v7.n5.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

MODELO DE APRENDIZADO PROFUNDO EM SEGMENTAÇÃO DE IMAGENS APLICADO A FOLHAS DE SOJA

DEEP LEARNING MODEL IN IMAGE SEGMENTATION APPLIED TO SOYBEAN LEAVES

MODELO DE APRENDIZAJE PROFUNDO PARA LA SEGMENTACIÓN DE IMÁGENES APLICADO A HOJAS DE SOJA

Maicon Aparecido Sartin[1], Benevid Felix da Silva[2], Ivan Luiz Pedroso Pires[3], Silvio Cesar Garcia Granja[4]

e757929
https://doi.org/10.47820/recima21.v7i5.7929

PUBLICADO: 05/2026

RESUMO

A segmentação de imagens é uma etapa de extrema relevância no processamento de imagens. Essa etapa define regiões de interesse em imagens para facilitar a identificação de objetos e o reconhecimento de padrões em imagens. Métodos tradicionais de segmentação de imagens são muito sensíveis a variação de ambiente e luminosidade, com isso os modelos de Aprendizado Profundo (*Deep Learning*) consistem em técnicas modernas de processamento de imagens para resolver tais problemas. As redes neurais convolucionais têm evoluído e se estabelecido como uma das grandes promessas na área de processamento de imagens baseada em Aprendizado Profundo. Neste trabalho, investigamos e modificamos uma rede neural deconvolucional com o objetivo de segmentar imagens em folhas de soja. Esta pesquisa propõe uma arquitetura de aprendizado profundo otimizada para a segmentação de folhas de soja com baixo custo computacional. Por meio de uma metodologia aplicada, quantitativa e uma configuração experimental, a proposta tem sua avaliação e comparação com modelos tradicionais e outras redes neurais convolucionais consolidadas. A validação utiliza métricas estatísticas e testes de estresse com ruídos para comprovar a robustez e a precisão da proposta. Os resultados são comparados com diversos modelos para a tarefa de segmentação de imagens. O desempenho foi avaliado pelas métricas de *Dice*, *Recall* e *Specificity*. A abordagem proposta alcançou valores promissores de acurácia acima de 95% em todos os *datasets* de teste, mesmo com inserção de alterações nas imagens.

PALAVRAS-CHAVE: Redes neurais convolucionais. Folhas de soja. Segmentação. Aprendizado profundo. Imagens.

ABSTRACT

Image segmentation is a highly relevant step in image processing. This step defines regions of interest in images to facilitate object identification and pattern recognition. Traditional image segmentation methods are highly sensitive to environmental and lighting variations. Consequently, Deep Learning models offer modern image processing techniques to address these issues.

[1] Doutor em Engenharia Elétrica e mestre em Ciência da Computação, graduado em Engenharia da Computação. Professor adjunto na Universidade do Estado de Mato Grosso (UNEMAT).
[2] Doutor e mestre em Ciência da Computação, graduado em Licenciatura em Computação. Professor adjunto na Universidade do Estado de Mato Grosso (UNEMAT).
[3] Doutor e mestre em Ciência da Computação, graduado em Licenciatura em Computação. Professor adjunto na Universidade do Estado de Mato Grosso (UNEMAT).
[4] Doutor em Engenharia Elétrica e mestre em Física, graduado em Física. Professor adjunto na Universidade do Estado de Mato Grosso (UNEMAT).

Convolutional neural networks have evolved and established themselves as one of the great promises in the field of Deep Learning-based image processing. In this work, we investigate and modify the deconvolutional neural network with the objective of segmenting images of soybean leaves. This research proposes a deep learning architecture optimized for soybean leaf segmentation with low computational cost. Through an applied, quantitative methodology and an experimental setup, the proposal is evaluated and compared with traditional models and other established convolutional neural networks. Validation uses statistical metrics and noise stress tests to prove the robustness and accuracy of the proposal. The results are compared with several traditional methods and with traditional supervised machine learning for the image segmentation task. Performance was evaluated using the Dice, Recall, and Specificity metrics. The proposed approach achieved promising accuracy values above 95% in all test datasets, even with the insertion of alterations to the images.

KEYWORDS: *Convolutional neural network. Soybean leaves. Segmentation. Deep learning. Images.*

RESUMEN

La segmentación de imágenes es un paso fundamental en el procesamiento de imágenes. Este paso define regiones de interés para facilitar la identificación de objetos y el reconocimiento de patrones. Los métodos tradicionales de segmentación son muy sensibles a las variaciones ambientales y de iluminación. Por consiguiente, los modelos de aprendizaje profundo ofrecen técnicas modernas de procesamiento de imágenes para abordar estos problemas. Las redes neuronales convolucionales han evolucionado y se han consolidado como una de las grandes promesas en el campo del procesamiento de imágenes basado en aprendizaje profundo. En este trabajo, investigamos y modificamos la red neuronal deconvolucional con el objetivo de segmentar imágenes de hojas de soja. Esta investigación propone una arquitectura de aprendizaje profundo optimizada para la segmentación de hojas de soja con bajo costo computacional. Mediante una metodología cuantitativa aplicada y una configuración experimental, la propuesta se evalúa y compara con modelos tradicionales y otras redes neuronales convolucionales establecidas. La validación utiliza métricas estadísticas y pruebas de estrés de ruido para demostrar la robustez y precisión de la propuesta. Los resultados se comparan con varios métodos tradicionales y con el aprendizaje automático supervisado tradicional para la tarea de segmentación de imágenes. El rendimiento se evaluó utilizando las métricas de Dice, Recall y Specificity. El enfoque propuesto alcanzó valores de precisión prometedores, superiores al 95%, en todos los conjuntos de datos de prueba, incluso con la inserción de alteraciones en las imágenes.

PALABRAS CLAVE*: Red neuronal convolucional. Hojas de soja. Segmentación. Aprendizaje profundo. Imágenes.*

INTRODUÇÃO

A segmentação de imagens é a base para muitos sistemas de processamento de imagens, devido ao rearranjo de uma imagem destacando uma ou mais regiões de interesse. Tais regiões são definidas conforme um determinado padrão entre seus pixels como cor, intensidade, textura, frequência, entre outros. Diversos métodos efetuam o processo de segmentação por meio de distintas características da imagem, pode-se citar alguns tipos como: detecção de bordas, limiarização, crescimento de região, agrupamento, variação do gradiente, entre outros. Métodos

tradicionais de segmentação de imagens geralmente se baseiam em poucas características das imagens, tornando-se frágeis em uma variação ambiental real. Em muitos casos, a diferença de iluminação do ambiente pode ser um fator crítico na segmentação, assim, a área de interesse da imagem é comprometida.

Na última década, o uso das Redes Neurais Convolucionais (RNC), na visão computacional, tem se tornado alvo de diversos pesquisadores pelos excelentes resultados de precisão obtidos em distintas áreas do conhecimento. No passado, o uso de RNCs estava particularmente relacionado a classificação de imagens. Atualmente, nota-se uma grande abrangência de tarefas visuais, destaca-se as tarefas relacionadas com conteúdo nas imagens como a análise artística de obras (Gatys *et al.*, 2016), a detecção de anomalias (Singh *et al.* 2020), a detecção (Tuggener *et al.*, 2018) e a localização (Johnson *et al.* 2016) de objetos, a segmentação semântica (Papandreou *et al.*, 2015) e a reconstrução de imagens baseadas em conteúdo (Rejeb *et al.*, 2017).

As Redes Neurais Convolucionais caracterizam-se pela disposição de seus neurônios em uma estrutura volumétrica tridimensional, composta pelas dimensões de largura, altura e profundidade (canais). Uma definição convencional da arquitetura da RNC é baseada em três tipos de camadas, a convolucional, a pool e a totalmente conectada. Dependendo da tarefa ou problema, diversos tipos e quantidades de camadas são utilizadas. Existem arquiteturas que realizam o processo inverso, como nas redes neurais deconvolucionais (DeconvNet) (Zeiler & Fergus, 2014; Long *et al.*, 2015; Noh *et al.*, 2015; Chen *et al.*, 2017; Badrinarayanan *et al.*, 2017). O processo convolucional é relacionado a contração da imagem definindo os seus hiperparâmetros e o deconvolucional é a expansão desses dados, conforme ilustrada na Figura 1, destacando sua região de interesse (Ronneberger *et al.,* 2015).

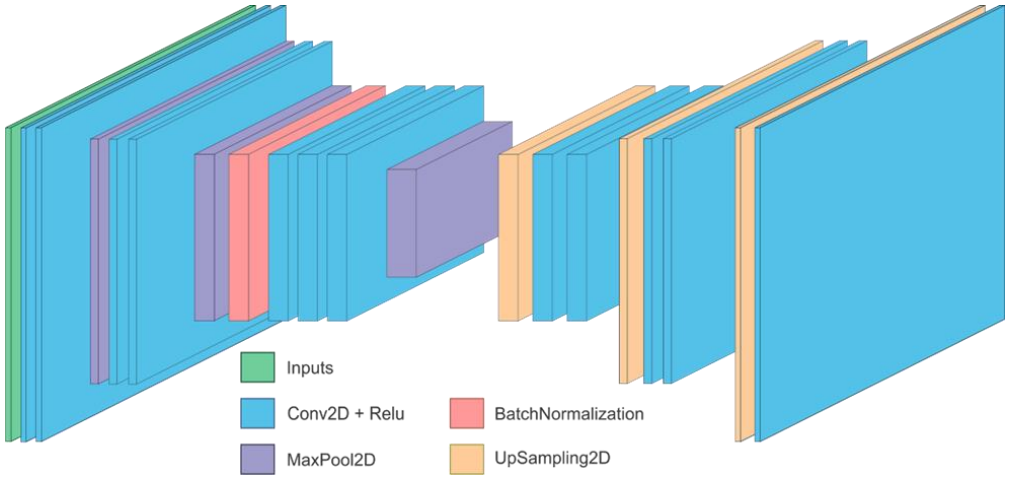
Além das DeconvNet, diversos tipos de modelos baseados em aprendizado profundo são utilizados na segmentação e na segmentação semântica, como as redes neurais artificiais (Chithambaram & Perumal, 2017), as redes recorrentes (Ren & Zemel, 2016), as redes adversárias generativas (Isola *et al.* 2017), as redes de memória de curto prazo (Chen *et al.*, 2015), entre outras. A grande quantidade de modelos de processamento de imagens traz um alto número de possibilidades para a análise e a avaliação da segmentação em imagens nas mais diversas áreas e aplicações.

Em aplicações como a fenotipagem, as folhas das plantas devido a tensões bióticas e abióticas podem ocorrer alterações na reflexão de luz no objeto, alterando a percepção de cores, ou seja, ocorrem mudanças na fluorescência de clorofila (Lin *et al.*, 2019). Tais alterações na folha podem ser associadas com alguma emergência na plantação como na deficiência de

macronutrientes, nas doenças ou nas pragas. Na segmentação de imagens é necessária a generalização do modelo para assimilar essas diferenças e considerá-las como pertencentes a região de interesse, neste caso, considera-se a folha como o objeto alvo.

As redes convolucionais e deconvolucionais consolidaram-se como arquiteturas de alto desempenho no processamento digital de imagens, apresentando aplicabilidade multidisciplinar em diversas áreas do conhecimento devido a flexibilidade na reconfiguração das arquiteturas. Diversos modelos de aprendizado profundo trabalham em tarefas baseadas em fenotipagem de plantas sobre diferentes espécies baseado em visão computacional. Destaca-se três tipos de tarefas, a classificação (Dyrmann, 2016; Sartin *et al.*, 2020), a detecção de objetos (Lin *et al.*, 2019), a segmentação para evidenciar regiões de interesse (Noh *et al.*, 2015; Scharr *et al.*, 2016) e a quantificação dessas regiões (Aich & Stavness, 2017). Dentre todas as tarefas, a segmentação e a segmentação semântica têm uma função especial, por ter característica de tarefa meio e as demais tarefas tornam-se dependentes dela.

Figura 1. Rede Neural Deconvolucional



Fonte: Elaboração Própria.

Este artigo investiga o problema de segmentação de imagens RGB de folhas de soja. A proposta baseia-se em uma abordagem baseada em redes deconvolucionais junto às redes totalmente conectadas com adição de máscara binária. Utilizou-se um *dataset* próprio para esta tarefa em diferentes configurações de imagens e modelos de segmentação de imagens. A seção 1 apresenta a metodologia, as ferramentas e as imagens utilizadas no trabalho. Os resultados e a comparação com diversos modelos de segmentação estão definidos na seção 2. A conclusão do trabalho e possibilidades futuras estão na seção 3.

1. MATERIAIS E MÉTODOS

Esta pesquisa trabalha com uma metodologia de natureza aplicada e quantitativa, além de uma configuração experimental. O objeto de pesquisa está na segmentação de folhas de soja com base em imagens de condições reais. O intuito é desenvolver uma ferramenta com base em aprendizado profundo e custo computacional reduzido para superar as limitações de métodos tradicionais. A pesquisa tem uma estratégia comparativa entre vários modelos de aprendizado profundo e métodos tradicionais de visão computacional. Na validação da proposta apresenta-se uma grande variação de ruídos em imagens e métricas estatísticas para avaliação, comparação e análise de robustez dos modelos.

1.1. Trabalhos relacionados

Na área de IA existem diversas ramificações dentre as suas subáreas, pode-se destacar quatro principais divisões baseadas no tipo de aprendizado: supervisionado, não supervisionado, aprendizado por reforço e aprendizado evolucionário (Marsland, 2015). Em (Deng & Yu, 2014), definem o aprendizado profundo em três classes: Redes profundas para aprendizado não supervisionado ou generativo, Redes profundas para aprendizado supervisionado, Redes profundas híbridas.

As Redes Neurais Convolucionais estabeleceram-se como arquiteturas fundamentais no desenvolvimento e na consolidação do aprendizado profundo, principalmente no reconhecimento e na detecção de objetos. Diversas arquiteturas de RNCs foram construídas com esse objetivo, pode-se destacar as seguintes redes: a LeNet (LeCun *et al.*, 1998), a AlexNet (Krizhevsky *et al.*, 2012), a VGG (Simonyan & Zisserman, 2014), a GoogleNet (Szegedy *et al.*, 2015) e a Inception-ResNet (Szegedy *et al.,* 2017). Essas arquiteturas proporcionaram um grande aumento no número de pesquisas relacionadas ao aprendizado profundo, além de diversos desafios e competições das melhores arquiteturas como: Imagenet, ISBI, Doenças cardíacas, DengAI etc.

As redes profundas para aprendizado supervisionado têm adquirido grande interesse da área de visão computacional. Mesmo com o aumento do número de camadas (dezenas), neurônios (centenas de milhares) e parâmetros (pesos e bias) (centenas de milhões) na arquitetura de uma RNC, diversos trabalhos associam-nas com outras técnicas de classificação como Máquinas de vetores de suporte em (Singh *et al.*, 2020). Geralmente, tais trabalhos utilizam técnicas de *fine tunning* e *learning transfer* para utilizar algumas das arquiteturas citadas. Porém,

nos trabalhos de (Ronneberger *et al.,* 2015; Milletari *et al.,* 2016) podemos observar uma tendência na área de segmentação de imagens.

A arquitetura U-Net (Ronneberger *et al.,* 2015) foi construída para segmentação de estruturas neuronais em pilhas microscópicas eletrônicas e foi a vencedora do desafio de rastreamento de células ISBI 2015. É constituída de uma organização de contração para capturar o contexto e um caminho de expansão simétrico que permite a localização precisa. Em uma RNC tradicional, é usada a técnica de contração, e a expansão é processo contrário. O processo de descida contração (*downsampling*) e de subida expansão (*upsampling*) são computados na rede, além disso, existem mapas de características do processo de contração, que são copiados e concatenados com as camadas de expansão. Os autores comentam ainda sobre o bom desempenho alcançado com imagens de dimensão 512 x 512 e a relevância no uso da técnica de *augmentation* no treinamento. A rede atingiu 92% de precisão e desempenho na segmentação menor que 1 segundo com uma GPU. Outra arquitetura baseada em contração e expansão é a V-Net (Milletari *et al*., 2016), arquitetura semelhante a U-Net. A arquitetura é refinada para o processo de segmentação volumétrica em 3D de exames de próstata em imagens de ressonância magnética. Considerando para cada conjunto convolucional (contração) tem ao menos 4 camadas, duas convolucionais, uma maxpooling e uma Relu.

Existe nas duas redes (U-Net e V-Net) a mesma quantidade de conjuntos convolucionais (contração) e de-convolucionais (expansão), 4 conjuntos. Obviamente, as arquiteturas tem diversas diferenças como: quantidades de filtros, tamanhos de máscaras e existência de outras camadas (Normalização e *Dropout*), etc. A precisão alcançada na segmentação foi de 87% e desempenho de 1 segundo com múltiplas GPUs. Em (Tran, 2016), o autor propõe uma arquitetura de rede completamente convolucional para o problema da segmentação automatizada do ventrículo esquerdo e direito. A arquitetura é adaptada por meio das técnicas de *fine tunning* e *transfer learning*. Como a U-Net, V-Net e a proposta deste artigo, tem o treinamento de ponta a ponta em um único estágio de aprendizado. Nesse caso, a parte de expansão não tem a mesma quantidade de conjunto de camadas como as demais, no artigo não comenta sobre esse conjunto de camadas e a Figura da arquitetura não deixa clara essa informação. Porém, mesmo sem essa parte da arquitetura, demonstra a viabilidade do aprendizado de ponta a ponta em estruturas anatômicas, evidenciando a precisão de 86 obtida com um caminho de expansão menos denso em relação a redes consagradas como a U-Net.

1.2. Modelo proposto

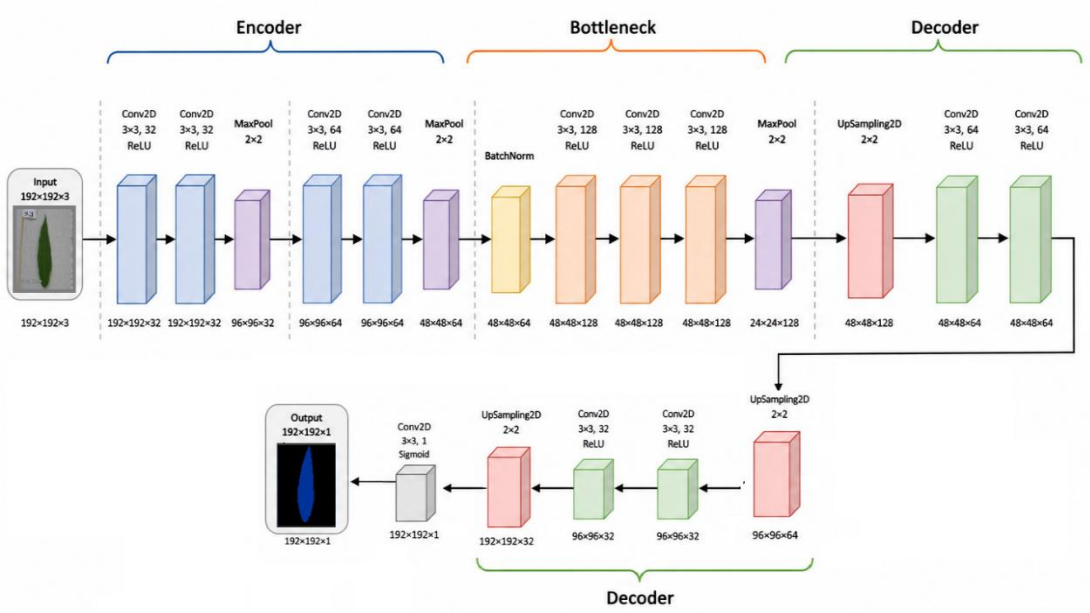
Algoritmos tradicionais em condições de variação de cor, nitidez, contraste e brilho falham na segmentação de imagens para ambientes reais. Embora arquiteturas como a U-Net demonstrem robustez em ambientes reais, sua complexidade e o custo operacional podem ser restritivos no processamento de determinadas aplicações. Nesse sentido, a utilização de uma versão modificada justifica-se pela necessidade de viabilizar uma arquitetura otimizada e com menor custo computacional. Este trabalho, se baseia em um conceito otimizado da U-Net para segmentação das imagens de folhas. Semelhante a V-Net, a proposta contém menos filtros nas camadas convolucionais e o número de camadas foi reduzido.

A arquitetura proposta no presente artigo contém 3 conjuntos de convolucionais (contração) e 3 conjuntos de deconvolucionais (expansão), como pode ser observado na Figura 2. Adicionalmente aos estágios convolucionais e deconvolucionais, a arquitetura proposta integra camadas totalmente conectadas (*Fully Connected layers*), fundamentadas no modelo de *Perceptron* Multicamadas (MLP), para a síntese de características no nível mais profundo da rede, conforme Figura 3. A proposta se diferencia da V-Net, U-Net e SegNet (Badrinarayanan *et al.,* 2017) em diversas caraterísticas como: otimização de camadas e conexões, normalização, organização convolucional e deconvolucional, reaproveitamento do resultado da DeconvNet em uma arquitetura totalmente conectada, etc.

As redes profundas para aprendizado supervisionado têm adquirido grande interesse da área de visão computacional. A arquitetura U-Net (Ronneberger *et al.*, 2015) foi construída para segmentação de estruturas neuronais em pilhas microscópicas eletrônicas. É constituída de uma organização com os processos convolucional e deconvolucional para a segmentação de imagens e foi uma das pioneiras na área. No auxílio a segmentação de imagens é comum o uso de características complementares as imagens como a caixa de delimitação, centroides, pontos chave e máscaras. A adição de uma máscara binária para complementar o processo de aprendizagem torna-se relevante na delimitação do objeto alvo (He *et al.*, 2017).

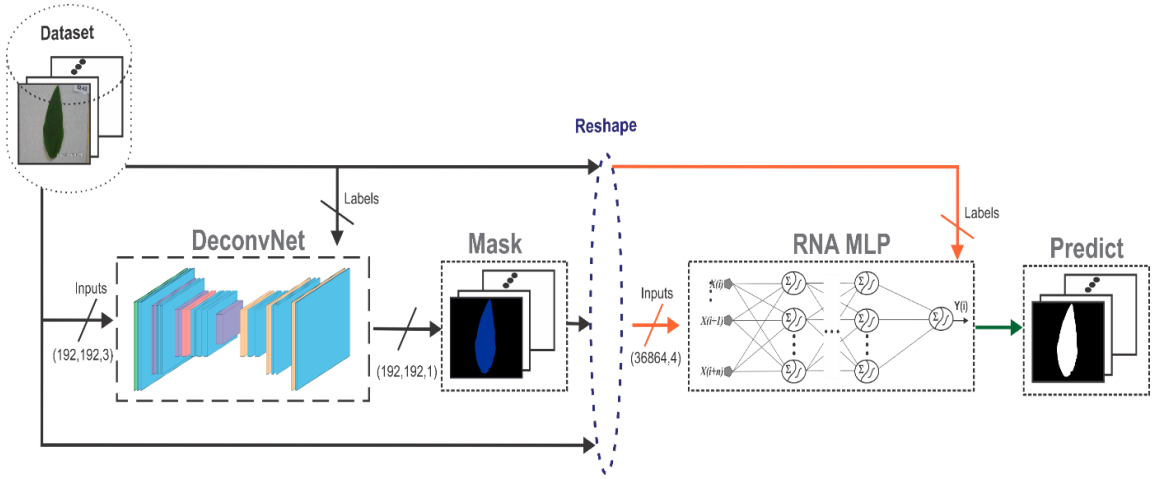
A construção da proposta baseia-se em três características básicas: junção de modelos (Aich & Stavness, 2017), redes deconvolucionais (Ronneberger *et al.*, 2015), adição de máscaras binárias (He *et al.,* 2017). Desta forma, utilizou-se uma rede deconvolucional simplificada para obtenção da região de interesse e uma rede totalmente conectada para o aprimoramento da acurácia do modelo, com a adição de uma máscara binária gerada pela primeira rede, conforme Figura 3.

Figura 2. Arquitetura DeconvNet proposta. A rede possui três etapas de codificação *(encoder),* um gargalo *(Bottleneck)* e três etapas de decodificação *(decoder)*



Fonte: Elaborada pelo autor com auxílio de inteligência artificial (OpenAI, 2026).

Figura 3. Organização da arquitetura DeconvNet proposta completa em conjunto com Rede Neural Artificial e adição de máscara binária



Fonte: Elaboração Própria.

A proposta criada tem uma arquitetura conforme a Figura 1, com três conjuntos convolucionais *(Encoder)* e deconvolucionais *(decoder).* Cada conjunto convolucional contém duas camadas convolucionais, junto com a função de ativação “Relu”, e uma camada de contração para cada conjunto convolucional. Adicionou-se apenas uma camada para normalização dos

dados. As camadas deconvolucionais têm duas diferenças gerais: na substituição da camada convolucional pela de deconvolucional e a última camada ser uma convolucional. A segmentação semântica é realizada com cada imagem contendo apenas uma folha de soja, conforme ilustrado na Figura 3.

Os três conjuntos convolucionais tem respectivamente 32, 64 e 128 filtros. O aumento da quantidade de filtros e de conjuntos convolucionais não trouxeram aprimoramentos relevantes nos resultados de precisão. Assim, priorizou-se uma arquitetura reduzida de parâmetros como um todo. A RNA *Perceptron* Multicamadas (MLP) contém 5 camadas, uma de entrada, três escondidas e uma de saída. Para o desenvolvimento de todo o trabalho utilizou-se linguagem *python* e a biblioteca *tensorflow*.

1.3. Conjunto de Dados (*Dataset*)

O *dataset* foi criado a partir de 32 imagens sementes próprias, coletadas em outro trabalho do grupo de pesquisa (Sartin, 2014; Sartin *et al*., 2022). Dessas 32 imagens de folhas de soja foram geradas mais 1600 imagens, total do *dataset* de 1632 imagens. A técnica de aumento de dados teve alteração em seis características das imagens: giro, rotação, cor, nitidez, contraste e brilho. O giro foi realizado apenas em 0º e 180º e a rotação de 90º em 90º graus. As quatro últimas características foram modificadas em quatro valores: 50%, 75%, 125% e 150%.

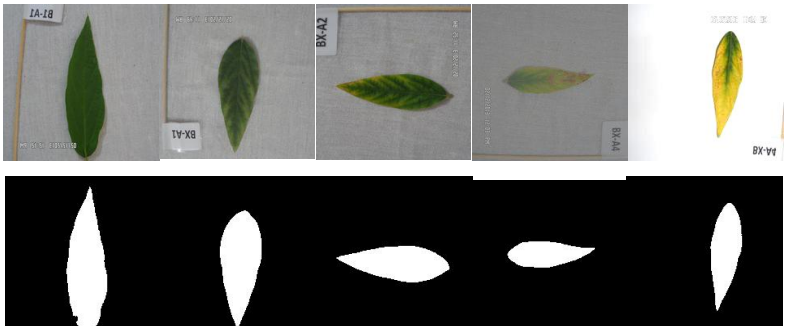
Outras 96 imagens (*labels*) serviram de máscara para o treinamento e a validação do modelo, devido ao giro e rotação realizado. A etapa de rotulagem das imagens consistiu na geração manual de máscaras binárias auxiliado por editor de imagem (*Corel Draw* - varinha mágica). A segmentação pixel-a-pixel foi realizada para cada imagem do *dataset*, o processo foi supervisionado para minimizar variações de sombreamento e ruídos de oclusão, a auto-oclusão não foi alterada. O critério de seleção das imagens sementes baseou-se na diversidade de condições de iluminância da folha da soja. As regiões correspondentes às folhas de soja foram rotuladas com valor unitário (1), enquanto o fundo e demais elementos foram atribuídos ao valor nulo (0), pós-normalização.

Na Figura 4 ilustra exemplos de imagens adquiridas em laboratório com uma câmera Samsung de 14 MP. Na primeira Linha tem as imagens originais e seus rótulos (*labels*) na segunda. As imagens originais permaneceram com datas e identificações para averiguar a robustez do método.

Todas as imagens (entrada e máscaras) foram mantidas com a mesma resolução das imagens originais, assegurando a integridade espacial dos dados durante os treinamentos dos modelos. Todas as imagens foram redimensionadas para a resolução 192x192 pixels, na leitura e

na manipulação das imagens utilizaram as bibliotecas PIL e *skimage* com antialias habilitado. as imagens de entrada no modelo de cores RGB, já as imagens *labels* tem apenas um canal. A resolução foi avaliada com outras quantidades, escolheu-se o pior caso dentre os avaliados, devido a relação entre resolução e filtros encontrada em outros trabalhos (Sartin *et al.*, 2020).

Figura 4. Conjunto de Dados de folhas de soja: Imagens originais (1º linha) e *labels* (2º linha)



Fonte: Elaboração Própria.

2. RESULTADOS E DISCUSSÃO

A metodologia experimental utilizada para a análise dos resultados está dividida em três partes a serem apresentadas, o treinamento dos modelos, as métricas de avaliação e os resultados alcançados. A arquitetura proposta deste trabalho foi comparada com mais 5 modelos de segmentação de imagens: (i) Um modelo com a alteração na DeconvNet (Pix2Pix), (ii) um modelo baseado na U-Net, (iii) uma RNA MLP e dois modelos tradicionais: (iv) Morphological Snakes (MS) e o (v) Gradiente Gaussiano Inverso (GGI).

2.1. Treinamento

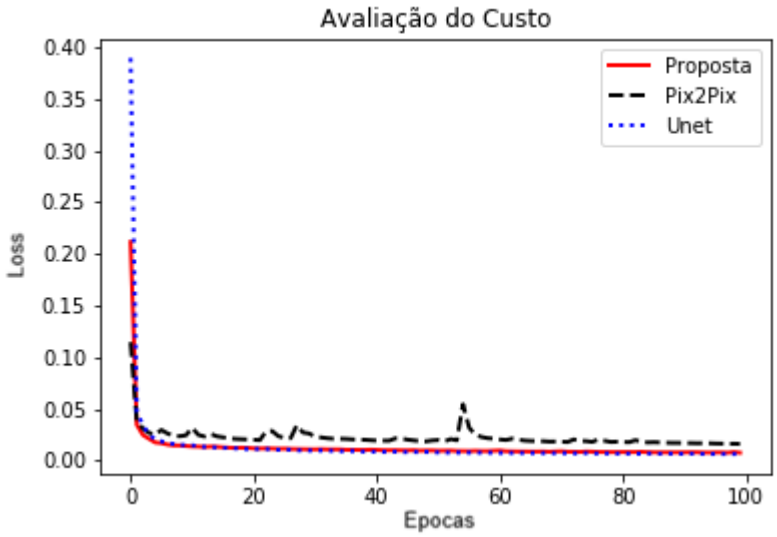
Durante a fase de experimentação, realizaram-se diversos testes variando-se os parâmetros arquiteturais e as configurações de treinamento. Dentre as variações testadas, a configuração final do modelo apresentado foi a que obteve os resultados mais satisfatórios. Todos os modelos utilizaram o mesmo conjunto de dados de treinamento e teste para validação dos resultados. O modelo da DeconvNet utilizou-se no treinamento, o otimizador Adam, tamanho do lote de 32, 100 épocas e o *dataset* foi dividido em 70% para o treinamento (10% de Validação) e 30% para o teste.

Três modelos foram avaliados em treinamentos distintos: modelo proposto, Pix2Pix e a U-Net. Na Pix2Pix foi utilizado a mesma configuração da Figura 1, porém a DeconvNet foi alterada para o gerador da Pix2Pix, baseada em (Isola *et al.,* 2017). A rede U-Net foi uma DeconvNet

original baseada em (Ronneberger *et al.*, 2015), devido ao underfitting com imagens de entrada RGB, implementou-se imagens em escala de cinza de entrada, como a original.

Nas Figuras 5 e 6 ilustram, respectivamente, a evolução do custo e da acurácia no treinamento dos três modelos. Já na Figura 7, contém apenas os dois primeiros modelos, cuja RNA é implementada com treinamento em 200 épocas. As figuras demonstram a eficiência dos modelos na fase de aprendizado. Comparando a proposta com a U-net, observa-se na Figura 5 o início do aprendizado com valores de erros menores, entretanto todas alcançaram a estabilidade próximo a 10 épocas. A Pix2Pix contém uma oscilação ao longo do treinamento, podendo ser um reflexo dos resultados posteriores de teste.

Figura 5. Resultado da avaliação do custo e do erro no treinamento dos três modelos de Redes Neurais Profundas (Deconvolucional)



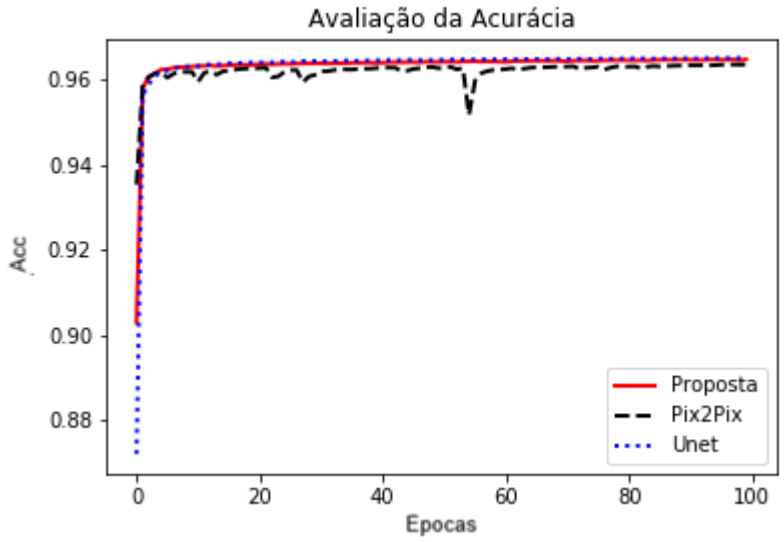
Fonte: Elaboração Própria.

O modelo com a RNA MLP (iii) teve diversos parâmetros (camadas, neurônios, entradas etc.) alterados e a melhor validação contém uma arquitetura (3-8-8-1) com as imagens no modelo RGB. As funções de ativação Relu em todas as camadas, exceto a saída, cuja função escolhida foi a tangente hiperbólica e o otimizador é o “RMSprop”.

Os modelos MS (iv) e GGI (v) são tradicionais em processamento de imagem. Ambos trabalham com operadores morfológicos para detecção das regiões de interesse, MS usa contornos ativos e o GGI utiliza uma gaussiana inversa. A fase de treinamento nos dois modelos avalia a evolução da precisão nas imagens com o aprendizado não supervisionado. Para o

desenvolvimento foi utilizado a biblioteca *skimage*. A definição do número de épocas nos modelos MS e CGI baseou-se em uma análise comparativa da convergência das curvas de perda (*loss*) e acurácia. Experimentos preliminares com uma quantidade de 200 época mostraram que ambos os modelos atingiam um platô de estabilidade, sem alterações significativas no erro residual. Diante desse comportamento, estabeleceram-se dois limiares: um próximo a estabilidade com 45 épocas e custo-benefício computacional, e outro com 100 épocas para avaliar o treinamento prolongado e a ausência de *overfitting*.

Figura 6. Resultado da avaliação da acurácia no treinamento dos três modelos de Redes Neurais Profundas (Deconvolucional)



Fonte: Elaboração Própria.

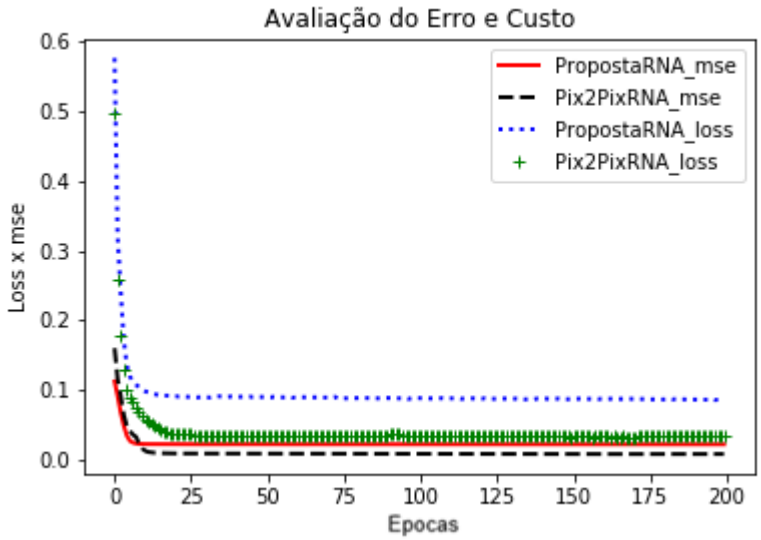
2.2. Métricas de avaliação

Para avaliação dos resultados utilizamos quatro métricas definidas a partir de grupos distintos não correlacionais, conforme (Taha & Hanbury, 2015). A relação Verdadeiro Positivo (VP) é quando a predição da RNC encontra a região de interesse na mesma posição espacial (x,y) da imagem chamada de *label*, e VN é o fundo da imagem. O cálculo de VP, VN, FP e FN é feito pixel a pixel. Com esses parâmetros definidos, as métricas definidas para avaliação dos modelos foram: *dice*, *recall* e *specificity*.

2.3. Discussões

Os resultados de precisão para todas as métricas e métodos estão definidos na Tabela 1. O modelo proposto obteve resultados promissores quando comparados aos demais métodos avaliados, chegando a valores acima de 98%. Os três modelos escolhidos com base em DeconvNet obtiveram excelentes resultados, acima de 91% para o *dataset* de folhas de soja.

Figura 7. Resultado da avaliação do custo e do erro no treinamento da Rede Neural Artificial em dois modelos



Fonte: Elaboração Própria.

Uma análise comparativa foi validada estatisticamente por meio do teste t de Student aplicado sobre as 490 imagens do conjunto de teste, nos grupos da arquitetura proposta e da U-Net. Pelo coeficiente *Dice*, a arquitetura proposta (*0,9816 ± 0,0041*) superou o da U-Net (*0,9781 ± 0,0106*) com significância de $p < 0,0001$ ($t = 8,868$). A proposta conseguiu alcançar um desvio padrão menor e uma maior robustez e estabilidade preditiva, indicando uma sensibilidade menor as variações das amostras em relação ao U-Net e aos demais modelos.

Em relação aos demais métodos tradicionais (GGI e MS) e a RNA. O método GGI, percebeu-se que a precisão na segmentação foi prejudicada em imagens com pouco brilho e contraste. A porcentagem em sua maioria ficava acima de 90%, mas na alteração do brilho e contraste esses valores passaram para 20% e 5% de precisão na métrica *dice*. Assim como o GGI, o método MS apresentou dificuldades em lidar com as alterações das características das imagens (rotação, cor, brilho etc.), resultando em métricas abaixo de 0,5. Na precisão com o método *dice*,

algumas imagens ficaram com 1% de precisão, tornando-se inviável o uso desses métodos em segmentações com altas variações ou ambientes reais. A RNA comportou-se de forma semelhante aos métodos MS e GGI, tiveram imagens com valores menores que 1% na precisão da segmentação. No treinamento da RNA averiguou-se que os valores de acurácia alcançavam um platô e não evoluíam, mesmo com variações em diversas características.

Tabela 1. Comparativo de desempenho entre a arquitetura proposta e modelos da literatura com base em métricas de segmentação

Arquitetura / Modelo	**Base de Dados (*Dataset*)**	**Métricas**		
		Dice	***Recall***	***Specificity***
GGI	Folhas (Própria)	0,8050	0,7482	0,9965
MS	Folhas (Própria)	0,4774	0,4902	0,4961
RNA	Folhas (Própria)	0,8175	0,8175	0,9796
U-Net	Microscópica	0,9203	*	*
U-Net	Folhas (Própria)	0,9796	**0,9991**	0,9945
Pix2Pix + RNA	Folhas (Própria)	0,9153	0,8948	0,9910
Proposta	Folhas (Própria)	**0,9821**	0,9886	**0,9966**

* Dados não reportados pelos autores

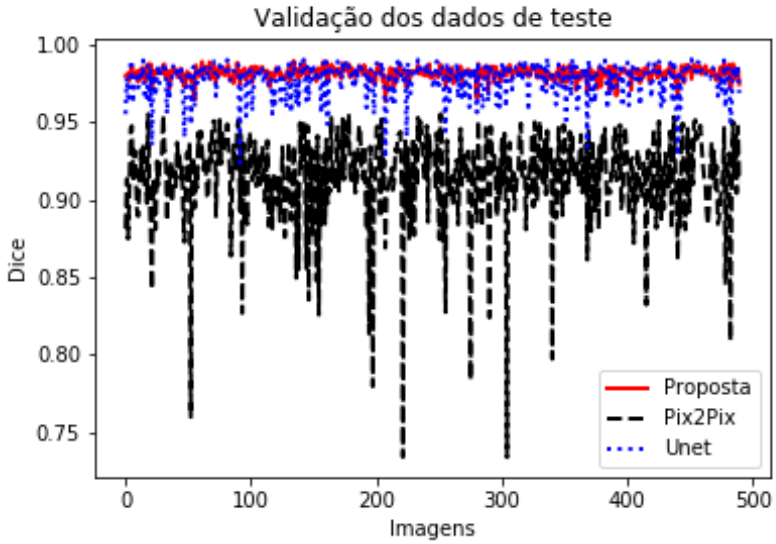
Fonte: Elaboração Própria.

Na análise da Figura 8 observa-se a consistência da segmentação ao longo das 490 imagens de teste em dois modelos, proposta e U-Net. O modelo Pix2Pix teve uma oscilação maior que nos outros dois modelos, em algumas imagens os valores ficam abaixo de 80%. Por outro lado, o modelo proposto obteve os resultados mais consistentes e baixa oscilação nos resultados. Na Figura 9 pode-se observar os resultados visuais das segmentações nas imagens selecionadas de forma aleatória com os três modelos. As diferenças nas segmentações entre os três modelos são sutis, mesmo com mudanças na fluorescência de clorofila, e apenas detalhes mínimos de pixels são revelados nesses casos.

Uma característica importante analisada para a segmentação, devido a ser uma tarefa meio, é a quantidade de parâmetros pertencentes ao modelo. Essa característica está diretamente relacionada ao tempo de treinamento e a relação com o desempenho do modelo. A Tabela 2

apresenta a quantidade de parâmetros e o tempo de treinamento completo de cada modelo no Google Colab com GPU e na mesma sessão.

Figura 8. Resultados da segmentação: Validação dos dados de teste



Fonte: Elaboração Própria.

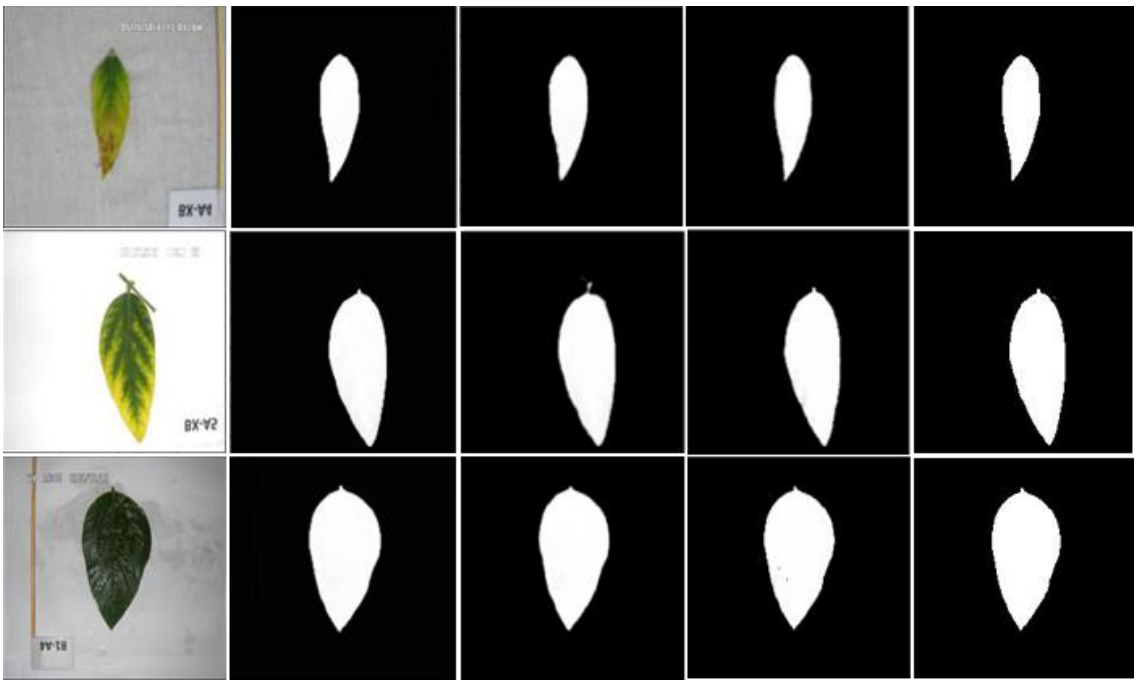
Tabela 2. Tempo de Treinamento e Total de Parâmetros dos Modelos de Aprendizado Profundo

Trabalhos	Treinamento (Seg	Total de Parâmetro
U-Net	8.124	31.031.685
Pix2Pix +RNA	2.710	29.248.094
Proposed	1.757	536.990

Fonte: Elaboração Própria.

Com a redução de camadas e filtros na DeconvNet proposta teve impacto na quantidade total de parâmetros do modelo. A quantidade de parâmetros está relacionada a proposta completa (DeconvNet+RNA), conforme Figura 3. Vale destacar, que o modelo Pix2Pix teve suas camadas e filtros reduzidos também. Nos modelos proposto e Pix2Pix, a quantidade de épocas na RNA poderia ser reduzida pela metade, conforme a Figura 5. Assim, na Tabela 2 o tempo de treinamento da RNA apenas corresponde a 698 segundos, a metade desse valor poderia ser reduzido do tempo total do modelo.

Figura 9. Resultados Visuais da segmentação: original (1º coluna), Proposta (2º coluna), U-Net (3º coluna), Pix2Pix (4º coluna) e *Label* (5º coluna)



Fonte: Elaboração Própria.

3. CONSIDERAÇÕES

O artigo apresentou uma proposta de segmentação de imagens baseada em aprendizado profundo pelas folhas da soja utilizando baseado em modelos Deconvolucionais atuais. A arquitetura proposta foi comparada com dois outros modelos baseados em aprendizado profundo, além da análise com diversos outros métodos de segmentação tradicionais e com aprendizado de máquina supervisionado.

A arquitetura apresentou desempenho satisfatório e robustez estatística para o problema de segmentação nas amostras testadas, mantendo a acurácia em patamares superiores a 98%. A arquitetura otimizada demonstrou resiliência mesmo com as variações nas características (luz, cor, intensidade, forma etc.) das imagens originais das folhas de soja, presentes nas amostras do *dataset*. As imagens modificadas foram distribuídas no subconjunto de treinamento e teste com o intuito de ampliar a variabilidade de imagens (aumento de dados) e permitir avaliar a robustez dos modelos em interferências ambientais típicas.

Os resultados obtidos demonstram estabilidade e consistência estatística dentro do escopo do conjunto de dados analisado. Os demais modelos avaliados apresentaram maior sensibilidade as variações nas condições das amostras do conjunto de dados. A exigência computacional do modelo teve um número reduzido de parâmetros em comparação às arquiteturas de aprendizado profundo convencionais, otimizando o custo de treinamento. Contudo, ressalta-se que a validade destas conclusões permanece vinculada às características do *dataset* próprio utilizado. Em trabalhos futuros pretende-se analisar a segmentação em tipos distintos de imagens com trifólios, folhagens e plantações, além de integrar mecanismos de localização de objetos.

REFERÊNCIAS

Aich, S., & Stavness, I. (2017). Leaf counting with deep convolutional and deconvolutional networks. In Proceedings of the IEEE international conference on computer vision workshops (pp. 2080-2089).

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.

Chen, X., Qiu, X., Zhu, C., Liu, P., & Huang, X. J. (2015, September). Long short-term memory neural networks for chinese word segmentation. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1197-1206).

Chithambaram, T., & Perumal, K. (2017, September). Brain tumor segmentation using genetic algorithm and ANN techniques. In *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)* (pp. 970-982). IEEE.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends& in Signal Processing*, *7*(3-4), 197-387.

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems engineering*, *151*, 72-80.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Johnson, J., Karpathy, A., & Fei-Fei, L. (2016). Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4565-4574).

Lin, K., Gong, L., Huang, Y., Liu, C., & Pan, J. (2019). Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Frontiers in plant science*, *10*, 155.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Marsland, S. (2015). Machine learning: An algorithmic perspective (2nd ed.). Chapman and Hall/CRC.

Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). Ieee.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520-1528).

OpenAI. (2026). ChatGPT com geração de imagens DALL·E [Software de inteligência artificial]. https://openai.com

Papandreou, G., Chen, L., Murphy, K., & Yuille, A. L. (2015). Weakly-and semi-supervised learning of a DCNN for semantic image segmentation. CoRR abs/1502.02734 (2015). *arXiv preprint arXiv:1502.02734*.

Rejeb, I. B., Ouni, S., & Zagrouba, E. (2017, October). Image retrieval using spatial dominant color descriptor. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)* (pp. 788-795). IEEE.

Ren, M. & Zemel, R. S. (2016). End-to-End Instance Segmentation and Counting with Recurrent Attention. CoRR, 2016. Disponível em: http://arxiv.org/abs/1605.09410. Acesso em: 11 out. 2025.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.

Sartin, M. A. (2014). Projeto e implementação de redes neurais artificiais em distintos níveis de abstrações para o reconhecimento de deficiências de diversos macronutrientes e cultivares. (2014). Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Engenharia de Ilha Solteira.

Sartin, M., Da Silva, A., Kappes, C., & S. Filho, T. (2020). Classifying the Macronutrient Deficiency in Soybean Leaf with Deep Learning. In Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional, (pp. 638-649). Porto Alegre: SBC. doi:10.5753/eniac.2020.12166

Sartin, M. A., da Silva, A. C. R., & Kappes, C. (2022). Recognizing Potassium Deficiency Symptoms in Soybean with ANN on FPGA. *Applied Engineering in Agriculture*, *38*(2), 445-453.
Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., Luengo, I., Pape, J. M., Polder, G., Vukadinovic, D., Yin, X., & Tsaftaris, S. A. (2016). Leaf segmentation in plant phenotyping: a collation study. Machine Vision and Applications, 27(4), 585–606.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, K., Rajora, S., Vishwakarma, D. K., Tripathi, G., Kumar, S., & Walia, G. S. (2020). Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets. *Neurocomputing*, *371*, 188-198.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Szegedy, C., Loffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, *15*(1), 29.

Tran, P. V. (2016). A fully convolutional neural network for cardiac segmentation in short-axis MRI. *arXiv preprint arXiv:1604.00494*.

Tuggener, L., Elezi, I., Schmidhuber, J., Pelillo, M., & Stadelmann, T. (2018). DeepScores--A dataset for segmentation, detection and classification of tiny objects. *arXiv preprint arXiv:1804.00525*.

Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Cham: Springer International Publishing.