Uso de redes neurais convolucionais para a segmentação de mancha anelar do mamoeiro com uso de imagens por UAV

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

9 de outubro de 2023

Uso de redes neurais convolucionais para a segmentação de mancha anelar do mamoeiro com uso de imagens por UAV

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Jefferson Rodrigo de Souza
Coorientador: Prof. Dr. Bilzã Marques de Araújo

9 de outubro de 2023

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

M929 Moura, Matheus José de, 1996-
2023 Uso de redes neurais convolucionais para a segmentação de mancha anelar do mamoeiro com uso de imagens por UAV [recurso eletrônico] : Redes de segmentação na detecção da doença mancha anelar no mamoeiro / Matheus José de Moura. - 2023.

Orientador: Jefferson Souza.
Coorientador: Bilzã Araújo.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.di.2023.550
Inclui bibliografia.
Inclui ilustrações.

1. Computação. I. Souza, Jefferson,1985-, (Orient.). II. Araújo, Bilzã,1986-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada "**Aprendizado de máquina: Uso de Redes neurais convolucionais para identificar a mancha anelar do mamoeiro com uso de imagens obtidas por UAV**" por **Matheus José de Moura** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação.**

Uberlândia, ___ de _____ de _____

Orientador: _____
Prof. Dr. Jefferson Rodrigo de Souza
Universidade Federal de Uberlândia

Coorientador: _____
Prof. Dr. Bilzã Marques de Araújo
Universidade Federal do Sul da Bahia

Banca Examinadora:

_____
Prof. Dr. Marcelo Zanchetta do Nascimento
Universidade Federal de Uberlândia

_____
Prof. Dr. Henrique Candido de Oliveira
Universidade Estadual de Campinas

Forget your lust for the rich man's gold,
all that you need is in your soul.

Agradecimentos

As conquistas sempre vem acompanhadas de pessoas que com pequenas palavras ou atitudes, ajudando no trajeto no qual estamos caminhando, e nesta jornada não posso deixar de agradecer.

Aqui agradeço primeiramente a Deus, pela oportunidade da vida e saúde, a minha esposa Regina que me acompanhando durante toda minha caminhada, apoiando, incentivando e sempre agarrando em minha mão em todas minha decisões, aos meu pais Raquel e Orlando deixo a minha eterna gratidão, pois nunca mediram esforços para que eu buscasse meus objetivos e sonhos.

Aos amigos do Centro de Tecnologia da Informação e Comunicação (CTIC) (Centro de Tecnologia da Informação e Comunicação) da Universidade Federal de Uberlândia (UFU), obrigado por me ajudarem desde o começo, este trabalho tem muito a contribuição de todos, ao Professor Doutor Bilzã Marques pela contribuição no meu aprendizado, a EMBRAPA pelo apoio na pequisa e ao Professor Doutor Jefferson Souza pela colaboração na orientação do meu mestrado.

Por fim agradeço minha Vó Márcia que sempre foi minha motivadora, me motivando a começar este sonho, este trabalho dedico inteiramente a você que não esta mais aqui, para ver eu conquistar meu objetivo.

"A emoção mais antiga e mais forte da humanidade é o medo, e o mais antigo e mais forte de todos os medos é o medo do desconhecido."

(H.P. Lovecraft)

Resumo

A agricultura de precisão é uma abordagem baseada em tecnologia que utiliza e analisa os dados para aperfeiçoar as práticas agrícolas. Ela permite aos agricultores tomarem as decisões mais informadas, maximizar a eficiência dos recursos e promover uma produção agrícola mais sustentável. Este trabalho visa integrar a agricultura de precisão com as técnicas de Aprendizado de Máquina (AM) para a detecção da doença mancha anelar em plantações de mamoeiro, a partir de um pequeno campo amostral, a fim de analisar o comportamento das redes de segmentação. As imagens foram capturadas usando um veículo aéreo não tripulado em plantios de mamão. As técnicas de AM utilizadas foram as redes neurais de segmentação: UNET, PSPNET, LINKNET e FCN combinadas com a arquitetura profunda VGG16, para o processo de treinamento, validação e teste, usou-se do dataset formado pelas imagens capturadas. As redes treinadas foram validades sobre os resultados de acurácia e F1-Score, no qual obtiveram resultados acima de 79%, porém esses indicativos não são afirmativos para uma validação conclusiva para a detecção da doença, ocorrendo variações de falsos positivos no resultado das imagens.

Palavras-chave: Aprendizado de máquina, agricultura, mamoeiro, mancha anelar.

Abstract

Precision agriculture is a technology-based approach that uses and analyzes data to improve agricultural practices. It enables farmers to make the most informed decisions, maximize resource efficiency and promote more sustainable agricultural production. This work aims to integrate precision agriculture with Machine Learning (ML) techniques for the detection of the ringspot disease in papaya plantations, from a small sampling field of papaya cultivation in the interior of Bahia, in order to analyze the behavior of the segmentation networks regarding the disease. The images were captured using an unmanned aerial vehicle on papaya plantations. The BF techniques used were segmentation neural networks: UNET, PSPNET, LINKNET and FCN combined with the VGG16 deep architecture, for the training, validation and testing process, the dataset formed by the captured images was used. The trained networks were validities on the results of accuracy and F1-Score, in which they obtained results above 79%, but these indications are not affirmative for a conclusive validation for the detection of the disease, with variations of false positives in the result of the images.

Keywords: Machine learning, agriculture, papaya, ring spot.

Lista de ilustrações

Lista de tabelas

Lista de siglas

AM Aprendizado de Máquina

AE Autoencoder

CTIC Centro de Tecnologia da Informação e Comunicação

CNN Convolutional Neural network

CAE Convolutional Autoencoders

DL Deep Learning

ESA Agência Espacial Europeia

FP False Positive

FCNs Fully Convolutional Networks - FCNs

GT Ground Truth

GPU Graphics Processing Unit

IoU Intersection over Union

IA Inteligência Artificial

JSON JavaScript Object Notation

MLP Multi Layer Perceptron

RGB Red, Green, Blue

RNA Redes Neurais Artificiais

SGD Stochastic gradient descent

TP True Positive

UFU Universidade Federal de Uberlândia

VANTs Veículos aéreos não tripulados

Sumário

Capítulo 1

Introdução

A produção de mamão desempenha um papel significativo no setor agrícola do Brasil, contribuindo para o consumo interno e internacional (GALEANO; MARTINS, 2015). Contudo, a iminência de pragas e doença, afeta o plantio e acarreta em perdas na produção e venda. Este capítulo apresenta o contexto deste estudo, mostrando as possibilidades de contribuição do desenvolvimento de recursos tecnológicos com uso de sensoriamento remoto por Veículos aéreos não tripulados (VANTs) (denominados UAV em inglês, *Unmanned Aerial Vehicle*), fundamentações teóricas e técnicas de aprendizado de máquina, que auxilia na detecção da doença mancha anelar no mamoeiro.

1.1 Contextualização

O mamoeiro é uma árvore que produz mamão, uma fruta tropical. O sul da Bahia, no Brasil, é conhecido por ser uma região propícia para o cultivo de diversas frutas tropicais, incluindo o mamoeiro, respondendo por 30% da produção nacional (BRIDI et al., 2023). As condições climáticas, como temperaturas mais altas e umidade adequada, proporcionam um ambiente favorável para o crescimento e o desenvolvimento do mamão.

A mancha anelar do mamoeiro ou conhecido como mosaico, causada pelo Papaya ringspot vírus é um dos principais problemas fitossanitários desta frutífera (OLIVEIRA; FILHO, 2022), sendo responsável pelo caráter itinerante da cultura. Essa doença pode causar sérios danos às plantas de mamão e também afetar a produção de frutas, a qual é considerada uma das mais destrutivas doenças do mamoeiro, sendo um dos fatores limitantes ao desenvolvimento da cultura (VENTURA; FERREGUETTI; MARTINS, 2015).

O vírus apresenta velocidade de disseminação muito rápida, a partir do primeiro foco da doença, podendo as plantas do pomar ser infectadas após um período de 3 a 7 meses. Os sintomas da mancha-anelar incluem o aparecimento de manchas de coloração mais clara ou amarelada nas folhas, que podem ter um formato de anel, como sugere o nome da doença. Essas manchas podem se apresentar de maneira anelar e formatos irregulares.

Conforme a infecção avança, as folhas infectadas podem ficar deformadas, murchas e desenvolver bolhas. A planta pode apresentar um crescimento lento e, em casos graves, pode levar à morte da planta. Uma maneira da não propagação da doença é a poda, identificando o vírus na planta, realizando a poda da árvore, reduz o avanço da doença.

A Figura 1 ilustra mamoeiro que mostra os sintomas da mancha-anelar, como o amarelecimento das folhas mais novas, a não formação da planta e enrugamento nas folhas.

Figura 1 – Mamoeiro com sintomas mancha-anelar (OLIVEIRA; FILHO; FILHO, 2011)

A Figura 2 apresenta uma lavoura de mamoeiro, a qual a doença mancha-anelar não é presente. É possível notar que a planta tem sua folhagem bem desenvolvida sem amarelecimento, sua copa mais desenvolvida em relação ao mamoeiro infectado.

Figura 2 – Plantio de mamoeiro sem pragas ou doença.

Atualmente, nas regiões produtoras, o controle da doença se faz por meio da identificação visual dos sintomas, por um operário rural que percorre toda a área, com posterior eliminação das plantas com sintomas da mancha anelar. No entanto, as plantas infectadas permanecem no campo sem sintomas visíveis, aumentando a disseminação do vírus.

(SANTOS; ESPERIDIÃO; AMARANTE, 2019) apresentma o avanço tecnológico no setor agrícola brasileiro e aponta o uso de UAV para a obtenção de imagens áreas do dossel de mamoeiros e o desenvolvimento de metodologia de identificação precoce de plantas infectadas com a virose por Aprendizado de Máquina (AM), reduzindo a taxa de disseminação da doença. Essa estratégia de agricultura auxilia a gestão de fruticultores e defesa vegetal dos Estados.

1.2 Motivação e Objetivos

Este projeto colabora no processo de melhorias em técnicas de detecção e combate ao vírus mancha anelar presente no mamoeiro, buscando auxiliar na precisão da identificação da doença de forma dinâmica, comparando com os recursos e métodos já utilizados atualmente, que são técnicas manuais em solo, realizado por agrônomos.

O objetivo geral é usar imagens aéreas de um VANT para a detecção automática de doenças de mamoeiro com técnicas de AM. Objetivos Específicos:

- ❏ Criar uma base de dados a partir das imagens capturadas para a detecção do vírus;
- ❏ Estudar e escolher o melhor modelo de redes de segmentação de imagens da pesquisa;
- ❏ Rotulação das imagens com agrônomos, treinamento dos modelos escolhidos e avaliação dos resultados obtidos; a fim de analisar a viabilidade da detecção;
- ❏ Escolha das métricas de avaliação da pesquisa;
- ❏ Analisar os resultados obtidos para melhoria na aplicação das redes escolhidas para a detecção da doença.

1.3 Fundamentação Teórica

As estratégias usando modelos e algoritmos de AM têm sido usadas em várias aplicações de imageamento agrícola. (SINGH et al., 2018) apresentam uma revisão das principais perspectivas no uso de *deep learning* na fenotipagem de estresse vegetal. Os resultados são satisfatórios com o uso de Redes Neurais Artificiais, Máquinas de Vetor de Suporte e Modelos Gaussianos com pré-processamento e seleção de características das imagens. A partir das imagens capturadas por VANTs, posteriormente rotuladas e sem a necessidade de extração de características, as RNC têm sido usadas para a segmentação

semântica de plantas e identificação de ocorrências de doenças. Características são extraídas e processadas por camadas não supervisionadas da rede e treinadas com transferência de aprendizagem. As últimas camadas são treinadas e foram obtidos bons resultados da identificação de estresse vegetal.

O trabalho apresentado por (RIBEIRO et al., 2019), tem como objetivo desenvolver um método de identificação e reconhecimento das nervuras presentes nas folhas de plantas utilizando técnicas de Redes Neurais Convolucionais (Convolutional Neural network (CNN)). As nervuras das folhas são estruturas importantes e únicas que desempenham um papel crucial na saúde e no desenvolvimento das plantas. Através do reconhecimento preciso das nervuras, é possível obter informações valiosas sobre a saúde da planta, seu crescimento e possíveis deficiências nutricionais ou problemas de saúde. O estudo utiliza Redes Neurais Convolucionais devido à sua eficiência em lidar com dados de imagem. As CNNs são capazes de extrair características relevantes das imagens de folhas, identificando padrões específicos que correspondem às nervuras, mesmo em casos de variações nas formas e tamanhos das folhas.

O processo de desenvolvimento do método inclui a construção de um conjunto de dados contendo imagens de folhas com suas respectivas nervuras previamente anotadas. Esse conjunto de dados é usado para treinar a rede neural, permitindo que ela aprenda a identificar as nervuras a partir de novas imagens não vistas anteriormente. Os resultados do trabalho demonstram que as Redes Neurais Convolucionais são altamente eficazes na tarefa de reconhecimento das nervuras de folhas em plantas. A precisão alcançada pelo método permite o uso dessa abordagem em aplicações práticas, como monitoramento da saúde das plantas em cultivos agrícolas, identificação de espécies vegetais e detecção precoce de doenças ou estresses que afetam as plantas. Em suma, o estudo representa um avanço significativo no campo da visão computacional aplicada à botânica, contribuindo para a automatização e melhoria da análise de folhas em estudos agrícolas e ambientais. A utilização de Redes Neurais Convolucionais abre portas para o desenvolvimento de tecnologias mais sofisticadas e precisas no monitoramento e manejo de plantas, beneficiando a agricultura e a pesquisa botânica.

1.4 Hipótese

A pesquisa tem como objetivo analisar uma pequena área amostral de plantação de mamoeiro, no qual existe a doença mancha anelar, e a partir de uso de VANTs e AM, verificar a possibilidade do uso de redes de segmentação para a detecção da doença e avaliar sua eficácia.

1.5 Organização da Dissertação

Este trabalho está estruturado da seguinte forma:

- ❏ Introdução
- ❏ Fundamentação Teórica
- ❏ Aprendizagem de Máquina
- ❏ Resultados
- ❏ Conclusão

1.6 Conclusão

O projeto visa ajudar os pequenos produtores de mamoeiros, no qual podem sofrer com a perda do seu plantio devido a doença se propagar rapidamente, diminuir os custos humanos para detecção da doença, uma vez que é feita de forma manual. Contribui também para o meio-ambiente, pois não requer uso de agrotóxicos para combate da doença e uso de maquinários de combustíveis poluentes.

Capítulo 2

Estado da Arte

Este capítulo apresenta o estado da arte, apresentando as definições sobre segmentação, e os trabalhos acerca das técnicas de segmentação.

2.1 Segmentação

A segmentação de imagem é um processo fundamental na área de processamento de imagens, que envolve a divisão de uma imagem em regiões ou segmentos significativos com base em determinadas características, como cor, intensidade de pixel, textura, forma ou qualquer outra propriedade visual. O objetivo principal da segmentação de imagem é simplificar a representação de uma imagem, tornando-a mais fácil de analisar, interpretar e extrair informações.

(CHENG et al., 2001) aponta várias técnicas e métodos para realizar a segmentação de imagem, e a escolha depende do contexto e dos objetivos específicos da aplicação. A seguir apresenta-se alguns dos métodos comuns de segmentação de imagem:

- Segmentação baseada em intensidade: Essa técnica considera os níveis de intensidade de pixel na imagem para dividir as regiões. Por exemplo, um método simples seria definir um limiar de intensidade e classificar os pixels como pertencentes a uma região ou outra com base nesse limiar.

- Segmentação por região: Nesse método, a imagem é dividida em regiões que compartilham características semelhantes, como cor, textura ou forma. Isso pode envolver técnicas de agrupamento, como segmentação por crescimento de região ou segmentação por k-means.

- Segmentação em cluster: É uma abordagem avançada para a segmentação de imagens que utiliza algoritmos de agrupamento para dividir a imagem em regiões ou clusters com base em características similares, como cor, textura ou intensidade de pixel.

- ❏ Segmentação de limite: Nela dividi-se uma imagem em regiões distintas com base na diferença de intensidade de pixel. Ela envolve a escolha de um valor de limite e a classificação dos pixels como parte do objeto de interesse ou do fundo, dependendo se sua intensidade é maior ou menor que o limite estabelecido. Embora seja sensível a variações de iluminação e ruído, a segmentação de limite é amplamente utilizada em tarefas como reconhecimento de caracteres, detecção de objetos e análise de imagens biomédicas, oferecendo uma abordagem rápida e intuitiva para a separação de regiões em imagens digitais.

- ❏ Segmentação de borda: Este método enfoca a detecção de bordas ou contornos na imagem. Geralmente, ele identifica transições abruptas na intensidade de pixel e cria uma imagem binária que destaca as bordas.

- ❏ Segmentação por detecção de objetos: Em algumas aplicações, o objetivo é identificar e segmentar objetos específicos na imagem. Isso pode envolver o uso de técnicas de detecção de características específicas de objetos.

- ❏ Segmentação semântica: Esse tipo de segmentação procura atribuir um rótulo semântico a cada pixel na imagem, identificando objetos ou regiões correspondentes a categorias específicas, como carros, árvores, pessoas, etc. Isso é comumente usado em visão computacional e aprendizado profundo.

A segmentação de imagem é uma etapa crucial em muitas aplicações de processamento de imagens, como reconhecimento de objetos, rastreamento de objetos, diagnóstico médico por imagem, visão computacional, análise de satélite, entre outros. Ela desempenha um papel importante na extração de informações relevantes de imagens e na tomada de decisões com base nessas informações.

2.2 Segmentação Panótica, Instância e Semântica.

A segmentação panótica, de instância e semântica são três abordagens cruciais em tarefas de processamento de imagens, (CARVALHO, 2022). A segmentação panótica visa oferecer uma visão completa de uma cena, combinando a segmentação semântica, que rotula cada pixel com uma classe específica (como carro, árvore), com a segmentação de instância, que distingue objetos individuais da mesma classe. Isso proporciona uma compreensão detalhada de objetos únicos em uma imagem. A segmentação semântica, por sua vez, atribui rótulos de classe a cada pixel, facilitando a análise de categorias gerais de objetos. Já a segmentação de instância identifica e diferencia objetos individuais da mesma classe. Essas técnicas desempenham papéis essenciais em aplicações como veículos autônomos, visão computacional e análise de imagens médicas, contribuindo para uma compreensão mais profunda e precisa de cenas visuais complexas.

2.3 Trabalhos correlatos

Nesta seção buscou-se, descrever trabalhos nos quais abordam técnicas de segmentação.

O trabalho desenvolvido por (CRISPI, 2022), buscou desenvolver um modelo de inteligência artificial capaz de detectar e estimar automaticamente a gravidade dos sintomas causados pelo ataque da mosca minadora em folhas de tomateiro. O conjunto de dados utilizado no estudo contou com 1932 imagens capturadas em condições de campo, com as folhas apresentando o sintoma da praga. Essas imagens foram manualmente anotadas com três classes: plano de fundo, folha do tomateiro e sintoma foliar da mosca minadora.

Foram realizadas comparações entre três arquiteturas e quatro espinhas dorsais (backbone) para a tarefa de segmentação semântica multiclasse, utilizando métricas como acurácia, precisão, revogação e Intersection over Union (IoU) (Intersect over Union). O modelo U-Net com a espinha dorsal Inceptionv3 alcançou o melhor resultado de IoU médio, atingindo 77,71%, seguido pelo modelo FPN com a espinha dorsal DenseNet121, com um IoU médio de 76,62%. Quando analisada separadamente a classe do sintoma em estudo, o modelo FPN com a espinha dorsal DenseNet121 obteve um resultado de 61,02%, seguido pelo modelo LinkNet, também com a espinha dorsal DenseNet121, com um resultado de IoU de 60,99%. Para a estimativa da gravidade do sintoma, o modelo FPN mostrou-se mais eficiente em relação aos demais. As espinhas dorsais ResNet34 e DenseNet121 apresentaram os menores valores de RSME (Raiz Quadrada do Erro Médio), confirmando os valores de IoU encontrados para esses modelos.

Os resultados dos experimentos computacionais realizados neste estudo mostraram-se promissores, especialmente em relação à capacidade dos modelos em segmentar automaticamente objetos pequenos em imagens com condições desafiadoras de iluminação e fundo complexo, mesmo com o uso de um banco de dados com desbalanceamento das classes.

(SANTOS, 2019), apresentou um estudo que tem como objetivo utilizar imagens do espectro visual (Red, Green, Blue (RGB)) coletadas por um Veículo Aéreo Não Tripulado (UAV) para a detecção autônoma de invasões biológicas no Cerrado, adotando técnicas de Aprendizado Profundo (Deep Learning). Para obter as imagens, foi escolhido um UAV (Quadricóptero) equipado com um sensor RGB, por sua maior acessibilidade e reprodutibilidade dos resultados. As redes Convolutional AutoEncoder (CAE) e U-Net foram adotadas por serem amplamente usadas em conjuntos de dados com poucas amostras, devido à sua capacidade de generalização, apesar de possuírem poucos exemplos para treinamento.

(SANTOS, 2019) criou um Conjunto de Dados original a partir da área de estudo, utilizando delineação manual, e posteriormente essa base foi ampliada com a técnica de Data Augmentation. Ao analisar o banco de dados inalterado, a rede Convolutional AutoEncoder superou a rede U-Net, com uma pontuação de 88% de F-score contra 84%. Com o segundo Conjunto de Dados com Data Augmentation, os resultados foram ainda melho-

res, com um F-score de 93% para o Convolutional Autoencoders (CAE), em comparação com 84% da U-Net, e maior precisão em ambos os cenários (85,4% CAE e 82% U-Net para o Conjunto de Dados original, e 93% CAE e 84% com Data Augmentation). Essas diferenças são relevantes devido à necessidade de precisão nos resultados para direcionar corretamente as equipes em suas tarefas de busca por invasões biológicas em todo o vasto território do Cerrado.

A Figura 3 foi retirada do trabalho de (SANTOS, 2019), a fim de apresentar o resultado da predição da rede U-net por ele treinada.

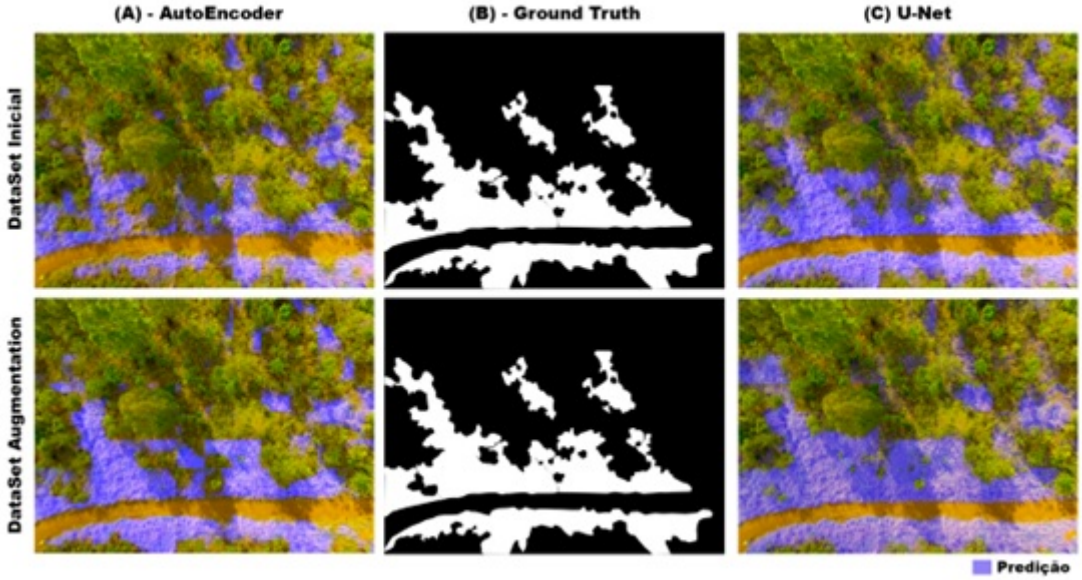


Figura 3 – Retirado do trabalho de (SANTOS, 2019)

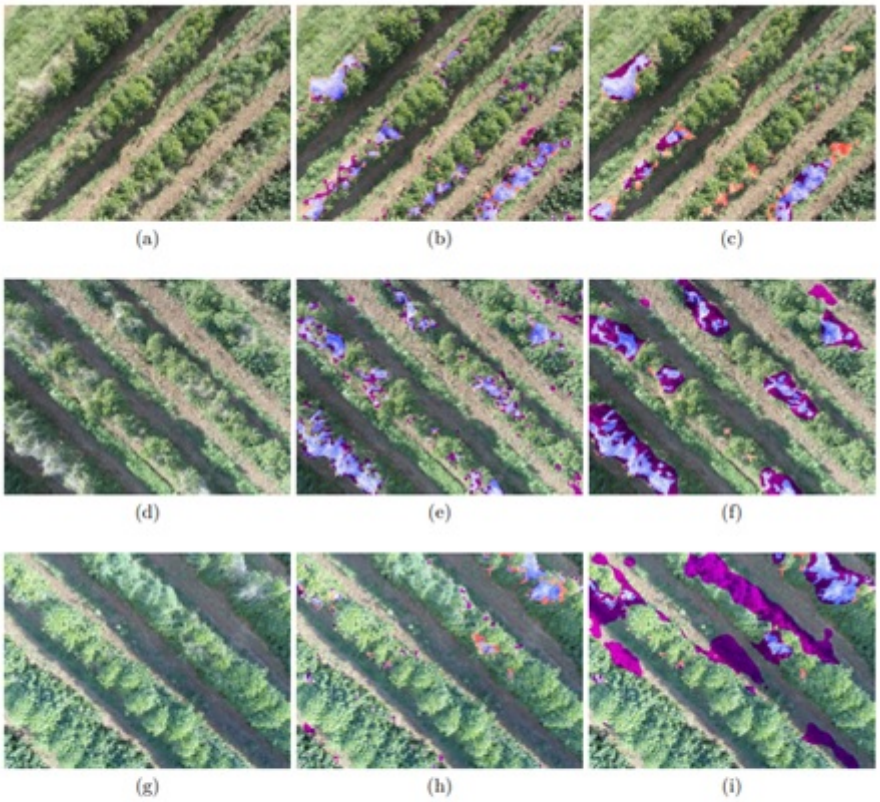
Além disso, enfatiza-se as características do CAE, considerando seu tamanho menor, com um número reduzido de camadas e neurônios, e melhores métricas para essa aplicação. Dessa forma, foi possível constatar que o modelo preditivo gerado pela rede AutoEncoder pode ser usado de forma eficiente, com grande potencial para outras bases de dados. Por fim, conclui-se que este trabalho representa um avanço no AM e em sua capacidade de auxiliar a vida cotidiana, expandindo as possibilidades de trabalhos futuros.

(MASOUD; PERSELLO; TOLPEKIN, 2019) apresentam uma pesquisa que aborda a detecção e delineação das fronteiras dos campos agrícolas em imagens capturadas pelo satélite Sentinel-2. O Sentinel-2 é uma missão espacial da Agência Espacial Europeia (Agência Espacial Europeia (ESA)) que fornece imagens de alta resolução e multiespectrais da superfície terrestre. Essas imagens são amplamente utilizadas para monitorar e gerenciar áreas agrícolas, proporcionando informações valiosas sobre o uso da terra, saúde das plantas e variações temporais nas plantações.

O trabalho em questão se concentra no desenvolvimento de um novo método para detectar e delinear as bordas dos campos agrícolas com alta precisão usando redes neurais convolutivas totalmente convolucionais (Fully Convolutional Networks - FCNs). A partir

de uma banco de dados de imagens definido, o autor define parâmetros distintos para treinamento das redes Fully Convolutional Networks - FCNs (FCNs), de modo que possa alcançar e comparar o melhor resultado obtido. Ao final do trabalho comparando os modelos treinados, o MD-FCN obteve um desempenho significativamente melhor do que os métodos de linha de base com um incremento de 31% a 32% no F-score médio, isso ocorreu porque o MD-FCN consegue aprender automaticamente recursos espaciais-contextuais discriminativos das imagens de entrada e gerar com precisão as saídas necessárias.

(OLIVEIRA et al., 2019) apresenta um método inovador e automatizado para detecção de nematóides em lavouras cafeeiras, utilizando imagens aéreas. O sistema proposto baseia-se no processamento avançado de imagens de alta resolução capturadas por drones ou outras plataformas aéreas, combinado com técnicas de aprendizado de máquina. Inicialmente, são coletadas imagens aéreas da lavoura usando VANTS, abrangendo diferentes estágios de desenvolvimento das plantas. Essas imagens são submetidas a um processo de pré-processamento, que inclui correções de iluminação, remoção de ruídos e segmentação da área de interesse. Em seguida, os nematóides são identificados e destacados por meio de técnicas de visão computacional, que permitem uma análise detalhada de suas características morfológicas. O sistema é feito sobre uma aplicação de algoritmos de aprendizado de máquina, que são treinados com um grande conjunto de imagens rotuladas manualmente. Os algoritmos utilizados PSPNet e U-Net são capazes de identificar padrões sutis e características distintivas dos nematóides, permitindo a classificação automática e a detecção de infestações mesmo em estágios iniciais. Os resultados experimentais demonstraram que o método proposto apresenta alta precisão na detecção de nematóides, superando as abordagens tradicionais de inspeção manual, que são demoradas e suscetíveis a erros humanos. Além disso, a utilização de imagens aéreas possibilita uma cobertura mais ampla das lavouras, tornando o processo de monitoramento mais eficiente e abrangente. A Figura 4, apresenta os resultados de predição obtidos nos treinos das redes U-Net e PSP-Net utilizados por (OLIVEIRA et al., 2019).



Resultado da predição para as imagens obtidas pela comparação da U-Net e PSPNet. (a) Imagem 1. (b) Resultado imagem 1 para U-Net. (c) Resultado imagem 1 para PSPNet. (d) Imagem 2. (e) Resultado imagem 2 para U-Net. (f) Resultado imagem 2 para PSPNet. (g) Imagem 3. (h) Resultado imagem 3 para U-Net e (i) Resultado imagem 3 para PSPNet

Figura 4 – Retirado do trabalho do autor (OLIVEIRA et al., 2019). Método automático para detecção de nematóides em lavoura cafeeira usando imagens aéreas.

Em suma, este trabalho oferece um avanço significativo no controle de nematóides em lavouras cafeeiras, apresentando um método automático e preciso de detecção por meio de imagens aéreas e aprendizado de máquina. Essa abordagem pode ser uma ferramenta valiosa para produtores e agrônomos, ajudando-os a tomar decisões informadas e implementar estratégias de manejo mais eficazes para proteger suas culturas e otimizar a produção de café.

2.4 Conclusão

Os trabalhos desta revisão apresentaram o uso da aprendizagem profunda para a identificação de doenças relacionadas à agricultura. Os trabalhos tem como ponto principal, a captura de imagens a partir de UAV, criação de um banco de dados e o tratamento da imagem para que sejam treinados em redes de segmentação. As avaliações dos resultados foram obtidas a partir de datasets, onde cada trabalho lida especificamente ao problema;

alguns usam uma quantidade menor de imagens (OLIVEIRA et al., 2019), outros uma quantidade maior de imagens (SANTOS, 2019). A partir dos trabalhos apresentados é possível entender que não existe um padrão específico de Redes Neurais Artificiais (RNA) para o tipo de plantio. Analisando os trabalhos, os resultados apresentam valores de acurácia bem distintos; mostrando que o pré-processamento é crucial antes do aprendizado da rede. Ainda é possível avançar com as técnicas de AM em várias aplicações, mostrando-se uma abordagem promissora para a classificação de doenças em plantios no setor agrícola brasileiro.

Capítulo 3

Aprendizagem de Máquina

Neste capítulo, são apresentados os princípios referentes ao processamento de imagens e aprendizado de máquina, além de abordar as redes de segmentação adotadas neste trabalho. A abordagem de AM é discutida a partir da diferença entre o aprendizado supervisionado e não supervisionado, incluindo uma explanação sobre a Aprendizagem Profunda (do inglês, *Deep Learning - DL*). Para aprofundar o tema Deep Learning (DL), são detalhadas as teorias relacionadas as RNAs (Redes Neurais Artificiais) e RNC (Redes Neurais Convolucionais).

3.1 Imagens digitais

A visão computacional é uma área que busca interpretar as imagens de forma semelhante à percepção humana, propondo sistemas para a compreensão de imagens capturadas por sensores. Embora os sistemas existentes tenham realizado progressos significativos nas últimas décadas e tenham diversas aplicações, ainda não alcançaram a capacidade de igualar a precisão da visão humana (PARKER, 2010). As imagens digitais são compostas por conjuntos de *pixels*, podendo conter milhões deles, dependendo do sensor utilizado na câmera. Cada pixel representa uma célula na matriz que compõe a imagem.

Uma imagem digital pode ser descrita como uma representação visual de uma cena ou objeto, o qual foi convertido em formato digital, composta por um conjunto de pixels organizados em uma matriz bidimensional. Cada pixel representa uma célula na imagem e possui informações de cor e intensidade luminosa. Existem diferentes tipos de imagens digitais, incluindo imagens em escala de cinza; onde cada pixel possui um valor representando tonalidade de cinza, imagens em cores e com bandas espectrais.

A resolução de uma câmera digital é determinada pelo número total de pixels que a compõem. Quanto maior a quantidade de pixels, maior será a resolução. Por exemplo, uma imagem com resolução de 1920x1080 possui 1920 pixels na largura e 1080 pixels na altura da imagem.

Uma imagem digital pode ser descrita como função bidimensional que mapeia as coordenadas espaciais (x, y) para os valores numéricos, os quais representam as intensidades ou cores. Matematicamente, pode-se representar uma imagem digital como uma matriz de valores, onde cada elemento da matriz corresponde a um pixel na imagem. Ao considerar uma imagem em escala de cinza, cada pixel é representado por um único valor, que varia de 0 a 255 (esses valores referem-se para sensores de 8bits), indicando a intensidade luminosa daquele ponto na imagem. E assim, a função bidimensional de uma imagem em escala de cinza pode ser escrita como:

$$f(x, y) \rightarrow I \tag{1}$$

Onde f é a função que mapeia as coordenadas (x, y) para a intensidade luminosa I.

Para imagens coloridas, a função f é pouco complexa, pois cada pixel possui três componentes de cor (vermelho, verde e azul). Nesse caso, pode-se representar uma imagem colorida como três funções separadas, uma para cada canal de cor:

$$f_r(x, y) \rightarrow R; f_g(x, y) \rightarrow G; f_b(x, y) \rightarrow B \tag{2}$$

As funções f_r, f_g e f_b mapeiam as coordenadas espaciais (x, y) para os valores dos componentes vermelho (R), verde (G) e azul (B) do pixel respectivamente. Essas funções são essenciais para processar, analisar e manipular as imagens digitais, que utilizam as informações dos pixels para realizar diversas tarefas, como filtragem, segmentação, reconhecimento de padrões, entre outras aplicações (DADWAL; BANGA, 2012).

3.2 Aprendizado de Máquina

O Aprendizado de Máquina (do inglês, *Machine Learning*) é um subcampo da Inteligência Artificial (IA) que se concentra no desenvolvimento dos algoritmos e técnicas que permitem aos computadores aprenderem e melhorarem o seu desempenho em tarefas específicas, sem serem programados em cada passo. O objetivo principal do AM é permitir que as máquinas aprendam a partir dos dados e experiências, de forma que possam tomar as devidas decisões e realizar as tarefas de maneira autônoma. O processo de aprendizado de máquina envolve geralmente os seguintes elementos abaixo:

- Dados: Eles são relevantes e podem ser estruturados ou não estruturados, e devem ser representativos do problema que se deseja resolver;
- Algoritmos: São selecionados e aplicados os algoritmos de AM que analisam os dados e criam modelos matemáticos que representam padrões em relação aos dados;
- Treinamento: Ele envolve ajustar os parâmetros do modelo para minimizar os erros e fazer com que o modelo seja capaz de fazer as previsões corretas;

- ❏ Avaliação: Os modelos são avaliados usando dados de teste para medir a sua capacidade de generalizar previsões em novos dados não vistos durante o treinamento;

- ❏ Melhoria: Os modelos podem ser ajustados e o processo de treinamento pode ser repetido para melhorar o desempenho do algoritmo em questão.

O AM é aplicado em várias áreas, como reconhecimento de fala, processamento de linguagem natural, visão computacional, sistemas de recomendação, detecção de fraudes, diagnósticos médicos, dentre outros como apresenta (SANTOS, 2019). As suas técnicas e algoritmos têm impulsionado avanços significativos em diversas indústrias, proporcionando maior eficiência, precisão e automação em muitos processos e tomadas de decisão (MONARD; BARANAUSKAS, 2003).

(GOLLAPUDI, 2016) enfatiza que a chave para facilitar a definição de uma plataforma de solução de problemas em AM está em um conceito. Essencialmente, esse conceito é um mecanismo que permite que a máquina pesquise padrões e desenvolva aprendizado, melhorando suas capacidades com base em suas experiências. O processo de busca de padrões é crucial para a maneira como as máquinas percebem o ambiente, aprendem a distinguir comportamentos e tomam decisões de categorização, algo que é feito a humanos.

O objetivo da implementação da AM é desenvolver um algoritmo capaz de resolver os problemas práticos. Durante esse processo, os aspectos importantes como dados, tempo e requisitos de espaço devem ser considerados. Ademais, a capacidade de aplicar esses algoritmos a uma ampla gama de problemas de aprendizagem é crucial. O objetivo desses algoritmos é produzir resultados com mais alto nível de precisão (GOLLAPUDI, 2016).

Existem várias maneiras de classificar os algoritmos de AM, sendo comum dividi-los em categorias com finalidade. Duas categorias amplamente mencionadas no estado da arte são aprendizagem supervisionada e aprendizagem não-supervisionada.

3.2.1 Aprendizado de Máquina Supervisionado

De acordo com (GOLLAPUDI, 2016), o aprendizado supervisionado é realizado por meio de algoritmos que estabelecem as relações entre os atributos de entrada e saída, usando um conjunto de dados conhecidos, denominados “rotulados”. A denominação feita, aborda a marcação dentro de uma imagem de modo que especifique o que deseja segmentar dentro daquela imagem, criando “rótulos”, no qual especifica cada marcação.

Nesse tipo de aprendizado, durante o treinamento, o algoritmo busca aprender a relação entre as entradas e saídas para fazer as previsões precisas em novos dados não vistos antes. A aprendizagem supervisionada é utilizada para problemas de classificação e regressão, onde o objetivo é prever a categoria ou o valor associado a novos dados com base no aprendizado realizado a partir dos dados rotulados de treinamento.

3.2.2 Aprendizado de Máquina Não-Supervisionado

(GOLLAPUDI, 2016) apresenta a Aprendizagem de Máquina Não-Supervisionada, qual lida com conjunto de dados não rotulados. Ao contrário da aprendizagem supervisionada, o algoritmo não recebe informações explícitas sobre as saídas desejadas dos dados de entrada. O objetivo da aprendizagem não supervisionada é descobrir padrões, estruturas ou relações intrínsecas nos dados sem a orientação de rótulos ou valores esperados. Os algoritmos são projetados para agrupar os dados em categorias ou *clusters*, identificar anomalias, reduzir a dimensionalidade dos dados ou encontrar representações latentes.

Esse tipo de aprendizado é utilizado em tarefas como agrupamento de dados, análise de componentes principais, redução de dimensionalidade, entre outras aplicações. A aprendizagem não supervisionada é útil quando se deseja explorar e extrair as informações significativas de grandes conjuntos de dados não rotulados, possibilitando a identificação de padrões ocultos para as análises e as tomadas de decisão das aplicações em questão.

3.2.3 Redes Neurais Artificiais

Outra forma de aplicar AM é com uso das Redes Neurais Artificiais (RNAs) são modelos matemáticos e computacionais inspirados no funcionamento do cérebro humano, elas também podem ser classificadas em dois tipos principais, com base no tipo de aprendizado que realizam: supervisionado e não supervisionado. Esses modelos são projetados para aprender padrões complexos a partir de dados e, assim, realizar tarefas como classificação, regressão, reconhecimento de padrões, entre outras. Uma rede neural artificial é composta por camadas de unidades interconectadas, conhecidas como neurônios artificiais. Cada neurônio recebe entradas ponderadas, aplica uma função de ativação e, em seguida, transmite o resultado para as próximas camadas. O processo de treinamento envolve ajustar os pesos das conexões entre os neurônios para que a rede seja capaz de fazer previsões precisas.

O neurônio artificial Perceptron é uma unidade básica de processamento que faz parte da família de RNAs simples. Foi proposto inicialmente em 1958 por Frank Rosenblatt, tornando-se um dos primeiros modelos de aprendizado de máquina. O Perceptron é usado para problemas de classificação binária, ou seja, tarefas em que o objetivo é dividir os dados em duas classes. Sua estrutura é composta por uma camada de entrada com entradas ponderadas e uma função de ativação, além de uma função de saída que toma uma decisão com base na soma ponderada das entradas (MALSBURG, 1986).

A seguir estão os principais componentes do neurônio Perceptron:

- Entradas: Cada neurônio Perceptron recebe um conjunto de entradas (x1, x2, ..., xn), representa as características ou atributos dos dados a serem classificados;

- Pesos: Cada entrada é associada a um peso (w1, w2, ..., wn). Os pesos representam a importância relativa de cada entrada na decisão final do neurônio;

- Função de Ativação: Após as entradas serem multiplicadas pelos pesos, o neurônio Perceptron soma esses produtos ponderados ($w_i * x_i$) e aplica uma função de ativação para determinar a saída do neurônio. A função de ativação é uma limiar, que retorna 1 se a soma for maior ou igual a um determinado limiar, ou 0 caso contrário. A saída y do neurônio é dada por: $y = f(w_i * x_i)$, onde f é a função de ativação;

- Função de Saída: O resultado da função de ativação é a saída do neurônio Perceptron, representa a decisão binária de classificação (0 ou 1);

- O treinamento do Perceptron envolve ajustar os pesos de forma iterativa. Esse algoritmo visa minimizar os erros de classificação nos dados de treinamento, e o processo continua até que o Perceptron alcance uma solução adequada ao problema.

(MALSBURG, 1986) aponta que o Perceptron só resolve problemas linearmente separáveis. Para superar essa limitação, foram desenvolvidas RNAs multi-camadas, que usam as várias camadas de neurônios para aprender as representações abstratas.

3.2.4 *Multi-Layer Perceptron*

O Multi-Layer Perceptron (MLP) é uma extensão do conceito de Perceptron simples, e é um dos tipos mais comuns de RNAs. O Multi Layer Perceptron (MLP) é conhecido como uma RNA feedforward, pois a informação flui em uma única direção, das camadas de entrada para as camadas de saída, sem ciclos de retro-alimentação.

A principal característica do MLP é a presença de uma ou mais camadas intermediárias, chamadas ocultas, entre a camada de entrada e a saída. Cada camada é composta por um conjunto de neurônios (conhecidos como unidades), e conexões entre as camadas têm pesos associados aprendidos durante o processo de treinamento da rede.

Segundo (HAYKIN, 2004), as RNAs têm sido adotadas em várias metodologias de AM, com destaque aprendizagem supervisionada e não supervisionada, proporcionando resultados bem-sucedidos, como classificação de imagens, tomada de decisões, diagnóstico médico e processamento de linguagem natural. Ao contrário do Perceptron, que possui apenas um neurônio de saída, uma rede MLP pode se conectar a vários neurônios de saída, permitindo a maior complexidade e capacidade de lidar com tarefas mais abrangentes.

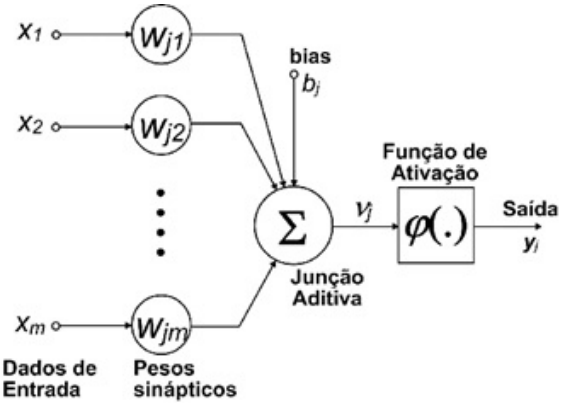


Figura 5 – Adaptação do autor (HAYKIN, 2004).

3.2.5 Função de Ativação

As funções de ativação são equações matemáticas essenciais que determinam a saída de cada neurônio em uma rede neural. Cada neurônio é associado a uma função de ativação que decide se ele deve ser ativado ou não, dependendo da relevância de sua entrada para as previsões do modelo. Além disso, as funções de ativação ajudam a normalizar as saídas dos neurônios, limitando-as a um intervalo entre 1 e 0 ou entre -1 e 1.

É crucial que as funções de ativação sejam eficientes, pois são aplicadas em milhares ou até milhões de neurônios para cada amostra de dados. Especialmente durante o treinamento do modelo com a técnica de retro-propagação (*backpropagation*), requer o cálculo de derivadas da função de ativação e eficiência computacional se torna útil.

(NORONHA; SILVA,) aborda que as RNAs modernas se apoiam na técnica de retro-propagação para ajustar os seus pesos e treinar o modelo, o que impõe uma carga computacional significativa na função de ativação quanto em sua derivada. Portanto, a escolha das funções de ativação eficientes e aprimoradas é de extrema importância para garantir o desempenho adequado e eficiente da RNA durante o treinamento e a inferência.

3.2.6 Função de Erro

A função de erro é usada para medir a discrepância entre a saída real produzida pelo modelo atual e a saída desejada. Durante o treinamento do modelo, o objetivo é minimizar essa função de erro, de forma a tornar a saída o mais próximo possível do valor correto.

Para avaliar a qualidade das predições feitas pela RNA e verificar a pertinência dos pesos e bias no sistema, utiliza-se a função de custo, conhecida como "Loss Function". Essa função é responsável por avaliar o quão bem a RNA está realizando as predições.

Em contextos supervisionados, a função de custo compara a saída correta de uma amostra com a saída gerada pela rede, revelando o erro estimado ou a discrepância entre ambas. O objetivo do treinamento é ajustar a rede de forma a minimizar a função de custo para todas as amostras, garantindo assim que o modelo se torne cada vez mais preciso em

suas predições. Segundo (MENDES et al., 2018), esse processo de minimização é essencial para obter um modelo bem ajustado e com boas capacidades de previsão.

(CRISPI, 2022) evidencia que dentre as funções de custo mais utilizadas em classificação, destaca-se a Entropia cruzada. Para um exemplo, sendo y a distribuição de classes reais e a predição dada por $f(x) = \hat{y}$, a soma das entropias cruzadas (entre a predição e a classe real) de cada classe j pode ser apresentada pela equação:

$$l^{(ce)} = -\sum_{j} yj.log(\hat{y_j} + \varepsilon) \tag{3}$$

3.2.7 Retro-Propagação

(GHAREHCHOPOGH, 2011) apresenta a técnica na implementação de RNAs que permite calcular o gradiente dos parâmetros, a fim de realizar a descida do gradiente e minimizar a função de custo. Após a definição de uma RNA com seus pesos iniciais, uma passagem é realizada para gerar a previsão inicial. Nesse ponto, uma função de erro é aplicada para determinar o quão distante o modelo está da previsão real. Existem vários algoritmos na literatura que podem minimizar essa função de erro. No entanto, para RNAs grandes, é necessário um algoritmo de treinamento eficiente em termos computacionais.

Assim, para descobrir os pesos ideais dos neurônios, executamos o processo de retrocesso (*backpropagation*), que consiste em retroceder da previsão da rede até os neurônios que geraram essa previsão. O backpropagation é a metodologia pela qual o modelo é avaliado após a definição de uma estrutura de previsão. Esse processo começa pela última camada da rede e vai retrocedendo até a primeira, identificando e ajustando os pesos dos neurônios para melhorar a eficácia da rede, utilizando a função de custo.

A retro-propagação acompanha as derivadas das funções de ativação em cada neurônio subsequente para encontrar os pesos que minimizem a função de custo, resultando na melhor previsão possível. (GHAREHCHOPOGH, 2011) descreve que a retro-propagação é a parte intensiva de uma RNA, desta forma a ideia básica é aplicar repetidamente uma regra matemática para calcular a influência de cada peso e sua relação com uma função de erro arbitrária. Esse procedimento é conhecido como descida de gradiente. Figura 6 mostra a retro-propagação por (GHAREHCHOPOGH, 2011).

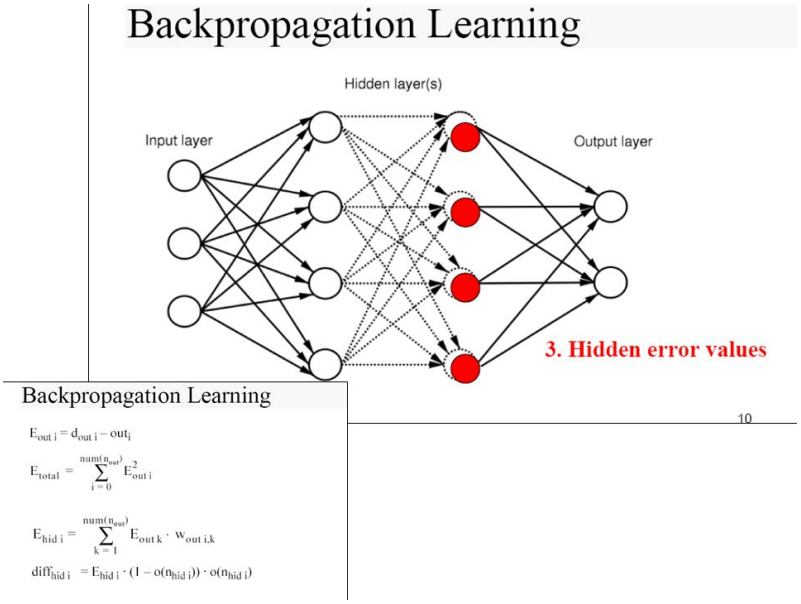


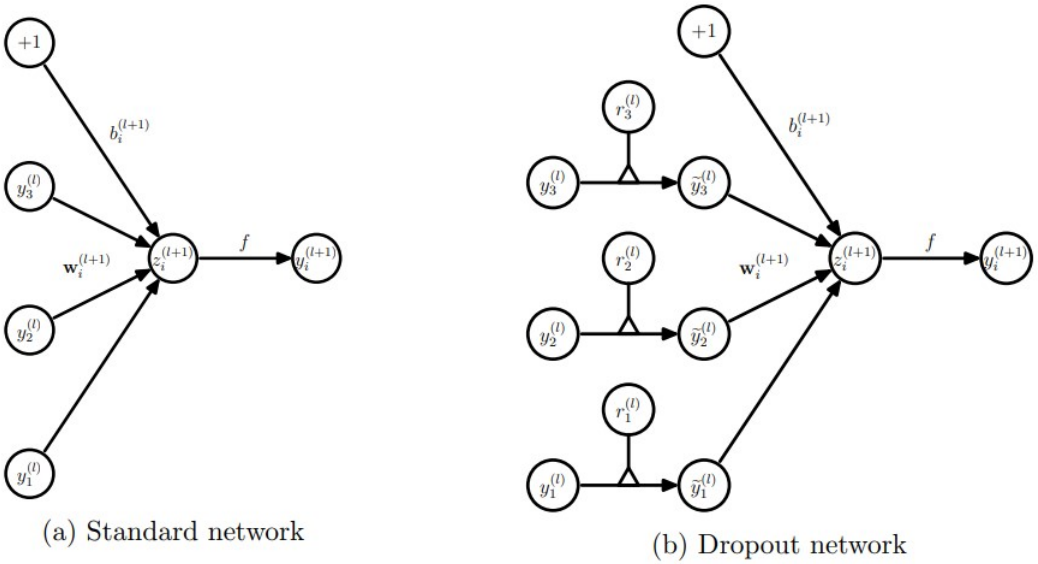
Figura 6 – Do autor (GHAREHCHOPOGH, 2011).

3.2.8 *Dropout*

É uma técnica de regularização utilizada em RNAs para evitar o overfitting e melhorar o desempenho de generalização, o primeiro a usar esta técnica foi (SRIVASTAVA et al., 2014). Em redes neurais, o overfitting ocorre quando o modelo se torna muito especializado nos dados de treinamento e tem um desempenho ruim em dados não vistos.

A técnica de dropout envolve desativar aleatoriamente, ou "eliminar", uma certa proporção de neurônios durante cada iteração de treinamento. Isso significa que, durante os avanços e retrocessos (forward e backward), alguns neurônios são excluídos da rede. A taxa de dropout é um hiperparâmetro que determina a fração de neurônios a serem eliminados, variando de 0,2 a 0,5. Ao usar o dropout, a RNA se torna mais robusta porque diferentes subconjuntos de neurônios são ativados em cada iteração, forçando a rede a aprender representações. Isso, por sua vez, ajuda a evitar que a rede dependa muito de neurônios específicos e a torna mais adaptável a diferentes entradas.

Durante a fase de teste ou inferência, o dropout geralmente é desativado e a RNA é usada para fazer previsões. O dropout é mais eficaz quando aplicado a RNAs profundas com muitos parâmetros, pois ajuda a evitar o overfitting e a melhorar a capacidade do modelo de generalizar para novos dados. Essa técnica se tornou utilizada no treinamento de RNAs de última geração em várias áreas e aplicações, como (SRIVASTAVA et al., 2014) aplicou e fez comparação das operações de uma rede neural padrão e com o uso do *dropout* (Figura 7).



(a) Standard network (b) Dropout network

Figura 7 – Do autor (SRIVASTAVA et al., 2014).

3.2.9 *Adam Optimization*

É um algoritmo de otimização usado no treinamento de RNAs. (ADAM et al., 2014) combina os conceitos do algoritmo de otimização RMSprop e do Momentum, buscando obter uma taxa de convergência mais rápida e eficiente durante o processo de treinamento.

O Adam Optimization calcula uma taxa de aprendizado adaptativa para cada parâmetro da RNA com base na média móvel dos gradientes anteriores e do momento dos gradientes. Essa abordagem ajuda a ajustar a taxa de aprendizado para cada parâmetro individualmente, o que permite que o algoritmo se adapte automaticamente a diferentes taxas de aprendizado para diferentes parâmetros (ZHANG, 2018).

O algoritmo Adam utiliza duas etapas principais durante o treinamento:

1. Cálculo do momento: Calcula o momento dos gradientes para cada parâmetro. Isso ajuda a acelerar a convergência do algoritmo, permitindo que ele "ganhe impulso"em direção às áreas ótimas mais rapidamente.

2. Cálculo da taxa de aprendizado adaptativa: Calcula uma taxa de aprendizado para cada parâmetro com base na média móvel dos gradientes anteriores. Essa abordagem permite que o algoritmo ajuste a taxa de aprendizado para cada parâmetro individualmente, melhorando a estabilidade e a eficiência do processo.

(ZHANG, 2018) aponta à sua eficácia em diferentes tipos de problemas, bem como sua capacidade de reduzir a necessidade de ajustar manualmente a taxa de aprendizado durante o treinamento em redes neurais.

3.3 *Deep Learning*

DL é uma subárea da inteligência artificial (IA) que se concentra no desenvolvimento de algoritmos e modelos de RNAs para realizar tarefas de aprendizado a partir de dados (DAMACENO; VASCONCELOS et al., 2018). (GOODFELLOW; BENGIO; COURVILLE, 2016) apresenta o treinamento de redes neurais com DL envolve da seguinte forma: Inicia-se com a preparação dos dados, onde os dados são coletados, esses dados podem ser áudios, imagens, texto, variando de acordo com o que será aplicado. Em seguida, escolhe-se a arquitetura da rede neural adequada à tarefa, inicializam-se os parâmetros, define-se a função de custo e o algoritmo de otimização. O treinamento é um processo iterativo, onde os pesos são ajustados para minimizar a função de custo ao longo de múltiplas iterações. O conjunto de validação é usado para ajustar hiper-parâmetros, e o conjunto de teste é usado para avaliar o desempenho final.

DL é uma abordagem de rede neural que se caracteriza por possuir múltiplas camadas ocultas e um maior número de operações. Algumas dessas camadas têm a função específica de extrair características relevantes, contribuindo significativamente para o resultado final do modelo. Nessa estrutura, cada camada sucessiva constrói representações mais complexas, permitindo que a rede aprenda características hierárquicas e abstrações dos dados de entrada. Com essa capacidade de extração e aprendizado de características, (DENG; YU et al., 2014) aponta que o DL tem se mostrado eficaz em tarefas como visão computacional, processamento de linguagem natural, reconhecimento de padrões e muitas outras aplicações complexas.

No campo do ML, o algoritmo é treinado usando uma quantidade significativa de dados, representando a base fundamental para a coleta e aprendizado a partir desses dados. Em seguida, o algoritmo é capaz de fazer determinações ou previsões sobre os eventos ou fenômenos no mundo, envolvendo os processos distintos e complementares.

Por outro lado, o conceito de DL é caracterizado por uma dimensão semelhante ao das RNAs humanas, com conexões e direções de propagação de dados, o que proporciona uma abordagem prática e eficiente ao ML. Dessa forma, o DL expande a existência e a capacidade do AM, permitindo a criação de modelos mais complexos e poderosos, capazes de aprender representações hierárquicas e abstratas a partir dos dados de entrada. A Figura 8 apresenta uma visão da forma de processamento e diferença entre ML e DL.

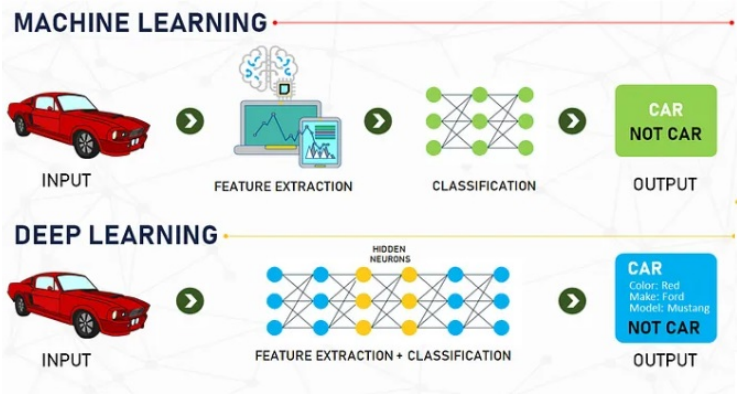


Figura 8 – Imagem adaptada e alterada de (LIMA; AQUINO; MILL, 2020), apresentando um comparativo entre os resultados de saída, machine learning e deep learning.

3.3.1 Redes Neurais Convolucionais

(VARGAS; PAES; VASCONCELOS, 2016) descreve que uma CNN é uma variação das Redes de Perceptrons de Múltiplas Camadas, sendo inspirada no processamento biológico de dados visuais. Assim como nos processos tradicionais de visão computacional, uma CNN é capaz de aplicar filtros em dados visuais, preservando a relação de vizinhança entre os pixels da imagem durante o processamento da rede. Essa abordagem permite que a CNN aprenda características visuais hierárquicas, tornando-a altamente eficaz em tarefas de reconhecimento e classificação de imagens. A Figura 9, ilustra uma CNN.

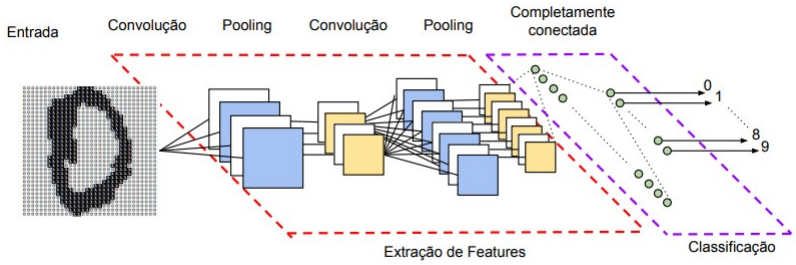


Figura 9 – Do autor (VARGAS; PAES; VASCONCELOS, 2016).

(CRISPI, 2022) apresenta que para a formação da arquitetura de uma rede de Aprendizagem profunda geralmente são escolhidos os seguintes tipos de camada:

- ❏ Camada convolucional e características da convolução;
- ❏ Camadas de agrupamento (Pooling layers);
- ❏ Camada Totalmente Conectada (Fully Connected layer);
- ❏ Camada de unidade Linear Retificada (ReLU);
- ❏ Unpooling e Camada de Convolução Transposta.

3.3.2 Camada Convolucional e Características da Convolução

Segundo (SILVA et al., 2018), as camadas de convolução têm a função de extrair as características através da aplicação de um kernel. Os filtros de convolução são aplicados nas imagens de entrada, sendo ajustados automaticamente para extrair as características relevantes para a classificação dos dados. A convolução, em termos matemáticos, é representada como uma operação linear que envolve a multiplicação de duas matrizes distintas, sendo a imagem e o filtro no caso específico do processamento de imagens. Os componentes dessa operação podem ser descritos como a entrada (a imagem), o detector de características (chamado de Kernel) e o resultado de saída (output) (CRISPI, 2022).

Esses filtros são responsáveis por detectar padrões específicos nas imagens, como bordas, texturas, padrões geométricos, ou outras características visuais importantes para a tarefa de classificação ou reconhecimento. Durante o treinamento da CNN, os filtros são ajustados automaticamente para aprender a reconhecer as características relevantes para a tarefa em questão. Assim uma vez que os detectores de características são aprendidos durante o treinamento, a CNN pode usar essas informações para fazer previsões e classificar novas imagens de entrada com base nas características extraídas pelos filtros.

3.3.3 Camadas de Agrupamento

As camadas de pooling, também conhecidas como camadas de redução de amostragem, são componentes das Convolutional Neural Networks (CNNs). Essas camadas são geralmente inseridas após as camadas de convolução e têm o objetivo de reduzir o tamanho espacial das representações das características geradas pelas convoluções, mantendo as informações mais relevantes (CRISPI, 2022).

O pooling é realizado aplicando uma janela (kernel) deslizante sobre os mapas de características. A operação mais comum é o max pooling, onde o valor máximo dentro da janela é selecionado e usado na nova camada reduzida. Outra operação é a média, que calcula a média dos valores dentro da janela. A Figura 10 é apresentada por (YANI; IRAWAN S; SETININGSIH ST, 2019), mostra o resultado do uso do max pooling.

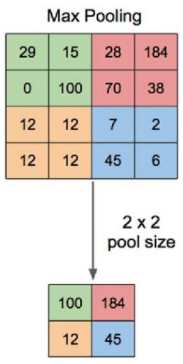


Figura 10 – Do autor (YANI; IRAWAN S; SETININGSIH ST, 2019).

As camadas de pooling também podem levar à perda de informações finas em algumas regiões da imagem, já que apenas os valores mais relevantes são preservados. Em alguns casos, isso pode ser desejável para aumentar a capacidade de generalização da rede, mas em outras situações, pode ser necessário ajustar o tamanho do pooling ou utilizar outras técnicas para mitigar essa perda de detalhes.

3.3.4 Unpooling

O unpooling é o processo inverso do pooling e tem como objetivo expandir a imagem, buscando retorná-la ao tamanho original, a Figura 11 apresenta o processo inverso do pooling. Essa técnica é frequentemente utilizada em combinação com camadas de convolução para restaurar as dimensões espaciais das representações após o processo de pooling. O unpooling ajuda a recuperar informações perdidas durante o pooling, permitindo que a rede mantenha detalhes importantes das características e retorne à resolução original da imagem (ESCALONA et al., 2019).

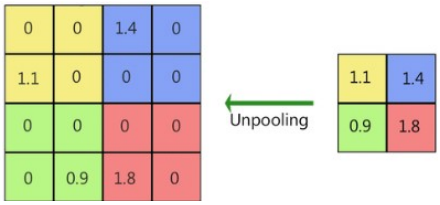


Figura 11 – Do autor (ESCALONA et al., 2019).

3.3.5 Camada de Unidade Linear Retificada

A função ReLU (Rectified Linear Unit) é uma função de ativação utilizada em redes neurais profundas, especialmente em arquiteturas de Convolutional Neural Networks (CNNs) (NAIR; HINTON, 2010). A função ReLU é definida matematicamente como:

$$f(x) = max(0, x) \tag{4}$$

Ou seja, a função retorna o valor de entrada x se ele for maior que zero, e retorna zero caso contrário. Em termos simples, a função ReLU mapeia valores negativos para zero e mantém os valores positivos inalterados.

3.3.6 *Convolutional Auto-Encoder*

(CRISPI, 2022) descreve que um Autoencoder (AE) pode ser desmembrado em duas partes distintas: um codificador, representado por "f", e um decodificador, representado por "g". A função do codificador é transformar uma entrada original em uma representação mais compacta, conhecida como código de representação. O decodificador, por sua vez, tem como objetivo reconstruir a entrada original a partir desse código de representação.

Dado que o código é uma representação limitada, o Autoencoder (AE) procura aprender um código que capture informações relevantes sobre a estrutura dos dados, pois não pode simplesmente copiar perfeitamente a entrada. Em vez disso, o Autoencoder (AE) tenta identificar as principais características dos dados para criar uma representação mais compacta e útil para o processo de reconstrução. A Figura 12 ilustra o processo do Autoencoder, dado uma entrada, passando pela compressão e gerando a reconstrução da entrada.

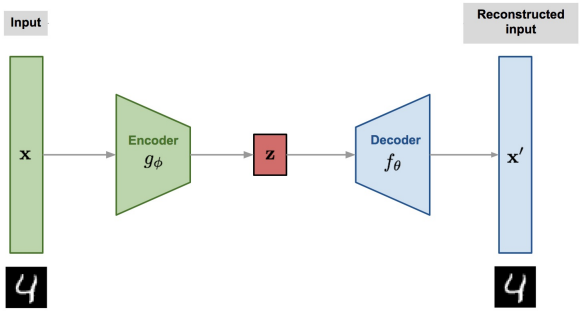


Figura 12 – Representação gráfica AE.

3.3.7 VGG-16

A VGG16 é uma arquitetura de rede neural convolucional que foi proposta pela Visual Geometry Group (VGG) da Universidade de Oxford. Essa arquitetura se tornou um marco no campo de Visão Computacional desde sua introdução em 2014.

(THECKEDATH; SEDAMKAR, 2020) descreve que a rede VGG-16 é composta por 16 camadas convolucionais e possui um pequeno campo receptivo de 3x3. Ela também possui uma camada de Max pooling de tamanho 2x2 e um total de 5 camadas desse tipo. Após a última camada de Max pooling, existem 3 camadas totalmente conectadas. Isso é seguido por três camadas totalmente conectadas. A rede usa o classificador softmax como a camada final. A ativação ReLU é aplicada a todas as camadas ocultas. Uma representação na Figura 13 em formato de diagrama, demostra sua estrutura.

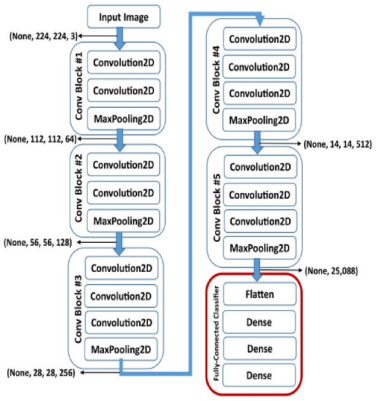


Figura 13 – Arquitetura VGG apresentada por (THECKEDATH; SEDAMKAR, 2020).

A VGG-16 pode ser usada como um componente dentro de um sistema de segmentação que inclui outras redes para realizar a tarefa de segmentação semântica. Uma abordagem comum é utilizar a VGG-16 como codificador (encoder) em uma arquitetura de rede de segmentação semântica, como o U-Net ou o LinkNet, por exemplo. O codificador é responsável por extrair características das imagens, e a VGG-16, com suas camadas convolucionais, pode cumprir essa função bem devido a sua profundidade. Já o decodificador (decoder) é a parte da rede que reconstrói a segmentação final da imagem a partir das características extraídas pelo codificador. O decodificador inclui camadas deconvolucionais (upsampling) e outras operações para recuperar a resolução da imagem segmentada.

Ao combinar a VGG-16 como codificador e decodificador, é possível criar uma arquitetura completa de segmentação semântica, que pode realizar tarefas como a identificação de objetos, detecção de bordas ou segmentação de classes de interesse em imagens.

Essa abordagem de combinar a VGG-16 com outras redes de segmentação permite aproveitar as vantagens de ambas as arquiteturas, tirando proveito da capacidade da VGG-16 em extrair características ricas e da capacidade da rede de segmentação em recuperar informações espaciais e realizar segmentação detalhada. O trabalho aqui apresentado foi feito usando a combinação entre a rede VGG-16 junto das seguintes redes: U-Net, PSPNet, LinkNet e FCN. Todas essas redes são descritas e apresentadas nas subseções seguintes.

3.3.8 U-Net

A U-Net é uma rede amplamente conhecida de Deep Learning que segue a estrutura de uma Convolutional Neural Network (CNN), tendo sido desenvolvida para a segmentação de imagens biomédicas. Essa arquitetura tem como objetivo permitir o uso eficiente de Datasets pequenos e possibilitar a classificação baseada na localização dos rótulos nas imagens, o que é essencial para a classificação de imagens.

O nome "U-Net"é derivado do formato em U que sua arquitetura possui. (RONNEBERGER; FISCHER; BROX, 2015) cita que a rede pode ser resumida em dois passos principais de tamanhos iguais: a contração e a expansão. Na fase de contração, a imagem é inserida na entrada da rede e é propagada através de camadas convolucionais, onde ocorre o aprendizado e a extração de características relevantes da imagem. Na fase de expansão, o resultado da contração é processado em camadas deconvolucionais (também conhecidas como camadas de upsampling) para produzir a segmentação final da imagem.

A U-NetVGG é a junção da U-Net quando combinada a VGG-16, possuindo conexões de salto (skip connections) que conectam as camadas correspondentes do caminho de contração ao caminho de expansão. Essas conexões de salto permitem que a informação de baixa resolução seja transmitida diretamente para o caminho de expansão, permitindo que a rede combine as informações contextuais de alto nível da VGG16 com os detalhes de baixo nível das camadas de contração. A Figura 14 apresenta sua arquitetura.

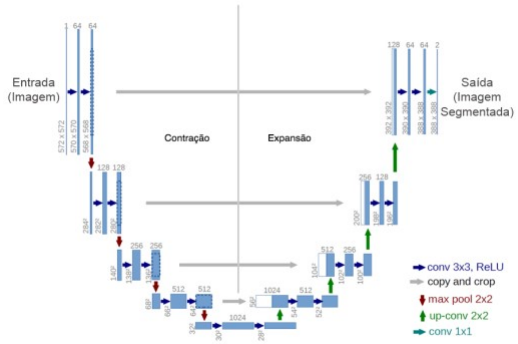


Figura 14 – Arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015).

A U-Net se destaca por sua capacidade de segmentar imagens, permitindo localizar e classificar características importantes em imagens, mesmo quando os conjuntos de dados são limitados em tamanho. Uma das principais distinções entre a rede U-Net e o Autoencoder reside na concatenação entre as camadas iniciais e finais, estabelecendo uma conexão direta entre a fase de contração e a fase de expansão da imagem. Essa concatenação possibilita que algumas características importantes da imagem original não sejam perdidas ao longo do processo e possam ser recuperadas durante a fase de expansão.

3.3.9 PSPNet

A PSPNet (Pyramid Scene Parsing Network) é uma rede de segmentação semântica que foi proposta para realizar a segmentação de cenas complexas em imagens (ZHU et al., 2021). Essa arquitetura foi desenvolvida para capturar informações contextuais em diferentes escalas, o que é crucial para entender a estrutura de uma cena em uma imagem.

A principal característica da PSPNet é o uso de uma pirâmide de pooling espacial (daí o nome Pyramid Scene Parsing Network). A pirâmide de pooling espacial consiste

em diferentes tamanhos de pooling que são aplicados a uma camada convolucional intermediária. Cada operação de pooling captura informações contextuais em uma escala diferente, o que permite que a rede considere detalhes finos e contextos mais amplos da imagem. A Figura 15 apresenta o processo da arquitetura PSPNet adaptada.

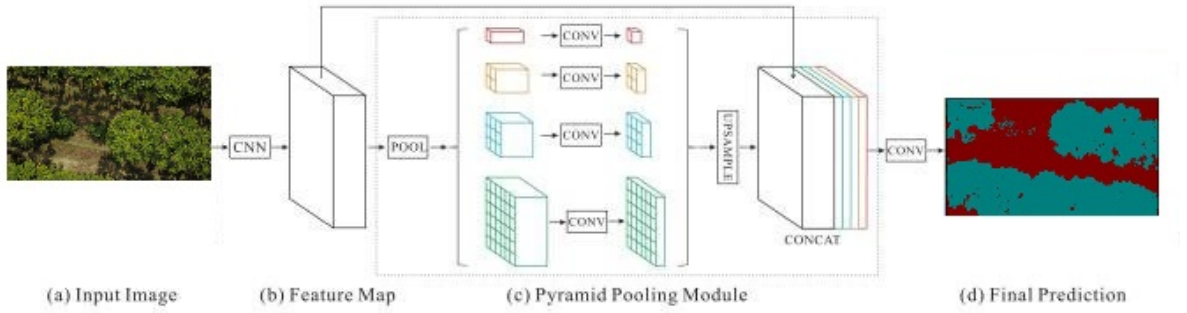


Figura 15 – Adaptação da arquitetura PSPNet apresentando imagens de entrada e resultados de saída do dataset criado.

A arquitetura da PSPNet segue os seguintes passos em seu processo de execução:

1. Extração de características: A imagem de entrada passa por várias camadas convolucionais para extrair características em diferentes níveis de abstração.

2. Pirâmide de Pooling Espacial: A camada convolucional intermediária é seguida por uma pirâmide de pooling espacial, onde múltiplos tamanhos de pooling são aplicados para capturar informações contextuais em diferentes escalas.

3. Upsampling: Após a pirâmide de pooling espacial, o resultado é processado por camadas de upsampling para restaurar a resolução original da imagem segmentada.

4. Camada de saída: A camada de saída é responsável por produzir a segmentação semântica da imagem, atribuindo uma classe a cada pixel da imagem, indicando a qual categoria ele pertence (por exemplo, pessoa, carro, árvore, etc.).

Essa rede é usada em aplicações como identificação de objetos em imagens de satélite, segmentação de cenas urbanas em imagens de ruas, entre outras tarefas que requerem uma compreensão abrangente do contexto da imagem para uma segmentação adequada.

A rede quando combinado com a VGG16 é empregada ao bloco de codificação inicial para extrair características de nível mais alto da imagem (encoder). Essas características de alto nível são então passadas para o PSPNet, que possui uma pirâmide de pooling espacial para capturar informações contextuais em diferentes escalas, desta forma permitindo que o PSPNet considere detalhes finos e contextos mais amplos da imagem, melhorando significativamente a qualidade da segmentação.

3.3.10 LinkNet

LinkNet é uma arquitetura de rede neural convolucional projetada para tarefas de segmentação semântica de imagens. A LinkNet propoe uma eficiente e de alta precisão para a segmentação de objetos em imagens, especialmente em cenários onde o número de classes a serem segmentadas é grande.

(RAMASAMY; SINGH; YUAN, 2023) apresenta que a principal característica distintiva do LinkNet é o uso de conexões "link"ou "atalhos"entre camadas, inspirado pelas conexões residuais (residual connections) da arquitetura ResNet. Essas conexões são adicionadas para ajudar no fluxo suave do gradiente durante o treinamento, o que facilita o treinamento de redes mais profundas e reduz o problema de degradação do desempenho, onde adicionar camadas pode levar a uma queda na precisão do modelo.

Além das conexões "link", o LinkNet utiliza blocos de codificação e decodificação para realizar a segmentação semântica (AE). A etapa de codificação compreende várias camadas convolucionais que vão reduzindo a resolução da imagem e capturando características de nível mais alto. Já a etapa de decodificação faz o caminho inverso, utilizando camadas deconvolucionais (ou upsampling) para reconstruir a resolução original da imagem segmentada, ao mesmo tempo em que realiza a fusão das características extraídas nas etapas anteriores. (RAMASAMY; SINGH; YUAN, 2023) apresenta na Figura 16 a LinkNet.

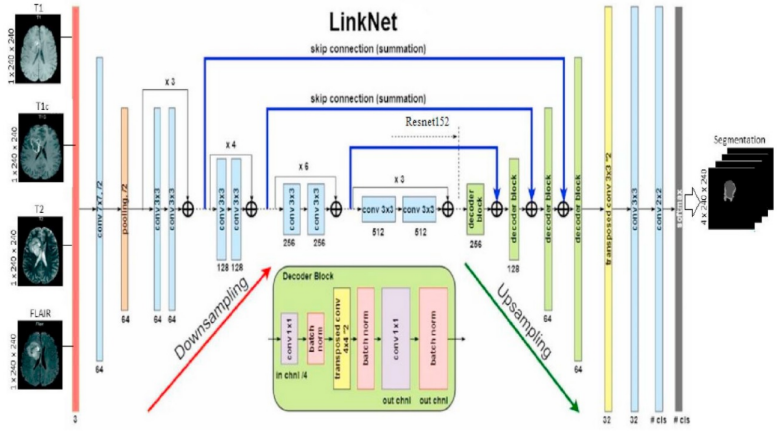


Figura 16 – Arquitetura LinkNet, adaptada por (RAMASAMY; SINGH; YUAN, 2023)

A arquitetura do LinkNet torna-a computacionalmente eficiente e facilmente treinável, proporcionando uma abordagem eficaz para segmentação semântica de imagens. Ela é aplicada em várias tarefas que requerem identificação e separação precisa de objetos em imagens, como detecção de objetos, reconhecimento de padrões e análise de cenas.

Neste trabalho combinou-se a arquitetura LinkNet com a VGG-16 que é uma abordagem comum para realizar a segmentação semântica de imagens. Nessa combinação, a VGG-16 é usada como o codificador (encoder) para extrair características ricas das imagens, enquanto o LinkNet atua como o decodificador (decoder) para fazer a segmentação.

3.3.11 FCN

FCN ("Fully Convolutional Network") em português (Rede Totalmente Convolucional) e é outra arquitetura de rede neural convolucional utilizada principalmente para tarefas de segmentação semântica em imagens. A FCN foi uma das primeiras arquiteturas a serem introduzidas especificamente para resolver o problema da segmentação semântica de imagens de forma end-to-end, ou seja, diretamente da entrada até a saída, sem a necessidade de etapas intermediárias de pré-processamento.

(XING; ZHONG; ZHONG, 2020) apresenta que diferente das arquiteturas convencionais de classificação de imagens, que normalmente possuem camadas densas totalmente conectadas após as camadas convolucionais, a FCN substitui essas camadas densas por camadas convolucionais transpostas (também conhecidas como camadas deconvolucionais ou upsampling layers). Essas camadas deconvolucionais ajudam a aumentar a resolução espacial das características extraídas pelas camadas convolucionais anteriores.

A arquitetura FCN é composta por três partes principais segundo (XING; ZHONG; ZHONG, 2020), podemos defini-las como:

- ❏ Codificação (Encoder): As camadas convolucionais realizam a codificação da imagem de entrada, capturando informações em diferentes níveis de abstração.
- ❏ Camadas deconvolucionais (Upsampling): As camadas deconvolucionais são usadas para restaurar a resolução original da imagem segmentada, ao mesmo tempo em que realizam fusão de características para obter segmentações mais detalhadas.
- ❏ Camada de saída: A camada final é uma camada convolucional que produz a segmentação semântica da imagem, atribuindo uma classe a cada pixel da imagem.

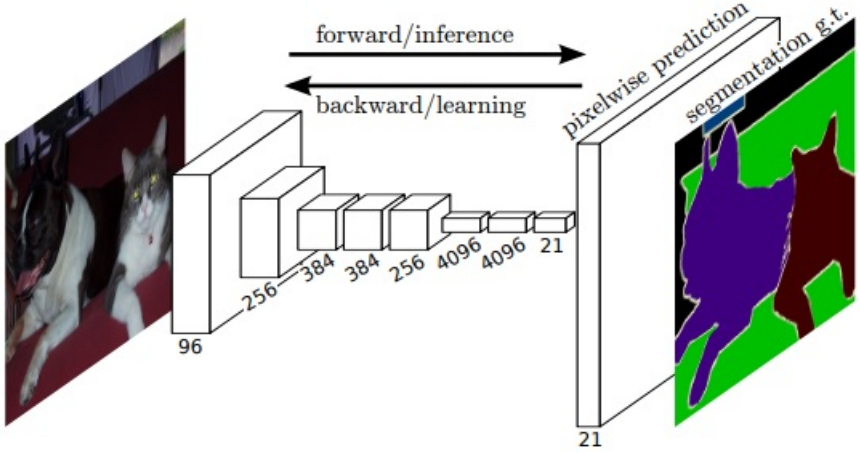


Figura 17 – Arquitetura FCN, adaptada por (LONG; SHELHAMER; DARRELL, 2015)

A Figura 17, é uma adaptação da arquitetura da rede FCN, apresentada por (LONG; SHELHAMER; DARRELL, 2015). Quando arquitetura da FCN se integra a da VGG16 ela adota a estratégia de upsample no caminho de expansão (decoder) para aumentar a resolução espacial das características. O upsample é realizado por meio de camadas de up-convolution, que são responsáveis por expandir a resolução das características obtidas no caminho de contração, ela também faz uso do recurso de (skip connection), estabelecendo assim conexões entre o caminha de contração e expansão.

No final do caminho de expansão, a FCNVGG usa camadas convolucionais 1x1 para mapear as características para o número de classes desejadas. Essas camadas convolucionais 1x1 reduzem a dimensionalidade das características e aplicam uma função de ativação softmax para gerar as probabilidades de segmentação para cada classe.

3.4 Veículos Aéreos Não Tripulados

UAV (Unmanned Aerial Vehicle) ou conhecido em português como VANT (Veículo Aéreo Não Tripulado) são termos equivalentes.. Ambos termos são usados para descrever aeronaves controladas remotamente que não requerem um piloto humano a bordo para operá-las. Neste trabalho usa-se as abreviações VANT.

Na última década, houve um notável aumento nas aplicações de VANTs para atividades civis, o que resultou em um rápido crescimento do mercado (LEAL, 2018). Os avanços nos sensores, eletrônica de controladores e sistemas de comunicação desempenharam um papel fundamental na versatilidade dos modelos de VANTs, além de acelerar consideravelmente o aprimoramento de suas funcionalidades.

3.4.1 Sistema de Imageamento

Um sistema de imageamento com VANT, refere-se à integração de uma aeronave não tripulada com um conjunto de sensores e câmeras que são utilizados para capturar imagens aéreas ou dados visuais de uma área específica, como descreve (FIGUEIRA; OLIVEIRA, 2013). O uso do VANT tem aplicações em diferentes setores, incluindo agricultura, mapeamento, inspeção, monitoramento ambiental, vigilância, entre outros. Abaixo lista-se os principais componentes e funcionalidades de um sistema de imageamento com VANT:

- Aeronave (Drone): O drone é a plataforma física que realiza o voo. Pode ser um quadricóptero, hexacóptero, octocóptero ou qualquer outro tipo de aeronave não tripulada. Ele é equipado com um sistema de propulsão, baterias, GPS e sistemas de controle de voo.

- Câmeras e Sensores: O coração do sistema de imageamento são as câmeras e sensores. Isso inclui câmeras RGB (para capturar imagens visíveis), câmeras multi-

espectrais (para análise de vegetação), câmeras térmicas (para detecção de calor), LIDAR (para mapeamento 3D), entre outros.

- ❏ Plataforma de Voo Autônomo: Um software de controle de voo autônomo permite que o drone voe de forma programada, siga trajetórias predefinidas e mantenha uma altitude constante para a captura de imagens estáveis.

- ❏ Estação de Controle: O operador controla o drone por meio de uma estação de controle, que pode ser um controle remoto ou um dispositivo móvel. Isso permite controlar o voo, câmeras e sensores remotamente.

- ❏ Software de Processamento de Dados: Após a coleta de dados, é necessário um software de processamento para transformar as imagens brutas em informações úteis. Isso pode envolver a criação de mapas, modelos 3D, imagens georreferenciadas ou análises específicas, dependendo da aplicação.

- ❏ Georreferenciamento: Muitas vezes, é essencial georreferenciar as imagens para que os dados possam ser usados com precisão em SIGs (Sistemas de Informações Geográficas). Isso envolve a captura de dados GPS precisos durante o voo e a combinação das coordenadas GPS com as imagens capturadas.

3.4.2 Legislação

A legislação (Lei nº. 7.565/86) brasileira sobre VANTs (Veículos Aéreos Não Tripulados), também conhecidos como drones, é regulamentada pela Agência Nacional de Aviação Civil (ANAC) e pelo Departamento de Controle do Espaço Aéreo (DECEA), com o objetivo de garantir a segurança das operações aéreas e proteger a privacidade das pessoas. (MARTINS et al., 2017) aborda sobre as principais diretrizes da legislação brasileira em relação a VANTs no pontuadas a seguir:

- ❏ Registro de VANTs: Todos os drones com peso superior a 250 gramas devem ser registrados na ANAC. Isso ajuda a rastrear a propriedade e a identificação dos operadores.

- ❏ Altitude Máxima de Voo: Os VANTs são restritos a voar a uma altitude máxima de 400 pés (aproximadamente 120 metros) acima do nível do solo. Isso visa evitar conflitos com o tráfego aéreo tripulado.

- ❏ Distância de Segurança: Os VANTs devem manter uma distância mínima de 30 metros de pessoas, veículos, edifícios e estruturas, a menos que haja autorização específica.

- ❏ Proibição de Voo em Áreas Restritas: É proibido voar VANTs em áreas restritas, como aeroportos e bases militares, a menos que haja autorização prévia das autoridades competentes.
- ❏ Seguro de Responsabilidade Civil: Operadores de VANTs devem contratar um seguro de responsabilidade civil para cobrir possíveis danos causados por suas operações.
- ❏ Comercialização e Uso Comercial: O uso comercial de VANTs requer autorização específica da ANAC e, em alguns casos, da ANATEL (Agência Nacional de Telecomunicações) e do DECEA.
- ❏ Privacidade e Proteção de Dados: A legislação visa proteger a privacidade das pessoas, proibindo o uso de VANTs para espionagem ou vigilância não autorizada.
- ❏ Restrições de Voos Noturnos: A maioria dos VANTs só pode operar durante o dia, a menos que tenham iluminação adequada e autorização específica para voos noturnos.

O não cumprimento dessas regulamentações pode resultar em penalidades legais e multas.

3.4.3 Resoluções

A resolução é uma medida da quantidade de informações visuais que uma imagem pode conter e é geralmente expressa em pixels, quanto mais altas as resoluções geralmente significam imagens mais detalhadas (SILVA; DALMOLIN, 1998).

A escolha da resolução adequada para imagens capturadas por VANTs depende da aplicação específica, por exemplo temos a resolução espacial, temporal, espectral, radiométrica, geoespacial e de câmera. (WOODS; GONZALEZ, 2008) traz o conceito sobre resoluções, assim apresenta-se a seguir nas sub-subseções os exemplos apresentados.

3.4.3.1 Resolução Espacial

A resolução espacial se refere à nitidez e ao nível de detalhes que uma imagem pode mostrar. Ela é medida em pixels por unidade de área, como pixels por polegada (PPI) ou pixels por metro (PXM). (WOODS; GONZALEZ, 2008) aborda que quanto maior a resolução espacial, mais detalhada será a imagem.

3.4.3.2 Resolução Temporal

A resolução temporal se refere à taxa de captura de imagens ao longo do tempo. VANTs podem ser configurados para capturar imagens em intervalos específicos, como a

cada segundo, minuto ou hora. Isso é importante para o monitoramento de mudanças ao longo do tempo, como o crescimento de culturas ou a evolução de fenômenos naturais. (ANTUNES; DEBIASI; SIQUEIRA, 2014)

3.4.3.3 Resolução Espectral

(ANTUNES; DEBIASI; SIQUEIRA, 2014) relaciona às diferentes faixas de comprimento de onda do espectro eletromagnético que uma imagem pode capturar. Alguns VANTs podem ser equipados com sensores multiespectrais ou hiperespectrais que capturam informações em várias bandas espectrais, o que é útil para análises agrícolas, detecção de vegetação e muito mais.

3.4.3.4 Resolução Radiométrica

A resolução radiométrica se refere à capacidade de uma imagem representar diferentes níveis de intensidade de luz ou cores. Ela é frequentemente medida em bits por canal, com resoluções mais altas permitindo uma representação mais precisa das variações de intensidade de luz.(ANTUNES; DEBIASI; SIQUEIRA, 2014)

3.4.3.5 Resolução Geoespacial

A resolução geoespacial segundo (SILVA, 2022), combina informações de resolução espacial com coordenadas geográficas, permitindo que as imagens sejam georreferenciadas. Isso é importante para mapeamento e análise geoespacial.

3.4.3.6 Resolução da Câmera

A resolução da câmera se refere ao número total de pixels em uma imagem. Por exemplo, uma câmera de 20 megapixels captura imagens com uma resolução de 20 milhões de pixels. Resoluções mais altas podem capturar detalhes mais finos, mas também resultam em arquivos de imagem maiores.(WOODS; GONZALEZ, 2008)

3.4.4 Conclusão

Para a construção desse trabalho, uma extensa base teórica foi utilizada, e esta foi abordada nesse capítulo. Atualmente, as técnicas de aprendizado de máquina têm sido amplamente empregadas em diversas finalidades, com especial destaque para as redes neurais artificiais e o Deep Learning. Essas abordagens têm se mostrado extremamente eficazes para resolver problemas complexos e realizar tarefas de alto nível em áreas como visão computacional, processamento de linguagem natural, reconhecimento de padrões e muito mais. O avanço contínuo nesse campo tem impulsionado significativamente o

progresso da inteligência artificial e possibilitado a criação de sistemas inteligentes cada vez mais sofisticados e capazes de tomar decisões autônomas em diversas aplicações práticas.

Capítulo 4

Metodologia

O capítulo apresenta uma descrição da metodologia deste trabalho. Inicia com uma visão geral apresentada por meio de um fluxograma, no qual são enumeradas as etapas realizadas. Depois, aborda a forma de aquisição de imagens no local de estudo, descrevendo cada passo desse processo. Por fim, apresenta a criação dos rótulos para as imagens e o conjunto de dados, métricas usadas e experimentos.

4.1 Metodologia Aplicada

A Figura 18 esquematiza a metodologia proposta para a detecção das folhas de mamoeiro com a presença da doença mancha anelar usando DL com as redes de segmentação.

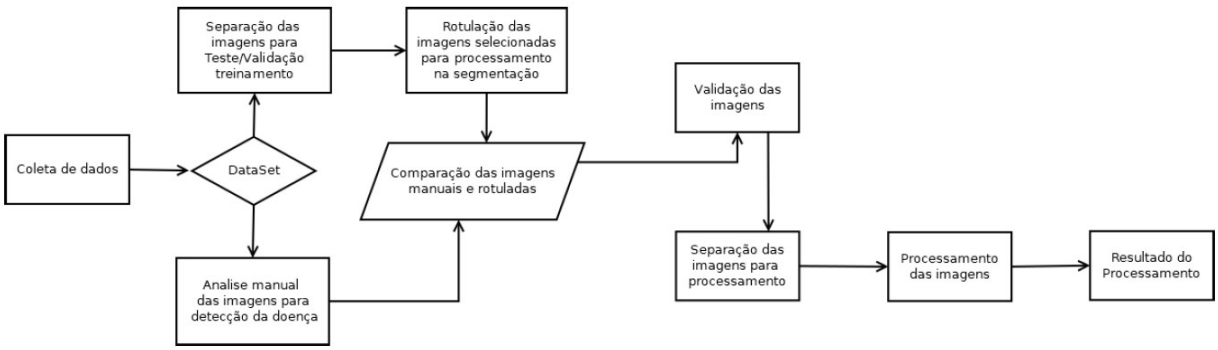


Figura 18 – Diagrama de fluxo do trabalho proposto.

Primeiramente, na Seção 4.1.1 descreve os passos que compõe o processo que inclui a aquisição de imagens, aplicação de fotogametria e ortomosaico, bem como a área escolhida e demarcada com/sem a presença da doença mancha anelar nos mamoeiros. Na Seção 4.1.2 apresenta-se os passos utilizados no pré-processamento dos dados, desde a criação dos exemplos para o treinamento da rede (rótulos), separação das imagens, criação do

conjunto de dados e o uso da técnica aumento de dados. Por fim, a Seção 4.1.3 mostra a arquitetura das RNC, com suas configurações, parâmetros e o motivo de sua escolha.

4.1.1 Método de Aquisição de Imagens

A coleta das imagens foram feitas na região sul do estado da Bahia, no qual foi escolhida por meio da parceria com a EMBRAPA a Fazenda Nossa Senhora Aparecida II, localizada em Eunápolis-BA, fazenda que mantém o cultivo e plantio do mamoeiro. Com a escolha do local já definido os agrônomos da EMBRAPA especialistas em mamoeiro demarcaram as áreas nas quais apresentam sintomas do vírus da mancha anelar na lavoura de maneira manual, ou seja, em solo. Após a definição da área escolhida pelos especialistas, foram feitas imagens aéreas da área demarcada por drone, modelo DJI Mavic Pro apresentado na seção anterior, a uma altitude de 12m relativa o solo no ponto de decolagem, com 80% sobreposição longitudinal e lateral e uma câmera na direção nadir, direcionada para o solo. A Figura 19 apresenta os agrônomos especialistas em solo identificando mamoeiros com a doença da mancha anelar e apresenta o drone usado em campo para a obtenção das imagens.

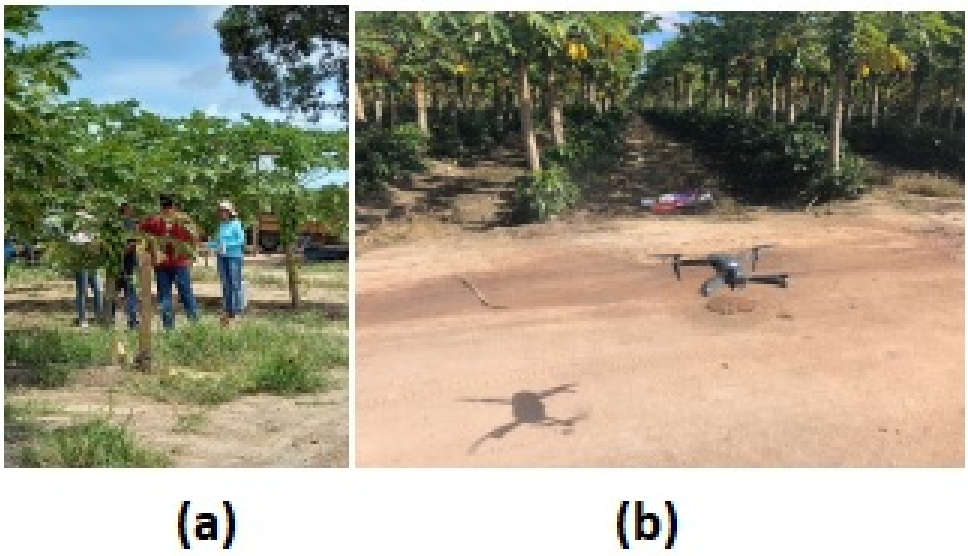


Figura 19 – (a) Agrônomos mapeando mamoeiros com a doença e (b) UAV usado.

Após a captura das imagens, foi gerado um ortomosaico (ou ortofoto) das imagens, o ortomosaico de imagens é uma imagem criada pela junção de várias fotografias aéreas ou imagens capturadas de uma região específica. O objetivo do ortomosaico é criar uma imagem com uma visão geral precisa e detalhada da área, onde cada pixel representa uma localização geográfica específica. Isso é alcançado por meio de técnicas de processamento de imagem e georreferenciamento. A Figura 20, apresenta o ortomosaico da área sobrevo-

ada e realizado a captura das imagens, enquanto a Figura 21 apresenta demarcado com linhas e sombreamentos circulares, no qual, o retângulo representa as fileiras do plantio no qual foi analisado e os sombreamentos apontam os mamoeiros com a possível presença da doença.

Figura 20 – Ortomosaico da área sobrevoada.

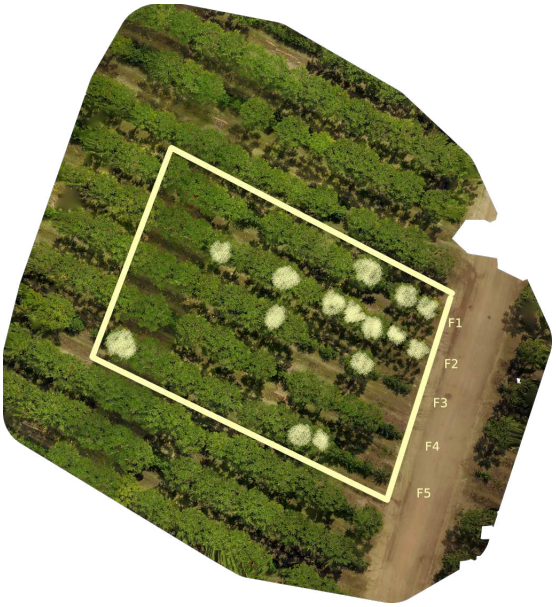


Figura 21 – Ortomosaico editado para demarcar o quadrante analisado na pesquisa e identificar os mamoeiros possivelmente doentes.

A ortofoto permite obter uma visão abrangente e precisa de uma área definida, facilitando as análises detalhadas. As imagens também foram obtidas com condições variadas de iluminação (variação de horário do dia e sombreamento por nuvens), em pelo menos

duas orientações (devido a mapeamento do VANT em “zig-zag”), com variação de posição das folhas devido ao vento, inclusive para o controle da mancha anelar. No entorno das plantas doentes, existem plantas saudáveis e solo. Além disso, inclui as plantas, frutos, folhas caídas, plantas de café em consórcio, arbustos e gramíneas.

4.1.2 Pré-Processamento

A base de dados de imagens aéreas de mamoeiros com sintomas de mancha anelar é composta por 40 imagens RGB com tamanho 4.000 x 2.240 píxeis, definidas pelos agrônomos. As imagens compreendem plantas com sintomas de mancha anelar e seu entorno em duas áreas selecionadas, demarcadas e mapeadas em um pomar comercial de mamoeiro Solo, localizado no sul do estado da Bahia.

A partir das imagens do UAV, foram realizadas as marcações aplicadas às imagens alvo, tais transposições de mapeamentos são realizadas pelos especialistas em solo e em seguida rotuladas em polígonos como apresenta a Figura 22 com forma arbitrária acompanhando o contorno das plantas, usando o software *labelme* exportando as suas rotulações no formato .JavaScript Object Notation (JSON), no qual contém as coordenadas rotuladas dentro da imagens e as suas classes definidas. Tal rotulação é importante para o treinamento dos algoritmos de AM.

Figura 22 – Rotulação em polígonos, feito a partir do software labelme.

Foi criado um único conjunto de dados e este mesmo conjunto foi replicado em pastas separadas como a Figura 23 ilustra. Esta separação é feita para que cada RNC possa gerar o seu resultado independente sem sobre-escrever o conjunto de dados original com as suas predições geradas ao final da execução.

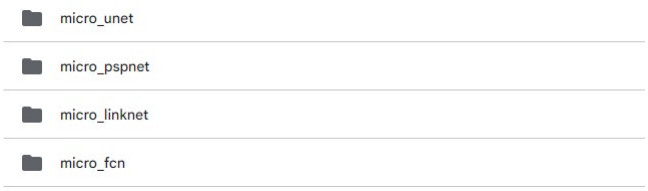


Figura 23 – Divisão das pastas do dataset.

Assim, o conjunto de dados possui a estrutura idêntica contendo 40 imagens no qual foi subdividida, 24 imagens (60%) para treinamento, 16 imagens (40%) para as avaliações semânticas dos resultados (teste), foram consideradas todas as imagens, incluem, além das plantas rotuladas, plantas na margem do mapeamento (MARTINS; VENTURA, 2022).

4.1.3 Arquiteturas

(BALAKRISHNA; DADASHZADEH; SOLTANINEJAD, 2018) propõe a construção da RNC de segmentação com base na U-Net com o codificador VGG16. Neste trabalho foi elaborado a junção VGG16 com as RNC de segmentação (Seção 3.3) para realizar as tarefas de segmentação semântica. A ideia é usar a parte convolucional da VGG16 para extrair as características das imagens e, conectar as camadas adicionais para realizar a segmentação pixel a pixel. A Figura 24 é um exemplo de uma RNC de segmentação sem o uso da VGG16, enquanto a Figura 25 a RNC de FCN é treinada junto da VGG16.

```
def fcn_model(input_shape, num_classes):
    inputs = Input(shape=input_shape)

    # Encoder
    conv1 = Conv2D(64, (3, 3), activation='relu', padding='same')(inputs)
    pool1 = MaxPooling2D((2, 2))(conv1)

    conv2 = Conv2D(128, (3, 3), activation='relu', padding='same')(pool1)
    pool2 = MaxPooling2D((2, 2))(conv2)

    conv3 = Conv2D(256, (3, 3), activation='relu', padding='same')(pool2)
    pool3 = MaxPooling2D((2, 2))(conv3)

    # Middle
    conv4 = Conv2D(512, (3, 3), activation='relu', padding='same')(pool3)

    # Decoder
    up1 = Conv2DTranspose(256, (2, 2), strides=(2, 2), padding='same')(conv4)
    merge1 = concatenate([conv3, up1], axis=-1)
    conv5 = Conv2D(256, (3, 3), activation='relu', padding='same')(merge1)

    up2 = Conv2DTranspose(128, (2, 2), strides=(2, 2), padding='same')(conv5)
    merge2 = concatenate([conv2, up2], axis=-1)
    conv6 = Conv2D(128, (3, 3), activation='relu', padding='same')(merge2)

    up3 = Conv2DTranspose(64, (2, 2), strides=(2, 2), padding='same')(conv6)
    merge3 = concatenate([conv1, up3], axis=-1)
    conv7 = Conv2D(64, (3, 3), activation='relu', padding='same')(merge3)

    # Output
    outputs = Conv2D(num_classes, (1, 1), activation='softmax')(conv7)

    model = Model(inputs=inputs, outputs=outputs)
    return model
```

Figura 24 – Rede FCN sem VGG16.

```
def fcn_vgg16_model(input_shape, num_classes):
    # Load VGG16 model with pre-trained weights (exclude fully connected layers)
    vgg16 = VGG16(include_top=False, weights='imagenet', input_shape=input_shape)

    # Freeze the VGG16 layers so that they are not trained during the FCN training
    for layer in vgg16.layers:
        layer.trainable = False

    # Get the output of the last convolutional block in VGG16
    x = vgg16.get_layer('block5_conv3').output

    # Decoder part of the FCN-VGG16
    x = Conv2D(4096, (7, 7), activation='relu', padding='same')(x)
    x = Conv2D(4096, (1, 1), activation='relu', padding='same')(x)

    # Upsampling layers to increase the resolution of the output
    x = Conv2DTranspose(256, (2, 2), strides=(2, 2), padding='same')(x)
    x = Conv2D(num_classes, (1, 1), activation='softmax')(x)

    model = Model(inputs=vgg16.input, outputs=x)
    return model
```

Figura 25 – Rede FCN com VGG16.

A ideia principal é transformar a arquitetura original da VGG16, que é projetada para classificação, em uma rede totalmente convolucional capaz de fazer segmentação pixel a pixel, usando as redes U-Net, PSPNet, LinkNet e FCN. O padrão de combinação segue em todas as redes e vão ser adaptadas de acordo com a rede, abaixo lista-se os passos que ocorre no processo da transformata da VGG16 com as redes de segmentação:

- ❏ Backbone VGG16: A VGG16 é carregada com os pesos pré-treinados e é utilizada como uma rede de extração de características. No entanto, a camada final de classificação conectada é removida, tornando a VGG16 uma rede convolucional.
- ❏ Encoder: A parte convolucional da VGG16 é usada como um encoder para extrair características da imagem de entrada. À medida que a imagem passa pelas camadas da VGG16, informações contextuais de diferentes níveis de detalhe são aprendidas.
- ❏ Decoder: Após a fase de extração de características (encoder), o decoder é responsável por realizar o upsampling (aumento da resolução) das características para que a saída final corresponda ao tamanho da imagem original.
- ❏ Skip Connections: Para melhorar a precisão da segmentação, as RNC de segmentação usam conexões de salto (skip connections) que combinam características de diferentes camadas do encoder com os mapas de características upsampling do decoder. Essas conexões permitem que a RNC use informações contextuais de diferentes escalas e locais para refinar a segmentação.
- ❏ Output layer:A saída final do decoder é passada por uma camada de convolução com filtro de tamanho $1x1$ e função de ativação softmax. Essa camada produzirá um mapa de probabilidade para cada classe de segmentação, atribuindo a probabilidade de cada pixel pertencer a uma classe.

A combinação do encoder VGG16 com o decoder e as conexões de salto permite que a U-Net, PSPNet, LinkNet e FCN realize a segmentação semântica de imagens, atribuindo a cada pixel uma classe específica com base nas características extraídas. Essa abordagem

possibilita a utilização de uma rede pré-treinada como a VGG16 e a obtenção de resultados precisos em tarefas de segmentação sem a necessidade de treinamento a partir do zero.

4.2 Avaliações

Nesta sub-seção apresenta-se as métricas de validação usadas nesta pesquisa, experimentos feitos e a avaliação dos resultados obtidos com a execução de cada experimento.

4.2.1 Métricas de Validação

Os modelos de redes de segmentação usando a VGG como transferência de aprendizado em 4.1.3 e entendendo cada modelo da seção 3, foi apresentado os parâmetros a serem avaliados a partir dos resultados, dentre eles acurácia e medida F.

(PANTALEÃO; SCOFIELD, 2009) descreve que a acurácia global é uma medida simples que avalia a proporção de pixels corretamente classificados (tanto positivos quanto negativos) em relação ao número total de pixels na imagem de teste. Essa métrica fornece uma visão geral da taxa de acerto da rede de segmentação, mas não leva em consideração informações sobre falsos positivos ou falsos negativos.

A máscara de referência (GT) foi feita manualmente por anotadores humanos e em seguida obtido o conjunto de dados rotulados apresentados na seção 4.1.2, ela define a localização e a forma precisa do objeto de interesse (doença, solo e saudável) no qual foi definido segmentar. A máscara de referência Ground Truth (GT) é comparada com a máscara predita pela rede e em seguida a sobreposição entre as duas máscaras é medida usando o F1-score.

Segunda (CHICCO; JURMAN, 2020) o F1-score é a medida que combina a precisão e a revocação em um único valor para obter uma avaliação geral do desempenho da rede de segmentação. A medida leva em consideração tanto os falsos positivos quanto os falsos negativos e é especialmente útil quando há um desequilíbrio entre as classes positivas e negativas.

Os parâmetros de avaliação seguiram a implementação original de cada rede inserindo a VGG como trasfer learning como apresentado na seção 4.1.3, as redes foram treinadas seguindo o mesmo padrão, este padrão é apresentado na seção seguinte 4.3. O uso do algoritmo de otimização escolhido foi o Stochastic gradient descent (SGD) ((Stochastic Gradient Descent)), no qual utiliza uma taxa de aprendizado fixa e atualiza os pesos com base no gradiente instantâneo.

Cada rede foi treinada sobre duas condições a primeira usando imagens menores que 1600 x 896 pixels e maiores que o valor assim podendo verificar o comportamento da rede, pontuasse também que a acurácia de cada rede foi feita sobre validação, uma vez que a base de dados é desbalanceada em detrimento de classes com menor área marcada, neste caso a doença em relação as demais classes.

4.3 Experimentos

Neste trabalho usou-se um UAV DJI Mavic Pro, um drone no qual possui em seu corpo uma câmera de alta resolução que pode capturar fotos de até 12 megapixels e gravar vídeos em 4K a 30 quadros por segundo acoplada para a captura de imagens, estabilização de imagem, tempo de voo prolongado no qual a bateria do Mavic Pro oferece um tempo de voo de aproximadamente 27 minutos, sistema de evitação de obstáculos possuindo sensores frontais e inferiores que ajudam a evitar colisões com obstáculos e a voar de forma mais segura. Por fim modos de voo inteligentes, o modo ActiveTrack que permite que o drone siga automaticamente um objeto em movimento, e o Waypoint, que permite criar trajetórias de voo pré-definidas.

Para execução dos experimentos foi utilizada a plataforma Google Colab com Graphics Processing Unit (GPU) integrando em PyTorch, para cada rede foi criado um arquivo do Colab dentro da pasta do dataset referida a cada rede (U-Net, PSPNet, LinkNet e FCN). Devido às limitações de memória de GPU e mediante avaliações de desempenho, as imagens tiveram seus tamanhos reduzidos para 1.600 x 896 pixels, como por (MARTINS; VENTURA, 2022).

Como técnica de segmentação, utilizou-se transfer learning para todas as redes com o algoritmo de gradiente descendente estocástico (SGD). Após análise paramétrica de performance, definiu-se taxas de treinamento da base em 0,01 e do núcleo 0,02, velocidade de decaimento 5e-4, momentum 0,9 e decaimento das taxas de treinamento a 20% a cada 30 épocas.(MARTINS; VENTURA, 2022) A base de dados usadas no momento de cada treinamento das redes, seguiu o esquema apresentado na seção 4.1.2.

A taxa de aprendizado e o momento foram definidos arbitrariamente, 0.000001 e 0.3, respectivamente, e a taxa de aprendizado diminuindo em 0.3 a cada 30 épocas. Para a convergência, considerou-se a estabilidade do desempenho do SGD, caracterizado por não mais que 1 (batch_size) de variação na precisão no subconjunto de validação para 30 épocas ou o máximo de 300 épocas totais. Pontuou-se também o pixel por pixel, no qual as classificações de produção para cada imagem são contrastadas com o conhecimento básico do agrônomo marcado e rotulado manualmente como máscaras de polígonos. A precisão é medida como a proporção de pixels corretamente classificados sobre o número total de pixels na imagem.

Capítulo 5

Resultados

Neste capítulo apresenta os resultados obtidos com esta pesquisa. Ela está dividida em avaliação dos resultados apresentando os valores das métricas definidas anteriormente e por fim uma breve conclusão sobre o capítulo.

5.1 Avaliação dos Resultados

Nas condições de desenvolvimento do estudo, observou-se que imagens menores que 1.600 x 896 pixels limitaram o desempenho da técnica, provavelmente não caracterizando suficientemente os sintomas da doença. Por sua vez, imagens maiores requerem maior complexidade dos modelos a serem comparados, sendo mais difícil a convergência.

A primeira condição avaliada foi a acurácia e execução gerada por cada modelo, a acurácia aqui avaliada é a de desempenho da rede neural após a conclusão do treinamento. Ela representa a acurácia alcançada pelo modelo nas imagens de teste ou validação após passar por todo o processo de treinamento ao final é calculada com base nas previsões feitas pelo modelo nessas imagens e na comparação com as máscaras de referência (ground truth).

Quanto à execução, redes de segmentação podem exigir um tempo de processamento significativo, principalmente para imagens de alta resolução. Isso ocorre porque a segmentação envolve a avaliação de cada pixel da imagem, o que pode ser computacionalmente intensivo.

A Tabela 1, apresenta o resultado apresentado em cada rede.

Modelo da rede	**Acurácia**	**Execução**
UNETVGG	0.8806	60.93 minutos
PSPNETVGG	0.8410	53.24 minutos
LINKNETVGG	0.8082	45.67 minutos
FCNVGG	0.8045	57.12 minutos

Tabela 1 – Tabela de acurácia e tempo de execução obtidos em cada modelo treinado.

Outro ponto avaliado no resultado foi o GT (ground truth) e prediction, que estão relacionadas aos conceitos de referência correta e estimativa feita pelo modelo em questão. O "ground truth"é a informação de referência ou os rótulos corretos associados a um conjunto de dados, ele representa a classe correta para cada exemplo do conjunto de dados. Já em tarefas de segmentação de imagem, o "ground truth"é uma máscara binária que indica a localização e a forma corretas do objeto de interesse na imagem. Enquanto a "prediction"é a estimativa ou previsão feita pelo modelo em relação às entradas fornecidas, segmentação, a "prediction"é a máscara gerada pelo modelo, indicando a localização e a forma estimadas do objeto na imagem.

A partir da Tabela 1, notou-se que o modelo U-NetVGG atinge a melhor precisão. Através dos resultados obtidos separadamente observou o comportamento do GT (ground truth) das imagens de cada modelo de rede, no qual sofreu alteração se comparado com a rotulação original. A Figura 26 apresenta os resultados do GT em relação a acurácia e é possível observa que quando menor a precisão da rede a identificação do GT acaba sendo afetada, a imagem original (a) apresenta vegetação rasteira junta ao solo e restante os mamoeiros, na imagem (b) rotulado temos em amarelo o que denominamos solo, ou seja, vegetação rasteira como caráter de solo, e em verde claro plantas saudáveis, aqui nessa ilustração não apresenta-se doença, a ideia aqui no GT é entender como as redes estão identificando vegetações rasteiras e planta (mamoeiro). Nota-se então que em (c) e (d) temos pouca identificação no solo como vegetação, em vermelho temos a identificação como solo e azul planta. Em (e) e (f) apresenta planta em locais de solo, porém observa-se em (a) que esses locais são apenas solo, vegetações rasteiras e frutos, sendo assim (e) e (f) tem pouca distinção de solo e planta saudável.

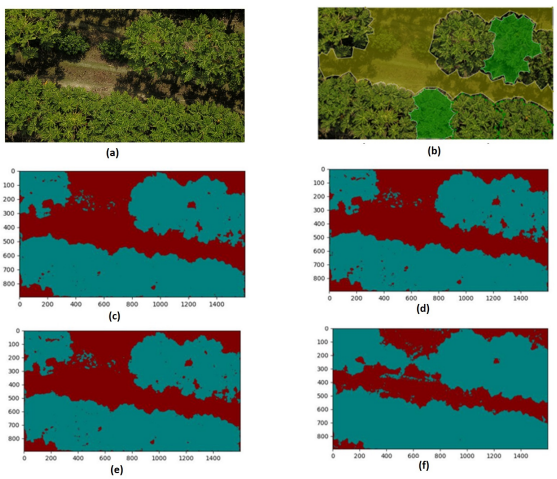


Figura 26 – Comparação entre os resultados de ground truth das redes treinadas. (a) imagem original; (b) rotulação da imagem; (c) resultado ground truth UNETVGG; (d) resultado ground truth PSPNETVGG; (e) resultado ground truth LINKNETVGG; (f) resultado ground truth FCNVGG.

Os resultados de prediction dos modelos treinados apresenta comparações significativas entre as redes, vide, a Figura 27 apresenta resultados gerados de cada modelo de rede e apresenta a imagem original, rotulação indicando em amarelo solo, vermelho plantas com a doença e verde plantas saudáveis, as imagens referidas dos resultados das redes, como identificado em legenda apresenta variações em seus resultados. Figura 27 apresenta em (a) imagem original não rotulada, (b) com as rotulações feita pelo software labelme no qual amarelo representa solo, verde saudável e vermelho doença, nota-se que não foi feito rotulação em toda imagem, somente em pontos escolhidos por especialistas, as imagens subsequentes (c), (d), (e) e (f) são os modelos de redes treinados, notou-se que o comportamento de coloração das imagens foi diferente em alguns modelos. Em (c) uma coloração verde fluorescente identifica a doença enquanto o verde não fluorescentes plantas saudáveis e vermelho solo. As imagens (d), (e) e (f) apresenta solo como vermelho e a um verde cintilante identificando como doença e o verde como saúdavel. A seguir pontua-se cada modelo observado. O modelo UNETVGG é maior ao centro das imagens, ocorrendo falsos positivos mais frequentemente em plantas nas extremidades das imagens, a PSPNETVGG apresenta comportamento em relação a coloração das folhas ela identifica como doença plantas com folhagem mais amareladas, no qual é um possível sintoma da doença, porém pode ser interpretada como folhas já antigas em processo de troca. A LINKNETVGG tem um contorno maior nas bordas identificando como doença e não ocorre como na UNETVGG um foco maior no centro das imagens, notavelmente é possível ver que a rede deu ênfase ao solo desta forma os resultados obtidos detectam os espaços existentes entre as folhas como solo, por fim, a FCNVGG apresentou baixos índices de presença de doença e tratou todo e qualquer tipo de vegetação seja rasteira em solo quanto mamoeiro como saudável.

A partir da prediction notou-se que nas redes com acurácia abaixo de 0.81, a vegetação rasteira que não são plantio de mamoeiro é confundido com solo, a Figura 28 apresenta essas inconsistências, no qual é possível visualizar que nas redes (a) UNETVGG a vegetação rasteira é totalmente identificada como solo, bem como frutos e folhas que estão juntos do solo, na (b) PSPNETVGG apresenta pequenas marcações em vegetações e frutos presente no solo, o modelo LINKNETVGG contorna e preenche todo e qualquer tipo de vegetação seja rasteira ou mamoeiro e identifica como saudável, por fim a (d) FCNVGG contorna e preenche como saudável locais de solo sem vegetação, ou seja, mesmo não contendo nenhum tipo de fruto, folha ou vegetação ela identifica partes do solo como plantas saudáveis.

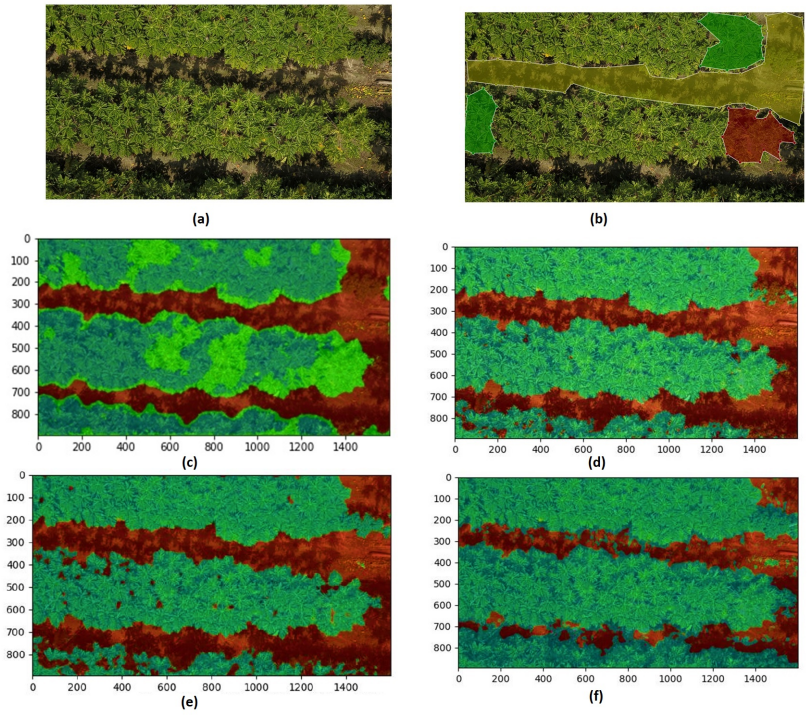


Figura 27 – C- Comparação entre os resultados de prediction das redes treinadas. (a) imagem original; (b) rotulação da imagem; (c) resultado prediction UNETVGG; (d) resultado prediction PSPNETVGG; (e) resultado prediction LINKNETVGG; (f) resultado prediction FCNVGG.

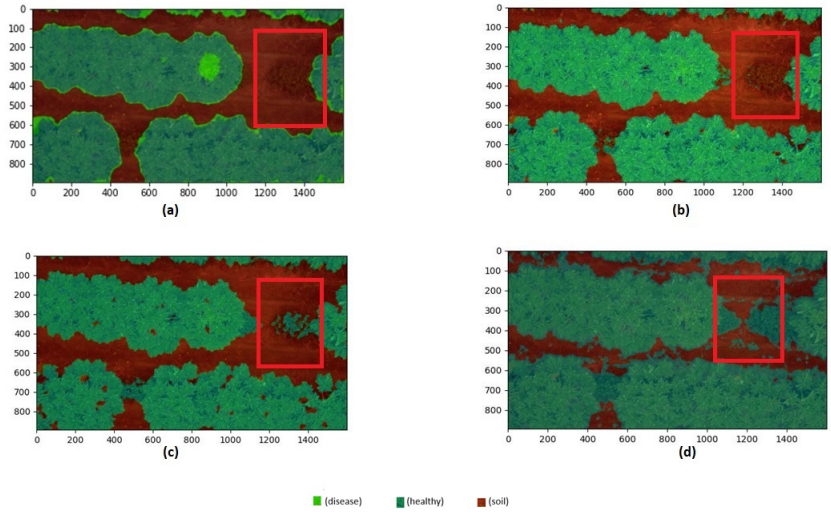


Figura 28 – Quadrado vermelho destaca vegetação rasteira no solo, (a) UNETVGG segue o modelo de rotulação e considera vegetação rasteira como solo; (b) PSPNETVGG apresenta pequenos pontos na vegetação rasteira confundindo com solo; (c) LINKNETVGG faz o contorno de toda vegetação rasteira destacando como planta saudável; (d) FCNVGG destaca até mesmo solo como planta saudável.

O F1-score também foi utilizado para avaliar o desempenho de redes de segmentação e é calculado a partir da combinação da precisão e da revocação. A precisão (ou acurácia) é a proporção de pixels corretamente classificados como parte do objeto de interesse em relação ao total de pixels classificados como parte do objeto de interesse pela rede (True Positive (TP) + False Positive (FP)) e a revocação (ou sensibilidade) é a proporção de pixels corretamente classificados como parte do objeto de interesse em relação ao total de pixels que deveriam ter sido classificados como parte do objeto de interesse, de acordo com a máscara de referência (TP + FN). Essa métrica combina a precisão e a revocação em um único valor, levando em consideração tanto os falsos positivos (FP) quanto os falsos negativos (FN). E é apresentado pela formula:

$$F_1 = 2\frac{precision.sensitivity}{precision.sensitivity} = \frac{2TP}{2TP + FP + TN} \tag{5}$$

A Tabela 2 apresenta a métrica F1-score de cada modelo treinado usando a fórmula padrão de cálculo:

Modelo da rede	**Acurácia**	**F1-Score**
UNETVGG	0.880	0.89
PSPNETVGG	0.841	0.85
LINKNETVGG	0.808	0.80
FCNVGG	0.804	0.78

Tabela 2 – Tabela de acurácia e F1-Score obtidos em cada modelo treinado.

Por fim com o objetivo de avaliar a eficácia das diferentes técnicas, é necessário comparar os resultados alcançados levando como consideração a detecção da doença mancha anelar do mamoeiro, a Figura 29 evidencia as diferenças entre as redes treinadas. A rede (a) UNETVGG destaca a doença em locais com maior concentração da doença, ressaltando o meio da planta onde ocorre uma redução do tamanho das folhas e o amarelamento, na (b) PSPNETVGG ela salienta como doença folhas com coloração amarela, no qual nem sempre indica sintomas da doença, a (c) LINKNETVGG salienta o meio da planta porém sem um destaque de coloração intenso para separação de saudáveis e doentes, por fim a (d) FCNVGG tem o contorno das plantas como saudável e a uma área significativa de folhagem como doente.

Mesmo todas as redes sendo de segmentação e tendo os mesmos parâmetros de treino e validação, observou-se que o comportamento para expressar a doença segue padrões distintos variando entre meio do mamoeiro, bordas e solo.

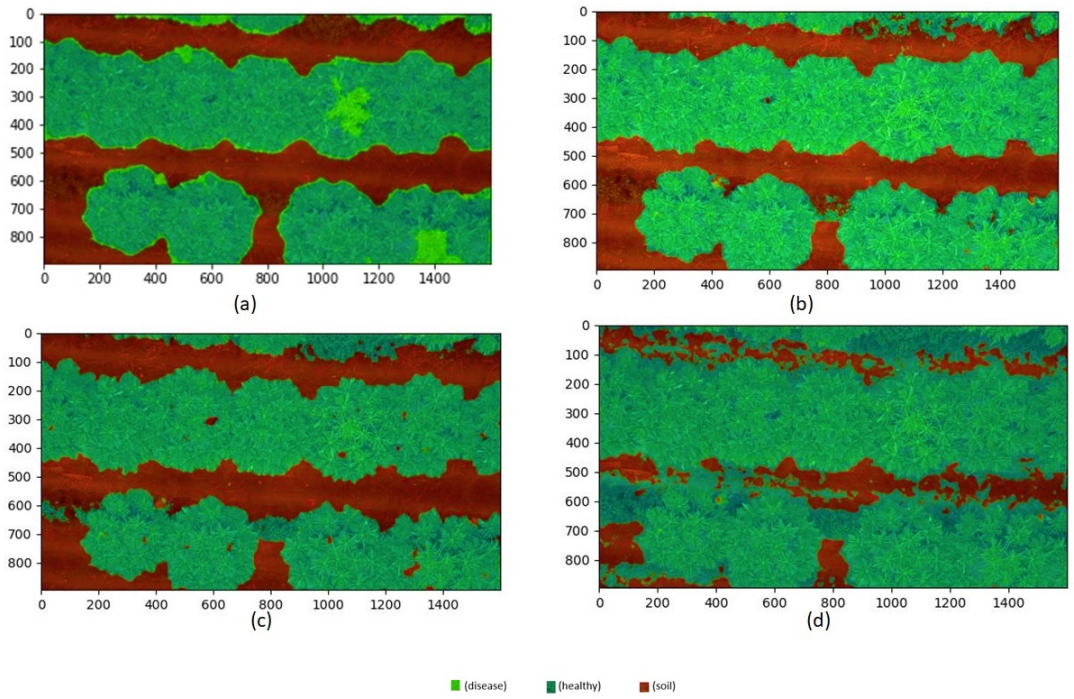


Figura 29 – Resultado de saída das redes de segmentação. (a) U-Net, (b) PSPNet, (c) Linknet e (d) FCN

5.1.1 Conclusão dos resultados

Neste trabalho foi demarcado uma área e um pequeno banco de imagens foi gerado a partir desta área, identificar o comportamento e resultados das redes de segmentação escolhidas, na identificação da doença do mamoeiro. Foi observado nos resultados que as redes teve uma acurácia e F1-Score significativas, porém não conclusiva uma vez que existe falsos positivos e áreas não demarcadas, como doença e locais identificados indevidos como plantas no local de solo e plantas saudáveis como doentes.

Capítulo 6

Conclusão

Neste capítulo, abordam-se as conclusões decorrentes da execução deste projeto, além de discorrer sobre os ganhos alcançados e os obstáculos enfrentados. Também se destaca as contribuições alcançada e possibilidades de trabalhos futuros.

6.1 Conclusões

Os resultados apontam para a viabilidade do uso da UNETVGG e PSPNETVGG na segmentação semântica de plantas de mamoeiro com sintomas de mancha anelar. Bons resultados visuais e numéricos (±84% à 88% acurácia nas imagens de teste) foram alcançados mobilizando poucas imagens de tamanho moderado (1.600 x 896 píxeis). Considerando o custo computacional do treinamento, o uso do modelo e os tamanhos dos arquivos resultantes, acredita-se que a manutenção do treinamento do modelo em processamento massivo, pode ser viável em sua integração (após treinado) em aplicação online a campo agrícola. Quanto as redes LINKNETVGG e FCNVGG a partir dos resultados é possível notar que as redes precisam de um maior refinamento, ou seja, um melhor treinamento para que elas possam ter uma acurácia similar as outras apresentadas.

É importante destacar que a detecção da doença mancha anelar é feita praticamente toda manual, demandando tempo, mão de obra especializada, recursos financeiros e veículos terrestres. Desta maneira, poluindo o meio ambiente. Esta dissertação mostrou a proposta com o uso da IA e VANTs, para agilizar a detecção da doença mancha anelar.

É importante destacar que durante o desenvolvimento do trabalho, ocorreu a combinação das redes de segmentação combinadas com a VGG16, problemas na plataforma de execução Google Colab, sendo necessário a aquisição da versão PRO, pois a versão free, possui limitações no qual foi incapaz de rodar as redes. Dificuldades na parceria com algumas fazendas para registro de imagem, deslocamento para as áreas para serem coletadas. Devido a esses contratempos, algumas etapas do projeto foram se estendendo e levando um tempo maior para sua conclusão.

Como trabalho futuro, ainda serão realizados estudos comparativos com outros modelos de redes de segmentação de imagens. Projeta-se a captura de imagens usando bandas espectrais e dados tridimensionais e realizar treinamento em cima desse tipo de imagem. O desenvolvimento de aplicação online para campo, tentando uma parceria com produtores. Na ocorrência de falsos positivos ou falsos negativos, o usuário poderá disponibilizar a imagem capturada e sugestão de classificação correta, possibilitando a ampliação da base de dados e o refinamento do modelo. Esta aplicação processa em tempo real e detecta a doença sem necessidade de extrair as imagens para o computador e fazer o processamento das imagens.

6.2 Contribuições na Produção

O trabalho pretende contribuir para o meio científico, disponibilizando as imagens coletadas e os códigos. Pretende-se ainda dentro do projeto a visita e coleta de imagens em outras fazendas da região do sul da Bahia, assim ao ser finalizado as visitas, gerar um conjunto de dados maior e disponibilizar aos usuários.

Referências

ADAM, K. D. B. J. et al. A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, v. 1412, 2014.

ANTUNES, M. A. H.; DEBIASI, P.; SIQUEIRA, J. d. S. Avaliação espectral e geométrica das imagens rapideye e seu potencial para o mapeamento e monitoramento agrícola e ambiental. **Revista Brasileira de Cartografia**, v. 66, n. 1, p. 105–113, 2014.

BALAKRISHNA, C.; DADASHZADEH, S.; SOLTANINEJAD, S. Automatic detection of lumen and media in the ivus images using u-net with vgg16 encoder. **arXiv preprint arXiv:1806.07554**, 2018.

BRIDI, S. et al. Substratos orgânicos no desenvolvimento de plântulas de mamoeiro cv. formosa mel. 2023.

CARVALHO, O. L. F. d. Deep learning & remote sensing: pushing the frontiers in image segmentation. 2022.

CHENG, H.-D. et al. Color image segmentation: advances and prospects. **Pattern recognition**, Elsevier, v. 34, n. 12, p. 2259–2281, 2001.

CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. **BMC genomics**, BioMed Central, v. 21, n. 1, p. 1–13, 2020.

CRISPI, G. M. Uso de redes neurais profundas na avaliação do ataque de mosca minadora ao tomateiro. Universidade Federal de Viçosa, 2022.

DADWAL, M.; BANGA, V. K. Estimate ripeness level of fruits using rgb color space and fuzzy logic technique. **International Journal of Engineering and Advanced Technology**, Citeseer, v. 2, n. 1, p. 225–229, 2012.

DAMACENO, S. S.; VASCONCELOS, R. O. et al. Inteligência artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular. **Caderno de Graduação-Ciências Exatas e Tecnológicas-UNIT-SERGIPE**, v. 5, n. 1, p. 11–11, 2018.

DENG, L.; YU, D. et al. Deep learning: methods and applications. **Foundations and trends® in signal processing**, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.

ESCALONA, U. et al. Fully convolutional networks for automatic pavement crack segmentation. **Computación y Sistemas**, Instituto Politécnico Nacional, Centro de Investigación en Computación, v. 23, n. 2, p. 451–460, 2019.

FIGUEIRA, N. M.; OLIVEIRA, L. C. d. Super-resolução: técnicas existentes e possibilidade de emprego às imagens do vant vt-15. **Revista Militar de Ciência e Tecnologia**, v. 30, p. 3–19, 2013.

GALEANO, E.; MARTINS, D. d. S. Evolução da produção e comércio mundial de mamão. In: SIMPÓSIO DO PAPAYA BRASILEIRO, 6., 2015, Vitória, ES. Tecnologia de ..., 2015.

GHAREHCHOPOGH, F. S. Neural networks application in software cost estimation: A case study. In: IEEE. **2011 International Symposium on Innovations in Intelligent Systems and Applications**. [S.l.], 2011. p. 69–73.

GOLLAPUDI, S. **Practical machine learning**. [S.l.]: Packt Publishing Ltd, 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.

HAYKIN, S. **Kalman filtering and neural networks**. [S.l.]: John Wiley & Sons, 2004.

LEAL, A. B. Criação de uma infraestrutura para o desenvolvimento de pesquisas sobre o uso de vants em aplicações de interesse para o estado de santa catarina. 2018.

LIMA, L. P. de; AQUINO, E. L. de C.; MILL, D. A influência dos algoritmos inteligentes no processo de aprendizagem autônoma. **Revista Eletrônica Internacional de Economia Política da Informação, da Comunicação e da Cultura (ISSN: 1518-2487)**, v. 22, n. 2, p. 128–147, 2020.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 3431–3440.

MALSBURG, C. V. D. Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. In: SPRINGER. **Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1–4, 1984**. [S.l.], 1986. p. 245–248.

MARTINS, D. dos S.; VENTURA, J. A. Papaya brasil. 2022.

MARTINS, M. et al. Viabilidade do uso de veículos aéreos não tripulados pela polícia militar de santa catarina no 19º bpm. Araranguá, SC, 2017.

MASOUD, K. M.; PERSELLO, C.; TOLPEKIN, V. A. Delineation of agricultural field boundaries from sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. **Remote sensing**, MDPI, v. 12, n. 1, p. 59, 2019.

MENDES, J. P. C. et al. Segmentação automática de lesões de mama em imagens de ultrassom utilizando redes neurais convolutivas. Universidade Federal do Amazonas, 2018.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: **Proceedings of the 27th international conference on machine learning (ICML-10)**. [S.l.: s.n.], 2010. p. 807–814.

NORONHA, V. G.; SILVA, J. C. da. A study of neural networks models applied to natural language inference.

OLIVEIRA, A.; FILHO, H. S.; FILHO, P. M. Manejo de doenças do mamoeiro. In: SIMPÓSIO DO PAPAYA BRASILEIRO, 5., 2011, Porto Seguro. Inovação e . . . , 2011.

OLIVEIRA, A.; FILHO, P. M. Mamoeiro do grupo solo: cultivo, colheita, pós-colheita e comercialização. Brasília, DF: Embrapa, 2022., 2022.

OLIVEIRA, A. d. J. et al. Método automático para detecção de nematóides em lavoura cafeeira usando imagens aéreas. Universidade Federal de Uberlândia, 2019.

PANTALEÃO, E.; SCOFIELD, G. B. Comparação entre medidas de acurácia de classificação para imagens do satélite alos. **XIV Simpósio Bras Sensoriamento Remoto [Internet].[accessed 2021 Mar 28]**, p. 7039–7046, 2009.

PARKER, J. R. **Algorithms for image processing and computer vision**. [S.l.]: John Wiley & Sons, 2010.

RAMASAMY, G.; SINGH, T.; YUAN, X. Multi-modal semantic segmentation model using encoder based link-net architecture for brats 2020 challenge. **Procedia Computer Science**, Elsevier, v. 218, p. 732–740, 2023.

RIBEIRO, A. S. et al. Reconhecimento de nervuras de folhas em plantas utilizando redes neurais convolucionais. Instituto Federal Goiano, 2019.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18**. [S.l.], 2015. p. 234–241.

SANTOS, I. A. D. Detecção de invasões biológicas no cerrado utilizando deep learning. Universidade Federal de São Carlos, 2019.

SANTOS, T. C. dos; ESPERIDIÃO, T. L.; AMARANTE, M. dos S. Agricultura 4.0. **Revista Pesquisa e Ação**, v. 5, n. 4, p. 122–131, 2019.

SILVA, E. M. D. C. da; DALMOLIN, Q. Avaliação da resolução de imagens digitais para cadastro. 1998.

SILVA, H. C. V. d. **Redes neurais de convolução na classificação de edifícios em imagens de alta resolução espacial**. Tese (Doutorado), 2022.

SILVA, I. R. et al. Utilização de redes convolucionais para classificação e diagnóstico da doença de alzheimer. **II Simpósio de Inovação em Engenharia Biomédica**, p. 73–76, 2018.

SINGH, A. K. et al. Deep learning for plant stress phenotyping: trends and future perspectives. **Trends in plant science**, Elsevier, v. 23, n. 10, p. 883–898, 2018.

SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

THECKEDATH, D.; SEDAMKAR, R. Detecting affect states using vgg16, resnet50 and se-resnet50 networks. **SN Computer Science**, Springer, v. 1, p. 1–7, 2020.

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the xxix conference on graphics, patterns and images**. [S.l.], 2016. v. 1, n. 4.

VENTURA, J. A.; FERREGUETTI, G.; MARTINS, D. d. S. Eficiência do roguing como estratégia de manejo da meleira e mosaico do mamoeiro. In: SIMPÓSIO DO PAPAYA BRASILEIRO, 6., 2015, Vitória, ES. Tecnologia de ..., 2015.

WOODS, R. E.; GONZALEZ, R. C. **Digital image processing**. [S.l.]: Pearson Education Ltd., 2008.

XING, Y.; ZHONG, L.; ZHONG, X. An encoder-decoder network based fcn architecture for semantic segmentation. **Wireless Communications and Mobile Computing**, Hindawi, v. 2020, 2020.

YANI, M.; IRAWAN S, S. M. B.; SETININGSIH ST, M. C. Application of transfer learning using convolutional neural network method for early detection of terry's nail. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2019. v. 1201, n. 1, p. 012052.

ZHANG, Z. Improved adam optimizer for deep neural networks. In: IEEE. **2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)**. [S.l.], 2018. p. 1–2.

ZHU, X. et al. Coronary angiography image segmentation based on pspnet. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 200, p. 105897, 2021.