UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Walysson Carlos dos Santos Oliveira

Segmentação Semântica de Áreas de Plantações Agrícolas via U-Net em Dois Estágios

São Luís - MA

2022

Walysson Carlos dos Santos Oliveira

Segmentação Semântica de Áreas de Plantações Agrícolas via U-Net em Dois Estágios

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Júnior

Coorientador: Prof. Dr. Daniel Lima Gomes Júnior

São Luís - MA

2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Carlos dos Santos Oliveira, Walysson.
Segmentação Semântica de Áreas de Plantações Agrícolas via U-Net em Dois Estágios / Walysson Carlos dos Santos Oliveira. - 2022.
75 p.

Coorientador(a): Daniel Lima Gomes Jr.
Orientador(a): Geraldo Braz Junior.
Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luís - Maranhão, 2022.

1. Agronegócio. 2. Deep Learning. 3. Monitoramento de Plantações. 4. Sensoriamento Remoto. I. Braz Junior, Geraldo. II. Lima Gomes Jr., Daniel. III. Título.

Walysson Carlos dos Santos Oliveira

Segmentação Semântica de Áreas de Plantações Agrícolas via U-Net em Dois Estágios

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho aprovado. São Luís - MA, 08 de Julho de 2022:

Prof. Dr. Geraldo Braz Júnior
Orientador
Universidade Federal do Maranhão

Prof. Dr. Daniel Lima Gomes Júnior
Coorientador
Instituto Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. Cláudio de Souza Baptista
Examinador Externo
Unversidade Federal de Campina Grande

São Luís - MA
2022

Dedico este trabalho a Henrique que até mesmo antes de nascer foi meu maior combustível para seguir em frente.

Agradecimentos

Os agradecimentos principais são direcionados à minha família, pelo amor e apoio incondicional, em especial a minha mãe Eliete e minha esposa Dayane, pelo suporte nas horas difíceis, de desânimo e cansaço.

A todos os meus professores, em especial, ao meu orientador Geraldo Braz Junior, pela orientação, pela paciência, por suas correções, incentivo, confiança e seu imenso empenho em ajudar.

Agradecimentos especiais são direcionados ao colega de Laboratório de Pesquisa Brenno Nascimento que auxiliou na árdua tarefa de anotação das imagens do conjunto de dados e aos meus colegas de trabalho Roberval Mariano, Gustavo Victorio, Victor Hugo e Adriano Rêgo que contribuíram com este trabalho me reservando uma parte do seu atarefado dia de trabalho e muito do seu conhecimento na área tributária.

"As ideias que defendo não são minhas. Eu as tomei emprestadas de Sócrates, eu as roubei de Chesterfield, eu as furtei de Jesus. E se você não gostar das ideias deles, quais seriam as ideias que você usaria?"

(Dale Carnegie)

Resumo

O imposto do agronegócio incide principalmente sobre a produção das safras agrícolas. Para reduzir a evasão fiscal no agronegócio, é possível monitorar o desenvolvimento dos plantios por meio da análise de imagens de satélite. Para isso, técnicas de *Deep Learning* podem ser aplicadas em imagens de satélite para segmentar a área plantada. A área segmentada, por sua vez, pode ser usada para estimar a produção dos estabelecimentos rurais monitorados. Este trabalho visa resolver a primeira etapa do problema, a segmentação da área plantada. Para isso, foi desenvolvida uma arquitetura de Rede Neural Convolucional para segmentação de áreas de plantação, a Two-stage U-net. Além disso, o trabalho também incluiu a criação de conjunto de dados de imagens de satélite com anotações de áreas de plantação. A arquitetura proposta foi treinada e seus hiperparâmetros foram ajustados considerando a rede *Encoder*, o Otimizador, a Função de Perda e o tamanho do Lote de imagens (*batch size*). Os resultados em $mIoU$ da Two-stage U-net se mostraram superiores aos resultados de outras arquiteturas utilizadas em trabalhos semelhantes.

Palavras-chave: Sensoriamento Remoto, Monitoramento de Plantações, *Deep Learning*, Agronegócio.

Abstract

The agribusiness tax is mainly levied on the production of agricultural crops. To reduce tax evasion in agribusiness, it is possible to monitor the development of plantations through the analysis of satellite images. For this, we can apply Machine Learning techniques to satellite images to segment the planted area, and the area, in turn, can be used to estimate the production of monitored plantations. This work aims to solve the first stage of the problem, the Segmentation of the Planted Area. For this, we developed a machine learning architecture for segmentation of plantation areas, the Two-stage U-net. In addition, the work also included the creation of a satellite image dataset with annotations for the segmentation of plantation areas. We trained the proposed model and we adjusted its hyperparameters considering the U-net Encoder, the Optimizer, the Loss Function, and the Batch Size of images. We selected the fitted model that performed best in tests with *Hyperopt* and *GridSearch*. The results in $mIoU$ of the Two-stage U-net were superior to the results of other architectures used in similar works.

Keywords: Remote Sensing, Plantations Monitoring, Deep Learning, Agribusiness.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ACC	Acurácia
ANN	*Artificial Neural Network*
API	*Application Programming Interface*
CNN	*Convolutional Neural Network*
ESA	*European Space Agency*
ESP	Especificidade
EVI	*Enhanced Vegetation Index*
FCN	*Fully Convolutional Network*
FN	Falso Negativo
FP	Falso Positivo
GEE	*Google Earth Engine*
IOU	*Intersection over Union*
MLP	*Multi-Layer Perceptron*
MSI	*MultiSpectral Instrument*
NASA	*National Aeronautics and Space Administration*
NDVI	*Normalized Difference Vegetation Index*
NIR	*Near Infrared*
ReLu	*Rectified Linear Unit*
RGB	*Red, Green, Blue*
ROI	*Region Of Interest*
SEFAZ-MA	*Secretaria de estado da Fazenda do Maranhão*
SEN	Sensibilidade
SGD	*Stochastic Gradient Descent*

SVM	*Support Vector Machine*
SWIR	*Short Wave Infrared*
USDA	*United States Departament of Agriculture*
VN	Verdadeiro Negativo
VNIR	*Visible and Near Infrared*
VP	Verdadeiro Positivo

Sumário

1 Introdução

O agronegócio brasileiro é uma atividade próspera e rentável. Área abundante, chuvas regulares, energia solar e clima diversificado fazem do Brasil um país com os requisitos naturais para a plantação. O agronegócio é responsável não somente por grande parte dos itens alimentares consumidos, como também, por uma cadeia de produção que envolve vários segmentos da economia. Em 2019, o setor representava 21,6% do PIB nacional, segundo o Ministério da Agricultura, Pecuária e Abastecimento (SILVA; CESARIO; CAVALCANTI, 2013).

Uma vez que o agronegócio é de suma importância na economia brasileira, é natural a participação da mesma quanto à contribuição para o funcionamento do Estado através de impostos e taxas advindas da sua atividade. Entretanto, o agronegócio apresenta um elevado índice de sonegação fiscal que ocorre em consequência da atual complexidade do sistema tributário brasileiro, além da dificuldade de fiscalização por parte das administrações tributárias, devido ao grande custo que essa fiscalização acarreta, inviabilizando sua execução (BRUGNARO; FILHO; BACHA, 2003).

O serviço público vem se modernizando. Não de forma tão ágil como o setor privado, mas projetos como o Governo Digital que visam acelerar a transformação digital no setor público, têm dado sua contribuição. Órgãos da Administração Tributária como Receita Federal, Secretarias da Fazendas Estaduais (SEFAZ) e Municipais (SEMFAZ) têm realizado cruzamento de grandes volumes de dados utilizando Big Data, Ciência de Dados e modelos de *Machine Learning* em dados como cartões de crédito, notas fiscais eletrônicas e outros documentos fiscais para identificar sonegações.

No contexto da tributação do agronegócio, a maior fatia da carga tributária é calculada sobre a produção das safras. A fim de reduzir a sonegação fiscal neste segmento do mercado, é possível estimar a produção das fazendas por meio do monitoramento e análise de imagens de satélite e comparar com os valores declarados pelo contribuinte. Uma abordagem proposta para atingir este fim é seguir os três passos seguintes: 1) aplicar técnicas de *Deep Learning* em imagens de satélite para segmentar a área cultivada das plantações; 2) aplicar técnicas de *Machine Learning* em imagens de satélite para classificar o tipo de cultura plantada na área segmentada; 3) utilizar a área segmentada e o tipo de plantação para estimar a produção das safras. Este trabalho irá se concentrar no primeiro passo deste processo, a segmentação da área de plantação. Para isso usaremos Segmentação Semântica com Redes Neurais Convolucionais.

Nos últimos anos, o aprendizado profundo com os algoritmos de redes neurais convolucionais, redes neurais recorrentes e redes adversárias generativas, tem sido amplamente

estudado e aplicado em vários campos com resultados promissores e grande potencial. Especificamente no agro, cada vez mais atenção tem sido dada à aplicação técnicas de *Deep Learning* na agricultura (ZHU et al., 2018) (KAMILARIS; PRENAFETA-BOLDÚ, 2018). Estudos recentes que se propõem a segmentar regiões da cobertura terrestre (RAKHLIN; DAVYDOW; NIKOLENKO, 2018), áreas de vegetação (YANG et al., 2020) e áreas de plantação (RUSTOWICZ et al., 2019) utilizando Redes Neurais Totalmente Convolucionais (LECUN; BENGIO; HINTON, 2015) para este propósito.

1.1 Objetivos

Este trabalho tem como objetivo analisar imagens de satélite de plantações e desenvolver uma arquitetura de aprendizagem de máquina para segmentar a área cultivada de plantações utilizando segmentação semântica com redes neurais convolucionais.

1.1.1 Objetivos específicos

Para alcançar o objetivo geral deste trabalho, alguns objetivos específicos deverão ser atingidos:

- Construir um conjunto de dados público de imagens de satélite com marcação de áreas plantações.
- Desenvolver uma arquitetura de Aprendizagem Profunda para segmentação da área cultivada de plantações em sua fase inicial, área de preparo para plantio, e na fase de crescimento da plantação.
- Comparar a arquitetura proposta com outras encontradas na literatura com respeito à métrica $mIoU$.

1.2 Contribuições

O presente estudo visa desenvolver uma arquitetura de aprendizagem de máquina para segmentar a área cultivada de plantações utilizando em imagens de satélite. Por se tratar de aplicação ainda pouco explorada no segmento do agronegócio, podemos elencar algumas contribuições que estão incluídas no presente trabalho, onde destacam-se como principais contribuições:

- Construção de um conjunto de dados de imagens de satélite, que pode auxiliar outros pesquisadores, com a anotação de oito classes identificadas na cobertura terrestre de regiões de plantações.

- Proposição de uma arquitetura de aprendizagem de máquina que auxilie governos, empresas e pessoas físicas na segmentação das áreas de plantação de estabelecimentos rurais.
- Possibilidade de aplicação da arquitetura proposta em outros problemas de contexto similar.

1.3 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 trata da fundamentação teórica das técnicas utilizadas. São abordados conceitos de Visão Computacional, Sensoriamento Remoto, Redes Neurais Artificiais, Aprendizagem Profunda com foco nas Redes Neurais Convolucionais e a compreensão a respeito da arquitetura de Rede Totalmente Convolucional U-net.
- O Capítulo 3 descreve trabalhos relacionados ao tema, os métodos, arquiteturas e conjunto de dados utilizados.
- O Capítulo 4 apresenta as etapas adotadas que compõem a metodologia proposta para este trabalho. São a Construção do *Dataset*, que está subdividida em Aquisição e Anotação das Imagens, Pré-processamento, Definição da arquitetura Two-stage U-net, Treinamento e Ajustes de Hiperparâmetros e Avaliação do modelo.
- O Capítulo 5 trata sobre os resultados obtidos e discussões em relação aos experimentos realizados na arquitetura proposta e comparação com resultados de outras arquiteturas.
- O Capítulo 6 apresenta as considerações finais sobre os resultados e trabalhos futuros e os artigos científicos desenvolvidos.

2 Fundamentação Teórica

Neste capítulo são apresentados os conceitos explorados para o desenvolvimento do estudo, eles são: Visão Computacional, Sensoriamento Remoto com Imagens de Satélite, Redes Neurais Artificiais, Aprendizagem Profunda com foco nas Redes Neurais Convolucionais e a compreensão a respeito da arquitetura de Rede Totalmente Convolucional U-net.

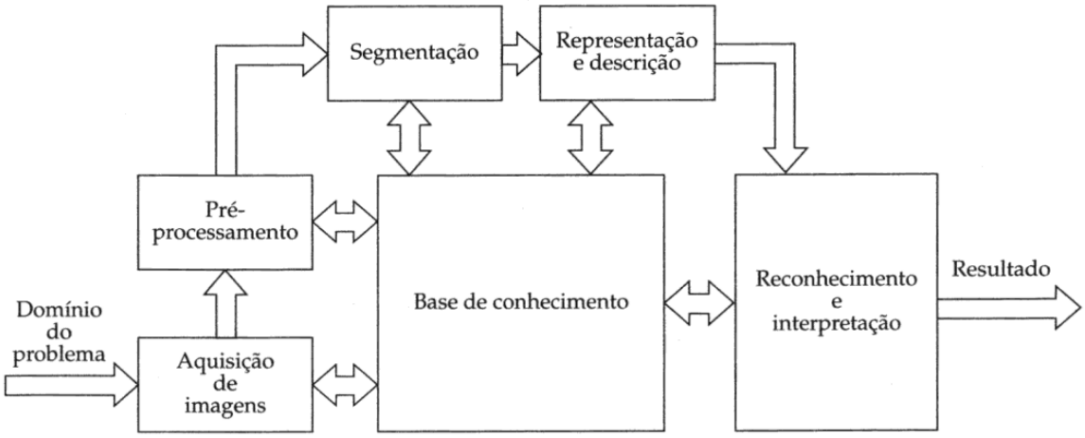
2.1 Visão Computacional e Processamento Digital de Imagens

Visão computacional é o campo da inteligência artificial dedicado à extração de informações a partir de imagens digitais (SANTOS et al., 2020). As técnicas de visão computacional são aplicadas em diversas áreas de pesquisa, resolvendo os problemas particulares de cada uma delas de forma específica. Tais sistemas são conhecidos como sistemas especialistas, que necessitam de conhecimento específico para a solução de um determinado problema (MILANO; HONORATO, 2014).

No contexto da agricultura digital, a visão computacional pode ser empregada na detecção de doenças e pragas, na estimação de safra e na avaliação não invasiva de atributos como qualidade, aparência e volume, além de ser componente essencial em sistemas robóticos agrícolas (SANTOS et al., 2020). Em visão computacional, geralmente são utilizadas diversas técnicas de processamento de imagens (CASTLEMAN, 1996). Gonzalez e Woods (2000) apontam que as etapas fundamentais para resolução de problemas de processamento de imagem são: aquisição das imagens digitais, pré-processamento, segmentação, representação, descrição, reconhecimento e interpretação conforme mostra a Figura 1.

Na etapa de Aquisição de Imagens ocorre a obtenção da representação digital da imagem, os seus *pixels*. Neste trabalho, a etapa de Aquisição de Imagens é realizada por meio de Sensoriamento Remoto utilizando imagens do Satélite Sentinel-2. A etapa de Pré-processamento consiste em melhorar a visualização ou realçar características presentes nas imagens. Neste trabalho, serão experimentados Índices de Vegetação para realçar caraterísticas das plantações. A Segmentação consiste no processo de isolar regiões de interesse específicas dentro da imagem. As regiões de interesse deste trabalho são áreas de plantações. A segmentação se dará, não por técnicas clássicas de processamento de imagens, mas sim por meio de Segmentação Semântica que é realizada por meio de Aprendizagem Profunda. Nas seções seguintes, estes conceitos serão detalhados. A iniciar pelo Sensoriamento Remoto.

Figura 1 – Passos fundamentais em processamento de imagens digitais



Fonte: (GONZALEZ; WOODS, 2000)

2.2 Sensoriamento Remoto

Sensoriamento remoto é definido como a aquisição de dados em diversas faixas do espectro eletromagnético obtidas a partir de sensores acoplados em aeronaves ou satélites, sendo que a análise dessa informação pode ser realizada visualmente ou via processamento digital de imagens (JENSEN; EPIPHANIO, 2009; GUIMARÃES, 2019).

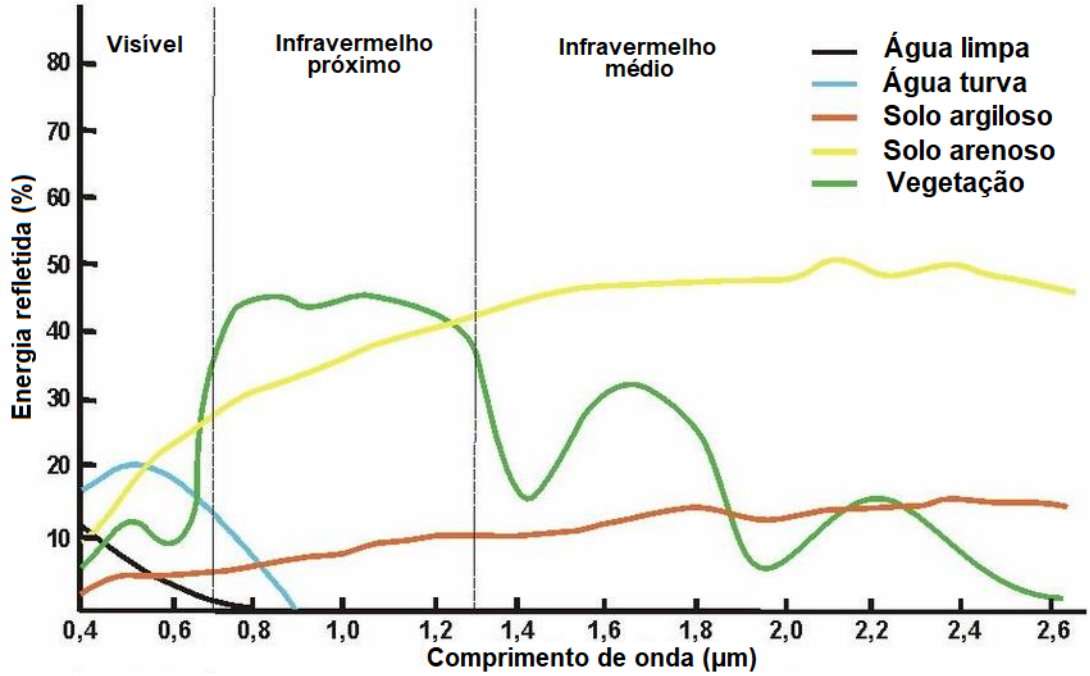
O sensoriamento remoto explora a relação entre as características do objeto alvo e a energia eletromagnética refletida, emitida ou espalhada por ele. Através da identificação do comportamento espectral de superfícies como o solo, água e vegetação, é possível relacionar informações qualitativas e quantitativas destes alvos (JENSEN; EPIPHANIO, 2009; GUIMARÃES, 2019). Assim, cada tipo de material ou superfície apresenta características únicas de comportamento espectral, ou seja, cada alvo é caracterizado pela sua assinatura espectral, a qual está diretamente ligada às suas propriedades físicas, químicas, biológicas ou geométricas (NOVO, 2010). A Figura 2 exemplifica o exposto, apresentando os diferentes comportamentos espectrais de alvos da superfície terrestre.

No contexto do Sensoriamento Remoto, dois temas são pertinentes para este trabalho que são: 1) As imagens obtidas via sensoriamento remoto pelo satélite Sentinel-2 e; 2) O pré-processamento utilizado para realçar a faixa do espetro relativo à vegetação, conforme a linha verde da Figura 2, onde são utilizados Índices de Vegetação.

2.2.1 Satélite Sentinel-2

As imagens de satélite se tornaram fontes importantes para o monitoramento da vegetação, fornecendo informações de forma rápida, principalmente em áreas de difícil acesso para pesquisas de campo devido a aspectos limitantes da topografia e vegetação densa (VARGA et al., 2015; MAIA, 2019). A maior vantagem da utilização das imagens

Figura 2 – Comportamento espectrais de alvos da superfície terrestre



Fonte: Adaptado de (FLORENZANO, 2002)

de sensores remotos, em especial na agricultura, está no fato da possibilidade de obtenção de informações e estimativas agronômicas sobre uma determinada cultura de interesse em extensas áreas e a baixos custos (ANTUNES; LAMPARELLI; RODRIGUES, 2015; MAIA, 2019). O satélite Sentinel-2 é um exemplo de satélite disponível publicamente.

A missão Sentinel (ESA, 2012) é uma frota de satélites designados especificamente para entregar uma riqueza de dados e imagens que são centrais para o programa Copernicus da Agência Espacial Europeia (ESA). Cada satélite integrante dessa missão possui diferentes tecnologias para atender diferentes demandas do programa Copernicus (SPOTO et al., 2012). Os satélites gêmeos, Sentinel-2A e Sentinel-2B, foram lançados em julho de 2015 e março de 2017, respectivamente, no foguete Vega na Guiana Francesa. O tempo de revisita de cada cena por cada satélite é de dez dias, e quando compostos pelo par o tempo de revisita diminui para cinco dias no equador. Como o satélite Landsat-8 dos EUA oferece imagens semelhantes e graças à cooperação entre a ESA e a NASA, o objetivo é reduzir isso para um tempo médio de revisitação de apenas três dias sobre o equador e gerar produtos de dados comparáveis (MIRANDA, 2019) .

A bordo de ambos os satélites Sentinel-2 está o instrumento MSI (MultiSpectral Instrument). Trata-se de um sensor multiespectral que abrange 13 bandas espectrais (443 nm - 2190 nm) com uma faixa de imageamento de 290 km. As quatro bandas do VNIR (visível e infravermelho próximo) possuem resolução espacial de 10m, as seis bandas do infravermelho de borda e de ondas curtas (SWIR) 20m e as três bandas de correção

atmosférica 60m (SPOTO et al., 2012). A Tabela 1 apresenta as principais características do sensor MSI abordo dos satélites Sentinel-2 (MIRANDA, 2019).

Tabela 1 – Sentinel-2 (MSI)

Banda	**Comprimento de Onda (nm)**	**Resolução Espacial**
B1 - Aerosol	433-453	60 m
B2 - Azul	458-523	10 m
B3 - Verde	543-578	10 m
B4 - Vermelho	650-680	10 m
B5 - Red Edge 1	698-713	20 m
B6 - Red Edge 2	733-748	20 m
B7 - Red Edge 3	773-793	20 m
B8A - Red Edge 4	785-899	20 m
B8 - NIR	855-875	10 m
B9 - Vapor D'água	935-955	60 m
B10 - Cirrus	1.360-1.390	60 m
B11 - SWIR 1	1.565-1.655	20 m
B12 - SWIR 2	2.100-2.280	20 m

Fonte: (ESA, 2012)

Neste trabalho, são usadas as bandas B2 - Azul, B3 - Verde, B4 - Vermelho e B8 - Infravermelho Próximo que correspondem aos canais de cor RGB e o canal NIR (*Near Infra Red*). Além disso, são aplicados processamentos combinando essas bandas através de índices de vegetação.

2.2.2 Índices de Vegetação

Nos estudos que se envolvem a vegetação, solo e água, destacam-se as transformações espectrais que dão origem a determinados índices (FERREIRA; FERREIRA; FERREIRA, 2008). Os índices de vegetação colorida são usados no sensoriamento remoto de plantações e florestas. Esses índices têm a função de acentuar uma cor específica, como o verde da plantação (WOEBBECKE et al., 1995). Neste trabalho são explorados dois índices de vegetação colorida:

- Normalized Difference Vegetation Index (NDVI).
- Enhanced Vegetation Index (EVI)

A seguir esses índices são explicados e suas fórmulas em termos dos canais RGB e NIR, onde R é o canal de cor vermelha, G é o canal de cor verde, B é o canal de cor azul e NIR é canal infravermelho próximo.

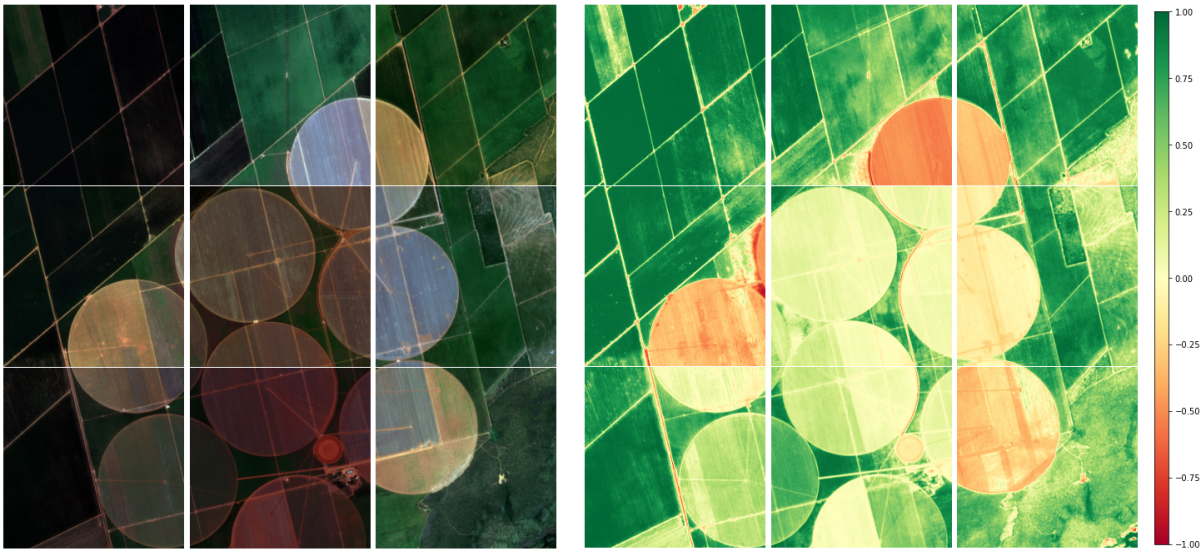
2.2.2.1 Normalized Difference Vegetation Index (NDVI)

O NDVI foi proposto por Rouse et al. (1973) e Jr et al. (1974) e é um índice de vegetação amplamente utilizado em diversas aplicações regionais e globais para monitoramento da vegetação (MAIA, 2019). O NDVI envolve a diferença e a soma entre as bandas do infravermelho próximo (NIR) e do vermelho (R). Fatores ambientais como tipo de solo, condições climáticas e de manejo são os principais aspectos que influenciam na magnitude dos valores de NDVI, uma vez que alteram as propriedades espectrais da área cultivada (BÉGUÉ et al., 2010). O índice de vegetação de diferença normalizada (NDVI) é definido pela equação a seguir:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{2.1}$$

A Figura 3 mostra o resultado da aplicação do NDVI em uma região de plantação. Após a aplicação do NDVI, cada pixel da imagem varia de -1 a 1. Normalmente para ilustrar visualmente o resultado do NDVI, utiliza-se um mapa de cores que varia do vermelho para o valor de pixel -1 até a cor verde para o valor do pixel próximo de 1. Um NDVI baixo pode significar baixa densidade de vegetação ou vegetação com saúde prejudicada.

Figura 3 – À esquerda, um mosaico de imagens de uma região de plantações em RGB. À direita, mosaico da mesma região com aplicação do índice de vegetação NDVI.



Fonte: O Autor

2.2.2.2 Enhanced Vegetation Index (EVI)

O Enhanced Vegetation Index (EVI) foi proposto por Huete et al. (2002) e modifica o NDVI através da utilização de coeficientes e da banda azul, minimizam o efeito da

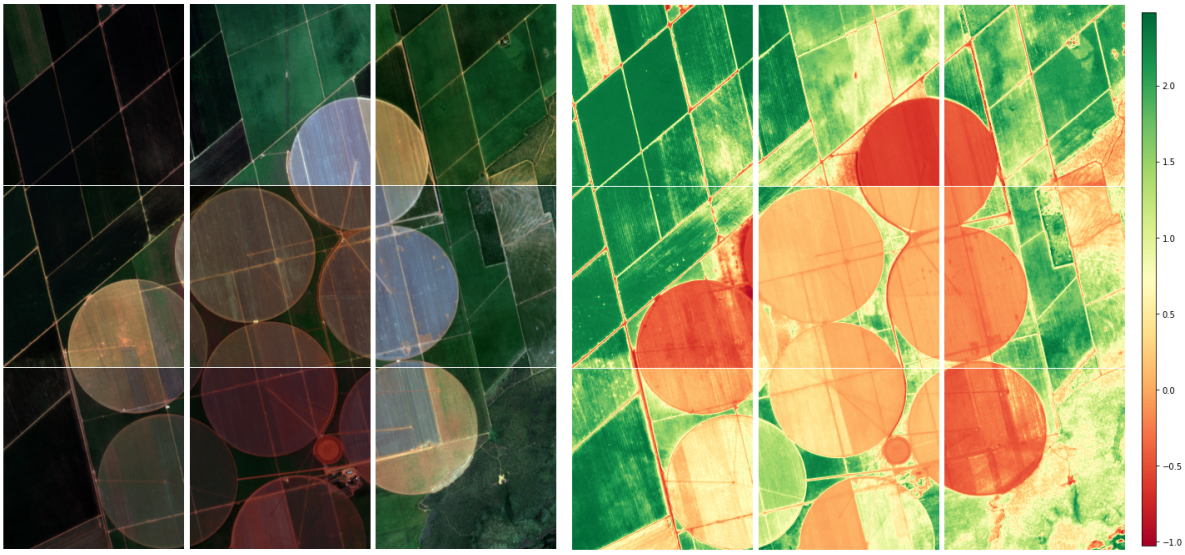
atmosfera e do solo dado por:

$$EVI = G\frac{NIR - R}{NIR + C_1R + C_2B + L} \tag{2.2}$$

onde C_1 coeficiente de correção dos efeitos atmosféricos para a banda do vermelho ($C_1 = 6$), C_2 é coeficiente de correção dos efeitos atmosféricos para a banda do azul ($C_2 = 7{,}5$); L é fator de correção para a interferência do solo (L = 1) e G é fator de ganho (G = 2,5).

O EVI é menos propenso a saturação em condições climáticas tropicais, o que o torna mais eficiente em relação ao NDVI. De acordo com Mondal (2011), o EVI tem sido mais efetivo no monitoramento sazonal e interanual de áreas de produção, detectando variações estruturais e de biomassa da vegetação. A Figura 4 mostra o resultado da aplicação do EVI em uma região de plantação.

Figura 4 – À esquerda, um mosaico de imagens de uma região de plantações em RGB. À direita, mosaico da mesma região com aplicação do índice de vegetação EVI.



Fonte: O Autor

2.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são modelos matemáticos biologicamente inspirados no sistema nervoso central humano. Esses modelos são representados por unidades computacionais, denominados neurônios artificiais, que são ligadas por uma grande quantidade de conexões, denominadas sinapses artificiais (SILVA; SPATTI; FLAUZINO, 2010). Esses modelos são amplamente utilizados em reconhecimento de padrões.

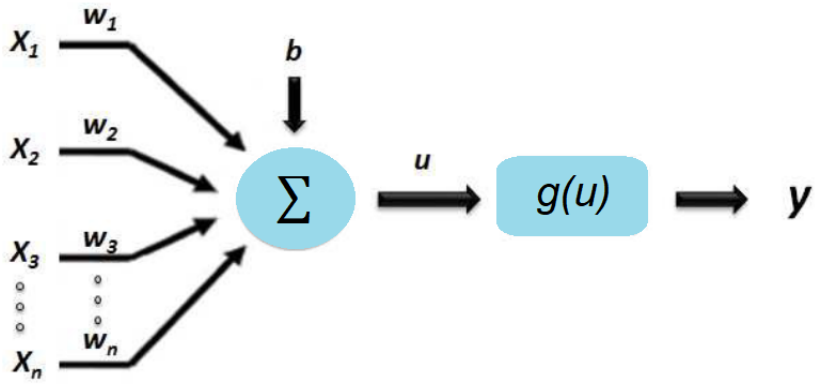
O neurônio artificial é a unidade básica de processamento das RNA e é fundamental para a construção de modelos mais complexos. Ele representa uma função matemática

em que, para um dado conjunto de entradas, é fornecida uma saída (HAYKIN, 2007). A função matemática que descreve um neurônio artificial é dada pela Equação 2.3,

$$y = g(\sum_{i=1}^{n} x_i w_i + b) \tag{2.3}$$

onde x_i representa um sinal de entrada do neurônio, w_i representa o peso sináptico associado a entrada i, b é o termo bias e g é a função de ativação. Um exemplo de neurônio artificial básico é apresentado na Figura 5.

Figura 5 – Neurônio Artificial



Fonte: Adaptado de (HAYKIN, 2007)

As funções de ativação podem ser lineares e não-lineares, onde as mais utilizadas são a função identidade, a sigmoide, a tangente hiperbólica e ainda a função ReLU (do inglês *Rectified Linear Units*) (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Experimentos realizados identificaram que a função de ativação ReLU apresenta uma velocidade de convergência seis vezes mais rápida que a demais função citadas (MAAS et al., 2013). A fórmula matemática da função de ativação ReLU é apresentada na Equação 2.4.

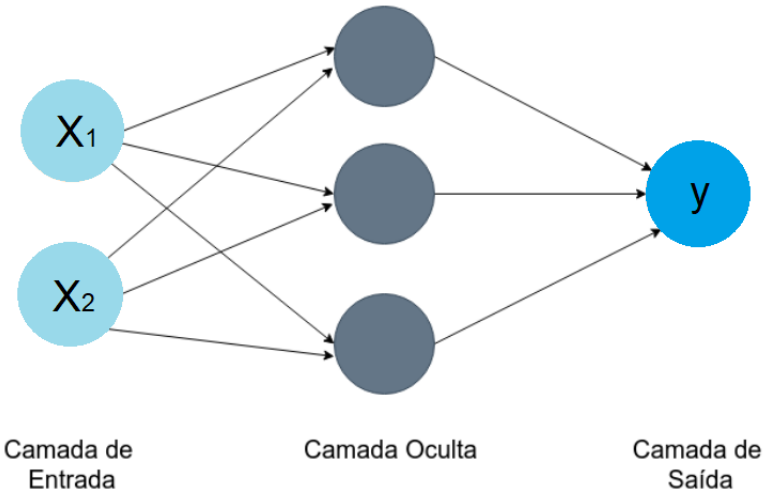
$$g(x) = max(0, x) \tag{2.4}$$

2.3.1 Redes *Multilayer Perceptron* (MLP)

As redes neurais MLP têm a estrutura baseada em camadas, cada uma delas completamente conectada à sua camada subsequente. A constituição básica das MLP são uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Com essa arquitetura, as MLP são capazes de aproximar qualquer função mensurável para qualquer grau de precisão desejado (HORNIK; STINCHCOMBE; WHITE, 1989). Essa rede tem a capacidade de aprender tarefas complexas extraindo progressivamente as características mais significativas dos padrões de entrada (SILVA; SPATTI; FLAUZINO,

2010). A arquitetura básica de uma MLP com apenas uma camada oculta é apresentada Figura 6.

Figura 6 – Neurônio Artificial



Fonte: Adaptado de (HAYKIN, 2007)

Existem vários algoritmos utilizados para fazer o treinamento das redes MLP, dentre eles, um dos mais populares é o *backpropagation*. Durante o treinamento, esse algoritmo realiza os ajustes dos pesos da rede com o objetivo de minimizar os erros utilizando a regra delta generalizada com aplicação do gradiente (SILVA et al., 2016). O *backpropagation* é caracterizado por dois passos distintos. Primeiro, dados de treinamento são inseridos na camada de entrada da rede. Eles seguem através da rede, camada por camada, sendo transformados por meio dos cálculos realizados até que, na camada de saída, seja produzida uma resposta. Em seguida, a saída obtida é comparada à saída desejada, obtendo-se um valor de erro. Esse erro é propagado partindo da camada de saída até a camada de entrada, modificando, dessa forma, os pesos das conexões nas camadas ocultas conforme a continuidade desse processo (SILVA et al., 2019).

A seguir são apresentados os passos do algoritmo *backpropagation* (ROJAS, 1996):

1. Inicializar os pesos sinápticos de cada neurônio com valores aleatórios.
2. Apresentar as entradas da rede em um vetor $x_1, x_2, ..., x_N$ de características e especificar um vetor $d_1, d_2, ..., d_N$ de saídas desejadas.
3. Calcular as saídas reais da rede y1, y2, ..., yN , baseadas na Equação 2.3.
4. Reajustar os pesos começando pelos neurônios da camada de saída, em direção aos neurônios da camada de entrada. Os pesos são ajustados através da Equação 2.5:

$$w_{ij}(t+1) = w_{ij}(t) + \eta\sigma_j x_i \tag{2.5}$$

onde w_{ij} é o peso do nerônio oculto j em um dado tempo t, x_i pode ser tanto um neurônio de saída quanto um de entrada, η é a taxa de aprendizagem e σ_j é um termo de erro para o neurônio j. Se j for um neurônio de saída, então σ_j é definido pela Equação 2.6:

$$\sigma_j = y_j(1 - y_j)(d_j - y_j) \tag{2.6}$$

onde d_j denota a saída desejada e y_j é a saída real da rede. Se o neurônio j for um neurônio oculto, então σ_j é definido pela Equação 2.7:

$$\sigma_j = x_j(1 - x_j) \sum k\sigma_k w_{jk} \tag{2.7}$$

onde k denota todos os neurônios acima do neurônio j.

5. Repetir o passo 2 até que uma dada condição seja satisfeita.

A Equação 2.5 se refere à descida do gradiente. O *backpropagation* também pode ser realizado por otimizadores estocásticos com uma equação diferente.

Durante o treinamento da rede, é importante configurar valores coerentes de hiperparâmetros. A taxa de aprendizado (η) influencia fortemente no desempenho do aprendizado da rede. Uma taxa grande normalmente produz grandes oscilações na curva de aprendizado da rede. Taxas muito pequenas podem conduzir os resultados a mínimos locais, retardando o aprendizado (SILVA; SPATTI; FLAUZINO, 2010).

2.4 Aprendizagem Profunda

Aprendizagem Profunda mais conhecida pelo seu termo em inglês, *Deep Learning*, é constituída de um conjunto de algoritmos que modelam abstrações partindo da identificação hierárquica de características, utilizando diversas camadas de processamento que executam operações de transformações lineares e não-lineares (AHMAD; FARMAN; JAN, 2019). Desse modo, a aprendizagem de características em vários níveis de abstração, do mais baixo até o mais alto, permite ao sistema computacional aprender e mapear funções complexas de forma independente das características criadas manualmente (BENGIO; COURVILLE; VINCENT, 2013).

Uma variedade de modelos de aprendizagem profunda tem sido aplicada em áreas como o Processamento de Linguagem Natural, Reconhecimento Automático de Fala, Visão Computacional, Reconhecimento de Áudio e Bioinformática com resultados que superam todas a técnicas anteriores, em muitos dos casos (AHMAD; FARMAN; JAN, 2019).

Este trabalho foca em aprendizagem de máquina profunda em imagens de satélite, mais especificamente em imagens de fazendas obtidas do Satélite Sentinel-2. Portanto,

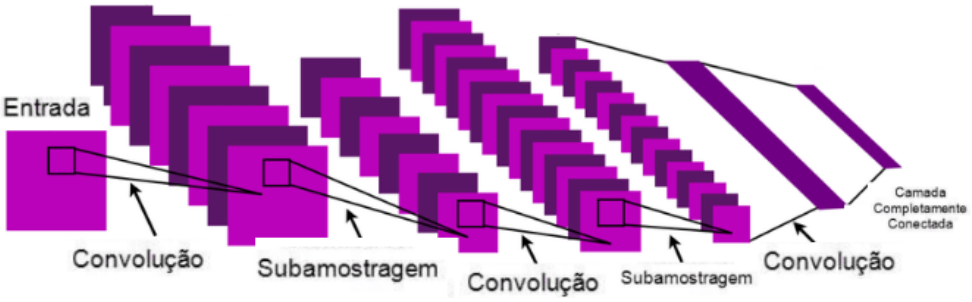
nas seções a seguir, são abordados os conceitos sobre as arquiteturas de Redes Neurais Convolucionais.

2.4.1 Redes Neurais Convolucionais

As redes neurais convolucionais (CNN - *Convolutional Neural Networks*) são modelos biologicamente inspirados que podem aprender características de forma hierárquica. São especialmente projetadas para lidar com a variabilidade em dados bidimensionais como as imagens em formato matricial (LECUN; BENGIO; HINTON, 2015). As redes neurais convolucionais tem como principal característica a capacidade de aprender padrões nas imagens.

As CNNs são constituídas, no mínimo, pelas três camadas a seguir: 1) Camada de Convolução que realiza operações de filtragem da imagem de entrada e os filtros convolucionais são constituídos de pesos que são aprendidos pela rede durante a etapa de treinamento. 2) Camada de Subamostragem que reduz a resolução da imagem a cada etapa e 3) Camada Totalmente Conectada que é composta por neurônios que fazem a classificação e saída final da rede (LECUN; BENGIO; HINTON, 2015). A Figura 7 ilustra a arquitetura de uma rede neural convolucional.

Figura 7 – Arquitetura da Rede Neural Convolucional



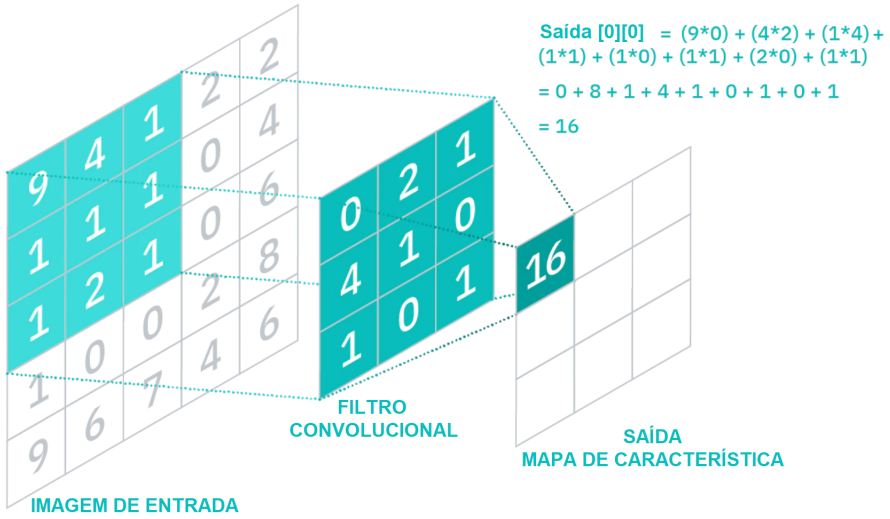
Fonte: Adaptado de (LECUN; BENGIO; HINTON, 2015)

As camadas de convolução servem de extratores de características e são constituídas por um conjunto de filtros treináveis representados em forma de matrizes de ativação pelos quais a imagem de entrada passa e resultado desse processo são denominados mapas de características (GUO et al., 2016). Após a rede ser treinada e os padrões aprendidos, cada filtro é responsável por detectar uma característica específica da imagem (HAFEMANN, 2014).

A convolução é uma operação matemática que é calculada a partir da sobreposição de duas matrizes. Uma das matrizes, denominada filtro, desliza sobre outra matriz, considerada entrada da convolução. Para cada valor de células correspondente entre a região do filtro e da entrada, uma operação matemática é realizada, cujo resultado é o

valor de saída para a célula localizada no centro do filtro (MAIA et al., 2019). A Figura 8 representa operação de convolução em uma Rede Neural Convolucional.

Figura 8 – Operação de Convolução na CNN



Há outros tipos de convolução como a Convolução Transposta que é detalhada na Subseção 2.4.3.

As camadas de *Pooling* ou Subamostragem reduzem a resolução espacial dos mapas de características usando funções do tipo máximo (*Max Pooling*) ou média(*Average Pooling*) de um conjunto de *pixels*. Semelhante à convolução, a Subamostragem é uma operação de janela deslizante, no entanto, sem treinamento de pesos. Os valores da janela são resumidos a um único valor de saída seja o valor máximo ou a média, de acordo com o tipo de *pooling* escolhido. Assim a resolução dos mapas de características é reduzida, há a redução de ruídos indesejados e a rede se torna mais robusta a variações de translação, rotação e escala (LECUN; BENGIO; HINTON, 2015; MAIA et al., 2019).

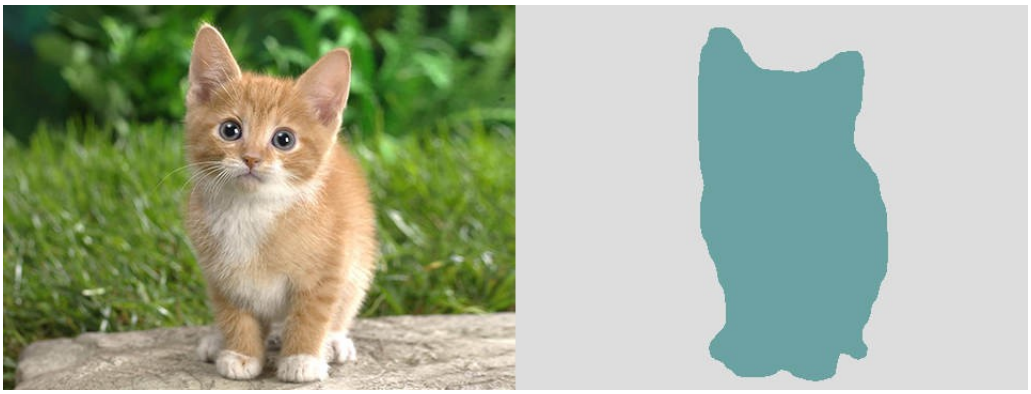
A última camada de uma rede neural convolucional geralmente é uma camada totalmente conectada. A camada totalmente conectada é uma camada de classificação que apresenta uma estrutura simular às MLPs, que recebe os *pixels* dos mapas de características da camada anterior em seus neurônios de entrada e geram a classificação dos padrões (SILVA et al., 2019).

2.4.2 Segmentação Semântica

No campo da visão computacional e de processamento de imagens, segmentação é o processo de decomposição de uma imagem digital em vários segmentos ou várias regiões que a formam (SALDANHA; FREITAS, 2009). Em resumo, a segmentação visa distinguir determinadas regiões da imagem. Em geral, existe uma região de interesse que é buscada na imagem. A segmentação tradicional é alcançada a partir de diversas técnicas de processamento de imagens como binarização, detecção de bordas, morfologia e

manipulação do histograma. Recentemente, surgiu a segmentação semântica onde cada pixel da imagem original é classificado a partir de técnicas de Aprendizado de Máquina resultando numa imagem de saída segmentada como no exemplo da Figura 9.

Figura 9 – Segmentação Semântica



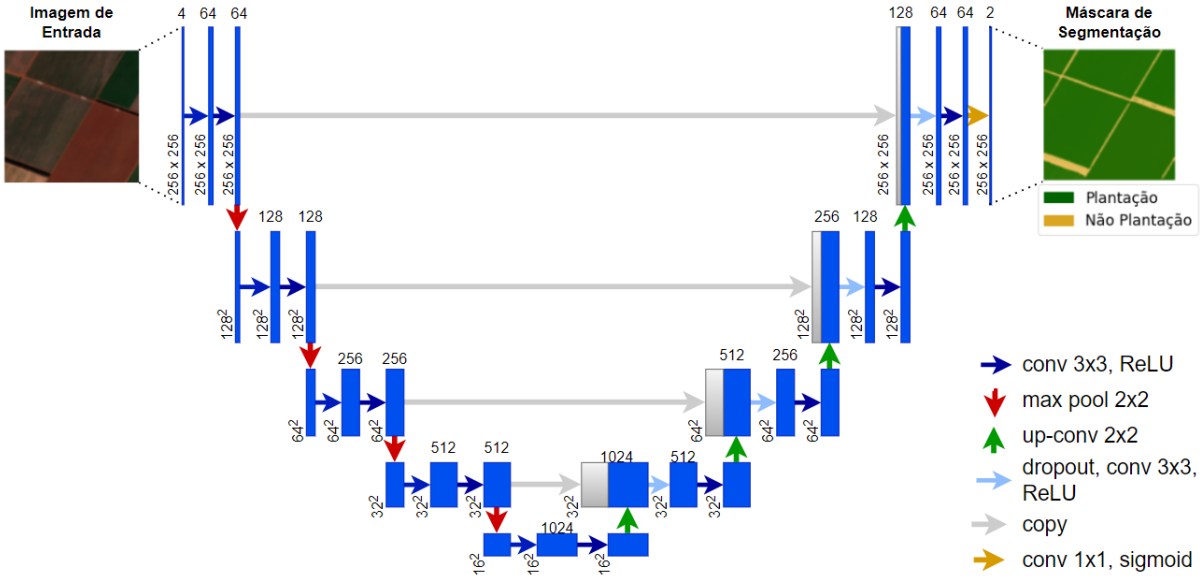
A arquitetura de rede neural convolucional apresentada na Subseção 2.4.1 estabelece o estado da arte para o problema de classificação de imagens. No entanto, para problemas de segmentação de imagens, é necessário que tanto a entrada da rede quanto a saída da rede sejam imagens, com isso foram propostas as Redes Totalmente Convolucionais, onde a rede produz uma saída nas mesmas dimensões da imagem de entrada, realizando uma segmentação pixel-a-pixel da imagem de entrada, como é o caso da U-net apresentada na Subseção 2.4.3.

2.4.3 U-net

A U-net é uma rede neural convolucional do tipo *Fully Convolutional* ou Totalmente Convolucional. No tipo de arquitetura das Redes Neurais Totalmente Convolucionais, as camadas de múltiplos perceptrons são substituídas por camadas convolucionais onde a saída é um mapa de ativação com resolução igual ou aproximada à resolução da imagem de entrada. A U-net (RONNEBERGER; FISCHER; BROX, 2015) foi proposta em 2015 com aplicação em imagens biomédicas e desde então ganhou popularidade por seu desempenho e por conseguir ser treinada com relativamente poucas imagens de teste. A arquitetura da U-net tem uma forma semelhante a letra U e seu nome é devido a essa característica. A arquitetura da U-net é apresentada na Figura 10.

Nessa arquitetura, a metade esquerda da U-net é chamada de caminho de contração, codificador ou *Encoder* dos mapas de características. É por onde as imagens entram na rede e passam pelas operações de convolução e subamostragem. Quanto mais profundo na rede, mais cresce a quantidade de mapas de características e a resolução desses mapas de características vai ficando cada vez menor. A metade direita da U-net é chamada de caminho de expansão, decodificador ou *Decoder* dos mapas de características. No *Decoder*, existem novas operações ainda não detalhadas neste trabalho como o *dropout* e

Figura 10 – Arquitetura da U-net.



Fonte: Adaptado de (RONNEBERGER; FISCHER; BROX, 2015)

up-convolution que fazem com que a quantidade de mapas de características seja reduzida, quanto mais esses recursos são passados para frente na rede. Por outro lado, faz com que a resolução aumente cada vez mais de forma que a saída seja uma máscara de segmentação que tenha a mesma dimensão das imagens de entrada.

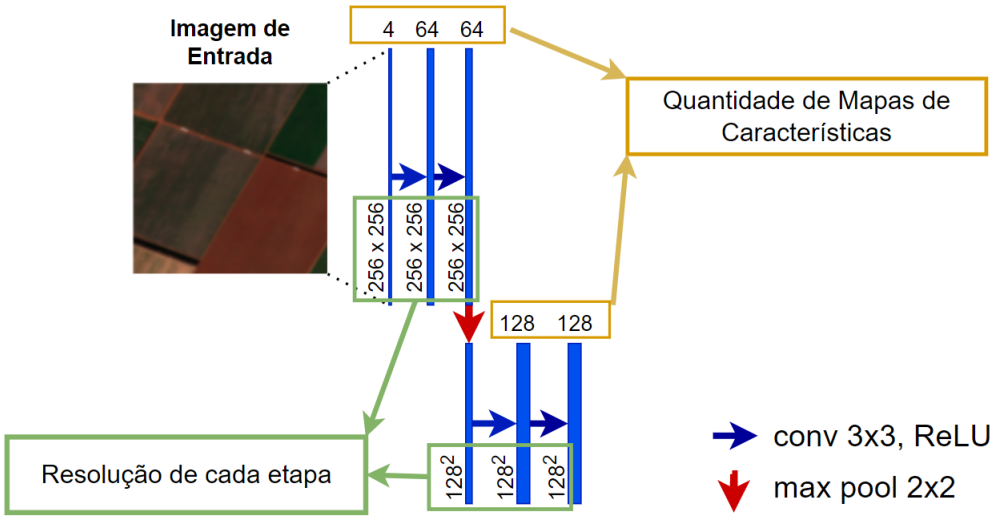
As operações que definem a arquitetura da U-net que são as operações de Convolução com Função de Ativação *ReLU*, Subamostragem (*max pooling*), Convolução Transposta (*up-convolution*), *Dropout*, Cópia dos mapas de ativação, Convolução 1x1 e Função de Ativação Sigmoide. Cada uma dessas operações é explicada a seguir.

2.4.3.1 Convolução com Função de Ativação ReLU

As operações de convolução e função de ativação ReLU já foram abordadas nas seções anteriores, contudo é importante contextualizar sua funcionalidade na U-net. Após entrar na U-net, cada imagem passa por filtros convolucionais com janelas de tamanho 3x3. Na arquitetura da U-net apresentada na Figura 10, são 64 filtros de convolução na primeira camada de convolução que geram 64 mapas de características. Mapas de Características ou Mapas de Ativação (*Feature Maps ou Activation Maps*) são os resultados de uma filtragem, por isso, os resultados das operações internas da U-net são chamados dessa forma. Na U-net, toda operação de convolução é seguida da função de ativação *ReLU* que deixa passar adiante apenas os valores positivos dos mapas de características, enquanto que os valores negativas são alterados para zero. Na ilustração da arquitetura da U-net na Figura 10, a entrada é uma imagem que contém 4 canais cada um com resolução de 256 *pixels* por 256 *pixels*, então a resolução total da entrada é 4 x 256 x 256. A Figura 11 detalha as

informações das primeiras camadas da arquitetura da U-net.

Figura 11 – Detalhes da Arquitetura da U-net



Fonte: Adaptada de (RONNEBERGER; FISCHER; BROX, 2015)

Após a sequência de duas operações de convoluções com ativação *ReLU*, a imagem de entrada com 4 canais, resulta em uma saída composta por 64 mapas de características e resolução total de 64 x 256 x 256. Não há perda de resolução após as operações de convolução seguida de ativação *ReLU*, apesar do tamanho do filtro de convolução 3x3 e passo igual a 1, pois a U-net do exemplo tem um preenchimento de zeros (*padding*) igual a 1. Com isso, não se perdem linhas e colunas de *pixels* das bordas após as operações de convolução.

As operações de convolução 3x3 seguida de ativação ReLU são realizadas diversas vezes na U-net, como após operações de subamostragem, que reduzem a resolução dos mapas de características e também após as operações de *up-convolution*, que ampliam a resolução dos mapas características. Dessa forma, as convoluções são aplicadas aos mapas de características em diferentes níveis de resoluções extraindo e aprendendo diferente níveis de recursos e características. Na etapa de treinamento da rede com o *backpropagation*, os filtros de convolução são atualizados e ajustados e assim a rede aprende a segmentar a região de interesse.

2.4.3.2 Subamostragem *Max Pooling*

A subamostragem da U-net é do tipo *Max Pooling*. O *Max Pooling* usado na U-net gera a perda de resolução da matriz de entrada passando apenas o valor máximo de uma janela deslizante. O resultado da subamostragem é um subconjunto representativo da entrada. A subamostragem é importante para reduzir o processamento que seria necessário

para computar os recursos na resolução da imagem entrada. Também é importante para que a rede identifique características em diferentes níveis de detalhes que podem ir de linhas e contornos na alta dimensionalidade (muita informação), a objetos inteiros na baixa dimensionalidade. Por fim, a subamostragem torna o modelo mais robusto a pequenas variações na posição dos objetos na imagem. Essa característica é chamada de invariância à rotação e translação dos objetos.

Como na operação de convolução, a operação de subamostragem tem uma janela deslizante na imagem com um passo definido. Na U-net a janela tem tamanho 2 x 2 e o valor do passo (*stride*) é igual a 2 tanto na horizontal quanto na vertical. Com isso, a janela de subamostragem sempre é aplicada em regiões diferentes da matriz sem sobreposição. Dessa forma, o resultado de saída da operação de *Max Pooling* da U-net tem a resolução reduzida pela metade.

Para explicar melhor como a operação ocorre na U-net, na Figura 11, observa-se que a entrada é uma imagem de dimensão 4 x 256 x 256, após as duas operações de convolução, o resultado são 64 mapas de características, apresentando uma resolução de 64 x 256 x 256. A operação seguinte é o *Max Pooling*, onde a quantidade de mapas de características se mantém e a resolução é reduzida pela metade. O resultado, que pode ser visualizado na Figura 11 como a entrada da segunda camada da U-net, tem dimensão 64 x 128 x 128. Nas próximas camadas do *Encoder* da U-net, a operação de *Max Pooling* continua reduzindo a dimensionalidade dos mapas de características pela metade.

2.4.3.3 *Up-convolution* Convolução Transposta

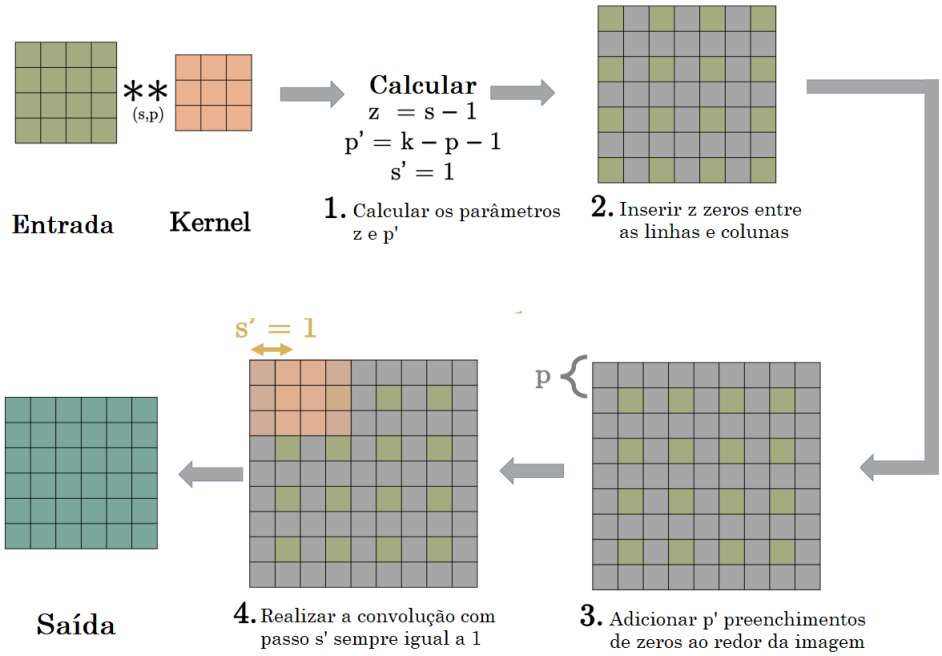
As próximas operações, ao iniciar pela Convolução Transposta, são mais específicas do tipo de arquitetura da U-net. São operações da metade direita da U-net, o *Decoder*, que após a codificação, expande os mapas de características até a resolução igual à resolução da imagem de entrada gerando a máscara de segmentação final da rede.

A convolução transposta, em contraponto com a subamostragem, é usada para ampliar a resolução de uma matriz de entrada para uma dimensão desejada usando parâmetros aprendidos pela rede, um kernel de convolução (ZHANG et al., 2021). A convolução transposta é a operação inversa da operação de convolução padrão, mas somente em relação às dimensões de entrada e saída, considerando os mesmos parâmetros de tamanho da janela de convolução (*kernel*), passo em que a janela desliza na horizontal e na vertical (*stride*) e o preenchimento de zeros nas bordas da entrada (*padding*). Considerando esses fatores, caso uma matriz de $I \times I$ passe por uma operação de convolução padrão gerando uma saída $O \times O$ e em seguida essa saída entre em uma operação de convolução transposta que tem o mesmo tamanho de janela de convolução, mesmo passo e mesmo preenchimento de zeros, a saída vai ter dimensões $I \times I$. No entanto, apesar das matrizes terem a mesma dimensão após a aplicação das duas operações, os seus valores não

necessariamente serão os mesmos, pois a convolução padrão e a convolução transposta não são operações inversas no diz respeito aos valores.

A saída da operação de convolução transposta pode obtida, em termos da operação de convolução padrão, seguindo as quatro etapas (DUMOULIN; VISIN, 2016) a seguir e ilustradas da Figura 12.

Figura 12 – Convolução Transposta



Fonte: Adaptada de (DUMOULIN; VISIN, 2016)

A Etapa 1 consiste em calcular os valores dos parâmetros z, p' e s' conforme as fórmulas apresentadas na Figura 12. Onde z é quantidade de linhas e colunas de zeros que serão inseridas na matriz de entrada de forma intercalada. O parâmetro p' é o contorno de zeros (*padding*) e s' é o passo da janela de convolução (*stride*). Podemos observar que que há tanto o parâmetro p quanto o p' para representar o *padding* e tanto o s quanto o s' para representar o *stride*. Os parâmetros s e p são os correspondentes às operações aplicadas na convolução padrão, já os parâmetros s' e p' são uma transformação dessas operações para a convolução transposta.

Na Etapa 2 são inseridas entre cada linha e coluna da entrada, o número z de linhas e colunas de zeros. Após essa operação, a matriz de entrada tem um incremento da sua resolução para $(2 * l - 1) \times (2 * c - 1)$, onde l é o número de linhas de entrada e c o número de colunas de entrada. Essa operação é bem similar ao contorno de zeros (*padding*), porém de forma interna à entrada e com linhas e colunas intercaladas.

Na Etapa 3 é realizada a operação de preenchimento de zeros na borda da mesma forma como ocorre na convolução padrão, contudo em vez de usar o parâmetro p, na

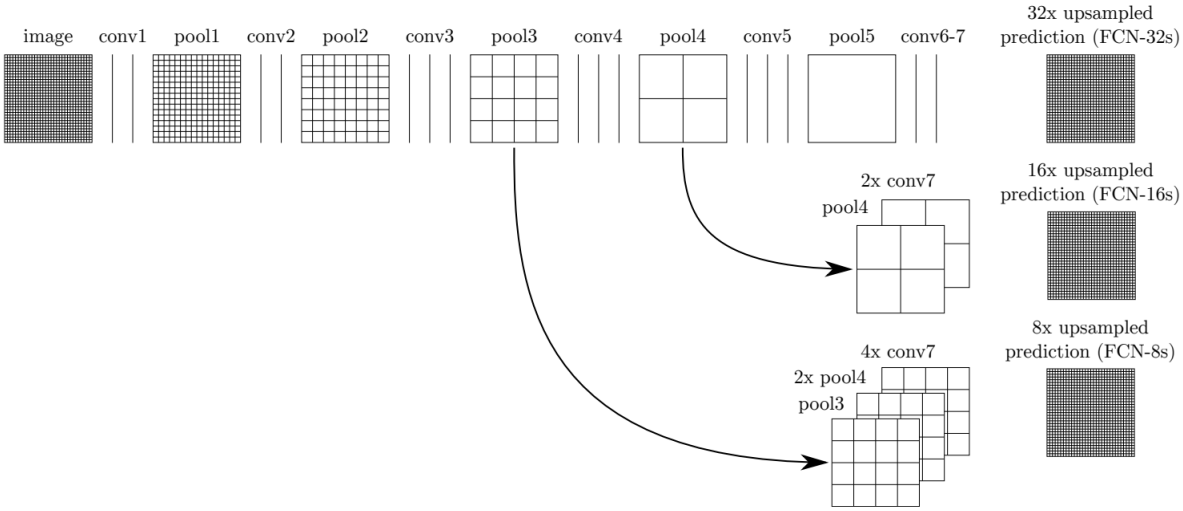
convolução transposta é usado o parâmetro p' calculado na Etapa 1. Por fim, na Etapa 4 é realizada a operação de convolução padrão sempre com o passo da janela de convolução (stride) equivalente a um. A saída dessa operação é o resultado final da convolução transposta.

Na U-net, a operação de *Up-convolution* pode ser realizada com a Convolução Transposta. Nessa arquitetura é utilizada uma operação de *Up-convolution* com tamanho de janela 2 x 2 e passo igual a 2 e sem preenchimento de bordas. Uma operação de convolução padrão com esses mesmos parâmetros de tamanho de janela, passo e bordas, resultaria em uma saída com tamanho de resolução igual metade da resolução de entrada. Então a saída da operação de *Up-convolution* com os mesmos parâmetros resultará em uma saída com o dobro da resolução de entrada. A operação de *Up-convolution* faz no *Decoder* o trabalho inverso que o *Max Pooling* faz no *Encoder*, com a grande diferença que os pesos da janela do *Up-convolution* são ajustados e aprendidos. Já o *Max Pooling* sempre deixa passar o valor máximo dentro da janela, não há aprendizado.

2.4.3.4 Cópia dos mapas de características

Uma particularidade da U-net é a cópia dos mapas de características do *Encoder* para as camadas correspondentes de mesma resolução no *Decoder*, isto é, os recursos da camada da metade esquerda da U-net são copiados para a camada com mesma resolução metade direita. Para explicar a importância de copiar os mapas de características, é necessário retornar aos estudos das primeiras Redes Totalmente Convolucionais (*Fully Convolutional Networks - FCN*). A Figura 13 apresenta a arquitetura das primeiras Redes Totalmente Convolucionais FCN-32s, FCN-16s e FCN-8s (LONG; SHELHAMER; DARRELL, 2015).

Figura 13 – Redes Totalmente Convolucionais FCN-32s, FCN-16s e FCN-8s

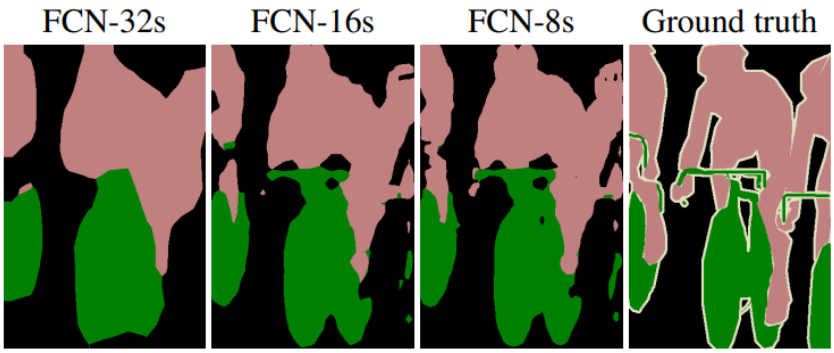


Fonte: (LONG; SHELHAMER; DARRELL, 2015)

Na arquitetura FCN-32s, primeira linha da Figura 13, a imagem entra na rede e vai passando por operações de convoluções e subamostragem (*pooling*) que vão reduzindo a resolução pela metade consecutivamente. Ao final da rede, após cinco camadas de subamostragem e sete de convolução, é aplicada uma ampliação de resolução (*upsample*) no mapa de características resultante da camada conv 7 que aumenta a resolução bruscamente em 32 vezes. O *upsample* foi de 32 vezes, porque a imagem de entrada passou por 5 operações de subamostragem em que uma reduziu a entrada pela metade, por isso para recuperar a resolução original da imagem o upsample foi de 2^5 vezes.

O resultado da segmentação da FCN-32s na primeira imagem da Figura 14 quando a máscara de segmentação correta (*Ground truth*) é a última imagem da mesma Figura. Com as consecutivas operações de subamostragem, a rede vai perdendo resolução, os mapas de características vão ficando cada vez menos nítidos e essa redução de informação acaba favorecendo o processamento para segmentar objetos inteiros, no entanto a rede vai perdendo informações sobre a localização dos pixels e características mais finas e detalhadas. Por isso, a FCN-32s conseguiu acertar onde há uma pessoa e uma bicicleta na Figura 14, porém sem segmentar de forma precisa os contornos dos objetos.

Figura 14 – Resultados FCN-32s, FCN-16s e FCN-8s



Fonte: (LONG; SHELHAMER; DARRELL, 2015)

Para solucionar o problema de falta de precisão da FCN-32s, foram propostas as redes FCN-16s e FCN-8s. Na FCN-16s, ilustrada nas duas primeiras linhas da Figura 13, a perda de resolução é exatamente igual à FCN-32s, no entanto há uma sutil diferença na etapa de *upsample*, em vez de haver uma recuperação direta de resolução de 32 vezes, a saída da camada conv7 passa por uma ampliação de 2 vezes e ainda é somada com o recurso pool4 de mesma resolução para só depois ampliar a resolução em 16 vezes. Dessa forma, a rede consegue encontrar os objetos através das convoluções e subamostragem e recupera informações de localização e detalhes de mais alta resolução ao remontar a imagem. Na rede FCN-8s, o processo se repete com um passo de *upsample* a mais e recuperando informações de um passo de subamostragem antes. Com essa recuperação de informação e ampliação de resolução de maneira gradual, a precisão da segmentação vai

melhorou conforme é apresentado na Figura 14.

Na U-net, a cópia dos mapas de características da camada correspondente é uma solução em escala maior para o mesmo problema enfrentado pela primeiras Redes Totalmente Convolucionais. Na U-net, todos os mapas de características que foram processados em etapas anteriores da rede são copiados para a camada correspondente de mesma resolução da metade final da rede, conforme ilustra as setas em cinza da Figura 10 e os blocos em branco no final das setas. Esses mapas de características são misturados com os mapas de características obtidos via convolução transposta, de modo a recuperar algumas informações e recursos.

2.4.3.5 *Dropout*

Após a cópia dos mapas de características da camada correspondente, a quantidade de recursos a ser processada dobra, por isso é aplicada uma operação que descarta aleatoriamente parte desses recursos em excesso. Essa operação se chama *dropout*. O *dropout* é uma técnica de regularização que desativa alguns neurônios da rede de forma aleatória (LABACH; SALEHINEJAD; VALAEE, 2019). No caso da U-net são filtros convolucionais. O *dropout* foi proposto para ajudar a resolver dois problemas recorrentes em redes neurais profundas, a perda de performance devido ao grande número de neurônios e o sobreajuste das redes, o *overfitting* (SRIVASTAVA et al., 2014).

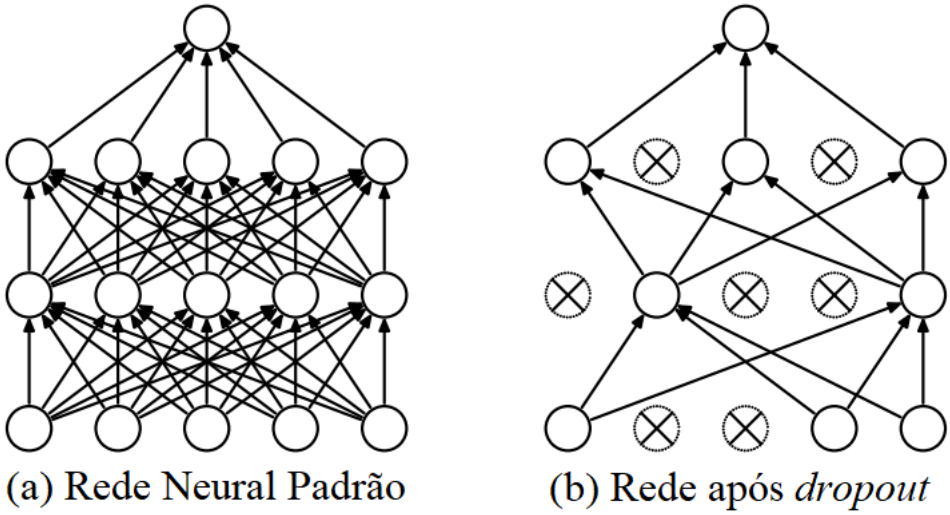
A ideia principal é descartar unidades aleatoriamente junto com suas conexões durante o treinamento. Isso evita que as unidades se adaptem demais (SRIVASTAVA et al., 2014). Como é ilustrado na Figura 15, alguns neurônios são desativados aleatoriamente. Durante a etapa de treinamento, em vez de treinar com todos os mapas de ativação, a rede é treinada apenas com os filtros ativos. Na próxima iteração da etapa de treinamento, os filtros ativados e desativados mudam devido ao seu comportamento probabilístico e isso permite que rede generalize mais (LABACH; SALEHINEJAD; VALAEE, 2019).

Na U-net a operação de *dropout* é aplicada no *Decoder* no início de cada camada após os mapas de mapas de características resultantes do *up-convolution* receberem o acréscimo dos mapas de características da operação de Cópia. O *dropout* aplicado desativa 50% dos mapas de características reduzindo a carga de processamento e garantindo maior variabilidade dos recursos no treinamento.

2.4.3.6 Predição

Por fim, na última camada da U-net são realizadas duas operações finais para gerar a máscara de segmentação de saída que são a convolução 1×1 (LIN; CHEN; YAN, 2013) e a função de ativação sigmoide. Na convolução 1×1, *kernels* de convolução de tamanho 1×1, ou seja, são pesos escalares, na quantidade de mapas características que chegaram até a última camada da U-net são combinados formando um *kernel* achatado que tem

Figura 15 – Dropout

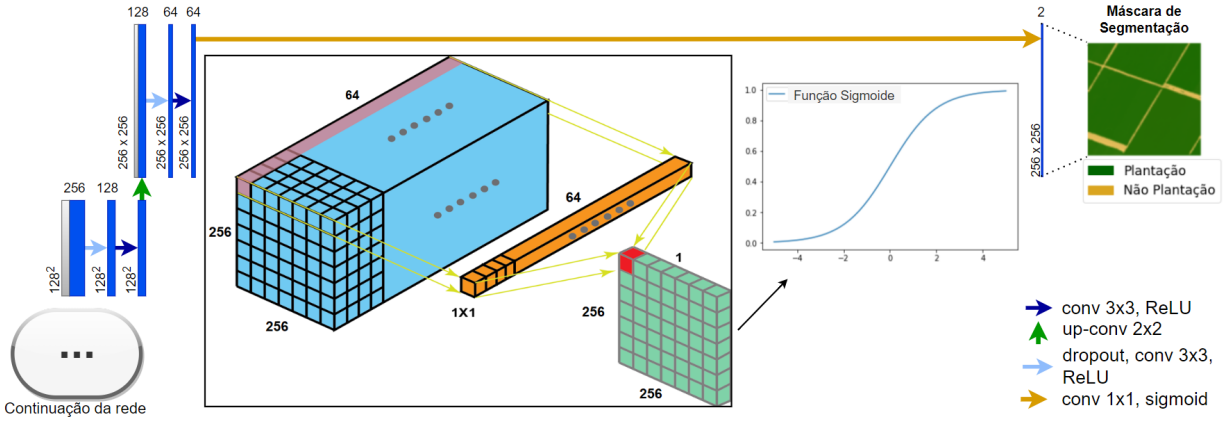


Fonte: Adaptado de (SRIVASTAVA et al., 2014)

dimensão 1x1 em uma perspectiva, mas tem uma profundidade alongada formando um vetor de pesos para a convolução. Na Figura 16, o *kernel* 1 x 1 tem dimensão 64 x 1 x 1, devido aos 64 mapas de características que chegaram ao final da rede.

É aplicada a operação de convolução padrão entre o *kernel* e os mapas características que chegaram até a última camada. A saída dessa operação é um único mapa de características de resolução igual ao da imagem de entrada. Por fim, a predição é feita aplicando uma ativação sigmoide para cada classe de segmentação. A função sigmoide varia de 0 a 1. Em algumas versões da U-net também é usada a função *Softmax* (GAO; PAVEL, 2017). O resultado final é a máscara de segmentação como ilustrado na Figura 16.

Figura 16 – Predição da U-net com convolução 1x1 e função sigmoide



Fonte: O Autor

2.5 Considerações Finais

Neste capítulo foram apresentados os conceitos explorados para o desenvolvimento do estudo. Explicar esses conceitos são importantes para o entendimento dos próximos Capítulos, uma vez que neste trabalho são seguidas etapas de Visão Computacional e Processamento de Imagens, onde a etapa de aquisição de imagens utiliza-se de Sensoriamento Remoto com o uso de imagens do Satélite Sentinel-2. A etapa de pré-processamento são experimentados índices de vegetação para realçar área de plantações. Para a tarefa de segmentação, são analisadas arquiteturas de Aprendizagem Profunda, mais precisamente a classe de Redes Neurais Totalmente Convolucionais. Por fim, para atingir o objetivo deste trabalho, é desenvolvida uma arquitetura baseada na rede U-net.

3 Trabalhos Relacionados

Para solucionar problemas de segmentação de imagens de satélite e imagens aéreas de plantações, as abordagens encontradas na literatura em geral fazem uso de técnicas de aprendizado profundo. Esses trabalhos se concentram em conjuntos de dados criados por empresas privadas, governos e universidades. Tais conjuntos de dados, normalmente são criados especificamente para as particularidades da pesquisa. Foi realizada uma pesquisa bibliográfica para conhecer os métodos e os conjuntos de dados utilizados em problemas semelhantes ao exposto neste trabalho. Os principais estudos levantados são discutidos nesta seção.

Os algoritmos de classificação e segmentação da cobertura terrestre e áreas de plantação são usados para prever áreas da superfície terrestre, onde as classes podem incluir tipos de cobertura como floresta, área urbana, água e agricultura. Historicamente, esses métodos usaram informações de uma imagem de satélite para prever os tipos de cobertura da terra (GÓMEZ; WHITE; WULDER, 2016). As entradas incluem recursos espectrais coletados pelo satélite, que geralmente se estendem além dos canais de cor vermelho, verde e azul que constituem uma imagem colorida típica. Outras bandas baseadas em índices de textura ou vegetação podem ser construídas, e as previsões geralmente são feitas pixel a pixel sem o uso de informações contextuais. Os estudos de classificação de terras cultivadas geralmente usam muitas observações temporais como entrada, uma vez que as propriedades espectrais das culturas mudam ao longo de uma estação de crescimento (SCHULTZ et al., 2015; HAO; WANG; NIU, 2015; FOERSTER et al., 2012). Alguns dos conjuntos de dados usados nessas pesquisas estão disponíveis publicamente.

Durante a pesquisa, foram encontrados diversos conjuntos de dados usados para avaliação de métodos de segmentação de áreas de plantações. Alguns deles tem foco exclusivo em áreas de plantações como o *Slovenia Dataset* e *Oregon Dataset* que apresentam regiões de vegetação em imagens de satélite em três classes sendo árvores, arbustos e grama (AYHAN; KWAN, 2020a), o *Agriculture-Vision Dataset* que é composto por seis classes que são sombra de nuvem, dupla plantação, falha na plantação, poça d'água, caminho de água e bloco de erva daninha (CHIU et al., 2020b) e o *PASTIS Dataset* (GARNOT; LANDRIEU, 2021a) que é um conjunto de dados para segmentação panóptica. Já outros conjuntos de dados, apresentam uma variedade de classes da cobertura terrestre que vai além de áreas de plantações, como o *BigEarthNet Dataset* (SUMBUL et al., 2019) que apresenta 44 classes da cobertura terrestre e o *CORINE Dataset* (BÜTTNER et al., 2004) que foi desenvolvido pela Agência Ambiental Europeia e apresenta uma variedade de classes.

A respeito dos métodos pesquisados, em um contexto mais amplo, técnicas de aprendizagem de máquina e de processamento de imagens tem sido empregadas para solucionar problemas na agricultura. Abdullahi, Sheriff e Mahieddine (2017) utilizaram Máquina de Vetores de Suporte (*Support Vector Machine* - SVM) para classificar e reconhecer diferentes classes de imagens de plantas de milho, detectar doenças de plantas e determinar a taxa de crescimento de plantações ao mesmo tempo. Tsai e Chen (2017), com o objetivo de segmentar campos de café e estimar sua produção, utilizaram um método baseado em Transformação de *Fourier* para extrair características estruturais no domínio espectral para segmentação de imagens. O método não segmenta por cor, porque o café tem a mesma cor da vegetação ao redor. Os resultados da pesquisa foram robustos para separar os campos de café plantados da floresta. Vogt et al. (2010), também utilizaram transformação de *Fourier* para identificar plantações a partir de imagens de satélite. Além da transformação de *Fourier*, eles usaram também o histograma da matriz Hessiana e o histograma do Tensor de Estrutura para identificar as plantações a partir de imagens de satélite.

Já Nunes e Conci (2007) aplicaram uma medida para quantificar a textura de uma região multibandas, denominada Coeficiente de Variação Espacial (CVE) e em seguida aplicaram o algoritmo de clusterização *K-Means*. As amostras do conjunto de treinamento são agrupadas utilizando-se o algoritmo de clusterização K-Means e as coordenadas dos centróides dos agrupamentos são utilizadas para classificação da imagem. Os resultados obtidos permitiram distinguir diferentes classes de texturas, além de localizar o contorno de regiões de interesse, mantendo sua complexidade e localização. Burgos-Artizzu et al. (2010) apresentam métodos de visão computacional para a estimativa de porcentagens de plantas daninhas, culturas e solo presentes em imagens de região de interesse do campo agrícola. O processamento da imagem foi dividido em três estágios diferentes, nos quais cada elemento agrícola é extraído: (1) segmentação da vegetação contra a não vegetação (solo), (2) eliminação da linha de colheita e (3) extração de ervas daninhas. Um algoritmo genético foi usado para encontrar os valores de parâmetros e foi usada uma combinação de métodos para diferentes conjuntos de imagens. Os testes resultaram em um coeficiente de correlação médio com dados reais (biomassa) de 84%.

Por fim, Vibha et al. (2009) apresentam uma abordagem para a segmentação automática de imagem de satélite em regiões distintas e extração da contagem de árvores da área vegetativa. Na pesquisa existem três regiões de interesse: a área de terra, a área residencial e a área de vegetação. Nela é utilizada um algoritmo de aprendizado não supervisionado. Comparando a área segmentada com a área real, o método teve uma acurácia 88%. Zhou et al. (2013) propõem uma técnica para mapear a densidade local de jovens e pequenos eucaliptos em uma grande plantação no Brasil. O estudo faz uma análise da estrutura de plantações de eucalipto ao longo de vários meses, coleta várias características como diâmetros de copas, altura, etc. É utilizado um método chamado

Marked Point Process (MPP) para detecção das copas de eucalipto. Os resultados foram sequências de pontos que representam as copas de eucalipto, essa sequência de pontos distribuídas é uma área específica permitiu calcular a densidade, objeto da pesquisa.

Em um contexto mais específico para a resolução dos problemas de segmentação, tipicamente são utilizadas Redes Neurais Convolucionais. As redes de aprendizado profundo baseados na arquitetura da U-net (RONNEBERGER; FISCHER; BROX, 2015) tem se destacado nos estudos de segmentação semântica. O vencedor do desafio Agriculture-Vision usou uma arquitetura baseada na U-Net e combinou uma Residual DenseNet com blocos Squeeze-and-Excitation (RD-SE) (CHIU et al., 2020a). O conjunto de dados Agriculture-Vision (CHIU et al., 2020b) contêm 21.061 imagens aéreas de fazendas dos Estados Unidos capturadas ao longo de 2019. O conjunto de dados é composto por seis classes que são sombra de nuvem, dupla plantação, falha na plantação, poça d'água, caminho de água e bloco de erva daninha. O Agriculture-Vision é um conjunto de dados para segmentação semântica de múltiplas classes, onde essas classes podem se sobrepor, como por exemplo, uma poça d'água sombreada por nuvem. No RD-SE, para compensar a perda espacial que surge durante a extração de características, são utilizados blocos densos residuais e conexões de salto. Além disso, os blocos Squeeze-and-Excitation (bloco SE) são usados para recalibrar as respostas dos recursos de canal. Cinco camadas de convolução com tamanho de kernel 3x3 e normalização em lote estão incluídas em um bloco denso residual. Também foram usadas Redes Especialistas para segmentar objetos de classe menos frequentes, classes com presentes em menos imagens, uma vez que o dataset é desbalanceado.

Outras redes baseadas na U-net foram usadas em estudos sobre segmentação de cobertura terrestre e plantações (ULMAS; LIIV, 2020; WANG et al., 2020; RUSTOWICZ et al., 2019). Estes estudos realizam modificações na arquitetura original da U-net, como (ULMAS; LIIV, 2020) que utilizou duas arquiteturas de rede neural, sendo um modelo ResNet50 (HE et al., 2016) a tarefa de classificação e no outro modelo, a ResNet50 pré-treinada na etapa de classificação foi usada como *Encoder* no modelo baseado em U-Net para resolver a tarefa de segmentação. Já (GARNOT; LANDRIEU, 2021b) utilizou uma U-Net com Temporal Attention Encoder que funciona em três etapas: (1) cada imagem na sequência é incorporada simultaneamente e independentemente por um codificador convolucional espacial multinível compartilhado, (2) um o codificador de Temporal Attention (HE; FANG; PLAZA, 2020) colapsa a dimensão temporal da sequência resultante de mapas de recursos em um único mapa para cada nível, (3) um decodificador convolucional espacial produz um único mapa de recursos com a mesma resolução que as imagens de entrada.

Além das redes baseadas em U-net, são usadas redes baseadas na arquitetura DeepLabV3+ (CHEN et al., 2017). A AgriSegNet que foi desenvolvida para solucionar o desafio Agriculture-Vision. Para resolver esse problema, Anand et al. (2021) propôs uma

nova arquitetura de rede neural totalmente convolucional que chamou de AgriSegNet. A AgriSegNet utiliza portas de atenção e combina o resultado da segmentação das imagens em resoluções diferentes chegando a 47,96% de IoU médio. A AgriSegNet utiliza duas escalas de imagem diferentes para alimentar a rede durante o treinamento. A rede consiste no *Encoder*, e duas camadas que o autor nomeou de *Semantic Head* e *Attention Head*. Os recursos do *Encoder* são extraídos do modelo *DeepLabV3+*. A *Attention Head* combina os recursos de várias escalas com proporção ponderada. Como estratégia de combinação para tratar todas as escalas como sendo igualmente importantes é usada uma subamostragem média ou *Average Pooling*. Características mais finas são obtidas em escalas mais altas, enquanto trechos maiores dentro da imagem são obtidos em escalas mais baixas. A AgriSegNet aprende sobre máscara de atenção relativa para combinar ou atender previsões para múltiplas escalas. O modelo de inferência faz uso de três diferentes escalas de imagem para fazer previsões.

Ayhan e Kwan (2020a) utilizaram uma arquitetura de rede neural convolucional baseada na DeepLabV3+ para segmentar regiões de vegetação em imagens de satélite em três classes sendo árvores, arbustos e grama no *Slovenia Dataset* e *Oregon Dataset*. Os resultados atingiram os valores de 78,0% de acurácia no Slovenia Dataset e 78,9% no Oregon Dataset. Já Sheng et al. (2020), utilizou uma arquitetura baseada na DeepLabV3+ para segmentar as seis classes de plantações do desafio Agriculture-Vision. Foram usadas a EfficientNet-B0 e a EfficientNet-B2 como encoders da DeepLabV3+. Os resultados em IoU médio das seis classes foram de 46,87%.

Além das arquiteturas baseadas em U-net e DeepLabV3+, outras arquiteturas foram encontradas como em Pena, Tan e Boonpook (2019) que propôs um modelo de segmentação semântica baseado em SegNet (BADRINARAYANAN; KENDALL; CIPOLLA, 2017) para detecção de tipos de cultura. Liu et al. (2020) implementaram um modelo baseado na arquitetura codificador-decodificador do UperNet (XIAO et al., 2018) por duas razões. Primeiro, é eficiente em termos de memória e é adequado para processar imagens de sensoriamento remoto de alta resolução. Em segundo lugar, o modelo UperNet aproveita a natureza hierárquica dos recursos e, portanto, pode capturar texturas de baixo nível e padrões complexos de imagens de sensoriamento remoto.

Na Tabela 2 compilamos a lista de Trabalhos Relacionados encontrados que utilizaram técnicas de aprendizagem de máquina para segmentação de áreas de plantações. Além da referência dos trabalhos, listamos o modelo de aprendizagem de máquina utilizado, o conjunto de dados e a tarefa resolvida.

Tabela 2 – Trabalhos Relacionados

Autor	**Modelo**	**Dataset**	**Tarefa**
Chiu et al. (2020a)	Baseado em U-net com portas de atenção	Agriculture-Vision	Segmentação de 6 classes de plantação
Ulmas e Liiv (2020)	Baseado em U-net com ResNet50 encoder	BigEarthNet e CORINE	Segmentação de áreas de plantações
Garnot e Landrieu (2021b)	Baseado em U-net com Temporal Attention	*PASTIS*	Segmentação de áreas de plantações
Abdani, Zulkifley e Mamat (2020)	Baseado em U-net com Spatial Pyramid Pooling	WiDS Datathon 2019	Segmentação de plantações de dendê
Wagner et al. (2019)	U-net	Dataset próprio com Satélite Worldview-3	Segmentação de florestas e plantações de eucalipto
Stoian et al. (2019)	Baseado em U-net - FG-UNET	Dataset próprio com Satélite Sentinel-2	Segmentação de cobertura terrestre
Kattenborn, Eichel e Fassnacht (2019)	Baseado em U-net	Dataset próprio com UAV imagery	Segmentação de áreas de plantações
Anand et al. (2021)	Baseado em DeepLabV3+ - AgriSegNet	Agriculture-Vision	Segmentação de 6 classes de plantação
Ayhan e Kwan (2020a)	Baseado em DeepLabV3+	Slovenia Dataset	Segmentação de árvores, arbustos e grama
Sheng et al. (2020)	Baseado em DeepLabV3+	Agriculture-Vision	Segmentação de 6 classes de plantação
Pena, Tan e Boonpook (2019)	Baseado em SegNet	Dataset próprio com Satélite Sentinel-2	Segmentação de áreas de plantações
Liu et al. (2020)	Baseado em UperNet	Agriculture-Vision	Segmentação de 6 classes de plantação

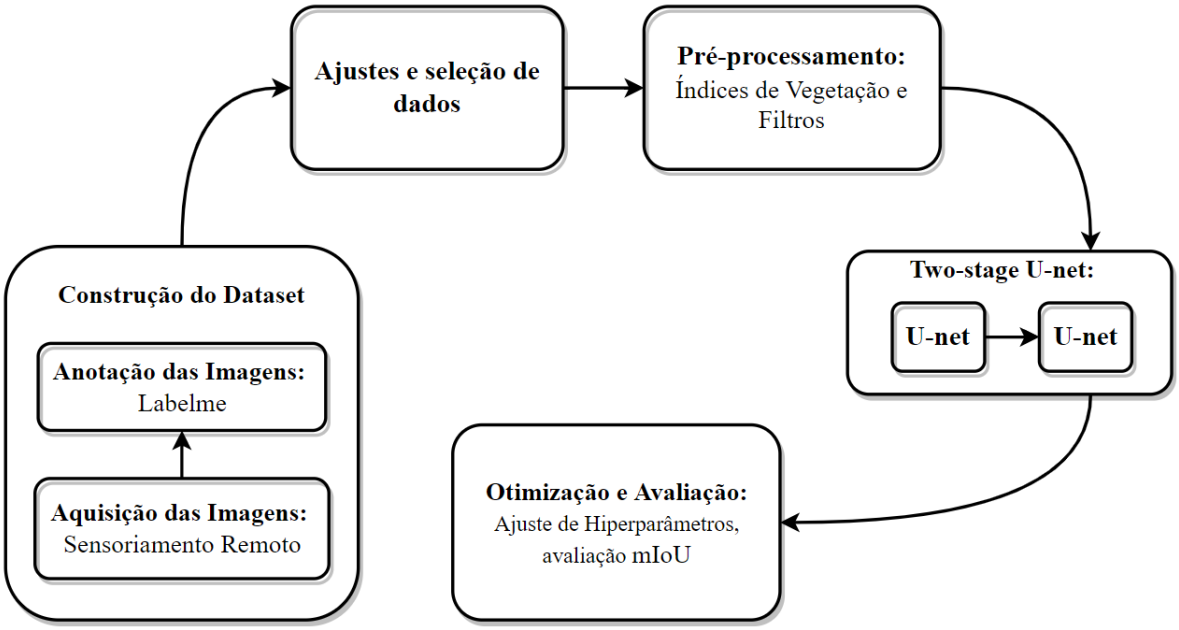
3.1 Considerações Finais

Neste capítulo foram apresentados os principais estudos encontrados na pesquisa bibliográfica, assim como seus métodos e os conjuntos de dados utilizados em problemas semelhantes ao exposto neste trabalho. O presente trabalho se difere dos trabalhos levantados neste capítulo, pois propõe uma arquitetura em dois estágios com o intuito de reduzir a complexidade da segmentação de múltiplas classes. No primeiro estágio são segmentadas classes que se diferem principalmente por características de textura e no segundo estágio as classes se diferem principalmente por cor. Além disso, neste trabalho é construído um conjunto de dados próprio com imagem de satélite de áreas de plantações do Brasil com suas particularidades locais.

4 Metodologia

Nesta seção, descrevemos a metodologia proposta para segmentar a área cultivada de plantações utilizando segmentação semântica com redes neurais convolucionais em imagens de satélite. Para isso, são seguidas as etapas de Construção do *Dataset*, que está subdividida em Aquisição e anotação das Imagens, seguida das etapas de Ajuste e seleção dos dados, Pré-processamento, Arquitetura proposta Two-stage U-net e Otimização e Avaliação do modelo. A Figura 17 ilustra a metodologia proposta. Cada uma dessas etapas é explicada nas seções a seguir.

Figura 17 – Metodologia Proposta



4.1 Construção do Dataset

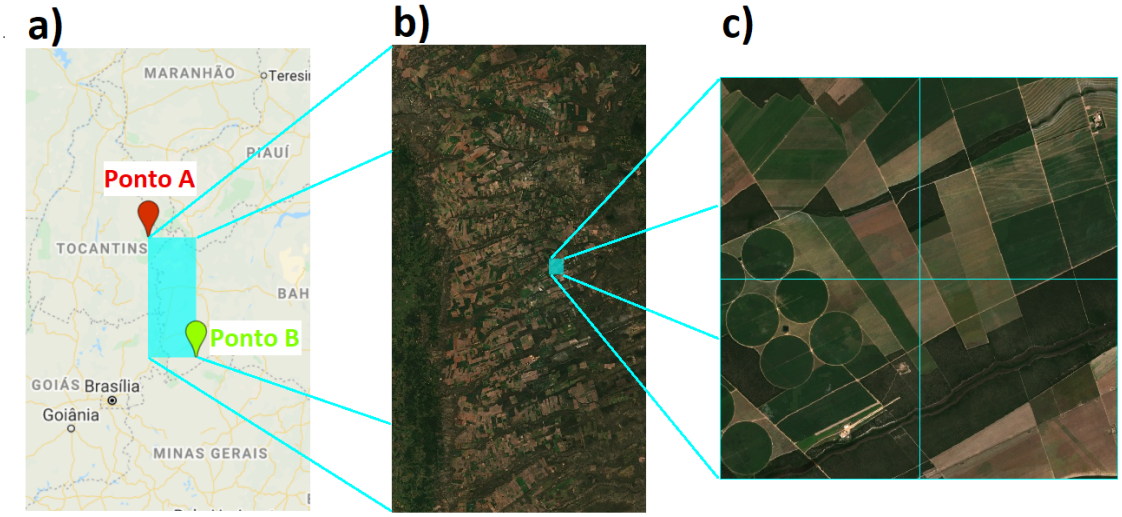
Conforme apresentado no Capítulo 3, foram levantados alguns conjuntos de dados de imagens utilizados em estudos que se propõem a segmentação de áreas de plantações. No entanto, nenhum dos conjuntos de dados encontrados apresenta o nível de detalhes desejado para o problema abordado e nem são de áreas brasileiras com sua vegetação, clima e plantações mais comuns. Dessa forma, optou-se por criar um conjunto de dados próprio.

A seguir são detalhados métodos para a construção do conjunto de dados próprio seguindo as etapas de Aquisição das Imagens que é realizada por meio de sensoriamento remoto com imagens do Satélite Sentinel-2 utilizando a plataforma *Google Earth Engine*; e a etapa de Anotação das Imagens que foi realizada com a auxílio da ferramenta *Labelme*.

4.1.1 Aquisição das Imagens

A aquisição das imagens é feita com o auxílio da plataforma de processamento geoespacial baseada em nuvem *Google Earth Engine* (GEE). O GEE é um catálogo de imagens de satélite e conjunto de dados geoespaciais que permite ao usuário visualizar, manipular, editar e criar dados espaciais (MUTANGA; KUMAR, 2019). Esta plataforma contém imagens de uma variedade satélites e apresenta uma interface de fácil integração com outros sistemas suportando códigos nas linguagens de programação de JavaScript e Python. O satélite utilizado neste estudo é o Sentinel-2 que apresenta resolução espacial de 10m por *pixel* nas suas 4 bandas principais (RGB-NIR) e atualização de imagens a cada 10 dias (RIBEIRO, 2019).

Figura 18 – a) Região de Aquisição. b) Imagem inteira. c) 4 blocos de imagem.

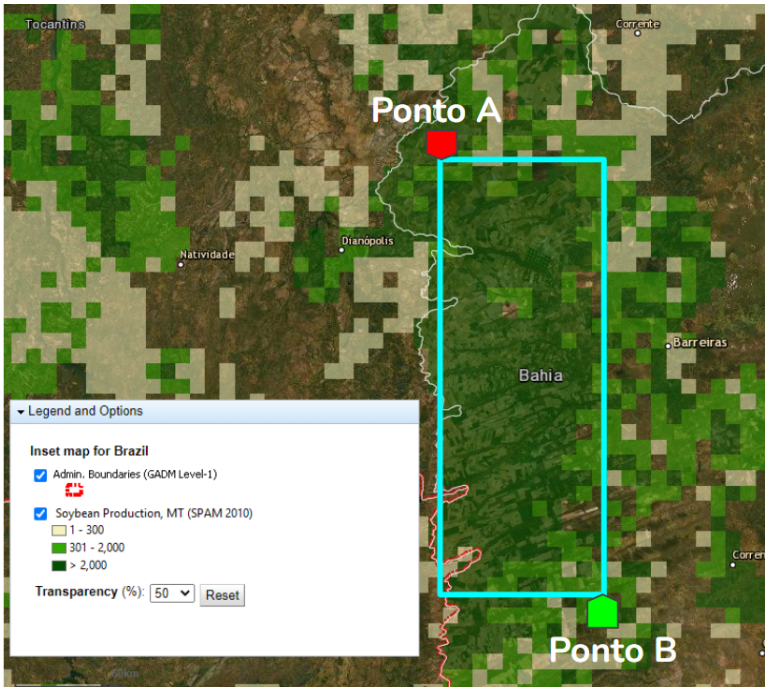


A Região de Estudo deste trabalho é uma área contínua de plantações no Nordeste do Brasil de 78.590,5 km^2 de área total e 1.233,8 km de perímetro como é ilustrado na Figura 18 a). O Ponto A da referida figura está localizado nas coordenadas 10°47'19.41"S e 46°73'36.57"W e o Ponto B nas coordenadas 14°32'85.88"S e 44°90'44.33"W. Essa região foi escolhida por ser uma vasta área de plantações com grande concentração de plantações de soja, milho e algodão. Essa alta concentração de plantações pode ser observada na Figura 19 com dados da USDA (*United States Departament of Agriculture*), o Ministério da Agricultura dos Estados Unidos.

A área de estudo também apresenta regiões de floresta, vegetação natural, áreas rochosas, rios, áreas alagadas, áreas planas, áreas irregulares, entre outras. Como se trata de uma área quase que exclusivamente de vegetação, é selecionada uma outra área para adquirir imagens urbanas e de mar.

Para adquirir as imagens que correspondem a essa Área de Estudo, são desenvolvidos agentes na linguagem Python que se integram com a API (*Application Programming Interface*) do GEE. A API permite definir a região de interesse passando a suas coordenadas,

Figura 19 – Concentração de plantação de soja na Região de Estudo.



Fonte: (USDA, 2022)

filtrar por data e pela probabilidade de nuvens para imagens com mais qualidade. Neste trabalho, após definir a região de estudo, é passado um período de 01 de janeiro de 2021 a 31 de dezembro de 2021 e filtrado por qualidade de imagem. Assim conseguimos maior qualidade de imagens e maior representatividade por adquirir imagens de todos as estações do ano.

O satélite Sentinel-2 possui uma variedade de bandas que capturam diferentes espectros eletromagnéticos da superfície terrestre. Para o propósito deste trabalho, são selecionadas 4 bandas de interesse que são as bandas B2 (Azul), B3 (Verde), B4 (Vermelho) e B8 (Near Infra Red) que representam o espectro visível RGB e o infravermelho próximo que tem alta reflexão em folhas de vegetação e por isso é muito usado neste tipo de estudo.

A região de aquisição das imagens cobre uma grande extensão de terra, dessa forma é necessário fragmentar a região em blocos de menor resolução. A área selecionada é fragmentada em forma tabelar (ou de grade) composta por vários blocos de imagens de resolução 256x256, como mostrado na Figura 18 c). A nomenclatura das imagens indicará sua posição na grade de imagens, de forma que seja possível remontar áreas maiores.

4.1.2 Anotação das Imagens

Após a etapa de aquisição e com posse das imagens na resolução apropriada, se inicia a etapa de Anotação das imagens. Neste processo, cada região da imagem é delimitada e recebe um rótulo que diz respeito à classe daquela região. O resultado do processo de

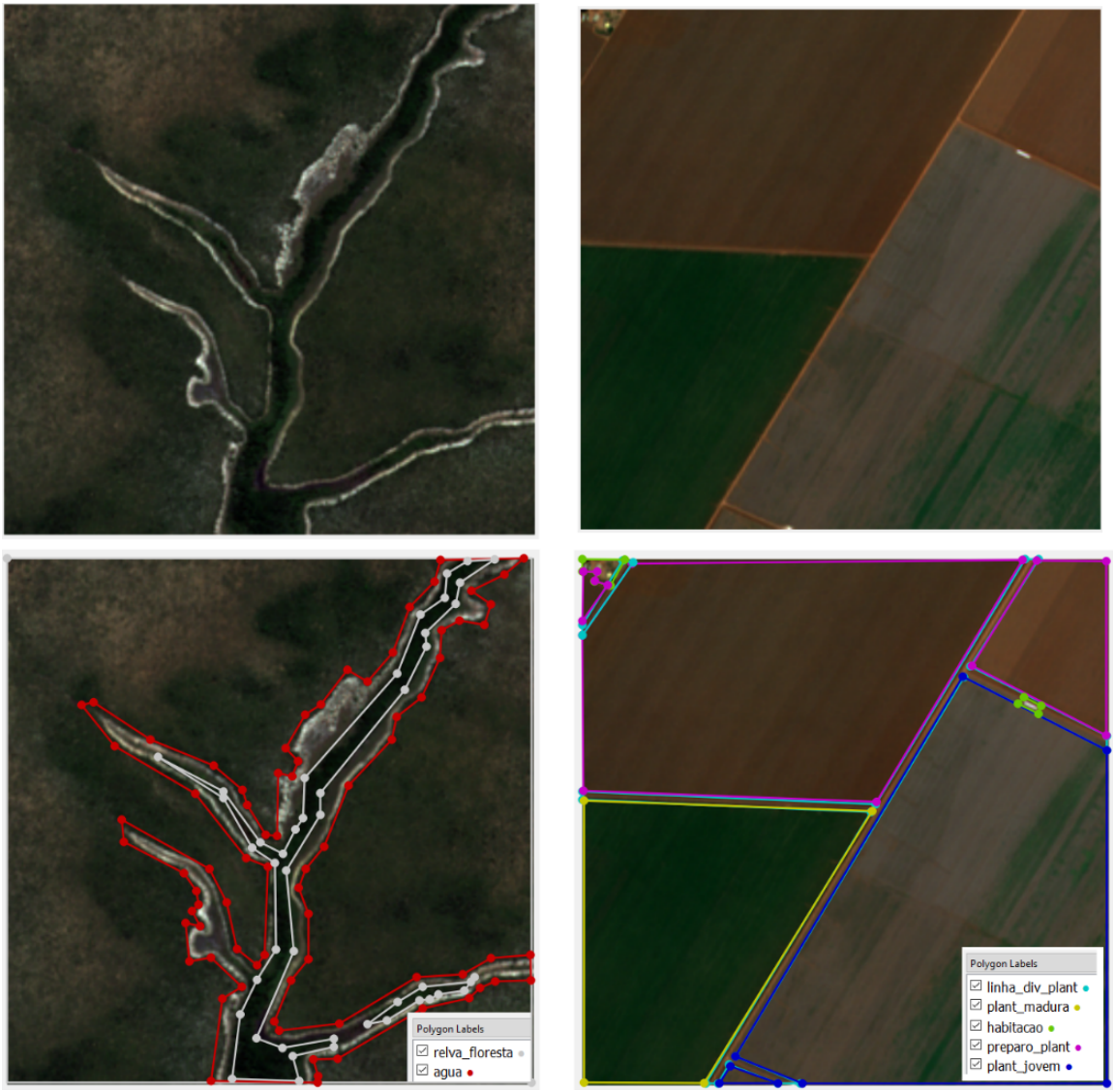
anotação são as máscaras de segmentação, onde cada pixel da região segmentada recebe um valor diferente para cada classe como ilustra a Figura 20. As máscaras de segmentação são utilizadas para realizar o treinamento e avaliação dos modelos aprendizagem profunda. Neste trabalho, o processo de anotação das imagens é realizado com o auxílio da uma ferramenta *Labelme* (WADA, 2016).

Figura 20 – Exemplo de máscara de segmentação de uma imagem, onde é possível ver que cada pixel representa uma classe.

A região de interesse deste trabalho é a área de cultivo das plantações. Para definição das classes, a área de cultivo de plantação está dividida em três classes de acordo com o estágio de desenvolvimento da plantação: Área de preparo de plantação (classe 1) que se trata da região de solo preparada para receber o plantio, Plantação jovem (classe 2) que é o estágio inicial da plantação que mescla áreas verdes com áreas ainda de solo e Plantação madura (classe 3) que são regiões onde a plantação está desenvolvida. Julgou-se importante anotar os caminhos que normalmente delimitam as plantações e foram nomeadas como Linhas de divisão das plantações (classe 4). A quinta classe representa as áreas de vegetação naturais, as áreas verdes que não são cultivadas, chamadas de Áreas de relva ou floresta (classe 5). As demais classes são Áreas de solo ou rochas (classe 6), Água (Áreas alagadas, lagos, rios e etc.) (classe 7) e Áreas de habitação (classe 8) que são casas ou construções artificiais nas regiões das fazendas. A definição das classes e anotação das áreas teve como referência outros conjuntos de dados com intuito similar: *DeepGlobe Land Cover* (DEMIR et al., 2018), *LandCoverNet* (ALEMOHAMMAD; BOOTH, 2020) e o *EOPatches Slovenia* (AYHAN; KWAN, 2020b). A Figura 21 mostra a anotação dessas classes com a ferramenta *Labelme*.

O processo de anotação das imagens é realizado em três etapas por pessoas diferentes, onde a primeira faz a anotação de todas a imagens, a segunda revisa todas as anotações feitas pela primeira pessoa e realiza as correções necessárias. A terceira pessoa é um especialista na área de agricultura que realiza a revisão de uma amostra das imagens anotadas e informa os ajustes necessários. Esse processo é ilustrado na Figura 22.

Figura 21 – Exemplos de anotação das imagens com o *Labelme*. As imagens à esquerda são de uma área de vegetação sem plantação e apresentam as classes Relva ou Floresta e Água. As imagens à direita são de uma área de plantação e apresentam todas as demais classes.

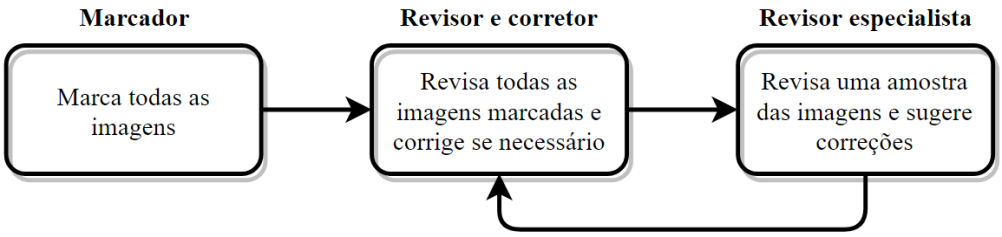


4.2 Ajustes e Seleção dos dados

O *dataset* construído contém 8 classes que podem ser usadas para uma variedade de tipos de aplicações. É uma das contribuições deste trabalho para a comunidade científica. No entanto, o problema a ser resolvido no estudo se limita a área cultivada das plantações. Dessa forma, para os experimentos deste trabalho, realizamos um processamento em que agrupamos as classes em áreas de Plantação e áreas de Não Plantação.

As classes agrupadas como Plantação são Plantação Madura, Plantação Jovem e Solo Preparado para Plantio que individualmente também são interessantes para este estudo, porque mostram o nível de maturação da plantação. No entanto, testes preliminares contendo as classes Plantação Madura e Jovem resultaram em um alto grau de confusão

Figura 22 – Etapas de Marcação das Imagens



entre as duas classes. Suspeitamos que seja porque os modelos assumam a cor como principal característica para definir a classe, no entanto nos experimentos deste trabalho são testados diferentes tipos de culturas de plantação que apresentam tons de cor diferentes durante seu desenvolvimento, dessa forma a cor causaria confusão entre culturas diferentes. Por outro lado, a área de Preparo para Plantio não apresenta este problema e é considerada neste estudo.

A arquitetura de rede proposta neste estudo realiza uma segmentação em dois estágios como será detalhado na próxima seção. O primeiro estágio realiza a segmentação Plantação/Não Plantação, o segundo estágio detalha a área de Plantação em Plantação Verde e Área de Preparo. Na Tabela 3 são apresentadas as classes de cada um desses estágios e sua correspondência com as classes originais do *dataset*.

Tabela 3 – Definição das classes agrupadas.

Classes originais do Dataset	**Classes estágio 1**	**Classes estágio 2**
Plantação Madura	Plantação	Plantação Verde
Plantação Jovem		
Solo Preparado para Plantio		Solo Preparado para Plantio
Caminho/Estrada na Plantação	Não Plantação	Não Plantação
Relva ou Floresta		
Areia ou Rocha		
Água		
Habitação		

4.3 Pré-processamento

Na etapa de pré-processamento são experimentados dois pré-processamentos nas imagens, sendo eles os dois índices de vegetação apresentados no Capítulo 2, o NDVI e o EVI. Os índices de vegetação são usados no sensoriamento remoto para de acentuar uma cor específica, como o verde da plantação (WOEBBECKE et al., 1995). As fórmulas dos dois índices de vegetação são descritas abaixo.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{4.1}$$

$$EVI = G\frac{NIR - R}{NIR + C_1R + C_2B + L} \tag{4.2}$$

Antes de iniciar o treinamento da arquitetura, há um processo de aplicar os índices de vegetação nas imagens de entrada, o resultado de cada índice é adicionado como mais um canal de recursos. São testadas algumas variações para avaliarmos o comportamento da rede.

4.4 Arquitetura Proposta - Two-stage U-net

Como explanado no Capítulo 2, as redes neurais convolucionais são modelos biologicamente inspirados que podem aprender características de forma hierárquica, especialmente projetados para lidar com uma variabilidade em dados bidimensionais, como as imagens em formato matricial (LECUN; BENGIO; HINTON, 2015). As redes neurais convolucionais resolvem muito bem problemas de classificação de imagens. No entanto, para problemas de segmentação de imagens, é necessário que tanto a entrada da rede quanto a saída da rede sejam imagens. Para este objetivo, foram propostas as Redes Totalmente Convolucionais, onde a rede produz uma saída nas mesmas dimensões da imagem de entrada, realizando uma segmentação pixel-a-pixel da imagem de entrada, como é o caso da U-net.

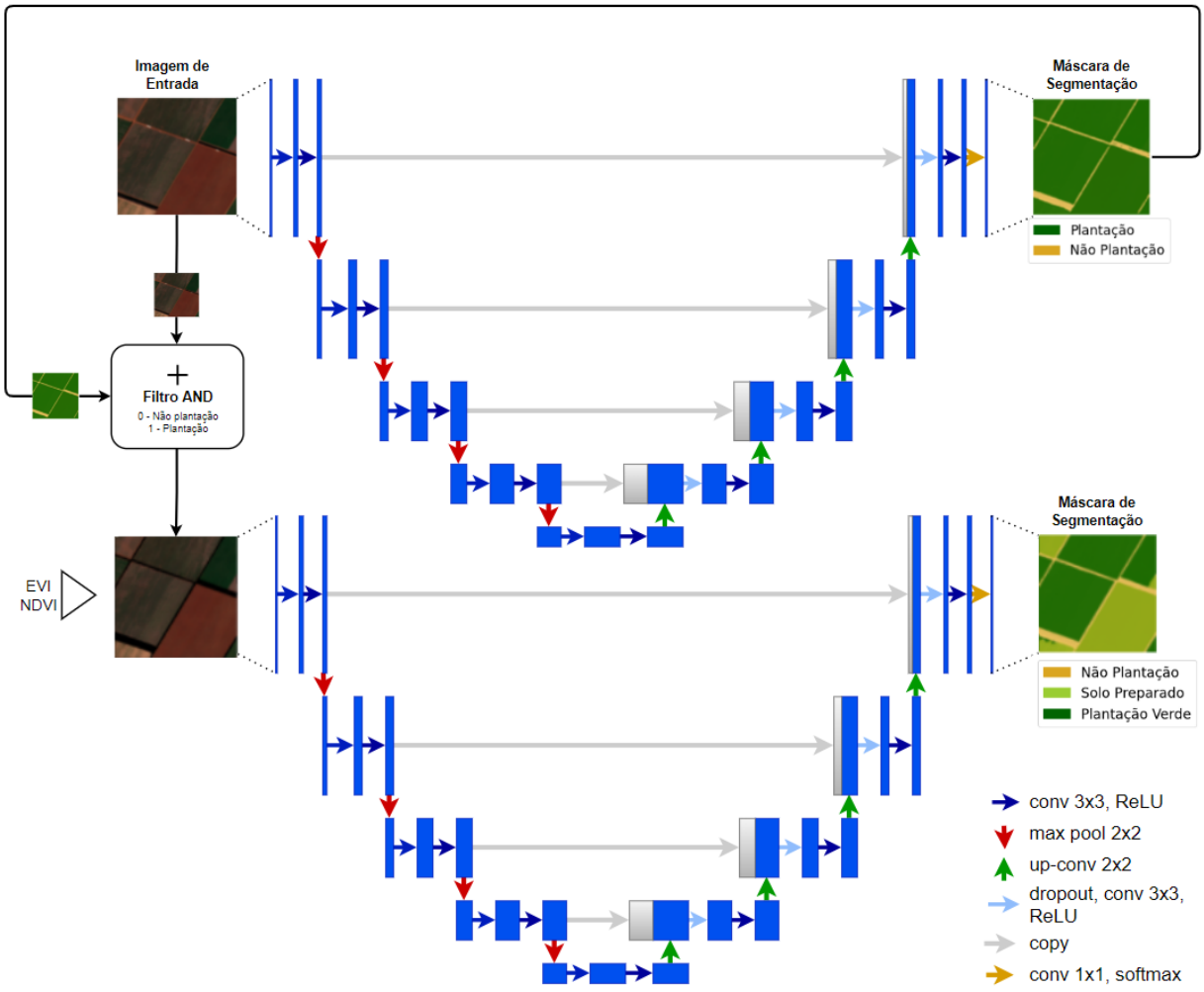
A arquitetura de rede proposta neste trabalho possui dois estágios, por isso o nome Two-stage e a arquitetura é baseada na U-net. A arquitetura completa da Two-stage U-net é ilustrada na Figura 23. No Estágio 1, a primeira rede baseada em U-net recebe as imagens de satélite com as bandas RGB-NIR e é treinada para aprender a segmentar as regiões de Plantação e Não Plantação. No Estágio 2, a segunda rede baseada em U-net também recebe as imagens RGB-NIR mais os índices de vegetação NDVI e EVI, porém com uma operação filtragem entre as imagens e suas máscaras de segmentação preditas no Estágio 1. O resultado desse processamento é a imagem RGB-NIR sem as regiões classificadas como Não Plantação no Estágio 1. Com isso, o Estágio 2 serve para corrigir erros de segmentação do Estágio 1 na predição da região de Não Plantação e serve para focar em aprender a diferenciar as sub-regiões de Plantação: Plantação Verde e Solo Preparado para o Plantio. Dividir as tarefas dessa forma faz com que a rede seja mais precisa.

A intenção ao propor uma arquitetura em dois estágios é reduzir a complexidade da segmentação de múltiplas classes considerando que há semelhanças de cor, por exemplo, nas classes Relva ou Floresta, que é da classe Não Plantação, com as classes Plantação jovem e Plantação madura que são classe Plantação. Todas essas classes são áreas de vegetação apresentam cores esverdeadas. O mesmo ocorre com a semelhança de cor entre as classes areia ou rocha (Não Plantação) com a classe Solo Preparado para Plantio (Plantação). O Estágio 1 aprende a resolver este tipo de problema, onde aspectos de textura, identificação

de linhas de arado, contornos regulares e outras características, são mais relevantes que aspectos de cor.

As áreas segmentadas como Não Plantação no Estágio 1, recebem o valor zero e ficam com a cor preta nas imagens entrada do Estágio 2. Naturalmente, a rede do segundo estágio aprende que cor preta significa Não Plantação. Eliminando essa complexidade inicial, o segundo estágio da arquitetura, além de melhorar a previsão da classe Não Plantação, consegue focar nas outras classes, inclusive levando em conta aspectos de cor, pois neste estágio, para diferenciar as classes Plantação Verde e Área de Preparo, as características cor são importantes. Por isso, no Estágio 2 também são incluídos os Índices de Vegetação *NDVI* e *EVI*.

Figura 23 – Arquitetura Proposta - Two-stage U-net



4.5 Ajuste de Hiperparâmetros

Hiperparâmetros são parâmetros que são escolhidas pelo desenvolvedor e que afetam diretamente no desempenho da rede, como a quantidade de épocas em que a rede é treinada,

o tamanho do lote de imagens, a resolução das imagens, taxa de aprendizado, função de perda, entre outras variáveis. Os hiperparâmetros se diferenciam dos demais parâmetros da rede, pois não podem ser estimados diretamente a partir do aprendizado das redes e devem ser configurados antes do início do treinamento de um modelo de Aprendizado de Máquina, uma vez que esses parâmetros definem a arquitetura do modelo (YANG; SHAMI, 2020).

Na etapa de ajuste dos Hiperparâmetro, a rede totalmente convolucional será treinada com variadas combinações de hiperparâmetros que serão ajustados para otimizar os resultados. Os experimentos se darão utilizando bibliotecas para este fim, como o *Hyperopt* que testa combinações aleatórias de hiperparâmetros e é a apropriada para problemas com espaço de testes muito grande. Também é usado o algoritmo *GridSearch* que testa todas as combinações do espaço de teste dado e por isso é um método mais completo que o anterior, no entanto pode levar muito tempo para concluir os experimentos. Todos os resultados dos testes são listados e é selecionado aquele com melhor desempenho em média de IoU das classes testadas. Serão testados os seguintes hiperparâmetros:

Tabela 4 – Conjunto de parâmetros para ajuste.

Hiperparâmetro	**Conjunto de parâmetros**
Encoder	VGG, ResNet, SeResNet, ResNext, SeResNext, SeNet, DenseNet, Inception, InceptionResNet, MobileNet e EfficientNet
Função de Perda	Jaccard, Dice, Categorical Focal, Categorical Cross Entropy e suas combinações.
Batche Size	2, 4, 6, 8, 10.
Otimizador	Adam, Ftrl, Adagrad, Adamax, RMSprop, SGD, Nadam.

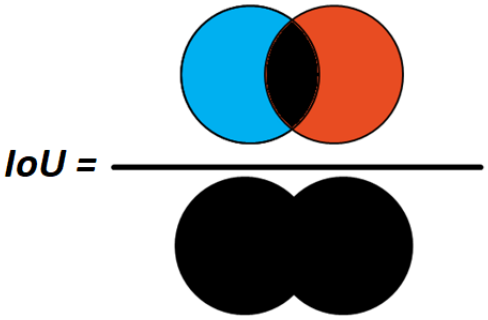
4.6 Avaliação da Desempenho

O desempenho da segmentação de imagens é medido a partir de da métrica de avaliação Interseção sobre União, conhecida como IoU. A métrica mIoU (IoU médio) é definida a seguir com sua formulação matemática na Equação 4.3, considerando que GT é a região real do objeto, que $Pred$ é a região predita pela rede e c é número de componentes. Na Figura 24 é apresentada uma representação visual da métrica IoU, onde regiões na cor preta representam a interseção e a união entre uma segmentação real e uma predita.

$$mIoU = \frac{1}{c}\sum_{i=1}^{c}\frac{Area(GT_i \cap Pred_i)}{Area(GT_i \cup Pred_i)} \tag{4.3}$$

Além da métrica IoU, em alguns momentos também são utilizadas métricas complementares, para isso são usadas as métricas de Precisão (PRE), Acurácia (ACU),

Figura 24 – Representação visual da métrica IoU.



Recall e F-score, cujas formulações estão descritas em função das medidas Verdadeiros Positivos (TP), Falsos Negativos (FN), Falsos Positivos (FP) e Verdadeiros Negativos (TN) (ZHU et al., 2010).

$$PRE = \frac{TP}{TP + FP} \tag{4.4}$$

$$ACU = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.5}$$

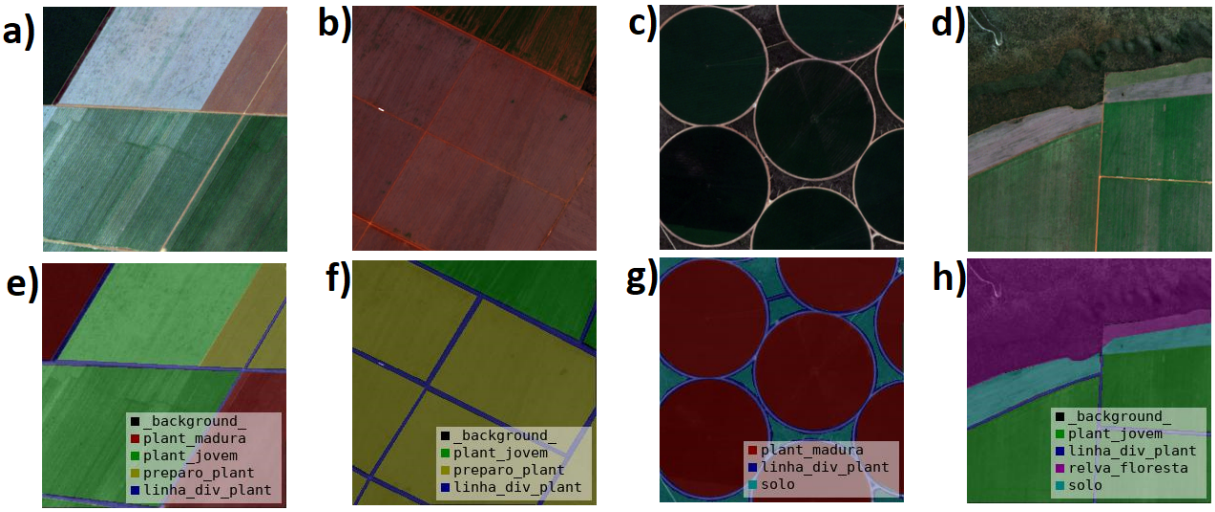
$$Recall = \frac{TP}{TP + FN} \tag{4.6}$$

$$Fscore = \frac{2 \times PRE \times Recall}{PRE + Recall} \tag{4.7}$$

5 Resultados

Neste capítulo são apresentados os resultados da avaliação do método proposto. Inicialmente, foi criado um conjunto de dados para segmentar as áreas de plantio. As imagens foram adquiridas do satélite *Sentinel-2* e obtidas através da plataforma *Google Earth* com a seleção das bandas necessárias e fragmentação da região de interesse em 9860 Blocos de imagem com resolução de 256x256. Foram marcadas 300 imagens que passaram pelas etapas de revisão e correção de anotações. O *dataset* está disponível publicamente no *GitHub*[1] e alguns exemplos de imagens marcadas são apresentados na Figura 25.

Figura 25 – a), b), c) e d) são imagens em RGB do conjunto de dados e e), f), g) e h) são, respectivamente, suas anotações.

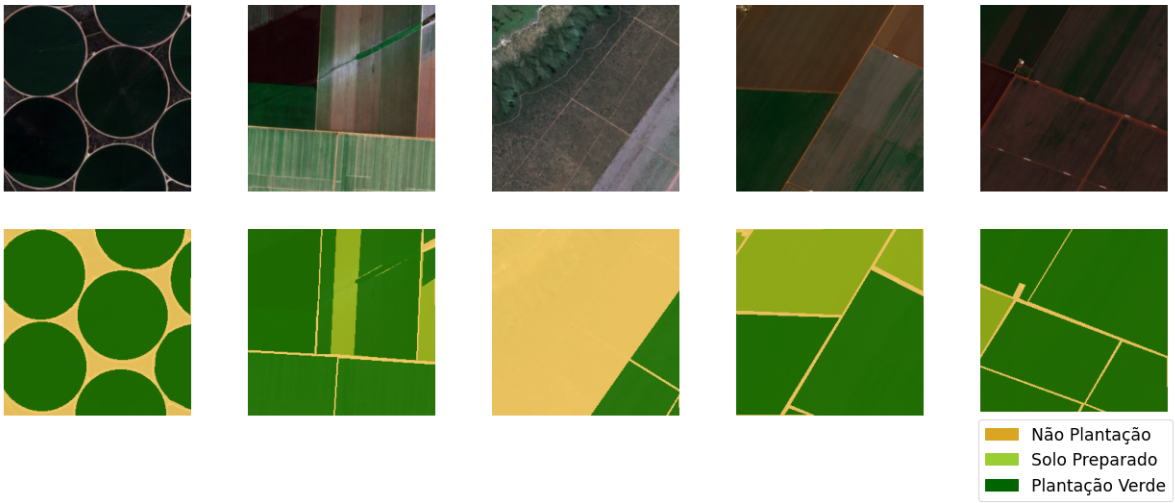


Após a etapa de Ajustes e Seleção de dados em que as classes são agrupadas para as classes de interesse do estudo, as imagens e as máscaras de segmentação geradas para o Estágio 1 apresentam duas classes e do Estágio 2 apresentam 3 classes. Exemplos de imagem e máscaras de segmentação do segundo estágio são apresentadas na Figura 26.

Realizadas as etapas iniciais que dizem respeito aos dados e a definição da arquitetura de aprendizado profundo, foram realizados os treinamentos da rede e ajustes mais específicos para obter a melhor configuração do modelo proposto. Dessa forma, foram realizados os testes de hiperparâmetros considerando a rede de *Encoder* da U-net responsável pela codificação da primeira metade da rede, a Função de Perda que é a função que o algoritmo de treinamento buscará minimizar, tamanho do Lote de imagens (*batch size*) que entrarão ao mesmo tempo na rede para treinamento e o otimizador que é a função

[1] Disponível em: <https://github.com/walysson21/AgriVis_Dataset>

Figura 26 – Imagens e máscaras de segmentação após a etapa de Ajustes e Seleção dos Dados para o Estágio 2 da arquitetura proposta.



que irá minimizar os erros. O espaço de teste do conjunto de parâmetros foi apresentado na Tabela 4 do Capítulo 4.

Para o treinamento inicial com o propósito de ajustar os hiperparâmetros do modelo, foi utilizada a biblioteca *Hyperopt* para gerar 100 combinações dos hiperparâmetros e realizar o treinamento dos modelos gerados para 250 épocas no conjunto de treinamento e taxa de aprendizado igual a 0,0001. Usamos uma validação de retenção em que o conjunto de dados foi embaralhado e dividido em 50% para o conjunto de treinamento, 20% para o conjunto de validação e 30% para o conjunto de teste. Usamos pesos pré-treinados da *imagenet* e os testes de cada combinação levaram em média 3 horas para serem concluídos rodando em uma GPU NVIDIA TESLA P100 (*Driver Version: 460.32.03, CUDA Version: 11.2*), o que implicou 14 dias de execução dos testes de hiperparâmetros.

Além disso, uma função de *callback* foi implementada para monitorar o desempenho da rede e salvar os pesos da rede no momento com o melhor *mIoU* no conjunto de validação. Ao final do treinamento das 250 épocas de cada combinação, os pesos da melhor rede foram recarregados na rede e aplicados ao conjunto de teste. Concluídas as 100 combinações, foi selecionada aquela com o melhor resultado no conjunto de teste. A combinação com melhor resultado para o teste utilizando o *Hyperopt* é apresentada na Tabela 5:

Tabela 5 – Melhor Combinação de hiperparâmetros com *Hyperopt*

Hiperparâmetro	**Resultado**
Encoder U-net	EfficientNetB7
Batch Size	8
Função de Perda	*Binary Focal Loss*
Otimizador	Adam

O número de combinações possíveis para o espaço do conjunto de hiperparâmetros selecionados é de $11 \times 4 \times 5 \times 7 = 1540$ combinações. Seguindo a média de 3 horas para o treinamento de cada combinação, levaria mais de 6 meses de testes ininterruptos. Para ter mais assertividade nos testes, foram realizadas novas rodadas de ajustes de hiperparâmetros. Dessa vez, foi utilizado o algoritmo *Grid Search* considerando apenas os Encoders para o Estágio 1 da arquitetura proposta e em seguida para a arquitetura completa com saída no Estágio 2. O histórico de teste dos modelos é ilustrado nas Figuras 27 e 28.

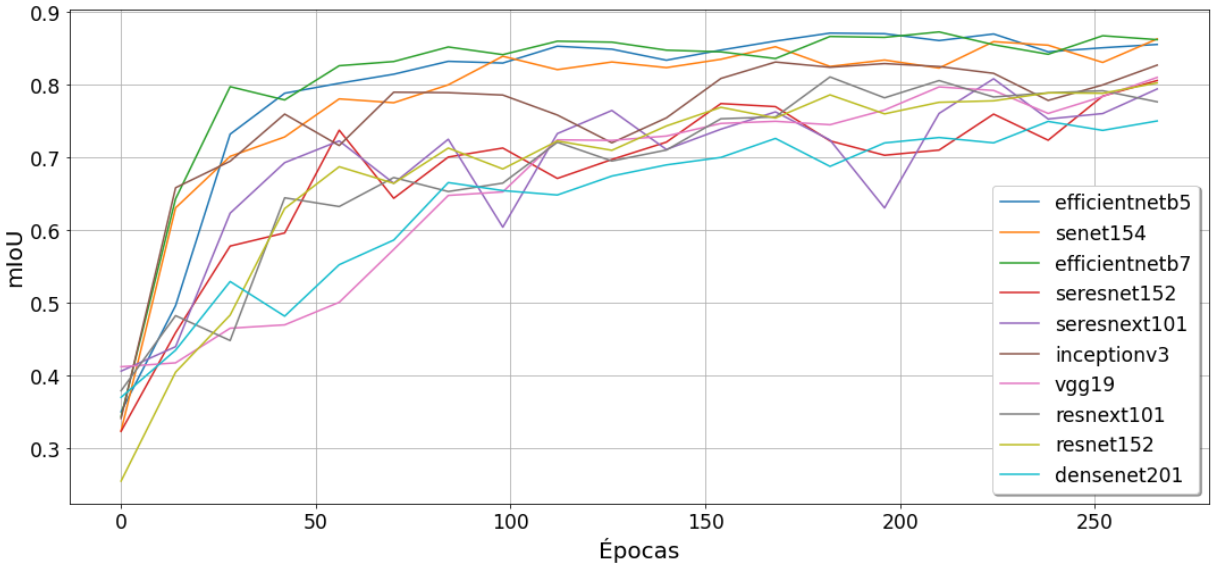


Figura 27 – Teste de Hiperparâmetros - Encoders U-net Estágio 1

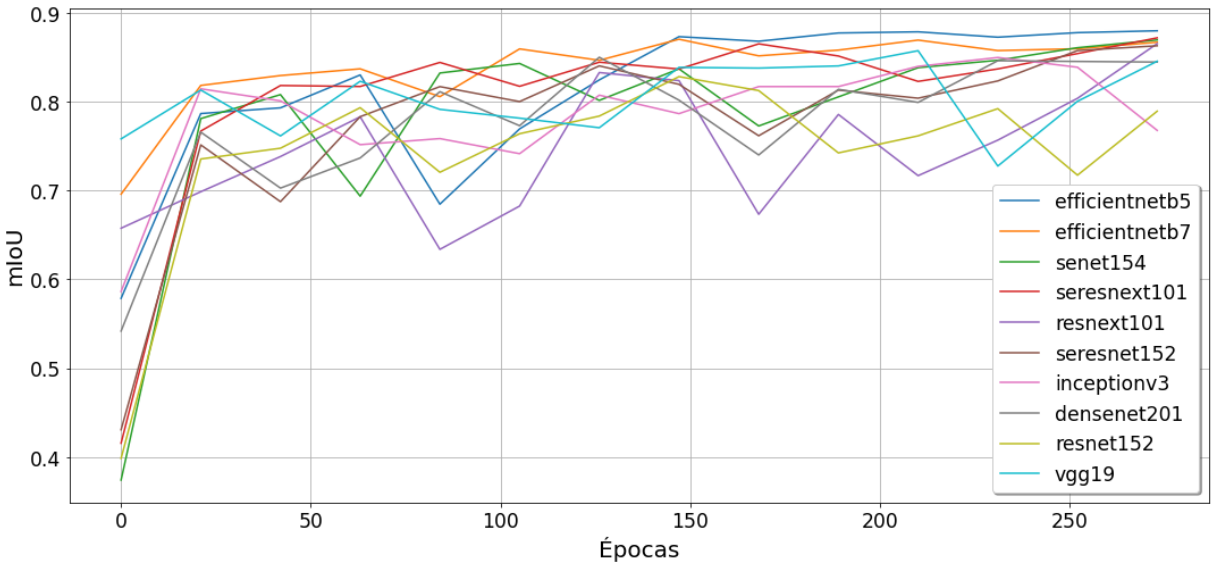


Figura 28 – Teste de Hiperparâmetros - Encoders U-net Estágio 2

Os testes com *Grid Search* para cada estágio da arquitetura Two-stage U-net foram realizados mantendo os demais hiperparâmetros selecionados no primeiro teste

com *Hyperopt*, variando apenas o *Encoder* para cada um dos dois estágios. Conforme as Figuras 27 e 28 mostram, o melhor *Encoder* em ambos os estágios foi o *EfficientNetB5* seguido de perto do *EfficientNetB7*. Sendo assim, a melhor combinação de hiperparâmetros selecionada após o teste dos *Encoders* com o *Grid Search* é apresentado na Tabela 6.

Tabela 6 – Melhor Combinação de hiperparâmetros com *Hyperopt* e *Grid Search*

Hiperparâmetro	**Resultado**
Encoder U-net	EfficientNetB5
Batche Size	8
Função de Perda	*Binary Focal Loss*
Otimizador	Adam

5.1 Segmentação das áreas de plantação

A segmentação é realizada em duas etapas seguindo a arquitetura proposta. A rede da etapa 1 teve como entrada quatro bandas das imagens de satélite sendo elas as bandas RGB e NIR. Os resultados visuais da etapa 1 são apresentados na Figura 29 com a segmentação das áreas de plantação em verde e áreas de não-plantação em dourado que mostra que as áreas de plantação das imagens de exemplo são bem delimitadas e visualmente próximas da segmentação real, porém com ruídos na segmentação.

Os resultados numéricos da segmentação da etapa 1 são apresentados na Tabela 7 e são coerentes com as observações dos resultados visuais. O IoU médio teve um resultado de 84,04%, o que pode ser considerado um resultado mediano para aplicações médicas devido a criticidade da aplicação, mas é um resultado razoavelmente consistente para uma aplicação da agricultura e com imagens de satélite. Os resultados da Precisão, Recall e F-score tiveram resultados acima de 90%.

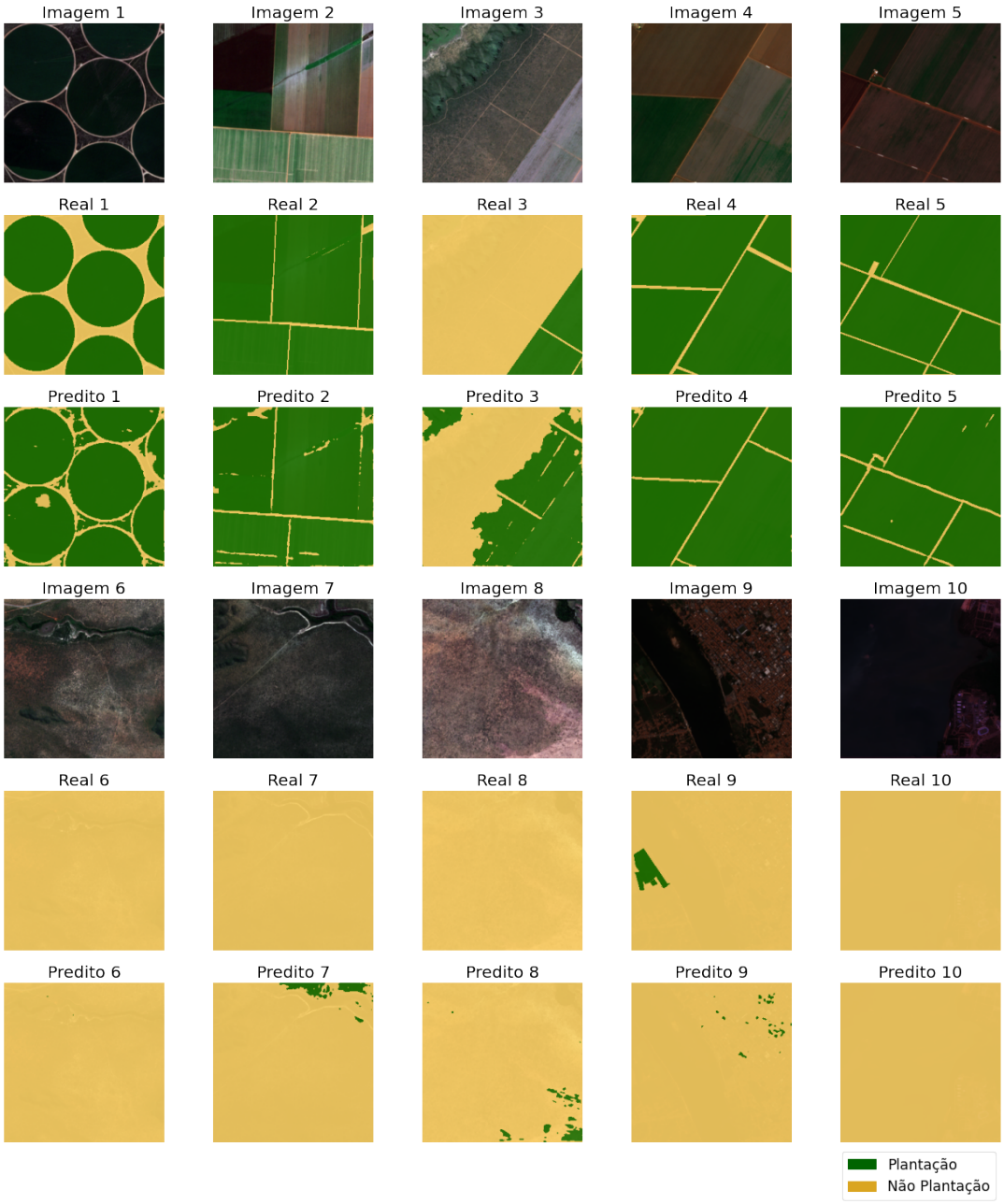
Tabela 7 – Resultado numérico das métricas para a primeira rede da arquitetura proposta.

Classe	**IoU**	**Precisão**	**Recall**	**F-score**
Plantação	88,26%	94,83%	92,57%	93,67%
Não Plantation	79,81%	86,92%	90,41%	88,57%
Média	84,04%	90,87%	91,49%	91,12%

Na Figura 30 são apresentados os resultados qualitativos da etapa 2 que apresenta um detalhamento maior nas áreas de plantação que foram subdivididas em áreas de Plantação Verde e de Solo Preparado para Plantio.

É possível observar na Figura 30 que a região de Não Plantação teve resultado muito similar à etapa 1, que era o que se pretendia com a segmentação em duas etapas,

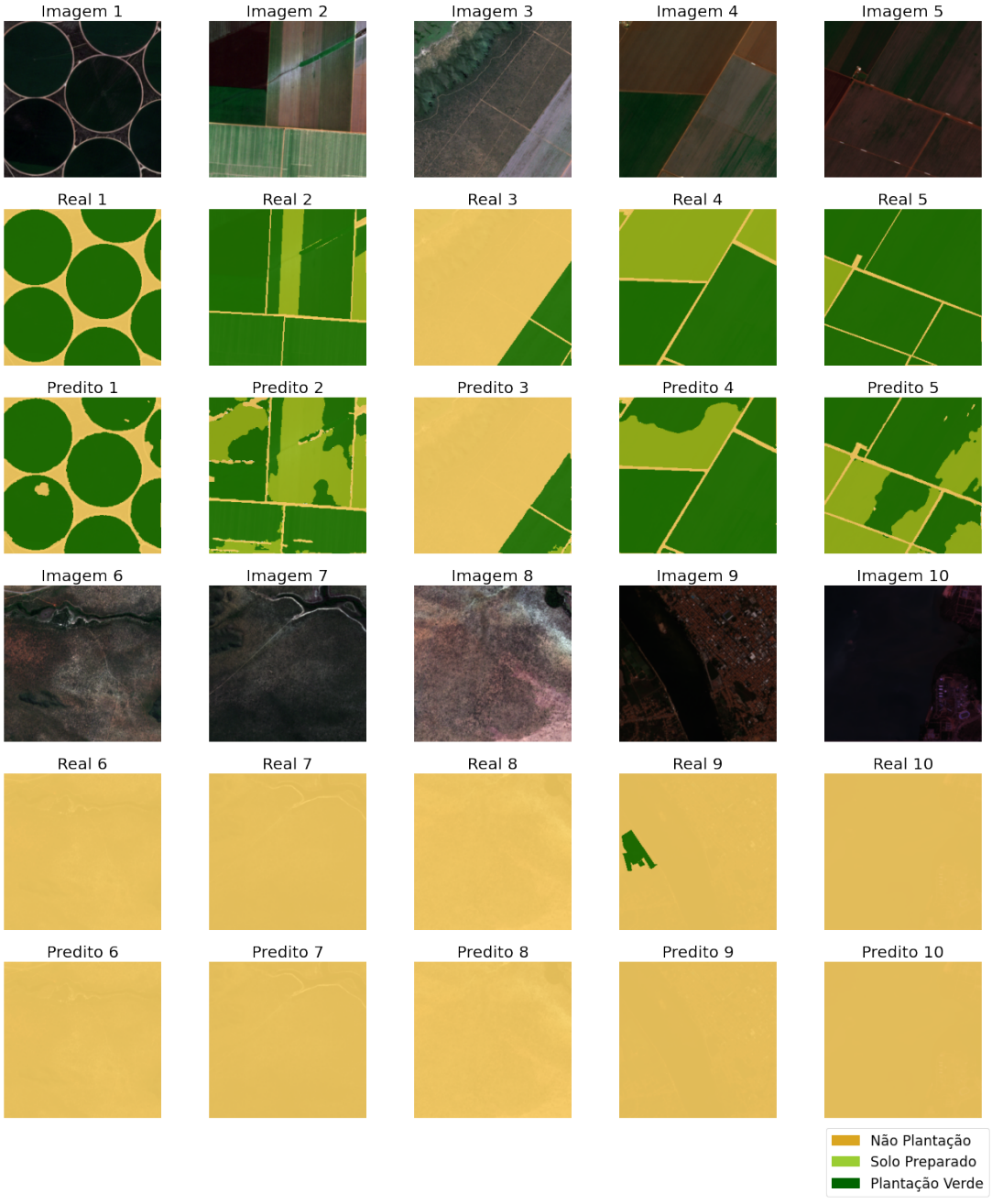
Figura 29 – Na primeira linha alguns exemplos em RGB, na linha central a máscara de segmentação verdadeira para as classes Não Plantação e Plantação e na última linha o resultado da segmentação da rede no Estágio 1. As linhas 4, 5 e 6 também são imagens RGB, máscaras de segmentação e predição da rede, mas as imagens são de área de que não são de plantação como áreas rochosas, rios, cidades e mar.



mantendo uma boa delimitação da área e sendo bem próxima da segmentação real. Um fato interessante que foi observado, é que os ruídos da segmentação observados na etapa 1 foram corrigidos na etapa 2 para a região de Não Plantação. Isso é um fato muito positivo e indica a região predita como Não Plantação na etapa 1 também é predita como Não Plantação na etapa 2, porém a região predita como Plantação na etapa 1 é predita como qualquer uma das três classes da etapa 2.

Por outro lado, as subáreas de Plantação: Solo Preparado e Plantação Verde não tiveram uma delimitação tão boas das áreas e percebe-se que os ruídos da etapa 1 na região de Plantação é transmitido como erro para a etapa 2.

Figura 30 – Na primeira linha alguns exemplos em RGB, na linha central a máscara de segmentação verdadeira para as classes Não Plantação, Solo Preparado para Plantio e Plantação Verde e na última linha o resultado da segmentação da rede no Estágio 2. As linhas 4, 5 e 6 também são imagens RGB, máscaras de segmentação e predição da rede, mas as imagens são de área de que não são de plantação como áreas rochosas, rios, cidades e mar.



Os resultados numéricos da etapa 2 são apresentados na Tabela 8 e mostram um resultado de IoU da classe Não Plantação de 92,61% que melhorou em relação ao resultado

de 79,81% da etapa 1 e se mostrou bem superior em comparação aos valores de IoU de 76,68% e 78,17% das classes Plantação Verde e Solo preparado, respectivamente, que em média são inferiores ao resultado de 88,26% da etapa 1. No geral, o IoU médio foi de 82,49% e os valores de Precisão, Recall e F-score ficaram em torno de 90%.

Tabela 8 – Resultado numérico das métricas para a arquitetura proposta completa

Classe	**IoU**	**Precisão**	**Recall**	**F-score**
Plantação verde	76,68%	83,41%	90,74%	86,77%
Solo Preparado	78,17%	93,07%	82,84%	87,65%
Não Plantação	92,61%	92,90%	99,67%	96,15%
Média	82,49%	89,79%	91,08%	90,19%

5.2 Discussão

Para comparação de resultados e para verificarmos se o método proposto de fato tem um bom desempenho, o mesmo problema com o mesmo conjunto de dados foi aplicado a outras redes de segmentação semântica, entre eles os principais modelos elencados nos Trabalhos Relacionados. Os treinamentos foram realizados com 250 épocas, otimizador *Adam* e Função de Perda *Jaccard Loss*. Os resultados são apresentados na Tabela 9.

Tabela 9 – Comparação dos Resultados - mIoU (%)

Modelo	**Média**	**Não Plant.**	**Plant. Verde**	**Solo Prep**	**Tempo**
Two-stage U-net	**82,61%**	**92,61%**	**76,68%**	**78,17%**	594 ms
Unet+efficientnetb7	76,17%	77,30%	75,21%	76,01%	428 ms
Unet+efficientnetb6	75,62%	75,46%	73,94%	77,47%	229 ms
DeepLabv3+	74,93%	78,60%	69,62%	76,57%	267 ms
Unet+efficientnetb5	74,59%	76,39%	72,91%	74,48%	297 ms
Unet+efficientnetb4	74,55%	75,68%	74,38%	73,58%	301 ms
Unet+inceptresnetv2	73,53%	76,93%	69,49%	74,16%	339 ms
Unet + senet154	73,42%	75,52%	71,31%	73,43%	713 ms
Unet+efficientnetb3	72,98%	74,15%	69,11%	75,67%	223 ms
SegNet	72,33%	75,71%	65,27%	76,01%	196 ms
Unet + vgg16	72,25%	71,97%	66,89%	77,89%	113 ms
Unet + vgg19	71,21%	70,93%	67,33%	75,38%	126 ms
Unet + seresnet152	71,16%	70,86%	68,76%	73,85%	198 ms
Unet + densenet201	70,81%	74,27%	64,63%	73,51%	366 ms
Unet + seresnet18	69,96%	66,61%	68,95%	74,33%	63 ms
Unet + seresnext101	69,87%	70,76%	65,51%	73,36%	238 ms
Unet + inceptionv3	68,51%	72,94%	64,06%	68,52%	192 ms
Unet + resnet18	66,21%	63,86%	61,27%	73,51%	60 ms
PSPNet	63,71%	59,31%	58,18%	73,66%	159 ms

Os resultados da Tabela 9 mostram que o método proposto apresenta desempenho superior em cada uma das classes em relação aos demais modelos testados. Como era de se esperar, a maior diferença é encontrada no desempenho da classe Não Plantação, pois no método proposto há um estágio exclusivo para segmentar essa região.

Os modelos mais encontrados nos Trabalhos Relacionais para segmentar áreas de plantação foram os modelos baseados em *SegNet*, baseados em *DeepLabV3+* e baseados em U-net. A *SegNet* original foi testada obtendo 72,33% de média de IoU entre as 3 classes. Já a *DeepLabV3+* obteve 74,93% de média de IoU entre as 3 classes, um resultado melhor que a *SegNet*, porém inferior ao método proposto e à U-net simples. O modelo que mais se aproximou do método proposto foi a U-net com uma *EfficientNetB7* no *encoder* com 76,17% de média de IoU.

A arquitetura Two-stage U-net divide o problema em dois e faz a segmentação em dois estágios. Isso tem um custo, nessa arquitetura há muitos parâmetros, o que implica em mais processamento e mais tempo de inferência. Os tempos de inferência do modelo proposto e de outros modelos são apresentados na última coluna da Tabela 9 para um Lote (*batch size*) de 8 imagens rodando em uma GPU NVIDIA TESLA P100 (*Driver Version: 460.32.03, CUDA Version: 11.2*). O tempo de inferência da Two-stage U-net de 594 ms ficou atrás apenas da U-net com *encoder* senet152 que levou 713 ms para processar as oito imagens do Lote. Em seguida vieram as redes U-net com encoder da família EfficientNet.

A *Two-stage U-net* tem como *Encoders* as redes da família *EfficientNet* que tem ótimos resultados de *mIoU*, porém são redes mais pesadas. Por isso, testamos uma versão mais leve da arquitetura proposta com *Encoders* da *resnet18* que foi o mais rápido entre os *Encoders* testados. Essa rede com menos parâmetros foi chamada de *Two-stage U-net Lite* e treinada seguindo o mesmo processo de treinamento da rede mais completa. Os resultados são descritos na Tabela 10 e mostram que *Two-stage U-net* em sua versão mais leve continua com resultados superiores aos demais modelos. Entre as cinco melhores redes, a *Two-stage U-net Lite* teve o menor tempo de inferência de 128 ms e em média de *mIoU* ficou atrás apenas da versão completa.

Tabela 10 – 5 melhores - Comparação dos Resultados - mIoU (%)

Modelo	**Média**	**Não Plant.**	**Plant Verde**	**Solo Prep**	**Tempo**
Two-stage U-net	**82,61%**	**92,61%**	**76,68%**	78,17%	594 ms
Two-stage U-net Lite	80,21%	84,74%	74,47%	**81,41%**	**128 ms**
Unet+efficientnetb7	76,17%	77,30%	75,21%	76,01%	428 ms
Unet+efficientnetb6	75,62%	75,46%	73,94%	77,47%	229 ms
DeepLabv3+	74,93%	78,60%	69,62%	76,57%	267 ms

5.2.1 Limitações

O método proposto apresenta resultados superiores a outros modelos testados para o problema proposto, no entanto também apresenta algumas limitações. A principal delas diz respeito à qualidade das imagens de entrada na rede e melhoramentos referentes aos campos de sensoriamento remoto e geoprocessamento.

O foco deste trabalho foi a proposição de uma arquitetura para segmentação de áreas de plantações, com isso, no período de duração de um mestrado, alguns aspectos importantes não puderam ser tão bem desenvolvidos. No campo do sensoriamento remoto, foi escolhido o satélite Sentinel-2 por ser gratuito e dentre eles, ser aquele com menor tempo de atualização de imagens. No entanto, há outros satélites com diferentes tipos de sensores que captam diferentes características das plantações.

No campo do geoprocessamento, existem processamentos de melhoramento da qualidade das imagens de satélite com redução de ruídos e interferências atmosféricas, além da aplicação de índices de vegetação que vão além dos clássicos NDVI e EVI. Otimizações nos aspectos citados melhorariam a qualidade das imagens de entrada na rede e isso resultaria em impactos positivos em seu desempenho.

5.2.2 Análise do Método para Estimar a Produção de Plantio

A partir dos resultados obtidos, podemos observar que foi possível segmentar bem as regiões de interesse. A segmentação das áreas de plantio é uma das etapas de um processo maior que visa estimar a produção de estabelecimentos rurais. Para isso, se faz necessário transformar o resultado da segmentação em uma métrica de área como m^2, km^2 ou hectare. Também é necessário classificar o tipo de cultivo da região segmentada. Conhecendo o tipo de cultivo e sua relação de peso por unidade área, é possível estimar a produção com uma métrica de peso como quilograma ou tonelada.

Naturalmente, há outros aspectos do campo agricultura e agronomia que dizem respeito à sazonalidade das plantações, ao período ideal para plantio e colheita, ao calendário agrícola, aos subtipos de culturas e a outros aspectos que sugerem a necessidade de uma análise série-temporal das imagens de satélite. Além de um *pipeline* de dados com atualizações constantes para adquirir as imagens mais recentes para um monitoramento proativo das plantações. No entanto, mesmo com muito trabalho pela frente, a segmentação da área de plantio é a base para todo o restante do processo. Os resultados obtidos já podem ser aplicados para análises de algumas fazendas específicas. A Figura 31 apresenta algumas possibilidades nesse sentido.

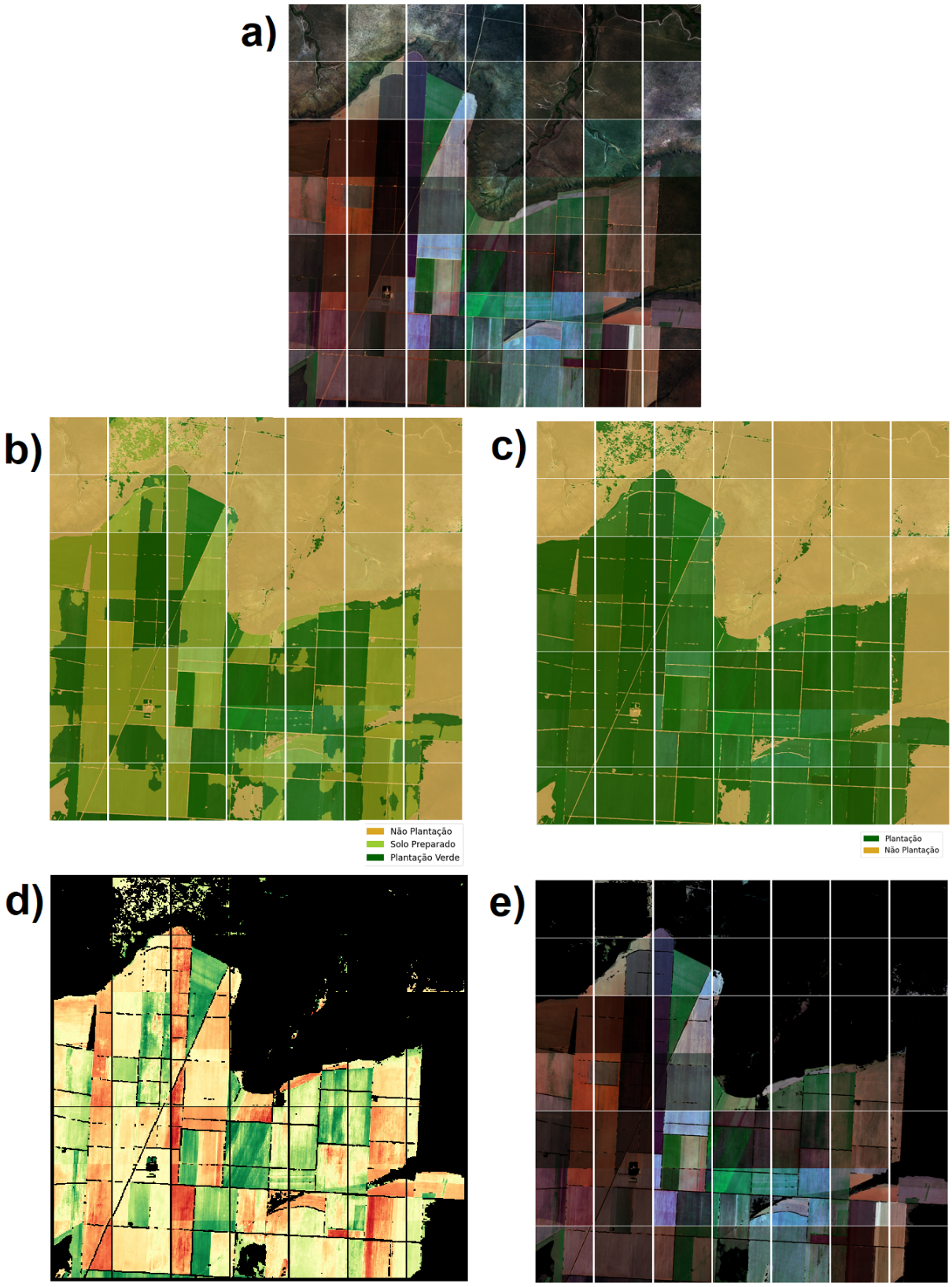


Figura 31 – a) Mosaico 7 x 7 de imagens do *dataset*. b) Segmentação Estágio 2 no modelo proposto. c) Segmentação Estágio 1. e) Isolamento da área de plantação. d) Isolamento da área de plantação com aplicação de NDVI.

6 Conclusão

Neste trabalho foi desenvolvida uma arquitetura de aprendizagem de máquina para segmentação de áreas de plantação, a Two-stage U-net. Além disso, o trabalho também incluiu a criação de conjunto de dados de imagens de satélite com anotações para segmentação de áreas de plantação. O modelo propostos foi treinado e seus hiperparâmetros foram ajustados considerando o Encoder da U-net, o Otimizador, a Função de Perda e o tamanho do Lote de imagens (*batch size*). Selecionamos o modelo ajustado que obteve o melhor desempenho nos testes com *Hyperopt* e *GridSearch*. Os resultados em $mIoU$ da Two-stage U-net se mostraram superiores aos resultados de outras arquiteturas utilizadas em trabalhos semelhantes.

A arquitetura Two-stage U-net foi desenvolvida para o problema específico deste trabalho e possui duas etapas de forma a reduzir a complexidade da segmentação de múltiplas classes. A Two-stage U-net retornou resultados de IoU médio acima de 80% em ambas as etapas. Houve uma melhoria significativa na segmentação da região de Não Plantação da Etapa 1, onde o resultado foi de 79,81%, para a Etapa 2, onde subiu para 92,61%, devido à remoção de ruídos de segmentação dessa classe na Etapa 2. Ao ser comparado com os resultados de outras arquiteturas como a *DeepLabV3+*, a *SegNet* e outras redes baseadas em U-net com estágio único, a arquitetura do método proposto foi melhor em todas as métricas avaliadas. Também foi desenvolvida uma versão mais enxuta da arquitetura proposta, Two-stage U-net Lite com *encoders* com menos parâmetros e mais rápidos. A Two-stage U-net Lite continuou apresentando resultados melhores que as demais arquiteturas testadas e com tempo de inferência quatro vezes menor que o da arquitetura completa e tempo similar ao das redes mais rápidas testadas.

Como principais contribuições alcançadas estão (1) a construção de um conjunto de dados público disponível do *GitHub* com 300 imagens de satélite com marcação de oito classes identificadas na cobertura terrestre de regiões de plantações que pode auxiliar outros pesquisadores em estudos e aplicações similares. (2) O desenvolvimento de uma arquitetura que apresentou, no problema abordado, resultados em $mIoU$ superiores a outras arquiteturas elencadas no Capítulo 3 - Trabalhos Relacionados. (3) A criação de um método para a segmentação de áreas de plantação que pode ser usado como uma das etapas iniciais de um processo automatizado de fiscalização de estabelecimentos rurais pelas Administrações Tributárias. A segmentação da área, pode ser usada para estimar a área em uma métrica de área como m^2, km^2 ou hectare que por sua vez pode ser usada para estimar a produção das safras em métricas de peso como quilograma ou tonelada. Processo semelhante pode ser aplicado por estabelecimentos rurais para estimar sua produção e realizar análises de perda de produtividade.

6.1 Trabalhos Futuros

A metodologia proposta certamente pode ser melhorada e para alcançar a estimativa de produção das safras, devem ser incluídas novas etapas. Assim, seguem sugestões para trabalhos futuros:

- Aprimorar o processo de aquisição da imagens avaliando a utilização de imagens de outros satélites.
- Realizar processamentos de melhoramento da qualidade das imagens e de redução de ruídos e interferências atmosféricas.
- Testar e avaliar arquiteturas mais recentes e que vierem a ser publicadas para a segmentação de imagens.
- Avaliar ou desenvolver modelos para classificação do tipo de cultura plantada.
- Ampliar o número de imagens do conjunto de dados, propor e testar outras possibilidades de pré-processamento e avaliar a necessidade de usar imagens série-temporais.

6.2 Produções Científicas

A Tabela 11 lista os artigos científicos publicados que possuem relação com o método proposto pelo presente trabalho.

Tabela 11 – Artigos publicados que possuem relação com o método proposto.

Título do Artigo	**Local de Publicação**	**Qualis**	**Ano**
A Two-stage U-net to estimate the Cultivated Area of Plantations	21st International Conference on Image Analysis and Processing (ICIAP)	A3	2022
Semantic segmentation of the cultivated area of plantations with U-net	Communications in Computer and Information Science (CCIS)	B2	2022

Referências

ABDANI, S. R.; ZULKIFLEY, M. A.; MAMAT, M. U-net with spatial pyramid pooling module for segmenting oil palm plantations. In: IEEE. *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*. [S.l.], 2020. p. 1–5. Citado na página 44.

ABDULLAHI, H. S.; SHERIFF, R.; MAHIEDDINE, F. Convolution neural network in precision agriculture for plant image recognition and classification. In: IEEE. *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*. [S.l.], 2017. v. 10. Citado na página 41.

AHMAD, J.; FARMAN, H.; JAN, Z. Deep learning methods and applications. In: *Deep learning: convergence to big data analytics*. [S.l.]: Springer, 2019. p. 31–42. Citado na página 27.

ALEMOHAMMAD, H.; BOOTH, K. *LandCoverNet: A global benchmark land cover classification training dataset*. 2020. Citado na página 48.

ANAND, T.; SINHA, S.; MANDAL, M.; CHAMOLA, V.; YU, F. R. Agrisegnet: Deep aerial semantic segmentation framework for iot-assisted precision agriculture. *IEEE Sensors Journal*, IEEE, 2021. Citado 2 vezes nas páginas 42 e 44.

ANTUNES, J. F.; LAMPARELLI, R. A.; RODRIGUES, L. H. Assessing of the sugarcane cultivation dynamics in são paulo state by modis data temporal profiles. *Engenharia Agrícola*, SciELO Brasil, v. 35, p. 1127–1136, 2015. Citado na página 21.

AYHAN, B.; KWAN, C. Tree, shrub, and grass classification using only rgb images. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 8, p. 1333, 2020. Citado 3 vezes nas páginas 40, 43 e 44.

AYHAN, B.; KWAN, C. Tree, shrub, and grass classification using only rgb images. *Remote Sensing*, v. 12, n. 8, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/8/1333>. Citado na página 48.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 12, p. 2481–2495, 2017. Citado na página 43.

BÉGUÉ, A.; LEBOURGEOIS, V.; BAPPEL, E.; TODOROFF, P.; PELLEGRINO, A.; BAILLARIN, F.; SIEGMUND, B. Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using ndvi. *International Journal of Remote Sensing*, Taylor & Francis, v. 31, n. 20, p. 5391–5407, 2010. Citado na página 23.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013. Citado na página 27.

BRUGNARO, R.; FILHO, E. D. B.; BACHA, C. J. C. Avaliação da sonegação de impostos na agropecuária brasileira. *Agric. São Paulo. SP*, n. 50, p. 15–27, 2003. Citado na página 16.

BURGOS-ARTIZZU, X. P.; RIBEIRO, A.; TELLAECHE, A.; PAJARES, G.; FERNÁNDEZ-QUINTANILLA, C. Analysis of natural images processing for the extraction of agricultural elements. *Image and Vision Computing*, Elsevier, v. 28, n. 1, p. 138–149, 2010. Citado na página 41.

BÜTTNER, G.; FERANEC, J.; JAFFRAIN, G.; MARI, L.; MAUCHA, G.; SOUKUP, T. The corine land cover 2000 project. *EARSeL eProceedings*, EARSeL Paris, v. 3, n. 3, p. 331–346, 2004. Citado na página 40.

CASTLEMAN, K. R. *Digital image processing*. [S.l.]: Prentice Hall Press, 1996. Citado na página 19.

CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 4, p. 834–848, 2017. Citado na página 42.

CHIU, M. T.; XU, X.; WANG, K.; HOBBS, J.; HOVAKIMYAN, N.; HUANG, T. S.; SHI, H. et al. The 1st agriculture-vision challenge: Methods and results. *arXiv preprint arXiv:2004.09754*, 2020. Citado 2 vezes nas páginas 42 e 44.

CHIU, M. T.; XU, X.; WEI, Y.; HUANG, Z.; SCHWING, A. G.; BRUNNER, R.; KHACHATRIAN, H.; KARAPETYAN, H.; DOZIER, I.; ROSE, G. et al. Agriculture-vision: A large aerial image database for agricultural pattern analysis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2020. p. 2828–2838. Citado 2 vezes nas páginas 40 e 42.

DEMIR, I.; KOPERSKI, K.; LINDENBAUM, D.; PANG, G.; HUANG, J.; BASU, S.; HUGHES, F.; TUIA, D.; RASKAR, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2018. Citado na página 48.

DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016. Citado na página 34.

ESA. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. SP-1322/2, 2012. Citado 2 vezes nas páginas 21 e 22.

FERREIRA, L. G.; FERREIRA, N. C.; FERREIRA, M. E. Sensoriamento remoto da vegetação: evolução e estado-da-arte. *Acta Scientiarum. Biological Sciences*, Universidade Estadual de Maringa, v. 30, n. 4, p. 379–390, 2008. Citado na página 22.

FLORENZANO, T. G. Imagens de satélite para estudos ambientais. In: *Imagens de satélite para estudos ambientais*. [S.l.: s.n.], 2002. p. 97–97. Citado na página 21.

FOERSTER, S.; KADEN, K.; FOERSTER, M.; ITZEROTT, S. Crop type mapping using spectral–temporal profiles and phenological information. *Computers and Electronics in Agriculture*, Elsevier, v. 89, p. 30–40, 2012. Citado na página 40.

GAO, B.; PAVEL, L. On the properties of the softmax function with application in game theory and reinforcement learning. *arXiv preprint arXiv:1704.00805*, 2017. Citado na página 38.

GARNOT, V. S. F.; LANDRIEU, L. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *ICCV*, 2021. Citado na página 40.

GARNOT, V. S. F.; LANDRIEU, L. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2021. p. 4872–4881. Citado 2 vezes nas páginas 42 e 44.

GÓMEZ, C.; WHITE, J. C.; WULDER, M. A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 116, p. 55–72, 2016. Citado na página 40.

GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Editora Blucher, 2000. Citado 2 vezes nas páginas 19 e 20.

GUIMARÃES, T. T. *Utilização de imagens de satélite para predição de clorofila-a e sólidos suspensos em corpos d\'água: estudo de caso da Represa do Lobo/SP*. Tese (Doutorado) — Universidade de São Paulo, 2019. Citado na página 20.

GUO, Y.; LIU, Y.; OERLEMANS, A.; LAO, S.; WU, S.; LEW, M. S. Deep learning for visual understanding: A review. *Neurocomputing*, Elsevier, v. 187, p. 27–48, 2016. Citado na página 28.

HAFEMANN, L. G. An analysis of deep neural networks for texture classification. 2014. Citado na página 28.

HAO, P.; WANG, L.; NIU, Z. Comparison of hybrid classifiers for crop classification using normalized difference vegetation index time series: A case study for major crops in north xinjiang, china. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 9, p. e0137748, 2015. Citado na página 40.

HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007. Citado 2 vezes nas páginas 25 e 26.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 42.

HE, N.; FANG, L.; PLAZA, A. Hybrid first and second order attention unet for building segmentation in remote sensing images. *Science China Information Sciences*, Springer, v. 63, n. 4, p. 1–12, 2020. Citado na página 42.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989. Citado na página 25.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, Elsevier, v. 83, n. 1-2, p. 195–213, 2002. Citado na página 23.

JENSEN, J. R.; EPIPHANIO, J. C. N. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. [S.l.]: Parêntese Editora São José dos Campos, 2009. Citado na página 20.

JR, J. R.; HAAS, R. H.; DEERING, D.; SCHELL, J.; HARLAN, J. C. *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. [S.l.], 1974. Citado na página 23.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, Elsevier, v. 147, p. 70–90, 2018. Citado na página 17.

KATTENBORN, T.; EICHEL, J.; FASSNACHT, F. E. Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery. *Scientific reports*, Nature Publishing Group, v. 9, n. 1, p. 1–9, 2019. Citado na página 44.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado na página 25.

LABACH, A.; SALEHINEJAD, H.; VALAEE, S. *Survey of Dropout Methods for Deep Neural Networks*. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1904.13310>. Citado na página 37.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado 4 vezes nas páginas 17, 28, 29 e 51.

LIN, M.; CHEN, Q.; YAN, S. *Network In Network*. arXiv, 2013. Disponível em: <https://arxiv.org/abs/1312.4400>. Citado na página 37.

LIU, C.; DU, S.; LU, H.; LI, D.; CAO, Z. Multispectral semantic land cover segmentation from aerial imagery with deep encoder-decoder network. *IEEE Geoscience and Remote Sensing Letters*, IEEE, 2020. Citado 2 vezes nas páginas 43 e 44.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Citado 2 vezes nas páginas 35 e 36.

MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. et al. Rectifier nonlinearities improve neural network acoustic models. In: CITESEER. *Proc. icml*. [S.l.], 2013. v. 30, n. 1, p. 3. Citado na página 25.

MAIA, F. C. d. O. Utilização de índices de vegetação para identificação de ambientes de produção de cana-de-açúcar. 2019. Citado 3 vezes nas páginas 20, 21 e 23.

MAIA, L. B. et al. Aprendizagem profunda aplicada ao diagnóstico de melanoma. Universidade Federal do Maranhão, 2019. Citado na página 29.

MILANO, D. de; HONORATO, L. B. *Visao computacional*. 2014. Citado na página 19.

MIRANDA, M. d. P. Imagens sentinel-2a (msi) aplicadas ao mapeamento geológico, região de itataia, santa quitéria, ce. 2019. Citado 2 vezes nas páginas 21 e 22.

MONDAL, P. Quantifying surface gradients with a 2-band enhanced vegetation index (evi2). *Ecological Indicators*, Elsevier, v. 11, n. 3, p. 918–924, 2011. Citado na página 24.

MUTANGA, O.; KUMAR, L. *Google earth engine applications*. [S.l.]: Multidisciplinary Digital Publishing Institute, 2019. Citado na página 46.

NOVO, E. M. de M. *Sensoriamento Remoto: princípios e aplicações*. [S.l.]: Editora Blucher, 2010. Citado na página 20.

NUNES, E.; CONCI, A. Segmentaçao por textura e localizaçao do contorno de regioes em imagens multibandas. *IEEE Latin America Transactions*, v. 5, n. 3, p. 185–192, 2007. Citado na página 41.

PENA, J.; TAN, Y.; BOONPOOK, W. Semantic segmentation based remote sensing data fusion on crops detection. *Journal of Computer and Communications*, Scientific Research Publishing, v. 7, n. 7, p. 53–64, 2019. Citado 2 vezes nas páginas 43 e 44.

RAKHLIN, A.; DAVYDOW, A.; NIKOLENKO, S. Land cover classification from satellite imagery with u-net and lovasz-softmax loss. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2018. Citado na página 17.

RIBEIRO, C. M. N. Classificação do uso e cobertura do solo do estado de goiás empregando redes neurais artificiais. 2019. Citado na página 46.

ROJAS, R. The backpropagation algorithm. In: _____. *Neural Networks: A Systematic Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996. p. 149–182. ISBN 978-3-642-61068-4. Disponível em: <https://doi.org/10.1007/978-3-642-61068-4_7>. Citado na página 26.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241. Citado 4 vezes nas páginas 30, 31, 32 e 42.

ROUSE, J.; HAAS, R.; SCHELL, J.; DEERING, D.; HARLAN, J. Monitoring the vernal advancement and retrogradation of natural vegetation [nasa/gsfct type ii report]. *Greenbelt, MD: NASA/Goddard Space Flight Center*, 1973. Citado na página 23.

RUSTOWICZ, R. M.; CHEONG, R.; WANG, L.; ERMON, S.; BURKE, M.; LOBELL, D. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 17 e 42.

SALDANHA, M. F.; FREITAS, C. Segmentação de imagens digitais: Uma revisão. *Divisão de Processamento de Imagens-Instituto Nacional de Pesquisas Espaciais (INPE), São Paulo*, 2009. Citado na página 29.

SANTOS, T. T.; BARBEDO, J. G. A.; TERNES, S.; NETO, J. C.; KOENIGKAN, L. V.; SOUZA, K. X. S. de. Visão computacional aplicada na agricultura. *Embrapa Agricultura Digital-Capítulo em livro científico (ALICE)*, In: MASSRUHÁ, SMFS; LEITE, MA de A.; OLIVEIRA, SR de M.; MEIRA, CAA ..., 2020. Citado na página 19.

SCHULTZ, B.; IMMITZER, M.; FORMAGGIO, A. R.; SANCHES, I. D.; LUIZ, A. J. B.; ATZBERGER, C. Self-guided segmentation and classification of multi-temporal landsat 8 images for crop type mapping in southeastern brazil. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 7, n. 11, p. 14482–14508, 2015. Citado na página 40.

SHENG, H.; CHEN, X.; SU, J.; RAJAGOPAL, R.; NG, A. Effective data fusion with generalized vegetation index: Evidence from land cover segmentation in agriculture. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2020. p. 60–61. Citado 2 vezes nas páginas 43 e 44.

SILVA, G. da; SILVA, A.; PAIVA, A. de; GATTASS, M. Classification of malignancy of lung nodules in ct images using convolutional neural network. In: SBC. *Anais do XVI Workshop de Informática Médica*. [S.l.], 2016. p. 2481–2489. Citado na página 26.

SILVA, I. F. S. d. et al. Detecção automática da presença de patologia na visão baseada em imagens do teste de brückner. Universidade Federal do Maranhão, 2019. Citado 2 vezes nas páginas 26 e 29.

SILVA, I. N. D.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas-curso prático. *São Paulo: Artliber*, 2010. Citado 3 vezes nas páginas 24, 26 e 27.

SILVA, M. da; CESARIO, A. V.; CAVALCANTI, I. R. Relevância do agronegócio para a economia brasileira atual. *Apresentado em X ENCONTRO DE INICIAÇÃO À DOCÊNCIA, UNIVERSIDADE FEDERAL DA PARAÍBA. Recuperado de http://www. prac. ufpb. br/anais/IXEnex/iniciacao/documentos/anais/8. TRABALHO/8C CSADAMT01. pdf*, 2013. Citado na página 16.

SPOTO, F.; SY, O.; LABERINTI, P.; MARTIMORT, P.; FERNANDEZ, V.; COLIN, O.; HOERSCH, B.; MEYGRET, A. Overview of sentinel-2. In: IEEE. *2012 IEEE international geoscience and remote sensing symposium*. [S.l.], 2012. p. 1707–1710. Citado 2 vezes nas páginas 21 e 22.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, JMLR.org, v. 15, n. 1, p. 1929–1958, jan 2014. ISSN 1532-4435. Citado 2 vezes nas páginas 37 e 38.

STOIAN, A.; POULAIN, V.; INGLADA, J.; POUGHON, V.; DERKSEN, D. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 11, n. 17, p. 1986, 2019. Citado na página 44.

SUMBUL, G.; CHARFUELAN, M.; DEMIR, B.; MARKL, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In: IEEE. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. [S.l.], 2019. p. 5901–5904. Citado na página 40.

TSAI, D.-M.; CHEN, W.-L. Coffee plantation area recognition in satellite images using fourier transform. *Computers and electronics in agriculture*, Elsevier, v. 135, p. 115–127, 2017. Citado na página 41.

ULMAS, P.; LIIV, I. Segmentation of satellite imagery using u-net models for land cover classification. *arXiv preprint arXiv:2003.02899*, 2020. Citado 2 vezes nas páginas 42 e 44.

USDA. *United States Departament of Agriculture - Soybean Production Brazil 2021*. 2022. Urlhttps://ipad.fas.usda.gov/cropexplorer/cropview/comm$_c$*hartview.aspx*?$fattributeid = 1cropid = 2222000sel_year = 2021startrow = 1ftypeid = 47regionid = brcntryid = BRAnationalGraph = Falsesubrgnid = br_bra016.Citadonapágina47$.

VARGA, Z.; CZÉDLI, H.; KÉZI, C.; LÓKI, J.; FEKETE, Á.; BÍRÓ, J. Evaluating the accuracy of orthophotos and satellite images in the context of road centrelines in test sites in hungary. *Research Journal of Applied Sciences*, v. 10, n. 10, p. 568–573, 2015. Citado na página 20.

VIBHA, L.; SHENOY, P. D.; VENUGOPAL, K.; PATNAIK, L. Robust technique for segmentation and counting of trees from remotely sensed data. In: IEEE. *2009 IEEE International Advance Computing Conference*. [S.l.], 2009. p. 1437–1442. Citado na página 41.

VOGT, K.; SCHEUERMANN, B.; BECKER, C.; BÜSCHENFELD, T.; ROSENHAHN, B.; OSTERMANN, J. *Automated extraction of plantations from ikonos satellite imagery using a level set based segmentation method*. [S.l.]: na, 2010. Citado na página 41.

WADA, K. *labelme: Image Polygonal Annotation with Python*. 2016. <https://github.com/wkentaro/labelme>. Citado na página 48.

WAGNER, F. H.; SANCHEZ, A.; TARABALKA, Y.; LOTTE, R. G.; FERREIRA, M. P.; AIDAR, M. P.; GLOOR, E.; PHILLIPS, O. L.; ARAGAO, L. E. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, Wiley Online Library, v. 5, n. 4, p. 360–375, 2019. Citado na página 44.

WANG, S.; CHEN, W.; XIE, S. M.; AZZARI, G.; LOBELL, D. B. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 2, p. 207, 2020. Citado na página 42.

WOEBBECKE, D. M.; MEYER, G. E.; BARGEN, K. V.; MORTENSEN, D. A. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, American Society of Agricultural and Biological Engineers, v. 38, n. 1, p. 259–269, 1995. Citado 2 vezes nas páginas 22 e 50.

XIAO, T.; LIU, Y.; ZHOU, B.; JIANG, Y.; SUN, J. Unified perceptual parsing for scene understanding. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 418–434. Citado na página 43.

YANG, L.; SHAMI, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, Elsevier, v. 415, p. 295–316, 2020. Citado na página 53.

YANG, M.-D.; TSENG, H.-H.; HSU, Y.-C.; TSAI, H. P. Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date uav visible images. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 4, p. 633, 2020. Citado na página 17.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021. Citado na página 33.

ZHOU, J.; PROISY, C.; DESCOMBES, X.; MAIRE, G. L.; NOUVELLON, Y.; STAPE, J.-L.; VIENNOIS, G.; ZERUBIA, J.; COUTERON, P. Mapping local density of young eucalyptus plantations by individual tree detection in high spatial resolution satellite images. *Forest Ecology and Management*, Elsevier, v. 301, p. 129–141, 2013. Citado na página 41.

ZHU, N.; LIU, X.; LIU, Z.; HU, K.; WANG, Y.; TAN, J.; HUANG, M.; ZHU, Q.; JI, X.; JIANG, Y. et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *International Journal of Agricultural and Biological Engineering*, v. 11, n. 4, p. 32–44, 2018. Citado na página 17.

ZHU, W.; ZENG, N.; WANG, N. et al. Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, v. 19, p. 67, 2010. Citado na página 54.