução. Geralmente, é permitido apenas entre 10 a 12 horas de processamento contínuo, dependendo do tipo de processo. Processos que exigem maior demanda de memória GPU podem ser interrompidos quando ocupam toda a memória disponível. Além disso, se um processamento requer alta demanda de memória devido à complexidade do processo ou uma grande

quantidade de interações, ele pode ser interrompido pela plataforma e impedido de ser retomado por 24 horas.

5.3 SERIE DE EXPERIMENTO 3

Essa série de experimentos foi realizado no workstation da UFPR, com a finalidade de verificação a relação do processamento em ambiente com menos limitantes e aumento dos números de épocas de treinamento.

Os resultados para as métricas calculadas com 300 épocas estão apresentados na Tabela 3. Notavelmente, aumentar o número de épocas não resultou em um aumento significativo nos valores das métricas de qualidade em comparação com os experimentos anteriores realizados com parâmetros semelhantes.

TABELA 3 – Resultantes para terceiro conjunto de testes usando workstation, 9760 amostras de treinamento e 1783 amostras de teste, 300 épocas de treinamento.

Amostras de Treinamento = 9760	TESTE – 1783 RECORTES			
		Serie 1	Serie 2	Serie 3
	Épocas	50	100	300
	Batch Size	32	64	32
	Neurônios	16	16	16
	Acurácia	91.85%	91.89%	91.90%
	Precisão	75.80%	76.22%	76.27%
	Recall	63.71%	63.97%	64.56%
	F1-Score	68.99%	67.93%	68.28%
	Mean IoU	53.94%	53.92%	54.21%
	Tempo	04h:53min	08h:32min	135h (05 dias, 15h)

Fonte: a autora (2023).

As séries de experimentos realizados com 16 neurônios são identificadas na Tabela 3 como série 1, série 2 e série 3 (resultante deste experimento). Ao comparar as métricas, observa-se o seguinte:

- A acurácia aumenta em 0,04% entre o 1° e 2° teste, em torno de 0,02% do 2° para o 3°, e a diferença é de 0,05% entre o 3° e o 1° teste.
- A precisão aumenta em 0,42% entre o 1° e 2° teste, em torno de 0,05% do 2° para o 3°, e a diferença é de 0,47% entre o 3° e o 1° teste.
- O recall aumenta em 0,26% entre o 1° e 2° teste, em torno de 0,59% do 2° para o 3°, e a diferença é de 0,85% entre o 3° e o 1° teste.

- O f1-score tem um decréscimo de 1,06% entre o 1° e 2° teste, um acréscimo de 0,35% do 2° para o 3°, e uma diminuição de 0,47% entre o 3° e o 1° teste, com o melhor valor de métrica para a 1° série;
- A mean IoU tem um decréscimo de 0,02% entre o 1° e 2° teste, um acréscimo de 0,29% do 2° para o 3°, e um aumento de 0,27% entre o 3° e o 1° teste, com o melhor valor de métrica para a 3° série.

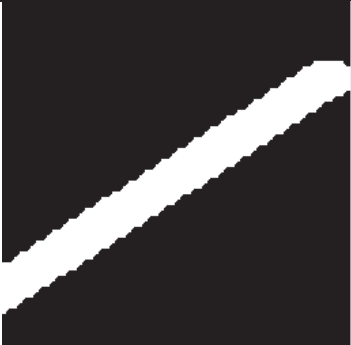
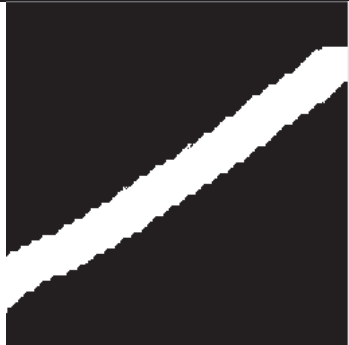
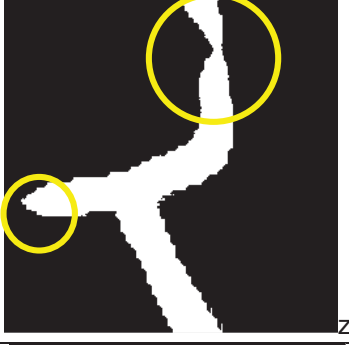
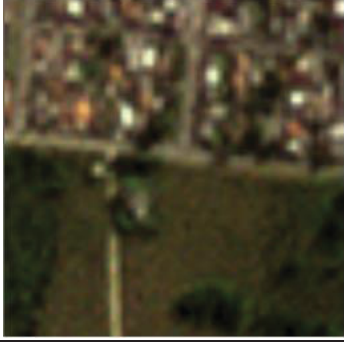
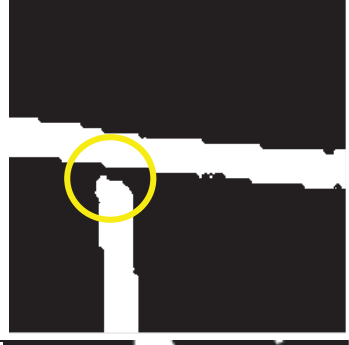
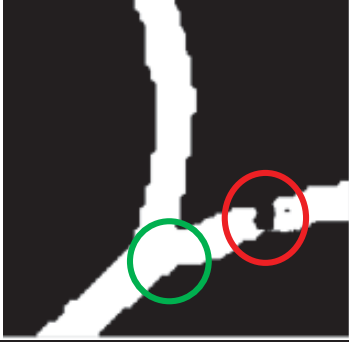
No entanto o fator tempo nesse experimento se mostrou extremamente problemático, pois o processo demandou em torno de cinco dias de processamento e obteve valores estatísticos semelhante aos obtidos nos experimentos anteriores, não sendo a melhor solução para o problema em questão. Considerando esses fatores, o modelo continua preciso, mantendo uma taxa de semelhança de pouco mais de 54% entre as máscaras previstas pelo modelo e a "verdade de campo".

No Quadro 7 são apresentados exemplos da saída produzida pela rede neural, composta por: (A) Um exemplo da região de entrada, (B) A máscara correspondente à verdade de campo para a região de entrada, e (C) O resultado produzido pela rede para esta série de experimentos.

Ao analisar a coluna C) do Quadro 7, pode-se notar exemplos bem-sucedidos para a terceira série de experimentos. Por exemplo, no exemplo 1 e 2, o resultado da rede coincide com a verdade de campo. No entanto, é notável que a rede falha em trechos que não estão bem delineados devido à presença de vegetação densa (exemplo 3, demarcados em amarelo). No exemplo 4, também demarcado em amarelo, a rede detectou a estrada na horizontal, mas falhou no entroncamento das estradas, possivelmente devido à vegetação sobre o trecho.

Outro exemplo (exemplo 5) mostra uma falha na demarcação dos pixels pertencentes à estrada no corpo da estrada secundária (demarcado em vermelho), no entanto, o entroncamento entre as estradas foi perfeitamente detectado (demarcado em verde), apesar da estrada secundaria não estar demarcada na verdade de campo.

QUADRO 7 – Resultado bem sucedido para terceira série de experimentos: A) exemplo de entrada; B) estrada; C) detecção resultante.

EX	A)	B)	C)
1			
2			
3			
4			
5			

Fonte: A autora (2023).

5.4 COMPARAÇÃO DOS RESULTADOS DOS EXPERIMENTOS

Ao comparar os três conjuntos de teste, nota-se diferenças estatísticas significativamente baixas entre os três experimentos. Em ordem decrescente, o maior resultado estatístico foi obtido usando 100 épocas e 32 neurônios, porém, os valores dos outros experimentos são muito próximos. Por exemplo a acurácia de menor valor, encontrado usando 50 épocas e 16 neurônios, é de 91,85%, sendo 0,2% inferior (Tabela 4).

Outra forma de avaliar a acurácia do produto é por meio da comparação do Índice de Jaccard (IoU). Neste caso, ao analisar a média IoU, o melhor resultado também corresponde ao Experimento 2 com 100 iterações e 32 neurônios. Essa estatística avalia a semelhança entre as respostas da detecção e a máscara de "verdade de campo" e apresenta um valor relativamente baixo na Tabela 4, indicando que os pixels detectados podem não estar bem demarcados.

TABELA 4 – Comparação das métricas resultantes para os experimentos.

	50 ÉPOCAS		100 ÉPOCAS		300 ÉPOCAS
Neurônios	16	32	16	32	16
Acurácia	91.85%	91.95%	91.89%	92.05%	91.90%
Recall	63.71%	65.92%	63.97%	66.30%	64.56%
F1-Score	68.99%	69.07%	67.93%	69.48%	68.28%
Precisão	75.80%	75.83%	76.22%	76.66%	76.27%
Mean IoU	53.94%	55.01%	53.92%	55.56%	54.21%
Tempo	04h:53min	06h:08min	08h:32min	10h:21min	135h (05 dias, 15h)

Fonte: A autora (2023).

Do ponto de vista estatístico, os resultados mais significativos foram alcançados após 100 épocas de treinamento e a utilização de 32 neurônios. A precisão atingiu 76,66%, uma melhoria de 0,86% em relação ao teste menos eficiente, que empregou 50 épocas e 16 neurônios. O mesmo padrão de melhoria se repetiu em outras métricas, como o Recall, que registrou uma taxa de 66,30% uma diferença considerável de 2,59%, em relação ao teste menos eficiente com 63,71%, indicando o quão bem as "estradas verdadeiras" foram identificadas.

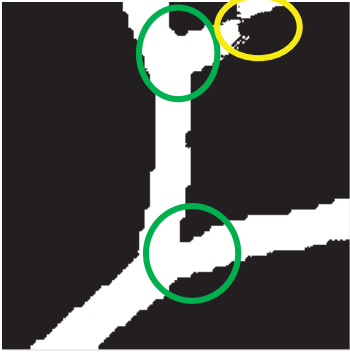
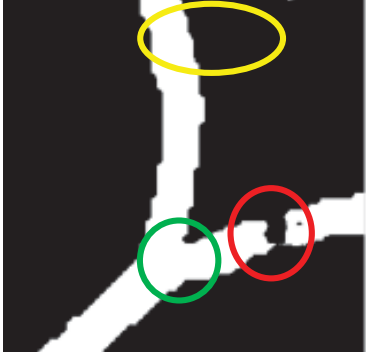
Entretanto, é importante mencionar que a diferença observada no melhor F1-score também foi notável. O teste com 100 épocas e 32 neurônios apresentou um F1-

score (69.48%) superior em 1,55% em relação ao resultado menos favorável obtido com 100 épocas e 16 neurônios (67,93%).

Ao contrário de outras estatísticas, à medida que a variável tempo aumenta, o processo torna-se mais moroso. Nesse sentido, a configuração mais eficiente foi utilizando 50 épocas e 16 neurônios, demandando apenas 4 horas para completar o processo. Em contrapartida, a abordagem mais demorada envolveu 300 épocas e 16 neurônios, mas não proporcionou resultados distintos em comparação com os demais experimentos (Tabela 4).

No entanto, ao analisar visualmente as imagens resultantes (Quadro 8), percebe-se que o valor estatisticamente baixo se justifica porque as estradas não foram marcadas na máscara da "verdade de campo", mas foram bem detectadas no processo.

QUADRO 8 – Primeira comparação visual entre as series de experimentos: exemplos de entrada na rede e máscaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos.

EXEMPLO DE ENTRADA		MASCARA DA VERDADE DE CAMPO	
			
A)	B)	C)	
			

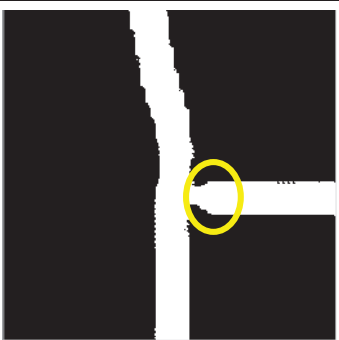
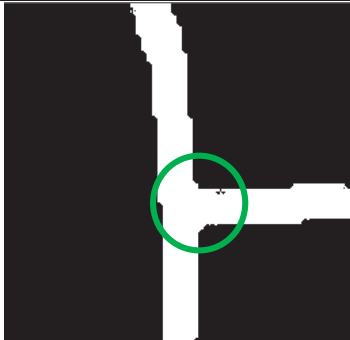
Fonte: A autora (2023).

No Quadro 8, as respostas visuais para as três séries de experimentos foram categorizadas da seguinte forma: detecções errôneas em vermelho, pixels detectados

corretamente, mas com alguma falha ou descontinuidade, em amarelo, e respostas bem-sucedidas em verde. A comparação visual das respostas obtidas pelos três conjuntos de teste revela que os resultados da segunda série de teste apresentaram uma melhor demarcação de pixels para a estrada, especialmente na região de entroncamento entre duas estradas.

O Quadro 8 apresenta a comparação visual para as respostas obtidas no experimento 1, é notável que os três produtos detectaram trechos de estradas secundárias não demarcados na máscara da verdade de campo, o que é considerado um ponto positivo. Entretanto, no experimento 1, apresentado em (A), ocorreu a detecção de uma pequena mancha (demarcada em vermelho) que não pertence à estrada, e isso pode ser atribuído à confusão com outras feições, como telhados de edificações ou solo exposto, presentes na imagem de entrada.

QUADRO 9 – Segunda comparação visual entre as series de experimentos: exemplos de entrada na rede e máscaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos.

EXEMPLO DE ENTRADA		MASCARA DA VERDADE DE CAMPO	
			
A)	B)	C)	
			

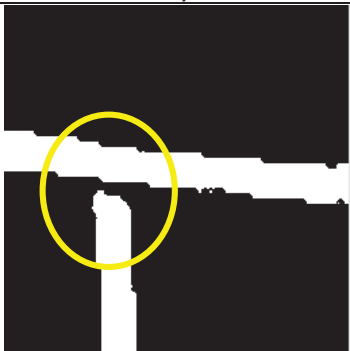
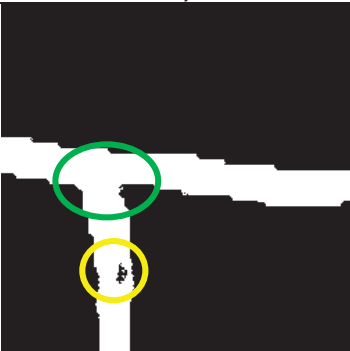
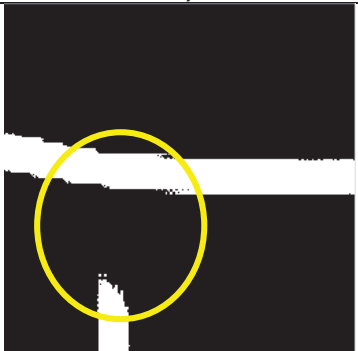
Fonte: A autora (2023).

Para o experimento 2 (B), apesar de detectar as estradas secundárias (demarcadas em verde), há uma descontinuidade em uma das estradas, sugerindo que o modelo ainda precisa de melhorias para uma detecção mais precisa. No

experimento 3 (C), uma das estradas secundárias não foi detectada (demarcada em amarelo), enquanto a outra, embora tenha sido detectada, apresentou descontinuidade nesse trecho (demarcado em vermelho). Esse resultado também indica a necessidade de aprimoramentos no modelo para melhorar a detecção desses trechos.

No Quadro 9, é apresentado um segundo exemplo para as três séries de experimentos. Observa-se que todas as séries detectaram a estrada secundária. No entanto, as séries de experimentos 1 (A) e 3 (C) apresentaram uma falha (demarcado em amarelo) na região do entroncamento das vias, enquanto somente a série de experimentos 2 (B) não apresentou um gargalo ou descontinuidade na detecção do entroncamento dos trechos.

QUADRO 10 – Terceira comparação visual entre as series de experimentos: exemplos de entrada na rede e marcaras da verdade de campo correlacionadas com exemplos de detectados, A) Detecção da primeira serie de experimentos; B) Detecção da segunda serie de experimentos; e C) Detecção da terceira serie de experimentos.

EXEMPLO DE ENTRADA		MASCARA DA VERDADE DE CAMPO	
			
A)	B)	C)	
			

Fonte: A autora (2023).

No Quadro 10, é apresentado um terceiro exemplo para as três séries de experimentos. Observa-se que todas as séries detectaram a estrada secundária, essa na vertical da imagem. No entanto, as séries de experimentos 1 (A) e 3 (C) apresentaram uma falha (demarcado em amarelo) na região do entroncamento das

vias, tendo como possível causa para esta falha a presença de vegetação sobre essa parte do trecho, enquanto somente a série de experimentos 2 (B) apresenta o entroncamento completo (demarcado em verde), mas apresenta uma pequena descontinuidade em uma parte da estrada secundária detectada.

Quanto à complexidade do modelo, ao aumentar o número de neurônios por camada de 32 para 64, ou aumentar o número de camadas, foram identificados problemas tanto no ambiente virtual do *Colab* quanto no processamento no workstation. Em ambos os ambientes, ocorreram mensagens de extrapolação do uso da memória.

No ambiente *Colab* ao tentar executar o processamento com mais neurônios e/ou mais camadas, a plataforma emite a mensagem de extrapolação do uso da GPU, ou seja, o processamento não é liberado no acesso gratuito. O aumento do número de neurônios o do número de camadas também causa problemas semelhantes quando se usa o workstation, estando relacionados ao kernel da máquina. Esse erro está associado ao uso da memória dedicada da GPU, condizente a extrapolação do uso da memória.

Quando analisado o tempo de processamento, usando a plataforma do *Google Colab*, com 50 épocas, o tempo total de processamento, incluindo treinamento e teste, variou entre quatro e cinco horas, ao aumentar o número de épocas para 100 o tempo variou entre nove e 10 horas para o processamento completo, se mostrando mais eficiente quando analisado o fator tempo de processamento. Enquanto no processamento em workstation foi muito mais lento, com duração em torno de 16 horas para 50 épocas e, entre 26 e 28 horas para 100 épocas, com GPU funcionando. Quando se usou apenas a CPU da máquina, esse processo passou para questão de dias, sendo 50 épocas em torno de 50 horas (2 dias), para 100 épocas foram gastos 71 horas (2 dias e 23 horas) e para 300 épocas o tempo de processamento salta para 135h (05 dias e 15 horas).

Apesar do uso de GPU, a workstation utilizava apenas de 5 a 12% de sua memória durante o processamento. Mesmo assim, o tempo demandado foi significativo, e os resultados obtidos foram semelhantes aos alcançados no ambiente *Colab*. Isso se apresenta como um ponto desafiador para a implementação da U-Net em um ambiente local.

5.5 AVALIAÇÃO DO MODELO

Em um último teste de avaliação o modelo *U-Net* foi usado para varrer uma imagem totalmente desconhecida para o treinamento ou verificação. Foi usada uma imagem Planet (L15-0744E-0875N) similar, e nela foi digitalizada a rede de estradas rurais.

No processo, a imagem foi varrida em intervalos regulares e a cada passo um recorte da imagem original foi obtido (tamanho 40 x 40). Este recorte poderia ou não conter estradas, e foi entregue à rede *U-Net* para verificar a presença de estradas. O resultado é um novo recorte, do mesmo tamanho, contendo a predição do modelo. Estes resultados foram juntados para compor um mosaico de resultados. Como a imagem completa é demasiado grande (4096 x 4096 pixels), apenas dois recortes de 400x700 serão apresentados para discutir alguns aspectos. Esses recortes são mostrados nas Figuras 17 e 19. A predição da rede pode produzir verdadeiros positivos, quando a estrada é detectada, falsos positivos, quando um trecho de estrada é detectado em uma região onde ela não existe, verdadeiros negativos correspondem aos recortes que não contem estradas e nos que a rede não detecta estrada, e falsos negativos, quando o modelo acha estrada em um local onde ela não existe.

A comparação da digitalização (em azul) e a predição (em Vermelho) do primeiro recorte é mostrado na Figura 18. Nesta Figura, preto demarca as áreas em comum, quando a predição bate com a digitalização.

Na Figura 17, alguns pontos foram demarcados para discutir a comparação. Por exemplo, quando a estrada está bem destacada do fundo, como é o caso “D” na Figura 17, a predição tem alta chance de coincidir com a digitalização, situação denotada em preto na Figura 18, embora se note falta de continuidade na predição.

Ocorrem erros perto de cidades ou áreas urbanizadas, visto que neste caso existem múltiplas possíveis soluções porque a rede de estradas é mais densa e o fundo é mais parecido com a cor das estradas. Outros erros, como o caso “A” na Figura 17, mostra situações em que a estrada não é bem visível devido à topografia ou à ocorrência de copas de árvores ou sombra cobrindo a faixa de rolagem. Este erro era já esperado, pois as amostras de treinamento mostravam principalmente situações em que a estrada se destaca em relação a sua vizinhança.

O modelo também caracteriza como estrada pequenas manchas de solo, mas isto não afeta significativamente o resultado, pois são áreas pequenas e isoladas na imagem (pequenas regiões em vermelho). Porém, nota-se que a rede foi capaz de detectar trechos de estrada que não foram digitalizados, como as áreas marcadas com “C” na imagem.

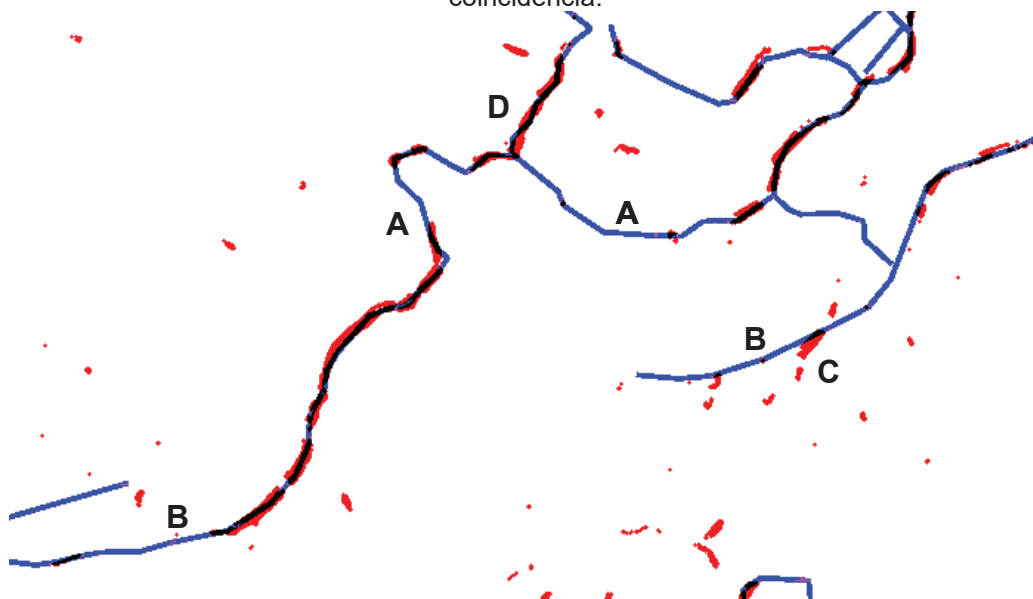
Um fato negativo é o caso “B”, pois neste caso a estrada está aparentemente bem delineada, mas a rede falha na predição.

FIGURA 17 – Primeiro Recorte de imagem Planet para teste.



Fonte: A autora (2023).

FIGURA 18 – Comparação do primeiro recorte: vermelho predição; azul digitalização; preto coincidência.



Fonte: a autora (2023).

Embora seja possível calcular as taxas de acertos e erros para compor a matriz de confusão, é importante analisar essa estatística com cautela, como evidenciado nas Figuras 17 (digitalização) e 18 (predição), nota-se que algumas das predições estão ligeiramente deslocadas em relação à digitalização.

Na Tabela 5 mostra a quantidade de pixels que foram classificados corretamente e falsamente, incluindo o fundo como uma classe. Com base nos valores da Tabela, a acurácia global atinge um valor consideravelmente alto de 0.981. No entanto, é importante ressaltar que esse valor é influenciado significativamente pela grande quantidade de pixels classificados corretamente como "fundo" (0), que não foram incluídos corretamente na classe "estrada" (1).

Analisando a acurácia do produtor para a classe estrada, observa-se que a taxa de acerto é baixa, beirando 30%. Embora a acurácia do usuário para a mesma classe ainda seja baixa, em torno de 40%, pode-se dizer que as predições são boas, a inclusão de pixels de "fundo" na classe de estradas ocorre de forma isolada e concentrada, como caracterizam as áreas em vermelho da Figura 18. Ou seja, as predições do modelo, quando produzem regiões maiores, são consistentes. Os erros ocorrem sob forma de regiões menores e geralmente isoladas.

TABELA 5 – Matriz de confusão da região L: 1501 - 1900, C: 1701 – 2400. AU. Acurácia do usuário. AP. Acurácia do produtor.

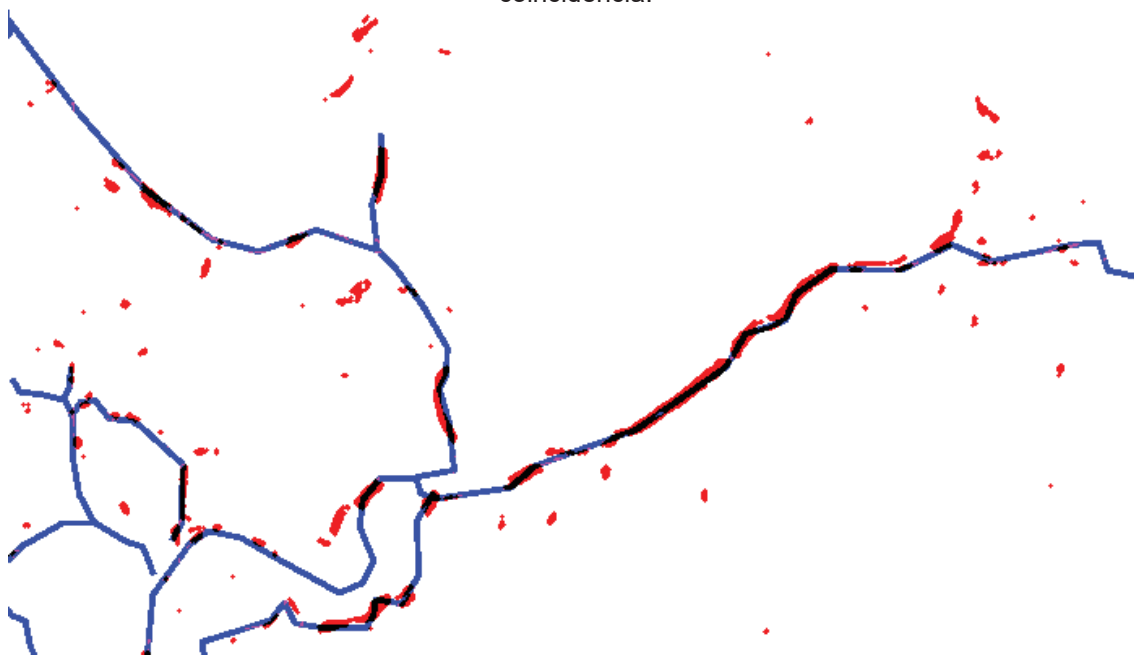
CLASSE	0	1	SOMA	AP
0	268718	3497	272215	0.987
1	5280	2505	7785	0.322
1	273998	6002		
AU	0.981	0.417		

Fonte: A autora (2023).

Uma situação similar pode ser constatada no segundo recorte, mostrado nas Figuras 19 e 20. Neste caso, a acurácia global que pode ser calculada da matriz de confusão respectiva (Tabela 2) também é alta (0.967) mas, novamente, este valor é causado pela grande quantidade de pixels de "fundo" em relação aos de estrada. Por ter menos pixels de estradas, a acurácia do usuário é ainda menor.

FIGURA 19 – Segundo recorte de imagem Planet para teste.

Fonte: A autora (2023).

FIGURA 20 – Comparação do segundo recorte: vermelho predição; azul digitalização; preto coincidência.

Fonte: a autora (2023).

TABELA 6 – Matriz de confusão da região L: 2301 - 3000, C: 501 - 900. AU. Acurácia do usuário. AP. Acurácia do produtor.

CLASSE	0	1	SOMA	AP
0	271002	2439	273441	0.991
1	6229	330	6559	0.050
1	277231	2769		
AU	0.979	0.368		

Fonte: a autora (2023).

Esta comparação mostra que o modelo é capaz de fazer boas previsões da classe estrada, mas que não é capaz, ainda, de mapear toda a rede de estradas devido aos problemas de oclusão ou até falta de amostras representativas. Tem que se considerar que a forma e aparência de trechos de estradas é muito variada e o modelo foi treinado com um restrito número de amostras. Aumentando a quantidade de amostras, a taxa de sucesso nestas situações reais pode aumentar.

6 CONCLUSÃO

A U-Net mostrou ser uma ferramenta viável para detecção de estradas rurais em imagens da constelação Planet, alcançando a taxa de acerto em torno de 92%. No entanto, a precisão foi relativamente baixa (76%) e f1-score em torno de 69%, a qualidade do modelo ainda necessita-se melhorar.

O sucesso da rede depende muito da disponibilidade de uma grande quantidade de amostras de treinamento e da melhor delimitação das amostras permitindo cobrir todos os cenários possíveis, refletindo as diferentes aparências das estradas. Para esse trabalho foram usadas cerca de nove mil amostras de trecho de estrada. No entanto ainda se nota a necessidade de mais amostras, vista que trechos retilíneos e cruzamentos são os com maior taxa de sucesso. Problemas podem ser esperados quando os trechos são sinuosos ou quando ocorrem outros elementos que escondem as estradas.

Nas áreas rurais predomina a vegetação, por isso, os erros mais frequentes são associados à presença de vegetação. A rede apresenta baixo desempenho quando ocorrem sombras causadas por árvores formando trechos retilíneos, pois tais situações são facilmente confundidos com trechos de estradas. Por outro lado, em regiões onde predomina apenas vegetação, a confusão é mínima.

Apesar das limitações causadas pelo uso limitado da memória GPU na plataforma Google Colab, este ambiente virtual se mostra mais eficiente para o uso das redes convolucionais em relação ao processamento em máquina local, quando o desktop não é munido de grandes recursos em termos de placa gráfica e memória. O uso do ambiente Colab também implica menor investimento em recursos de informática, em vista que a plataforma é virtual e de acesso remoto via web, enquanto a uma máquina local precisa ser reforçada com placas de vídeo e memória suficiente, aumentando assim o custo da implementação de um laboratório adequado.

No entanto, o ambiente Colab também tem algumas limitações de uso, especialmente quando se usa a opção gratuita. Quando se aumenta a complexidade da rede com a mudança de parâmetros do número de neurônios por camadas, números de camadas e número de operações de convolução, ou quando ocorre o gasto das células de processamento liberadas para uso gratuito ocorre a interrupção do uso da memória GPU por parte da plataforma por 24 horas. Vale ressaltar que o aumento no tamanho do dataset também aumenta a demanda de tempo, que também

ocasiona na interrupção do uso da plataforma, pois apenas 12 horas por dia são disponíveis de maneira gratuita. Tais limitações não ocorrem quando se usa uma máquina local.

6.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Para trabalhos futuros, recomenda-se efetuar novos experimentos destinados a avaliar formas de otimizar o tempo de processamento, selecionando criteriosamente os parâmetros de entrada. É recomendado também incluir métodos de pós-processamento, como filtros espaciais ou análise de contexto, a fim de refinar o resultado da rede.

Uma extensão interessante deste estudo seria aplicar o modelo treinado em diferentes cenários geográficos, considerando variações na topografia, uso da terra e meio ambiente. Isso permitiria uma avaliação mais abrangente do desempenho em diferentes contextos. Recomenda-se também avaliar este tipo de rede com outros conjuntos de dados, por exemplo imagens aéreas ou orbitais com resolução próxima daquela usada neste estudo. Amostras de trechos de estradas oriundos de outros locais e regiões refletiriam outras situações comuns na superfície da Terra, para uma análise mais abrangente.

É também recomendado comparar o desempenho da rede U-Net com outras arquiteturas de redes neurais convolucionais usando diferentes conjuntos de dados, para se verificar suas vantagens e desvantagens. Isto não foi possível no presente estudo devido às limitações de tempo.

Em termos do esforço computacional, para melhorar a eficiência de tempo de processamento e a capacidade de processamento, vale a pena investigar o uso de recursos de hardware mais avançados, como GPUs mais avançados e com alto desempenho ou sistemas distribuídos, para acelerar o treinamento e a inferência da rede.

REFERÊNCIAS

- ABDOLLAHI, A. et al. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. **Remote Sensing**, v. 12, n. 9, 2020.
- AHMED, A. S. Comparative Study Among Sobel, Prewitt and Canny Edge Detection Operators Used in Image Processing. **Journal of Theoretical and Applied Information Technology**, v. 96, n. 19, p. 6517–6525, 2018.
- ANIL, P. N.; NATARAJAN, S. **A novel approach using active contour model for semi-automatic road extraction from high resolution satellite imagery**. ICMLC 2010 - The 2nd International Conference on Machine Learning and Computing. **Anais...**2010. Disponível em: <DOI 10.1109/ICMLC.2010.36>. Acesso em: 1 maio. 2022
- ANIL, P. N.; NATARAJAN, S. A novel approach using active contour model for semi-automatic road extraction from high resolution satellite imagery. **ICMLC 2010 - The 2nd International Conference on Machine Learning and Computing**, p. 263–266, 2010b.
- BENGIO, Y.; GOODFELLOW, I.; COURVILLE, A. **Deep Learning**. Academia ed. [s.l.: s.n.]. v. 1
- BORDE, H. S. DE O.; SONDAK, D.; PROTOPAPAS, P. Convolutional Neural Network Models and Interpretability for the Anisotropic Reynolds Stress Tensor in Turbulent One-dimensional Flows. 29 jun. 2021.
- CAZES, T. B. et al. **Deteccão automática de estradas não pavimentadas em imagens de média resolução**. Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, INPE. **Anais...**2005a.
- CAZES, T. BROERMAN. et al. Deteccão automática de estradas não pavimentadas em imagens de media resolução. **Anais XII Simpósio Brasileiro de Sensoriamento Remoto. Goiânia. Brasil. 16-21 abril 2005. INPE. p. 3511-3518.**, p. 3511–3518, 2005b.
- CHAMBON, S. Detection of points of interest for geodesic contours: Application on road images for crack detection. **VISAPP 2011 - Proceedings of the International Conference on Computer Vision Theory and Application**, p. 210–213, 2011.
- CHEN, J.; RAN, X. Deep Learning With Edge Computing: A Review. **Proceedings of the IEEE Xplore**, p. 1–19, 2019.
- CHEN, J.; XUKAN, R. Deep Learning With Edge Computing: A Review. **Proceedings of the IEEE**, n. 1558- 2256 c? 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See, p. 1–19, 2019.
- CHENG, G. et al. Urban road extraction via graph cuts based probability propagation. **2014 IEEE International Conference on Image Processing, ICIP 2014**, v. 1, p. 5072–5076, 2014.

CHENG, G. et al. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 6, p. 3322–3337, 1 jun. 2017.

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

CHOI, J. H.; SONG, G. Y.; LEE, J. W. Combustion and Emission Characteristics of Wood Pyrolysis Oil-Butanol Blended Fuels in a Di Diesel Engine. **International Journal of Automotive Technology**, v. 13, n. 6, p. 941–948, 2012a.

CHOI, J. H.; SONG, G. Y.; LEE, J. W. ROAD IDENTIFICATION IN MONOCULAR COLOR IMAGES USING RANDOM FOREST AND COLOR CORRELOGRAM. **International Journal of Automotive Technology**, v. 13, n. 6, p. 941–948, 2012b.

DAL POZ, A. P.; ZANIN, R. B.; VALE, G. M. DO. EXTRAÇÃO AUTOMÁTICA DE FEIÇÕES RODOVIÁRIAS EM IMAGENS DIGITAIS. **Revista Controle & Automação/Vol.18 no.1/Janeiro, Fevereiro e Março**, 2007.

DANG, K. B. et al. A Convolutional Neural Network for Coastal Classification Based on ALOS and NOAA Satellite Data. **IEEE Access**, v. 8, p. 11824–11839, 2020.

DAS, S.; MIRNALINEE, T. T.; VARGHESE, K. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 49, n. 10, p. 3906–3931, 2011.

DUNG, C. V.; ANH, L. D. Autonomous concrete crack detection using deep fully convolutional neural network. **Automation in Construction**, v. 99, p. 52–58, 1 mar. 2019.

ELBASHIR, M. K. et al. Lightweight convolutional neural network for breast cancer classification using RNA-Seq gene expression data. **IEEE Access**, v. 7, p. 185338–185348, 2019.

Fan, L.; Zhao, H.; Li, Y.; Li, S.; Zhou, R.; Chu, W. RAO-UNet: a residual attention and octave UNet for road crack detection via balance loss. *IET Intell. Transp. Syst.* 16, 332– 343 (2022). <https://doi.org/10.1049/itr2.12146>.
<https://ietresearch.onlinelibrary.wiley.com/doi/epdf/10.1049/itr2.12146>

FAKHRI, S. A.; SHAH-HOSSEINI, R. Improved Road Detection Algorithm Based on Fusion of Deep Convolutional Neural Networks and Random Forest Classifier on VHR Remotely-Sensed Images. **Journal of the Indian Society of Remote Sensing**, p. 1–13, 1 ago. 2022.

FIORIN, D. V. et al. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, v. 33, n. 1, 2011.

FLAMIA, R. D. et al. **DETERMINAÇÃO DA QUALIDADE DE FORMAÇÃO DO PAPEL RECICLADO UTILIZANDO REDES NEURAS CONVOLUCIONAIS**. [s.l.: s.n.].

GALLIS, R. B. D. A. et al. EXTRAÇÃO SEMI-AUTOMÁTICA DE RODOVIAS ATRAVÉS DE OTIMIZAÇÃO POR PROGRAMAÇÃO DINÂMICA NO ESPAÇO OBJETO : CASO MONO Semi-automatic road extraction by dynamic programming optimization in the object space : single image case Universidade Estadual Paulista – Un. **RBC - Revista Brasileira de Cartografia** Nº, v. 57, n. 03, p. 168–179, 2005a.

GALLIS, R. B. DE A. et al. EXTRAÇÃO SEMI-AUTOMÁTICA DE RODOVIAS ATRAVÉS DE OTIMIZAÇÃO POR PROGRAMAÇÃO DINÂMICA NO ESPAÇO OBJETO: CASO MONO. **RC - Revista Brasileira de Cartografia**, v. 57, n. 3, p. 168–180, 2005b.

GRUEN, A.; LI, H. Road extraction from aerial and satellite images by dynamic programming. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 50, n. 4, p. 11–20, 1995.

Giurgi, D. -V. .; Josso-Laurain, T.; Devanne, M.; Lauffenburger, J. -P. "Real-time road detection implementation of UNet architecture for autonomous driving," 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 2022, pp. 1-5, doi: 10.1109/IVMSP54334.2022.9816237. <https://ieeexplore.ieee.org/abstract/document/9816237>

GUIDICI, D.; CLARK, M. L. One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. **Remote Sensing**, v. 9, n. 6, 1 jun. 2017.

HE, K.; LU, Y.; SCLAROFF, S. Local Descriptors Optimized for Average Precision. **IEEE Xplore**, 2018.

HOSSAIN, M. D.; CHEN, D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 150, p. 115–134, 2019.

HOU, Y. et al. C-U-Net: Complement U-Net for remote sensing road extraction. **Sensors**, v. 21, n. 6, p. 1–21, 2 mar. 2021.

HUANG, B.; ZHAO, B.; SONG, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. **Remote Sensing of Environment**, v. 214, p. 73–86, 1 set. 2018.

JIA, F. et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. **Mechanical Systems and Signal Processing**, v. 110, p. 349–367, 15 set. 2018.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and Electronics in Agriculture**, v. 147, n. July 2017, p. 70–90, 2018.

KASS, M.; WITKIN, A.; TERZOPOULOS, D. Snakes: Active contour models. **International Journal of Computer Vision**, v. 1, n. 4, p. 321–331, 1988.

LI, H.; HE, H.; WEN, Y. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. **Optik**, v. 126, n. 24, p. 4817–4822, 2015.

LIAN, R. et al. Road extraction methods in high-resolution remote sensing images: A comprehensive review. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 5489–5507, 2020a.

LIAN, R. et al. **Road extraction methods in high-resolution remote sensing images: A comprehensive review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**Institute of Electrical and Electronics Engineers Inc., , 2020b.

LIAN, R.; HUANG, L. DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 1905–1916, 2020a.

LIAN, R.; HUANG, L. DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 1905–1916, 2020b.

LIMA, K. K. S.; MENEZES, M. DA S. **DESENVOLVIMENTO E COMPARAÇÃO DE REDES NEURAIS CONVOLUCIONAIS PARA CLASSIFICAÇÃO DE OBJETOS.** [s.l: s.n.].

LIU, Y. et al. **Relation-shape convolutional neural network for point cloud analysis.** Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Anais...**2019a.

LIU, Y. et al. **Relation-Shape Convolutional Neural Network for Point Cloud Analysis.** IEEE Xplore . **Anais...**2019b.

LUNA, D. et al. **XI ENCONTRO ACADÊMICO COMPUTACIONAL MODELAGEM Reposicionamento para um Robô de Sondagem utilizando Redes Convolucionais.** [s.l: s.n.].

MA, L. et al. Deep learning in remote sensing applications: A meta-analysis and review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 152, n. April, p. 166–177, 2019.

MEDHI, A.; SAHA, A. K. Rural Road Extraction using Object Based Image Analysis (OBIA): A case study from Assam, India. **INTERNATIONAL CARTOGRAPHIC ASSOCIATION**, v. 1, p. 1–8, 2019a.

MEDHI, A.; SAHA, A. K. Rural Road Extraction using Object Based Image Analysis (OBIA): A case study from Assam, India. **INTERNATIONAL CARTOGRAPHIC ASSOCIATION**, v. 1, p. 1–8, 2019b.

MEI, A. VAN DER; DOOMERNIK, J. P. **Artificial intelligence potential in power distribution system planning.** CIRED - Open Access Proceedings Journal. **Anais...**Institution of Engineering and Technology, 1 out. 2018.

MENDES, T. . S. G; FAZAN, A. J.; POZ, A. P. D. Delineamento De Estradas Vicinais Usando Técnicas De Correlação De Imagens. **Boletim de Ciências Geodésicas Ciênc Geod sec**, v. 10, n. 2, p. 159–176, 2004.

MENDES, T. S. G.; FAZAN, A. J.; DAL POZ, A. P. DELINEAMENTO DE ESTRADAS VICINAIS USANDO TÉCNICAS DE CORRELAÇÃO DE IMAGENS. **Boletim de ciências Geodésicas**, v. 10, n. 2, p. 159–176, 2004.

MIAO, Z. et al. A semi-automatic method for road centerline extraction from VHR images. **IEEE Geoscience and Remote Sensing Letters**, v. 11, n. 11, p. 1856–1860, 2014.

MIAO, Z. et al. An Object-Based Method for Road Network Extraction in VHR Satellite Images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 10, p. 4853–4862, 2015.

MOKHTARZADE, M.; ZOEJ, M. J. V. Road detection from high-resolution satellite images using artificial neural networks. **International Journal of Applied Earth Observation and Geoinformation**, v. 9, n. 1, p. 32–40, 2007.

OZKAN, C. et al. THE ARTIFICIAL BEE COLONY ALGORITHM IN TRAINING ARTIFICIAL NEURAL NETWORK FOR OIL SPILL DETECTION. **ICS**, v. 6, n. 11, p. 473–493, 2011.

REICHSTEIN, M. et al. Deep learning and process understanding for data-driven Earth system science. **Nature**, v. 566, n. 7743, p. 195–204, 14 fev. 2019.

REN, Y. et al. Full convolutional neural network based on multi-scale feature fusion for the class imbalance remote sensing image classification. **Remote Sensing**, v. 12, n. 21, p. 1–21, 1 nov. 2020.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional networks for biomedical image segmentation**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais...** Springer Verlag, 2015a.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf. **Computer Science Department and BIOS Centre for Biological Signalling Studies, University Of Freiburg, Germany.**, p. 234–241, 2015b.

SENADEERA, K. U. H. P. T. S. **AUTOMATION OF ROAD FEATURE EXTRACTION FROM HIGH RESOLUTION IMAGES.**, 2021.

SHAO, Z.; YANG, K.; ZHOU, W. **Correction: A benchmark dataset for performance evaluation of multi-label remote sensing image Retrieval [Remote Sens., 10, (2018), (964)] DOI:10.3390/rs10060964.** **Remote Sensing** MDPI, 1 ago. 2018.

SHARMA, A. et al. A patch-based convolutional neural network for remote sensing image classification. **Neural Networks**, v. 95, p. 19–28, 1 nov. 2017.

SHARMA, N.; SHARMA, R.; JINDAL, N. Machine Learning and Deep Learning Applications-A Vision. **Global Transitions Proceedings**, v. 2, n. 1, p. 24–28, jun. 2021.

SHEN, C. et al. **From machine learning to deep learning: Advances in scoring functions for protein–ligand docking.** Wiley Interdisciplinary Reviews: Computational Molecular Science Blackwell Publishing Inc., 1 jan. 2020.

SILVA, C. R.; CENTENO, J. A. S. Extração semiautomática de estradas vicinais com base em algoritmos genéticos. **Acta Scientiarum - Technology**, v. 32, n. 2, p. 137–145, 2010a.

SILVA, C. R.; CENTENO, J. A. S. EXTRAÇÃO AUTOMÁTICA DE ESTRADAS VICINAIS COM BASE NA Automatic Rural Road Extraction with Based in Radon Transformation Universidade Federal do Piauí – UFPI Universidade Federal do Paraná – UFPR. **Revista Brasileira de Cartografia (2010)**, v. Edição esp, n. 62, p. 441–452, 2010b.

SILVA, C. R.; CENTENO, J. A. S. Extração semiautomática de estradas vicinais com base em algoritmos genéticos. **Acta Scientiarum - Technology**, v. 32, n. 2, p. 137–145, 2010c.

SILVA, C. R. DA; CENTENO, J. A. S. **Determinação de larguras de estradas vicinais em imagens de alta resolução.** Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril. **Anais...**2009.

SILVA, C. R. DA; CENTENO, J. A. S. EXTRAÇÃO AUTOMÁTICA DE ESTRADAS VICINAIS COM BASE NA TRANSFORMADA RADON Automatic Rural Road Extraction with Based in Radon Transformation. **RBC - Revista Brasileira de Cartografia**, v. 62, n. Edição Especial, p. 441–452, 2010d.

SILVA, F. et al. Unidade Estadual do Rio Grande do Sul. n. 2000, p. 5001–5007, 2009.

SILVEIRA, G. P.; SHIGUEMORI, E. H.; GOMES, V. C. F. **USO DE REDE NEURAL ARTIFICIAL PARA A CLASSIFICAÇÃO DE VIAS TERRESTRES EM IMAGENS OBTIDAS POR VANTS.** An. Simp. Cient. Tecnol. IEAv. **Anais...**2016.

SONG, M.; CIVCO, D. Road extraction using SVM and image segmentation. **Photogrammetric Engineering and Remote Sensing**, v. 70, n. 12, p. 1365–1371, 2004.

STATELLA, T.; SILVA, E. A. DA. **MORFOLOGIA MATEMÁTICA: EXTRAÇÃO DE FEIÇÕES A PARTIR DE IMAGENS ORBITAIS.** [s.l: s.n.].

STATELLA, T.; SILVA, E. A. DA. Morfologia matemática: extração de feições a partir de imagens orbitais. **Anais do Simpósio Brasileiro de Geomática, Presidente Prudente**, p. 287–291, 2002b.

STATELLA, T.; SILVA, E. A. DA. Morfologia matemática: extração de feições a partir de imagens orbitais. **Anais do Simpósio Brasileiro de Geomática, Presidente Prudente**, p. 287–291, 2002c.

TAN, Y. Q. et al. **Vecroad: Point-based iterative graph exploration for road graphs extraction.** Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Anais...**IEEE Computer Society, 2020.

VENTURIERI, A.; ROBERTO, J. DOS S. CAPÍTULO 18 TÉCNICAS DE CLASSIFICAÇÃO DE IMAGENS PARA ANÁLISE DE COBERTURA VEGETAL. Em: **Técnicas de classificação de imagens para análise de cobertura vegetal**. [s.l.: s.n.]. v. inep. 351–371.

Wahbi, M.; Bakali, I. E.; Ez-zahouani, B.; Azmi, R.; Moujahid, A.; Zouiten, M.; Alaoui, O. Y.; Boulaassal, H.; Maatouk, M.; Kharki, O. E. A deep learning classification approach using high spatial satellite images for detection of built-up areas in rural zones: Case study of Souss-Massa region - Morocco, *Remote Sensing Applications: Society and Environment*, Volume 29, 2023.

WAN, Y. et al. A road extraction approach based on fuzzy logic for high-resolution multispectral data. **Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007**, v. 2, n. Fskd, p. 203–207, 2007.

Wang, L.; MA, X. -h.; Ye, Y. "Computer vision-based Road Crack Detection Using an Improved I-UNet Convolutional Networks," 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 2020, pp. 539-543, doi: 10.1109/CCDC49329.2020.9164476. <https://ieeexplore.ieee.org/abstract/document/9164476>

WANG, W. et al. **A review of road extraction from remote sensing images**. **Journal of Traffic and Transportation Engineering (English Edition)** Periodical Offices of Chang- a University, 1 jun. 2016.

WANG, Y. et al. Dense semantic labeling with atrous spatial pyramid pooling and decoder for high-resolution remote sensing imagery. **Remote Sensing**, v. 11, n. 1, 1 jan. 2019.

WEGNER, J. D.; MONTOYA-ZEGARRA, J. A.; SCHINDLER, K. A higher-order CRF model for road network extraction. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 1698–1705, 2013.

XU, M. et al. Anomaly Detection in Road Networks Using Sliding-Window Tensor Factorization. **IEEE Transactions on Intelligent Transportation Systems**, v. 20, n. 12, p. 4704–4713, 1 dez. 2019.

Yang, X.; Li, X.; Ye, Y.; Lau, R. Y. K.; Zhang, X.; Huang, X. "Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7209-7220, Sept. 2019, doi: 10.1109/TGRS.2019.2912301. <https://ieeexplore.ieee.org/abstract/document/8714072>

YANG, C.; ROTTENSTEINER, F.; HEIPKE, C. Towards better classification of land cover and land use based on convolutional neural networks. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives**, v. 42, n. 2/W13, p. 139–146, 2019.

YUAN, J. et al. A robust road segmentation method based on graph cut with learnable neighboring link weights. **2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014**, p. 1644–1649, 2014.

YUAN, Q. et al. Deep learning in environmental remote sensing: Achievements and challenges. **Remote Sensing of Environment**, v. 241, n. February, p. 24, 2020.

ZEDADRA, O. et al. Swarm intelligence-based algorithms within IoT-based systems: A review. **Journal of Parallel and Distributed Computing**, v. 122, p. 173–187, 2018.

ZHANG, L.; ZHANG, L.; DU, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. **IEEE Geoscience and Remote Sensing Magazine**, v. 4, n. 2, p. 22–40, 2016.

ZHANG, X.; ZHOU, X.; LIN, M. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. **IEEE Xplore**, p. 6848–6856, 2018.

Zhenqing Liu, Yiwon Cao, Yize Wang, Wei Wang, “Computer vision-based concrete crack detection using U-net fully convolutional networks, *Automation in Construction*, Volume 104, 2019, Pages 129-139.
<https://www.sciencedirect.com/science/article/pii/S0926580519301244>