

Segmentação de doenças em plantas usando U-Net

Gabrielle Lira Dantas Wanderley, Henrique Santana Diniz, Joaquim de Moura Thomaz Neto,
Maria Eduarda Chrispim Santana, Saulo Pereira Vasconcelos

Instituto de Ciências Exatas e Tecnologia, Pontifícia Universidade Católica de Minas Gerais
Belo Horizonte, Brasil

{gabrielle.lira, hsdiniz01}@sga.pucminas.br, jota45@outlook.com,
mariae.chrispims@gmail.com, saulovasconcelos07@gmail.com

Abstract—A rede neural convolucional U-Net tem como objetivo a segmentação precisa de imagens de alta resolução, sendo bastante utilizada na área médica para identificar tumores, células e órgãos. Devido à sua precisão, essa rede também se mostra eficaz para identificar áreas afetadas por doenças em folhas de plantas, o que permite agir mais cedo para evitar prejuízos na produção agrícola. Utilizando a base de dados *PlantVillage*, o modelo foi treinado e avaliado com foco em precisão pixel a pixel. A versão modificada da rede apresentou desempenho superior em métricas como IoU e BCE Loss, demonstrando como ferramentas computacionais servem de suporte à agricultura de precisão.

Index Terms—machine learning, U-Net, segmentação de imagens, plantas, doenças, redes neurais

I. INTRODUÇÃO

Machine learning é uma técnica que ensina sistemas a aprender com base em dados, utilizando algoritmos para analisar informações e identificar padrões. Nos últimos anos, o avanço dessas técnicas tem proporcionado soluções eficazes para problemas complexos em diversas áreas. Entre elas, destacam-se as redes neurais convolucionais (Convolutional Neural Networks – CNNs), utilizadas para analisar e interpretar imagens, realizando tarefas como identificação, localização e segmentação de objetos. Essas redes conseguem reconhecer formas e texturas em imagens, sendo ideais para a análise automática de imagens com muitos detalhes. A segmentação de imagens consiste em identificar e separar diferentes regiões dentro de uma imagem em nível de pixel, o que permite uma análise detalhada e precisa das áreas de interesse. Essa técnica é fundamental para aplicações que exigem localização exata de objetos ou padrões visuais.

Dentre os modelos baseados em CNN, um que se destaca nesse contexto é a U-Net, proposta por Ronneberger et al. (2015) para segmentação de imagens biomédicas. Sua arquitetura em formato de "U", composta por etapas de contração e expansão conectadas por skip connections, permite realizar segmentações precisas em nível de pixel, preservando detalhes importantes da imagem original. Devido à sua alta precisão, a U-Net tem sido utilizada em aplicações como segmentação de tumores, detecção de vasos sanguíneos e identificação de estruturas celulares.

Essa capacidade de segmentação detalhada pode ser explorada em outras áreas, como na identificação de doenças em plantas, uma vez que essas doenças geralmente se manifestam por meio de padrões visuais específicos em folhas, como manchas, descolorações ou alterações na textura — características

que exigem precisão para serem corretamente detectadas e diferenciadas de áreas saudáveis. No caso da detecção de doenças em plantas, a segmentação detalhada permite não apenas reconhecer se uma folha está doente, mas também identificar com precisão quais partes dela estão afetadas, possibilitando um diagnóstico mais completo e eficaz. Diferentemente de técnicas de classificação simples, que indicam apenas a presença ou ausência da doença, a segmentação ajuda a analisar a região das infecções, o que é essencial para o controle e prevenção da disseminação.

Doenças em plantas podem ameaçar toda a produção agrícola, pois, quando não identificadas a tempo, tendem a se espalhar rapidamente e afetar grandes áreas. A detecção em estágio inicial dessas doenças é fundamental para evitar que se espalhem rapidamente, o que pode comprometer colheitas inteiras e afetar a economia dos produtores. No entanto, a identificação manual nem sempre é eficaz, por estar sujeita a erros, demandar tempo e depender de conhecimento especializado. Por isso, torna-se necessário o uso de abordagens automatizadas. Nesse sentido, o uso da U-Net para segmentar folhas doentes permite diferenciar com precisão as regiões afetadas das saudáveis, auxiliando no monitoramento das plantas e na prevenção do avanço das doenças.

Este trabalho propõe o uso da arquitetura U-Net para segmentar imagens de folhas doentes, com o objetivo de identificar, de forma precisa e automatizada, os padrões visuais associados a infecções.

II. METODOLOGIA

O método adotado neste trabalho baseou-se no uso de um subconjunto de imagens inspirado na base *PlantVillage*, simulando um cenário real de folhas doentes. As imagens foram organizadas em duas pastas principais: uma contendo imagens RGB de folhas saudáveis ou afetadas por doenças, e outra contendo máscaras binárias feitas manualmente, nas quais os pixels com valor 255 representam regiões doentes e os pixels com valor 0 representam áreas saudáveis. Para padronização, todas as imagens foram redimensionadas para 256×256 pixels.

Durante o pré-processamento, as imagens RGB foram normalizadas para o intervalo $[-1, 1]$ e convertidas em tensores utilizando a biblioteca *Torchvision*. As máscaras, por sua vez, foram convertidas para escala de cinza e binarizadas com um limiar de 0.5, resultando em valores de 0 e 1.

O carregamento dos dados foi gerenciado por uma classe personalizada de *Dataset*, implementada em *PyTorch* sob o nome *LeafDataset*. Essa classe foi responsável por associar corretamente cada imagem RGB à sua máscara binária correspondente, garantindo integridade nos pares. O *DataLoader* foi utilizado para fornecer os dados em mini-lotes (*batches*) de tamanho 4, com embaralhamento automático a cada época para promover maior generalização durante o treinamento.

A. Arquiteturas Utilizadas e Justificativa

Para resolver o problema de segmentação, foram comparadas três arquiteturas convolucionais distintas: Autoencoder, U-Net e ResUNet. A escolha dessas redes visa avaliar diferentes níveis de complexidade arquitetural e sua influência no desempenho da segmentação de folhas doentes.

O **Autoencoder convolucional** foi adotado como base-line. Trata-se de uma rede simétrica composta por blocos de codificação e decodificação, mas sem conexões de atalho entre eles. Essa simplicidade permite avaliar o desempenho mínimo esperado em tarefas de reconstrução semântica, ainda que apresente limitações na recuperação de detalhes espaciais, especialmente nas bordas das regiões segmentadas.

A **U-Net**, originalmente proposta para segmentação biomédica, foi escolhida por sua capacidade de preservar detalhes importantes da imagem por meio de suas *skip connections*, que ligam diretamente os blocos de codificação aos blocos de decodificação. Isso permite recuperar informações perdidas durante a compressão, tornando a U-Net especialmente eficaz em tarefas que exigem precisão na localização de regiões específicas, como as áreas doentes nas folhas.

Por fim, a **ResUNet** combina a estrutura da U-Net com blocos residuais, inspirados na ResNet. Esses blocos introduzem conexões de identidade que facilitam o treinamento de redes profundas, mitigando problemas como o desaparecimento do gradiente. Além disso, os blocos residuais tornam o modelo mais robusto e preciso na identificação de contornos e regiões de difícil distinção, características comuns em manchas de doenças.

As três arquiteturas foram implementadas de forma modular e treinadas sob as mesmas condições experimentais, o que permite uma comparação justa de desempenho.

III. EXPERIMENTOS E RESULTADOS

A. Configuração Experimental

Os experimentos foram realizados utilizando a base de dados PlantVillage, restrita à classe *Tomato - Late Blight*. Para garantir comparação justa, 50 imagens foram separadas para treinamento e validação, todas com máscaras binárias anotadas manualmente.

O pipeline de pré-processamento incluiu redimensionamento das imagens para 256×256 pixels, normalização para o intervalo $[0, 1]$, e uso de uma classe personalizada em *PyTorch* (*LeafDataset*) para garantir consistência no carregamento das imagens e máscaras.

Os três modelos – Autoencoder convolucional, U-Net clássica e ResUNet – foram treinados sob as mesmas condições: 50 épocas, tamanho de lote 4, otimizador Adam ($\text{lr} = 1 \times 10^{-4}$) e função de perda *Binary Cross Entropy* (BCELoss). A cada época, exemplos de predições foram salvos para análise qualitativa.

B. Exemplo de Entrada e Máscara de Referência

A Figura 3 ilustra um par de dados utilizados no experimento: à esquerda, a imagem original da folha de tomate afetada pela doença; à direita, a máscara binária criada manualmente, onde as regiões brancas representam áreas anotadas como doentes. Esse processo de anotação serviu de base para o treinamento e avaliação dos modelos de segmentação.



Fig. 1. *
(a) Imagem original

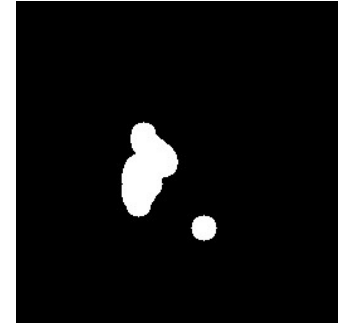


Fig. 2. *
(b) Máscara de referência

Fig. 3. Exemplo de par de dados utilizado: (a) folha de tomateira afetada, (b) máscara binária indicando a região doente (em branco).

C. Métricas Quantitativas

A Tabela I apresenta as métricas de desempenho das três arquiteturas. A avaliação quantitativa do desempenho dos modelos foi feita por meio de duas métricas: a função de perda *Binary Cross Entropy* (BCE Loss) e a métrica *Intersection over Union* (IoU), também conhecida como Jaccard Index. A BCE Loss mede a diferença entre os valores preditos e os valores reais de cada pixel, sendo apropriada para tarefas de segmentação binária. Valores menores indicam predições mais próximas da máscara real. Já a IoU é definida como a razão entre a interseção e a união dos pixels preditos como positivos e os pixels positivos reais: $\text{IoU} = \frac{A \cap B}{A \cup B}$, onde A representa a predição binária e B a máscara real. Quanto mais próximo de 1, melhor o desempenho. A IoU é amplamente utilizada em tarefas de segmentação por refletir a sobreposição entre as regiões previstas e as reais..

TABLE I
DESEMPENHO FINAL DOS MODELOS APÓS 50 ÉPOCAS

Modelo	BCE Loss (ép. 50)	IoU Médio (ép. 50)
Autoencoder	0.2143	0.3652
U-Net	0.2859	0.7019
ResUNet	0.0760	0.6892

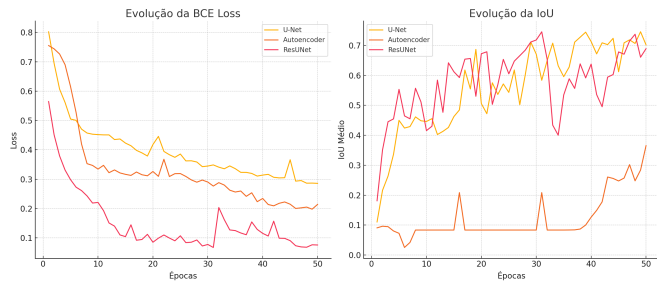


Fig. 4. Evolução da BCE Loss (à esquerda) e da IoU (à direita) ao longo das épocas de treinamento para os três modelos.

A análise dos valores revela que a ResUNet alcançou a menor BCE Loss, indicando grande proximidade entre as segmentações previstas e as máscaras reais. Apesar disso, o maior valor de IoU final foi obtido pela U-Net, o que sugere que, ao final do treinamento, ela apresentou uma ligeiramente melhor sobreposição com as regiões reais da doença, mesmo que suas predições fossem, em termos de valor absoluto, um pouco menos confiáveis.

O Autoencoder, como esperado por sua simplicidade estrutural e ausência de mecanismos de recuperação espacial (como skip connections ou blocos residuais), apresentou o pior desempenho nas duas métricas. Isso reforça seu papel como baseline e evidencia sua limitação em tarefas de segmentação de alta precisão.

O gráfico da Figura 4 corrobora esses achados, demonstrando que o Autoencoder manteve valores de IoU abaixo de 0.4 durante praticamente todo o treinamento, enquanto U-Net e ResUNet alcançaram picos superiores a 0.70. No entanto, é notável que a curva da ResUNet apresenta maior oscilação, sugerindo sensibilidade maior à variação dos dados ou a possíveis instabilidades durante o aprendizado.

Esses resultados indicam que, embora a ResUNet seja superior em minimizar a perda global, a U-Net foi capaz de generalizar ligeiramente melhor no critério de sobreposição, possivelmente devido à preservação mais eficaz da estrutura espacial da folha.

D. Análise Qualitativa

A Figura 8 apresenta as predições geradas por cada uma das três arquiteturas testadas, com o objetivo de comparar visualmente a qualidade da segmentação produzida. As diferenças entre as abordagens são notáveis tanto na definição das áreas afetadas quanto na separação entre regiões saudáveis e doentes.

O modelo Autoencoder, por ser o mais simples, apresenta predições visivelmente mais difusas, com baixa definição de bordas e pouca distinção entre a área da folha e o fundo. Isso é consequência direta da ausência de conexões de atalho (skip connections) e da compactação extrema na codificação, que reduz significativamente a capacidade da rede de preservar detalhes espaciais durante a reconstrução da saída.

A predição gerada pela ResUNet demonstra avanços significativos em relação ao Autoencoder, especialmente na clareza

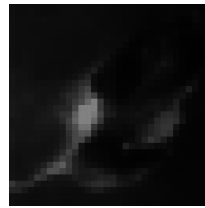


Fig. 5. *
(a) Autoencoder

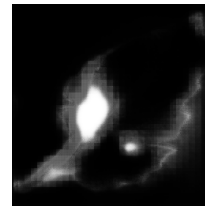


Fig. 6. *
(b) ResUNet

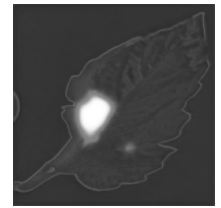


Fig. 7. *
(c) U-Net

Fig. 8. Máscaras preditas pelos modelos: (a) Autoencoder, (b) ResUNet, (c) U-Net.

das regiões classificadas como doentes. Observa-se um fundo praticamente preto e homogêneo, o que sugere maior confiança do modelo ao identificar as regiões saudáveis, ou seja, os pixels onde não há doença. No entanto, a silhueta da folha é menos evidente do que na U-Net, o que pode indicar uma menor preservação de detalhes estruturais finos.

Por sua vez, a U-Net clássica apresentou resultados visualmente mais equilibrados. A silhueta da folha é nitidamente visível, e as regiões classificadas como afetadas coincidem, em boa parte, com a máscara de referência. Esse desempenho se deve à presença das conexões de atalho entre as fases de codificação e decodificação, que permitem a recuperação mais precisa de padrões espaciais e de contorno ao longo da imagem. A U-Net também mantém melhor contraste entre áreas segmentadas e não segmentadas, o que contribui para uma visualização mais interpretável dos resultados.

Em resumo, a ResUNet demonstra maior precisão negativa (confiança nas áreas não afetadas), enquanto a U-Net se sobressai em precisão positiva e riqueza de detalhes. O Autoencoder, apesar de captar padrões gerais, não se mostra adequado para segmentações exigentes em precisão morfológica.

E. Discussão dos Resultados

A análise dos resultados evidencia o impacto das escolhas arquiteturais. O Autoencoder, usado como baseline, mostrou-se limitado para segmentação precisa, sendo útil apenas como ponto de partida. A U-Net, por sua vez, equilibra simplicidade e boa performance, enquanto a ResUNet se destaca como a abordagem mais robusta e estável, especialmente para delimitação de contornos.

Visualmente, a ResUNet apresenta melhor separação entre regiões doentes e saudáveis, corroborando os valores superiores de IoU e BCE Loss. Em aplicações práticas, essa diferença pode resultar em diagnósticos mais confiáveis e intervenções mais assertivas na agricultura de precisão.

F. Limitações e Próximos Passos

Apesar dos bons resultados, o tamanho reduzido do conjunto de dados anotados limita a generalização dos modelos. Para trabalhos futuros, pretende-se ampliar o número de amostras anotadas, explorar outras métricas (como *f1-score* e precisão/recall pixel a pixel), bem como testar arquiteturas com mecanismos de atenção ou imagens multispectrais.

IV. CONCLUSÃO

Este trabalho propôs e avaliou o uso de três arquiteturas convolucionais (Autoencoder, U-Net e ResUNet) para a tarefa de segmentação de áreas afetadas por doenças em folhas de tomate, utilizando imagens simuladas baseadas no padrão da base PlantVillage. A partir da comparação entre as arquiteturas, foi possível verificar o impacto direto da complexidade e profundidade dos modelos no desempenho das segmentações.

Os resultados quantitativos evidenciam que a ResUNet apresentou os menores valores de *Binary Cross Entropy Loss* e IoU mais estáveis ao longo do treinamento, destacando-se especialmente nas primeiras épocas. Essa performance inicial pode ser atribuída à presença dos blocos residuais, que favorecem a propagação do gradiente e aceleram a convergência do modelo. Entretanto, a U-Net demonstrou uma evolução mais constante da métrica IoU ao longo do tempo, ultrapassando a ResUNet nos estágios finais. Isso sugere que, embora a ResUNet seja mais eficiente em termos de aprendizado rápido, a U-Net pode ser mais robusta em relação à estabilidade e generalização, especialmente quando o volume de dados é limitado.

O Autoencoder, utilizado como baseline, demonstrou claras limitações. Sua estrutura mais simples, sem *skip connections* ou blocos residuais, resultou em uma capacidade reduzida de capturar os padrões espaciais necessários para segmentações precisas. A métrica IoU permaneceu baixa mesmo após 50 épocas de treinamento, evidenciando sua ineficácia para essa aplicação específica.

Na análise qualitativa, observou-se que as predições da U-Net mantêm maior fidelidade à silhueta da folha e conseguem representar com mais precisão as regiões afetadas. A ResUNet, por sua vez, se destacou por produzir máscaras com fundo mais escuro e menos ruído, o que sugere maior confiança na exclusão de regiões saudáveis. Ambas as arquiteturas superaram o Autoencoder tanto em precisão visual quanto em definição de bordas.

Portanto, conclui-se que a escolha da arquitetura tem impacto significativo na qualidade das segmentações. A U-Net apresenta um equilíbrio entre simplicidade estrutural e desempenho eficaz, sendo uma opção viável para aplicações com recursos computacionais moderados. Já a ResUNet, embora mais complexa, é altamente promissora em cenários que exigem alta acurácia e respostas mais rápidas.

Entre as limitações do trabalho, destaca-se o uso de um conjunto de dados relativamente pequeno e de anotações manuais. Esse fator pode ter afetado a capacidade de generalização dos modelos, especialmente nas regiões mais sutis das imagens. Além disso, os experimentos foram conduzidos com apenas uma classe de doença, o que restringe a aplicabilidade dos resultados a outros contextos agrícolas.

Como trabalho futuro, propõe-se a ampliação da base de dados, inclusão de múltiplas culturas e tipos de doenças, uso de imagens multiespectrais e investigação de arquiteturas modernas com mecanismos de atenção, como UNet++ e Attention U-Net. Além disso, a adoção de estratégias de

aumento de dados (data augmentation) e validação cruzada podem contribuir para melhorar a robustez dos modelos. Espera-se, assim, avançar no desenvolvimento de ferramentas computacionais confiáveis e eficientes para a agricultura de precisão, contribuindo diretamente para a sustentabilidade e produtividade no setor agropecuário.

REFERENCES

- [1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.
- [2] Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.
- [3] Paszke, A., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*.
- [4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *CVPR*.