UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Karla Gabriele Florentino da Silva

Uma Base de Imagens de Folhas de Feijão e uma Rede Neural Profunda para Estimativa Não-Destrutiva de Área Foliar

Juiz de Fora
2023

Karla Gabriele Florentino da Silva

Uma Base de Imagens de Folhas de Feijão e uma Rede Neural Profunda para Estimativa Não-Destrutiva de Área Foliar

Dissertação apresentada ao Programa de Pós Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Marcelo Bernardes Vieira

Coorientador: Prof. Dr. Luiz Maurílio da Silva Maciel

Juiz de Fora

2023

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

da Silva, Karla Gabriele Florentino.
Uma Base de Imagens de Folhas de Feijão e uma Rede Neural Profunda para Estimativa Não-Destrutiva de Área Foliar / Karla Gabriele Florentino da Silva. – 2023.
74 f. : il.

Orientador: Marcelo Bernardes Vieira
Coorientador: Luiz Maurílio da Silva Maciel
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós Graduação em Ciência da Computação, 2023.

1. Rede neural. 2. Aprendizado profundo. 3. Área foliar. 4. Base de imagens. 5. Método não destrutivo. I. Vieira, Marcelo Bernardes, orient. II. Maciel, Luiz Maurílio da Silva, coorient. III. Título.

Karla Gabriele Florentino da Silva

Uma base de imagens de folhas de feijão e uma rede neural profunda para estimativa não-destrutiva de área foliar

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação, da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Ciência da Computação.

Aprovada em 21 de dezembro de 2023.

BANCA EXAMINADORA

Prof. Dr. Marcelo Bernardes Vieira - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Luiz Maurílio da Silva Maciel - Coorientador

Universidade Federal de Juiz de Fora

Prof. Dr. Saulo Moraes Villela

Universidade Federal de Juiz de Fora

Prof. Dr. Hélio Pedrini

Universidade Estadual de Campinas

Juiz de Fora, 29/11/2023.

Documento assinado eletronicamente por **Marcelo Bernardes Vieira**, **Professor(a)**, em 21/12/2023, às 13:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Helio Pedrini**, **Usuário Externo**, em 21/12/2023, às 14:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luiz Maurílio da Silva Maciel**, **Professor(a)**, em 21/12/2023, às 14:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Saulo Moraes Villela**, **Professor(a)**, em 21/12/2023, às 14:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1598639** e o código CRC **A8094857**.

AGRADECIMENTOS

Agradeço a Deus pela força e consolo que me deste para continuar seguindo em frente.

Aos meus pais e irmãs pelos incentivos, cuidados e esforços para que eu pudesse me dedicar aos estudos. Ao meu namorado, André, por estar sempre ao meu lado me apoiando e dando suporte emocional. A todos que me encorajaram a embarcar nesta jornada acadêmica.

Aos meus orientadores, Marcelo e Luiz pela confiança, aconselhamento, paciência e dedicação em me orientar durante estes anos. Agradeço, por desde, o início se fazerem presentes auxiliando no que foi preciso.

Aos meus colegas de mestrado Aleksander e Arthur por alegrarem momentos difíceis e tornarem o desafio do ensino remoto mais fácil.

Aos alunos de graduação Igor, Paulo, Lucas, Kaio, Rafael, Alemilson e Artur pela dedicação na anotação das imagens, sem a qual este trabalho não seria possível. Agradeço também aos alunos Ana Luíza, Gabriel, João e Igor que trabalharam comigo em outras tarefas durante meu mestrado.

Ao professor Dr. Leandro Elias Morais pela colaboração e conhecimento compartilhado sobre fisiologia vegetal.

Ao professor Dr. Helio Pedrini pelas sugestões valiosas que engrandeceram esta pesquisa.

Ao IFMG campus Ouro Branco MG pela parceria, e em especial ao técnico do laboratório Alex, pela ajuda com a medição das folhas.

Aos professores do PPGCC pelos seus ensinamentos e aos funcionários da UFJF que contribuíram de algum modo para minha formação.

Agradeço à CAPES e FAPEMIG pelo suporte financeiro.

"*In our obscurity, in all this vastness, there is no hint that help will come from elsewhere to save us from ourselves.*" (SAGAN, 1994)

RESUMO

As folhas desempenham papel fundamental para o corpo vegetal ao realizarem fotossíntese e as características morfológicas (*e.g.*, área foliar) associadas a sua superfície são parâmetros que podem contribuir para explicar respostas a diversos processos, como mudanças climáticas, relações ecológicas e produtividade agrícola. Porém, a maioria dos métodos para medição das dimensões de superfícies foliares existentes são trabalhosos e onerosos. Além de utilizarem muitas das vezes abordagens destrutivas que impossibilitam acompanhar o crescimento da planta. Neste contexto, construiu-se uma nova base de imagens anotadas para estimativa não destrutiva das dimensões (área, largura, comprimento e perímetro) de superfícies de folhas de feijão, com base em um marcador de realidade virtual adicionado na cena. A construção do conjunto de dados envolveu um processo de plantio, aquisição de imagens, colheita das folhas, medição manual das dimensões reais, segmentação semi-automática e estimativa de pose do marcador. Além disso, desenvolveu-se uma nova rede neural profunda que receba uma imagem de entrada contendo uma folha saliente acompanhada de um marcador, e retorne a estimativa da área foliar pela comparação entre as proporções dos dois objetos na imagem. O modelo proposto é baseado na arquitetura de uma rede neural de segmentação semântica. A hipótese principal é que é possível adaptar uma rede neural convolucional para realizar a regressão da área dos *pixels* da imagem. Assim, propõe-se um novo módulo decodificador para a rede, utilizado para remapear a representação da imagem na estimativa da área relativa dos objetos de interesse, folha e marcador. O modelo apresentado é composto por um codificador e dois decodificadores, que estimam a segmentação da imagem e a área dos *pixels* dos objetos de interesse. Também define-se uma forma para calcular a perda deste decodificador e critérios para seleção do melhor modelo. Para determinar a viabilidade da proposta realiza-se uma análise extensiva, em termos quantitativos e qualitativos, do comportamento das predições do modelo para 1033 imagens de 90 folhas distintas. Os resultados obtidos evidenciam que o modelo é capaz de aprender a estimar a área dos objetos de interesse tendo apenas uma imagem de entrada.

Palavras-chave: Rede neural. Aprendizado profundo. Área foliar. Base de imagens. Método não destrutivo.

ABSTRACT

Leaves perform a fundamental role for the plant body doing photosynthesis and morphological characteristics (e.g., leaf area) associated with its surface are parameters that could help explain responses to various processes, such as climate change, ecological relationships, and agricultural productivity. Most of the existing methods for measuring leaf surface dimensions are expensive and often complicated. In addition, it commonly uses destructive approaches that make it impossible to monitor plant growth. In this context, a new database with annotated images was constructed for non-destructive estimation of bean leaf dimensions (area, width, length, and perimeter), based on a virtual reality marker added to the scene. The construction of the database involved a process of planting, image acquisition, leaf picking, manual measurement, semi-automatic segmentation, and marker pose estimation. Furthermore, a new deep neural network was developed that receives one input image containing a salient leaf accompanied by a marker and provides relative leaf area by comparing the proportions of both objects in the image. The proposed method is based on the architecture of a semantic segmentation neural network. The main hypothesis is that is possible to adapt a convolutional neural network to regress image pixels area. Thus, a new decoder module is proposed for the network, used to remap image representation on relative area estimation of the objects of interest, leaf and marker. The model presented is composed of one encoder and two decoders, which estimate the image segmentation and the pixels area of the objects of interest. A way of calculating the loss of the decoder and selection criteria for the best model are also defined. For proposal viability determination, an extensive quantitative and qualitative analysis is performed with the model's predictions on 1033 images of 90 different leaves. Results indicate that the model is capable of learning to estimate the area of the objects of interest with only one input image.

Keywords: Neural network. Deep Learning. Leaf area. Image database. Non-destructive method.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

1D	Unidimensional
2D	Bidimensional
3D	Tridimensional
AF	Área foliar
ASPP	*Atrous Spatial Pyramid Pooling*
CNN	*Convolutional Neural Network*
CRF	*Fully Conditional Random Field*
DL	*Deep Learning*
GD	*Gradient Descent*
GT	*Ground truth*
JPEG	*Joint Photographic Experts Group*
Mask R-CNN	*Mask Region Convolutional Neural Network*
MDF	*Medium Density Fiberboard*
mIOU	*Mean Intersection over Union*
MSE	*Mean Squared Error*
ReLU	*Rectified Linear Unit*
RER	*Relative Error Rate*
SCC	Sistema de Coordenadas da Câmera
SCI	Sistema de Coordenadas da Imagem
SCM	Sistema de Coordenadas do Mundo
SCP	Sistema de Coordenadas em *Pixel*
XML	*Extensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO

As folhas são responsáveis por converter a energia solar em energia química através do processo de fotossíntese (EVERT; EICHHORN, 2013), e seu crescimento influencia na determinação de produtividade da planta (KOESTER et al., 2014). Além disso, a morfologia da superfície foliar, em especial a área foliar (AF), desempenha um papel crucial no desenvolvimento vegetal, sendo a AF o principal fator no acúmulo de biomassa e produtividade (TAIZ; ZEIGER, 2010). É importante observar também que o tamanho e a forma das folhas apresentam variações intra e interespecíficas que impactam diversos parâmetros fisiológicos (e.g. fotossíntese e respiração foliar) (WRIGHT et al., 2004).

A análise das respostas de vegetais a diversas condições ambientais, como mudanças climáticas (SRINIVASAN; KUMAR; LONG, 2017), disponibilidade de água no solo ou períodos de seca (WELLSTEIN et al., 2017), disponibilidade de luz (POORTER et al., 2019), poluição (JANHÄLL, 2015), fertilidade do solo (LAUGHLIN, 2011), efeito de defensivos agrícolas (MAREK et al., 2018) e presença de fitopatógenos (LU et al., 2018) é frequentemente conduzida com base na superfície foliar (determinada pela área). Desse modo, ferramentas para a aferição da AF (e derivações a partir desta) são de grande relevância para profissionais que estudam plantas, como melhoristas vegetais, botânicos e agrônomos.

O dimensionamento (área, largura, comprimento e perímetro) da superfície foliar pode ser realizado por vários métodos. Dentre eles estão equipamentos de alto custo, como o aparelho integrador LI-COR® (LI-COR, 1996), muitas vezes referenciado na literatura, que mede a área da folha pelo princípio de células de grade de área conhecida. Uma alternativa menos custosa é a estimativa realizada manualmente com base no peso de um contorno da folha, traçado sobre papel milimetrado de gramatura conhecida (PANDEY; SINGH, 2011). Nota-se que este método está sujeito a erros cometidos no traçado da folha, no recorte do contorno e na pesagem. Outra possibilidade são as soluções computacionais que utilizam imagens para medição de superfícies foliares. Alguns desses métodos realizam a estimativa com base na resolução de captura da imagem (*dots per inch* – dpi) (KAUR; DIN; BRAR, 2014) ou um padrão de escala na cena (JADON, 2018) por meio de um algoritmo que integra o tamanho dos *pixels* que compõem a folha. A maioria dos métodos de medição mencionados anteriormente são destrutivos, ou seja, é preciso remover fisicamente as folhas da planta. Além disso, é comum que a folha seja planificada (*e.g.*, utilizando-se uma placa de vidro para pressioná-la), uma vez que sua superfície é irregular.

Uma possível abordagem não destrutiva para estimativa da AF é ajustar um modelo alométrico para cada espécie de cultura. Fanourakis, Kazakos e Nektarios (2021) desenvolveram um modelo de regressão para medição da AF em função da largura e do comprimento de folhas individuais de crisântemo. Para isso, os parâmetros morfológicos

de uma elevada quantidade de folhas devem ser analisados. Os autores usaram 2625 folhas no total, tornando o processo lento e trabalhoso. Há, portanto, o interesse em desenvolver metodologias para medição não invasiva e que possibilitem acompanhar o desenvolvimento da folha. Segundo Koubouris et al. (2018), medições sucessivas da AF auxiliam na otimização do cultivo, através do monitoramento do crescimento e produtividade de culturas.

Com os avanços em visão computacional e processamento de imagens, o aprendizado profundo (*Deep Learning* – DL) tem sido amplamente estudado e aplicado para solucionar problemas em várias áreas do conhecimento, alcançando resultados que superam modelos anteriores (SARKER, 2021). Ao longo dos últimos anos, técnicas de DL têm sido aplicadas em diversas tarefas do campo da agricultura (DHANYA et al., 2022), que vão desde auxílio aos agricultores na preparação do solo até o momento da colheita. Focando especificamente em aplicações com plantas, algumas soluções que tiveram resultados promissores são segmentação de áreas de vegetação (YANG et al., 2020), detecção de doenças e/ou anomalias em plantas (ALE et al., 2019; WU et al., 2021), classificação de plantas (ISLAM et al., 2021), predição do crescimento de plantas (KIM; LEE; KIM, 2022) e contagem de folhas (FAN et al., 2022).

As redes neurais convolucionais (*Convolutional Neural Network* – CNN), em especial, fornecem grande auxílio para análise foliar. Estudos, como os de Lee et al. (2015), Dyrmann, Karstoft e Midtiby (2016) e Liu et al. (2017), demonstraram resultados com altos níveis de acurácia no uso de CNN para identificação de características das folhas para reconhecimento de espécies e doenças em plantas. Além disso, estudos recentes, como de Triki et al. (2021) e Li et al. (2023), atestaram a eficácia das CNNs para segmentação de folhas como base para medição posterior de suas superfícies. Inspirando-se no sucesso obtido por trabalhos como estes na literatura, neste trabalho é proposta uma adaptação da arquitetura de uma rede de segmentação semântica (CHEN et al., 2018), para estimativa não destrutiva da AF utilizando imagens da folha ainda na planta e em seu ambiente natural. A meta é obter uma nova arquitetura capaz de prever a medida relativa da folha em comparação com um marcador conhecido adicionado na cena.

1.1 MOTIVAÇÃO

O feijão-comum (*Phaseolus vulgaris L.*) é uma planta dicotiledônea (SAKHRAVI; DEHDARI; FAHLIANI, 2023), ou seja, é caracterizada por apresentar dois cotilédones ou folhas embrionárias durante a germinação da semente. Existem várias espécies conhecidas como feijoeiro, sendo a sua morfologia foliar semelhante à folhagem da soja, uma das *commodities* de maior relevância para a economia mundial (SONG et al., 2021).

Devido à facilidade de acesso ao grão na região de Ouro Branco, Minas Gerais, o conjunto de imagens proposto nesta dissertação é composto por imagens de folhas

de feijão-comum. Adicionalmente, o feijão é considerado um alimento essencial para a população brasileira (OLIVEIRA et al., 2018). Além de ser utilizado como suplemento alimentar em alguns países (ONAKPOYA et al., 2011), o que reforça o interesse pela espécie.

Visão computacional e processamento de imagens podem oferecer diversos recursos para apoiar o processo de análise foliar através da estimativa da área de superfícies. A abordagem proposta neste trabalho objetiva estimar não destrutivamente a área foliar. Este é um parâmetro diretamente relacionado a produtividade agrícola, que auxilia no monitoramento da saúde de plantas e contribui para estudos de processos fisiológicos vegetais.

Embora haja soluções relacionadas na literatura que utilizem redes neurais para realizar a tarefa de dimensionamento de superfícies foliares, como em (ZHANG et al., 2020; TRIKI et al., 2021; LI et al., 2023), o campo ainda é relativamente novo. Dessa forma, este trabalho propõe uma nova abordagem para o problema. A proposta consiste em uma arquitetura de rede neural para a tarefa de segmentação e regressão da área dos *pixels* da imagem de entrada. Deste modo, estima-se a área da superfície foliar sem a exigência de pós-processamento dos elementos de imagem.

1.2 DEFINIÇÃO DO PROBLEMA

O problema principal deste trabalho é estimar a área relativa de superfícies com a topologia do plano e imersas em um espaço tridimensional (3D), representadas projetivamente por uma imagem cujo suporte é bidimensional. As superfícies de interesse são aquelas que representam folhas isoladas, acompanhadas de um marcador de realidade virtual e aumentada cujas dimensões reais são conhecidas.

Este trabalho aborda a seguinte pergunta de pesquisa: “É possível obter uma rede neural capaz de comparar dois objetos, um de dimensões conhecidas, e outro cuja dimensão será calculada em relação ao primeiro, prevendo assim uma estimativa de área?”.

1.3 OBJETIVOS

O objetivo deste trabalho é analisar a viabilidade de desenvolver uma nova arquitetura de rede neural profunda capaz de estimar a área relativa da superfície foliar tendo apenas uma imagem como entrada. Esta imagem em questão possui uma folha saliente próxima de um marcador fiducial de tamanho fixo conhecido. Para realizar o treinamento do modelo de rede neural, propôs-se construir uma base de dados com imagens de folhas de feijão-comum ainda na planta, acompanhadas por um marcador fiducial.

1.4 CONTRIBUIÇÕES

As contribuições deste trabalho são:

- Construção de uma base de dados com anotações de duas máscaras de segmentação, bem como as dimensões esperadas de cada folha (área, largura, comprimento e perímetro), informações da situação de captura (plana ou curva, com ou sem recortes, ao sol ou à sombra) e de calibração (matriz extrínseca homogênea, matriz perspectiva, entre outras).
- Desenvolvimento de uma nova arquitetura de rede neural profunda para estimativa da área foliar de amostras coletadas, adaptando a rede de segmentação DeepLabv3+ (CHEN et al., 2018).

1.5 ORGANIZAÇÃO

Além deste capítulo introdutório, esta dissertação está estruturada do seguinte modo: O Capítulo 2 introduz os fundamentos que se relacionam com este trabalho. O Capítulo 3 apresenta trabalhos da literatura relacionados à medição de superfícies de objetos utilizando imagens. O Capítulo 4 descreve a construção da base dados. O Capítulo 5 apresenta o método proposto. O Capítulo 6 descreve os resultados e discussões. Por fim, o Capítulo 7 apresenta as conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os fundamentos que embasam esta dissertação. Na Seção 2.1, são descritos alguns conceitos relacionados com geometria de câmera. A Seção 2.2 aborda sobre redes neurais convolucionais. Por fim, a Seção 2.3 apresenta a arquitetura original da rede DeepLabv3+.

2.1 MODELO DE CÂMERA

O processo de formação da imagem pode ser representado por uma câmera *pinhole* ou câmera de furo. A imagem neste modelo é formada quando raios luminosos atravessam um pequeno furo e incidem sobre o fundo de uma caixa. No entanto, como a imagem obtida é invertida, considera-se a existência de um plano à frente deste orifício ou centro óptico $\mathbf{c}$. O funcionamento geométrico de uma câmera *pinhole* é constituído exclusivamente pelo processo de projeção perspectiva. Isso faz com que esse modelo seja muito importante, em teoria, para compreensão desse processo. Além disso, na prática, ele é utilizado para modelar diversas câmeras disponíveis no mercado.

A geometria deste modelo define que, para cada ponto $\mathbf{p} = (x_c, y_c, z_c)$ do espaço existe um ponto $\mathbf{q} = (x_i, y_i)$ correspondente no plano π situado a uma distância f, conforme ilustra a Figura 1. Esta projeção é definida através da interseção do plano π com a reta que liga os pontos $\mathbf{c}$ e $\mathbf{p}$, formada por todos os pontos da forma $\alpha(x_c, y_c, z_c)$. Para simplificar, pode-se considerar que a imagem é formada em $z_c > 0$. Desta forma, o ponto sobre a imagem possui $\alpha = f/z_c$, com coordenadas dadas por:

$$\begin{aligned} x_i &= \frac{f}{z_c} x_c \\ y_i &= \frac{f}{z_c} y_c. \end{aligned} \tag{2.1}$$

As correspondências entre os pontos do espaço e pontos da imagem podem ser expressas através de transformações de coordenadas. No sistema de coordenadas do mundo (SCM) tridimensional, utiliza-se coordenadas (x_m, y_m, z_m) para representar um ponto. Enquanto, no sistema de coordenadas da câmera (SCC), cuja origem é o centro óptico da câmera $\mathbf{c}$, as coordenadas são denotadas por (x_c, y_c, z_c).

Para realizar a mudança de coordenadas entre os dois referenciais tridimensionais, tem-se $\mathbf{t}$ o vetor que representa a origem do mundo no SCC e $\mathbf{R}$ uma matriz ortogonal composta por vetores unitários correspondentes aos eixos do SCM. Dado um ponto $\mathbf{s} = (x_m, y_m, z_m)$, o ponto $\mathbf{p}$ no SCC é dado por:

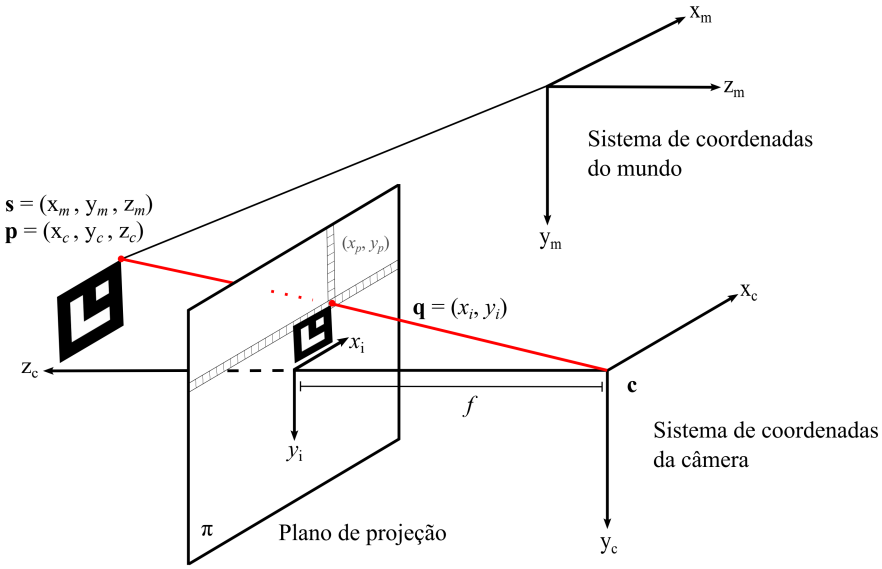
$$(x_c, y_c, z_c) = \mathbf{t} + x_m \mathbf{r_1} + y_m \mathbf{r_2} + z_m \mathbf{r_3}, \tag{2.2}$$

onde $\mathbf{r_1}$ $\mathbf{r_2}$ e $\mathbf{r_3}$ são as colunas da matriz $\mathbf{R}$. Em coordenadas homogêneas, tem-se:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix} . \tag{2.3}$$

Além disso, a rotação $\mathbf{R}$ também pode ser representada por 3 números reais que caracterizam a rotação do espaço. Isto pode ser feito utilizando os ângulos de Euler, quatérnions, ou ainda pela fórmula de Rodrigues (CARVALHO et al., 2005). A matriz $\mathbf{R}$ e o vetor $\mathbf{t}$ correspondem aos parâmetros extrínsecos da câmera e estão relacionados ao seu posicionamento em relação às coordenadas do mundo.

Figura 1 – Projeção perspectiva.



Fonte: Elaborada pela autora (2023).

Além dos parâmetros extrínsecos, uma câmera possui parâmetros intrínsecos relativos a fatores que influenciam na formação da imagem, como a distância focal, tamanho do *pixel* e distorções de lente. A projeção perspectiva do modelo *pinhole*, no qual o único parâmetro intrínseco é a distância focal f, pode ser expressa como uma transformação do SCC para o sistema de coordenadas da imagem (SCI):

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \simeq \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} , \tag{2.4}$$

onde (x_i, y_i) são coordenadas bidimensionais (2D) no SCI, cujas coordenadas situam-se sobre o plano de projeção com origem na projeção ortogonal de $\mathbf{c}$. É importante observar que a transformação dada por esta equação não é invertível.

O sistema de coordenadas em *pixel* (SCP) também 2D, por sua vez, possui coordenadas (x_p, y_p) que representam a matriz de *pixels*, cuja origem é no canto superior esquerdo. Pode-se definir a transformação dos pontos de SCI para o SCP, como:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & \tau & u_c \\ 0 & s_y & v_c \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \tag{2.5}$$

onde s_x e s_y representam o número de *pixels* por unidade de comprimento, nas direções horizontal e vertical, respectivamente, τ é a tangente do ângulo que as colunas de *pixels* formam com as linhas (idealmente $\tau = 0$). Os valores u_c e v_c fornecem a posição em *pixels* da projeção da origem sobre o plano de projeção. Contudo, os valores individuais de f, s_x e s_y não são possíveis de estimar, somente o fabricante da câmera pode disponibilizá-los. Sendo assim, a matriz intrínseca da câmera $\mathbf{M}_{cam}$ ou matriz de calibração, normalmente é definida pelos produtos destes parâmetros fs_x e fs_y :

$$\mathbf{M}_{cam} = \begin{bmatrix} fs_x & f\tau & u_c \\ 0 & fs_y & v_c \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.6}$$

podendo ser expressa por:

$$\mathbf{M}_{cam} = \begin{bmatrix} f_x & c & u_c \\ 0 & f_y & v_c \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.7}$$

A calibração é um procedimento realizado para estimar os parâmetros intrínsecos e extrínsecos de uma câmera. Existem diversos métodos de calibração presentes na literatura, sendo um dos mais populares, introduzido por Zhang (2000). De forma simplificada, o método necessita de um conjunto de imagens de um padrão conhecido, para realizar uma busca por pontos $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ que correspondam a pontos conhecidos $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n$ do padrão no espaço tridimensional. Desta forma, a partir das correspondências encontradas, obtém-se um conjunto de valores $\{f_x, f_y, c, u_c, v_c\}$ para a matriz de câmera (Equação 2.7) e outro para modelar a distorção ótica. Quando as configurações da câmera não são ajustadas, estes parâmetros não variam entre diferentes imagens capturadas.

O modelo *pinhole* considera a existência de um único raio luminoso para cada ponto da cena. Na prática isso é inviável, em razão do tempo de exposição da câmera ter que ser muito grande, pois a luz atinge o plano com intensidade extremamente baixa. Além disso, para garantir nitidez na imagem, a abertura precisaria ser muito pequena, o que possivelmente causaria difração e, consequentemente, borrões na imagem. Para contornar este problema, utiliza-se um conjunto de lentes que concentram a entrada dos feixes de luz. A inclusão de lentes pode resultar em raios luminosos desviando-se da trajetória, causando

distorções radiais (efeito de curvatura em linhas retas) na imagem. A distorção radial impacta a transformação dos pontos de SCC ao SCI, ao invés do ponto (x, y) dado pela Equação 2.4, a imagem se forma em (x', y'). Uma forma de modelagem desta distorção é dada por:

$$d = k_1 r^2 + k_2 r^4 + k_3 r^6 + ..., \tag{2.8}$$

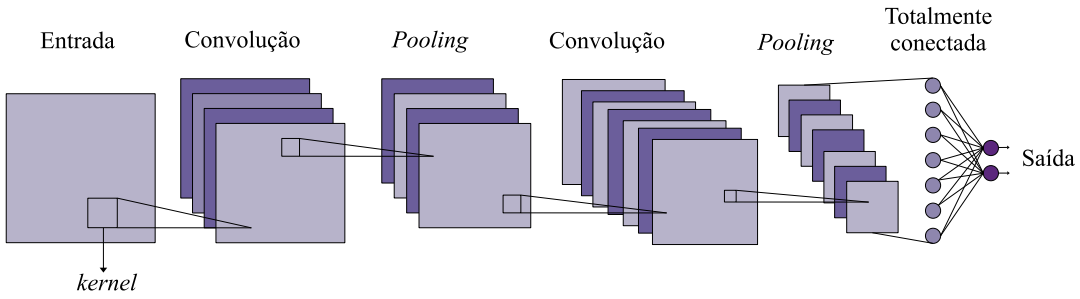
onde $r = \sqrt{x'^2 + y'^2}$ e k_i, com $i \in \{1, 2, 3, ...\}$ são os coeficientes de distorção radial. Outra forma de distorção que pode ocorrer é a distorção tangencial, causada pela descentralização das lentes, que provoca rotação e inclinação do plano da imagem.

2.2 REDES NEURAIS CONVOLUCIONAIS

Em matemática, a convolução é um operador linear sobre duas funções, resultando em uma terceira função, que expressa o quanto a forma de uma é modificada pela outra (GOODFELLOW; BENGIO; COURVILLE, 2016). Sua definição contínua é uma integral do produto de uma das funções por uma cópia deslocada e invertida da outra. No contexto de processamento de imagens, é comumente utilizada a sua versão discreta. Isso possibilita aplicar a convolução em duas matrizes que modelam imagens, normalmente a primeira é uma imagem discreta e a segunda uma matriz menor chamada de filtro ou núcleo (ou em inglês *kernel*) (GOODFELLOW; BENGIO; COURVILLE, 2016). Desse modo, o filtro desliza sobre a imagem, enquanto multiplicações *pixel* a *pixel* em toda a região de sobreposição são realizadas e a soma dos produtos resulta no valor do *pixel* da saída. O valor do deslocamento percorrido por um filtro nesta operação é chamado de tamanho do passo ou *stride*.

As redes neurais convolucionais são modelos que utilizam múltiplas operações de convolução para extrair mapas de características da imagem. O primeiro uso notável deste modelo foi descrito por (LECUN et al., 1998) para reconhecimento de dígitos escritos à mão em envelopes. As redes convolucionais profundas (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) exploram a propriedade de composição de muitos sinais naturais, usando um padrão neural tridimensional em que características de níveis superiores são obtidas através da composição de características de níveis inferiores. Nas imagens, por exemplo, as combinações locais de arestas resultam em partes, as quais formam objetos (LECUN; BENGIO; HINTON, 2015). Estas arquiteturas, conforme ilustra a Figura 2, consistem em pelo menos três tipos de camadas: 1) camadas convolucionais que realizam filtragem e detectam características locais a partir da camada anterior, gerando um mapa de recursos. 2) camadas de agrupamento (*pooling*) que agrupam características semanticamente semelhantes do mapa de recursos e 3) camadas totalmente conectadas compostas pelos neurônios, pesos e vieses que conectam as informações extraídas por diferentes camadas, gerando a saída (*logits*) da rede (LECUN; BENGIO; HINTON, 2015).

Figura 2 – Exemplo de uma rede neural convolucional.



Fonte: Elaborada pela autora (2023).

Existem diferentes tipos de operações de agrupamento, sendo os dois mais utilizados *max pooling* (o maior elemento do filtro é passado para a saída) e *average pooling* (retorna a média dos elementos do filtro). Outro elemento importante de uma CNN são as chamadas funções de ativação, as quais decidem como os neurônios serão ativados. Em outras palavras, são usadas para aplicar transformações não lineares à saída do neurônio, adicionando viés a ele. Existem várias funções de ativação, sendo uma das mais tradicionais a função ReLU, usada quando se quer preservar valores positivos e transformar os negativos em zero:

$$\sigma(z) = max(0, z), \tag{2.9}$$

essa característica leva a uma convergência mais rápida da rede, por outro lado, alguns neurônios podem nunca ser ativos.

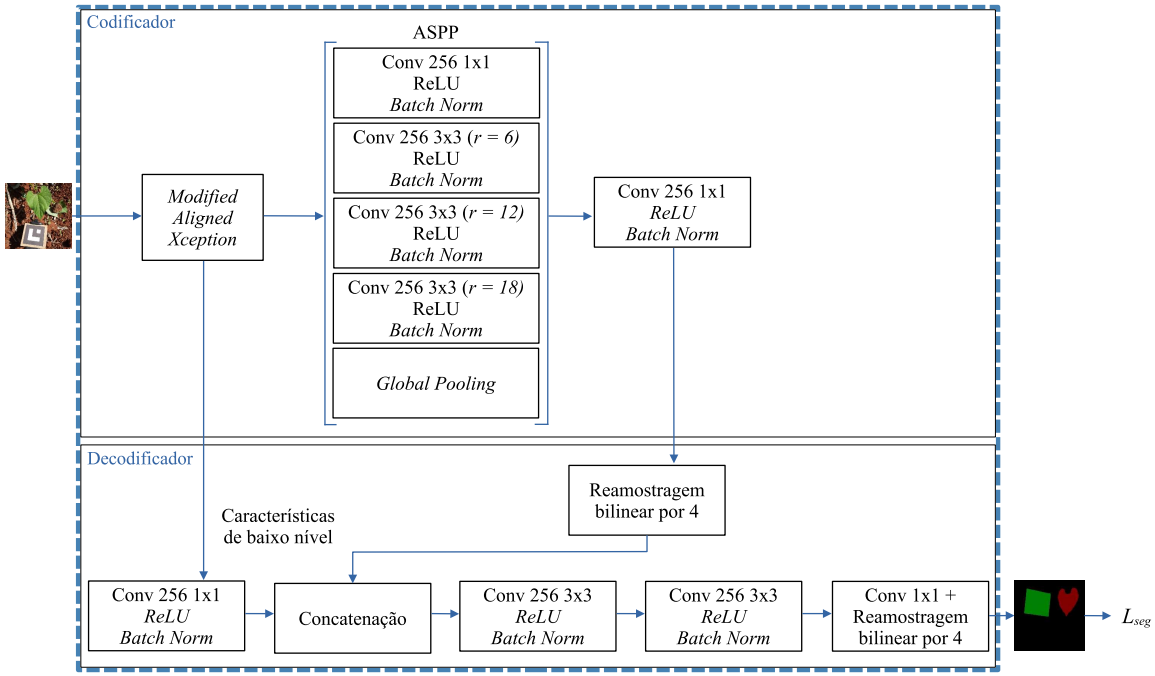
2.3 DEEPLABV3+

A segmentação, no campo da visão computacional, é o processo de decomposição de uma cena em suas partes constituintes (JAIN, 1989). Em síntese, a segmentação de imagens visa determinar um conjunto finito de regiões homogêneas que não se sobrepõem. Desta forma, pode-se dizer que uma imagem é formada por regiões conectadas pelas bordas, onde cada *pixel* recebe um rótulo indicando a região à qual ele pertence (ACHARYA; RAY, 2005). Para problemas de segmentação, espera-se que a entrada e a saída da rede sejam imagens, de forma que a saída possua as mesmas dimensões da entrada. Em geral, este processo é realizado como passo inicial para análise de imagens em busca de uma região de interesse. A segmentação convencional é realizada a partir de métodos de processamento de imagens envolvendo detecção de descontinuidades ou de similaridades. Ao longo dos últimos anos, modelos de segmentação semântica vêm sendo propostos para esta tarefa, cujo objetivo é realizar previsões refinadas da localização espacial de classes presentes em uma imagem de entrada.

DeepLab (CHEN et al., 2014) é um modelo de aprendizado profundo concebido e

disponibilizado pela Google, para segmentação semântica de imagens usando convoluções dilatadas (Subseção **2.3.1**). As duas primeiras versões da DeepLab (CHEN et al., 2014; CHEN et al., 2017a) contavam com aprimoramento de detalhes dos objetos através de um campo aleatório condicional totalmente conectado (*Fully Conditional Random Field* – CRF) (KRÄHENBÜHL; KOLTUN, 2011). Além disso, a segunda atualização do modelo passou a contar com uma camada de agrupamento de pirâmide espacial dilatada (*Atrous Spatial Pyramid Pooling* – ASPP) que permite segmentar paralelamente uma imagem em múltiplas escalas e em múltiplas taxas de amostragem. No terceiro aprimoramento, DeepLabv3 (CHEN et al., 2017b), o CRF foi removido e o módulo ASPP aprimorado com recursos ao nível da imagem, conforme descrito na Subseção **2.3.2**. Finalmente, a extensão atual nomeada DeepLabv3+ (CHEN et al., 2018), conta com uma arquitetura codificador-decodificador utilizando a DeepLabv3 como codificador, conforme mostra a Figura 3. Na Subseção **2.3.3** são apresentados detalhes sobre o decodificador do modelo. A Subseção **2.3.4** descreve a sua função de perda. Para o extrator de características os autores testaram alguns modelos, dentre eles, uma versão modificada do modelo Xception (CHOLLET, 2017), o qual se empregou no presente trabalho e será apresentado na Subseção **2.3.5**.

Figura 3 – Fluxograma DeepLabv3+.



Fonte: Adaptado de Chen et al. (2018).

2.3.1 Convolução dilatada

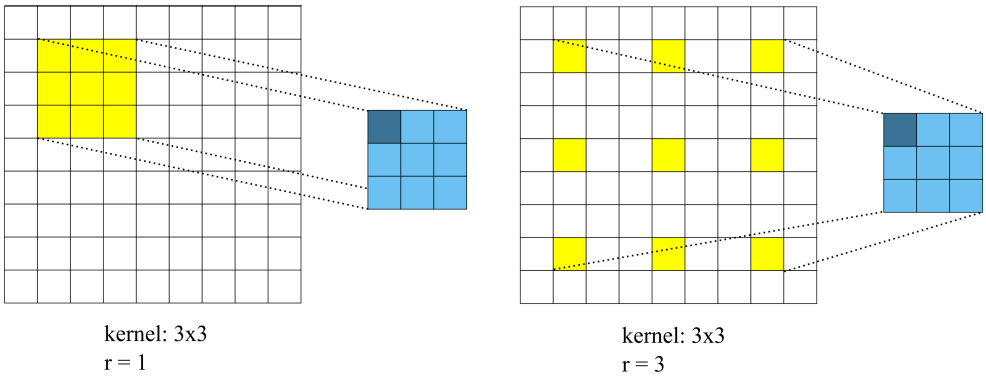
Convolução dilatada é uma implementação para modelos de DL, do algoritmo conhecido como *hole algorithm*, desenvolvido por Mallat (1999), para o cálculo da transformada *wavelet* não-decimada. O processo de uma convolução dilatada se difere da

convolução tradicional quanto aos filtros. Enquanto a convolução tradicional processa *pixels* contíguos, a dilatada analisa *pixels* espaçados entre si. Dessa forma, é possível escolher a resolução com a qual a resposta de uma camada é calculada. Considerando um sinal unidimensional (1D), a equação para esta convolução é definida como:

$$y_i = \sum_{k=1}^{n} x_{i+rk} w_k, \tag{2.10}$$

onde y_i é a saída para o sinal x_i 1D, com um filtro w_k de tamanho n. O parâmetro r define a taxa de dilatação da convolução ou tamanho do passo na qual o sinal é amostrado, sendo que, quando $r = 1$ o efeito é igual à convolução tradicional. Desse modo, este mecanismo oferece uma forma de controle e equilíbrio entre localização (filtro pequeno) e captura do contexto (filtro grande) sem aumentar o número de parâmetros ou custo computacional. Contornando-se assim o problema da redução da amostragem do sinal e da invariância espacial das CNNs. A Figura 4 ilustra um comparativo entre o comportamento dos filtros em uma convolução tradicional 3×3 e uma convolução dilatada, também 3×3, com $r = 3$. É possível observar que quanto maior o valor de r, maior será a área considerada pelos campos receptivos dos filtros, visto que $r - 1$ zeros são adicionados entre as entradas da convolução.

Figura 4 – Comparativo entre os filtros de uma convolução tradicional (a) e uma convolução dilatada com $r = 3$ (b).



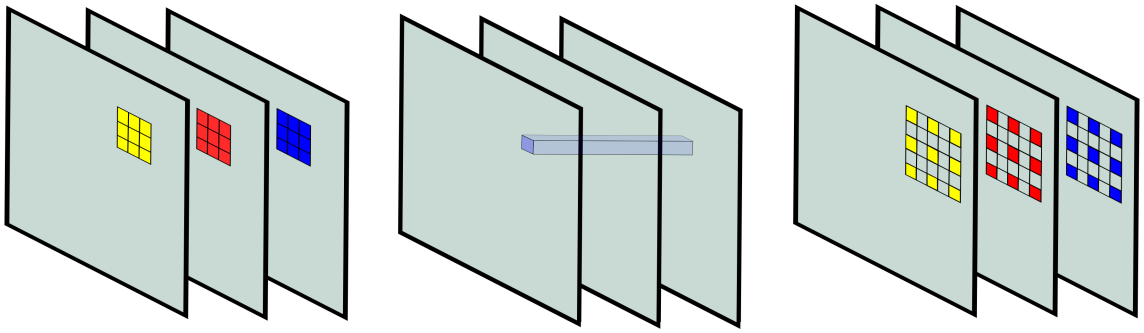
(a) convolução tradicional (b) convolução dilatada

Fonte: Elaborada pela autora (2023).

Para DeepLabv3+, adicionou-se o conceito de *group convolution* apresentado por Krizhevsky, Sutskever e Hinton (2012) ou convolução separável em profundidade como descrita por Sifre (2014), que consiste em decompor uma convolução tradicional em: uma convolução espacial que aplica um único filtro para cada canal de entrada (*depthwise convolution*) e outra que combina os resultados da primeira, em profundidade nos canais (*pointwise convolution*). A Figura 5 ilustra um comparativo entre as duas operações. Ao adaptar a convolução dilatada na convolução em profundidade (Figura 5c), surge um novo tipo de operação chamada por Chen et al. (2018) de convolução separável dilatada (*atrous*

separable convolution). Os autores conseguiram reduzir significativamente a complexidade de computação do modelo com esta proposta, mantendo-se um desempenho semelhante aos trabalhos anteriores.

Figura 5 – Decomposição de uma convolução separável em profundidade em (a) *depthwise convolution* e (b) *pointwise convolution*. A Figura (c) ilustra uma convolução dilatada em profundidade, com $r = 2$.



(a) *depthwise convolution* (b) *pointwise convolution* (c) convolução dilatada em profundidade

Fonte: Adaptado de Chen et al. (2018).

2.3.2 ASPP

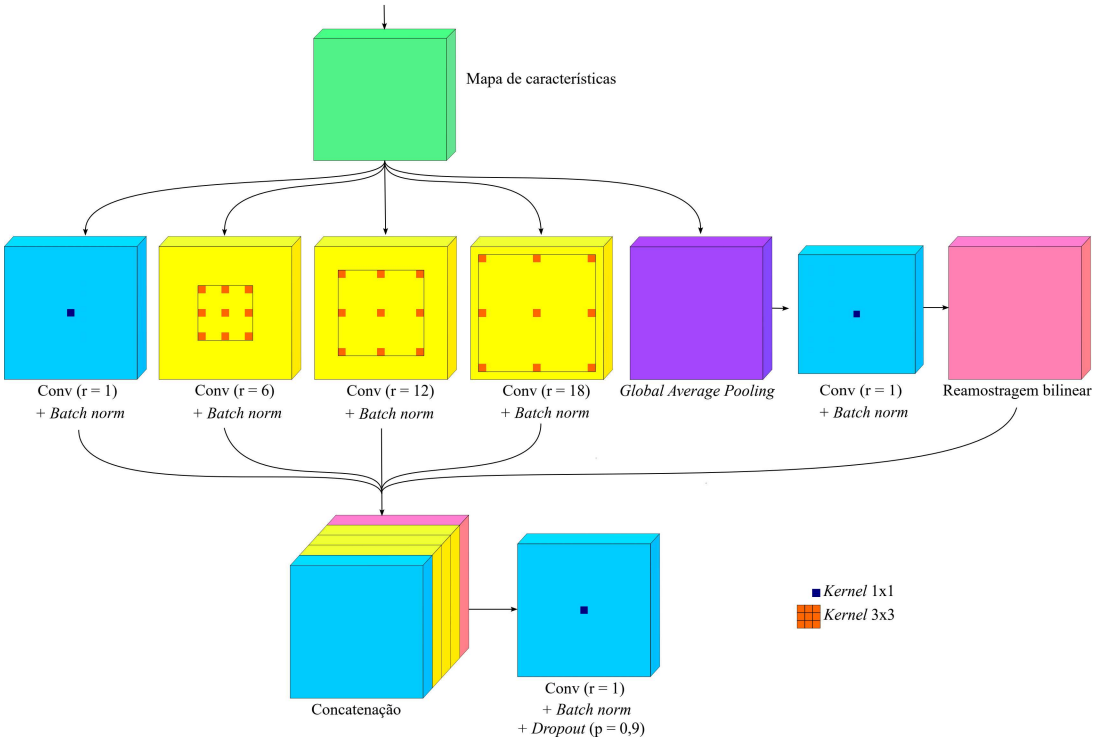
O módulo ASPP, inspirado no trabalho de He et al. (2015), foi projetado para o mapeamento do último mapa de características do extrator a partir de diferentes campos receptivos. A codificação das informações contextuais em múltiplas escalas é motivada pela possibilidade de existirem objetos na cena da mesma classe e em tamanhos diferentes. Para isso, conforme ilustra a Figura 6, são aplicadas paralelamente uma convolução 1×1 e três convoluções 3×3 separáveis dilatadas com diferentes taxas de dilatação (as taxas padrões são 6, 12 e 18 respectivamente).

Porém, descobriu-se que à medida que a taxa de dilatação aumenta, o número de pesos de filtro aplicados à região de características válidas diminui, devido à quantidade de zeros do preenchimento. Para contornar este problema, adicionou-se no ASPP um agrupamento da média global (*global average pooling*), explorado no modelo ParseNet (LIU; RABINOVICH; BERG, 2015), proposto para substituir camadas totalmente conectas em CNNs. Esta operação consiste em calcular a saída média de cada mapa de características (*average pooling*) na camada anterior. Na sequência, é aplicada outra convolução tradicional 1×1 e realizada uma reamostragem bilinear para recuperar o tamanho do mapa anterior.

A interpolação bilinear, realiza uma média ponderada dos quatro vizinhos mais próximos de um *pixel* para estimar o seu novo valor na imagem de saída. É importante observar que todas as convoluções mencionadas possuem 256 filtros. Além disso, camadas de normalização do lote (*batch normalization*) foram inseridas entre as camadas convolucionais.

Após aplicadas todas as operações paralelas, os resultados de todas as ramificações são concatenados e aplica-se uma terceira convolução 1×1 seguida de uma camada de *dropout* com probabilidade $p = 0.9$, para definir quais nós da rede serão descartados. Desse modo, gera-se o último mapa de características do codificado (ou DeepLabv3).

Figura 6 – Fluxograma *Atrous Spatial Pyramid Pooling*.



Fonte: Elaborada pela autora (2023).

2.3.3 Decodificador

Para o decodificador, Chen et al. (2018) propuseram primeiro realizar uma reamostragem bilinear do mapa de características gerado pelo codificador por um fator de 4. Na sequência, as características aumentadas são concatenadas com as características de baixo nível correspondentes, de mesma resolução espacial, geradas pelo extrator de características. Isto foi inspirado pelo trabalho de Hariharan et al. (2015) que demonstrou as vantagens de se explorar as informações de interesse distribuídas em todos os níveis de uma CNN.

Antes da concatenação, uma convolução 1×1 é aplicada nas características de baixo nível para reduzir o número de canais para 256, para que não se tornem mais importantes para a rede do que as características aprendidas pelo codificador. Uma vez concatenadas, duas convoluções separáveis dilatadas 3×3 com 256 filtros são aplicadas para refinar a segmentação e definir quais recursos do codificador devem ser usados.

Os *logits* finais são obtidos aplicando-se outra convolução 1×1, com o número de filtros igual ao número de classes do conjunto de imagens no qual a rede é treinada,

seguida de outro aumento bilinear com fator 4. Os autores definem a proporção entre a resolução espacial da imagem de entrada e a resolução de saída final como passo de saída (*output stride*). Além de determinar a resolução do mapa de características de saída, este valor está associado ao tamanho do lote necessário para o treinamento da rede com ajuste fino da normalização do lote. Isso ocorre porque é necessário um lote maior para se obter estatísticas razoáveis.

O passo de saída da DeepLabv3+ é o mesmo que na DeepLabv3, porém, ao aplicar dois aumentos separadamente, o resultado alcançado é mais eficiente do que aplicar uma única reamostragem por um fator de 16.

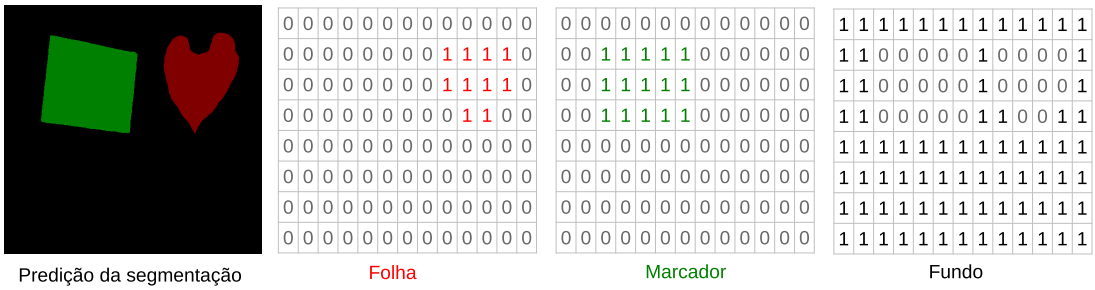
2.3.4 *Função de perda*

Para o cálculo da perda do modelo, primeiro cria-se uma codificação *one-hot* nos *logits* aumentados, conforme ilustra a Figura 7. Esta codificação se trata de uma representação binária, onde o valor 1 no vetor representa a classe correspondente e os demais valores são iguais a 0. Posteriormente, são atribuídos pesos para os *pixels* conforme suas classes (*e.g.*, peso 1,0 para todas as classes). Adicionalmente, a rede DeepLabv3+ possibilita que *pixels* não rotulados sejam desconsiderados. Para isso, atribui-se no GT um rótulo de classe ao qual o modelo deve ignorar (*e.g.*, ignorar rótulo = 255). Dessa forma, no cálculo da perda utiliza-se uma máscara com pesos, gerada de modo que os elementos referentes aos *pixels* válidos recebem o peso pré-definido e os *pixels* ignorados recebem peso 0,0. Por fim, pode-se definir a função de perda L_{seg} como uma soma, ponderada pela máscara de pesos, dos termos da entropia cruzada *softmax* para cada posição da saída:

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} t_{ij} \cdot \log\left(\frac{e^{y_{ij}}}{\sum_{k=1}^{C} e^{y_{ik}}}\right) \cdot w_{ij}, \tag{2.11}$$

onde N é o total de *pixels*, C é o número de classes, w_{ij} é peso atribuído ao *pixel* i da classe j, t_{ij} e y_{ij} são, respectivamente, o valor real e a predição.

Figura 7 – Exemplo de codificação *one-hot* para as classes folha, marcador e fundo.



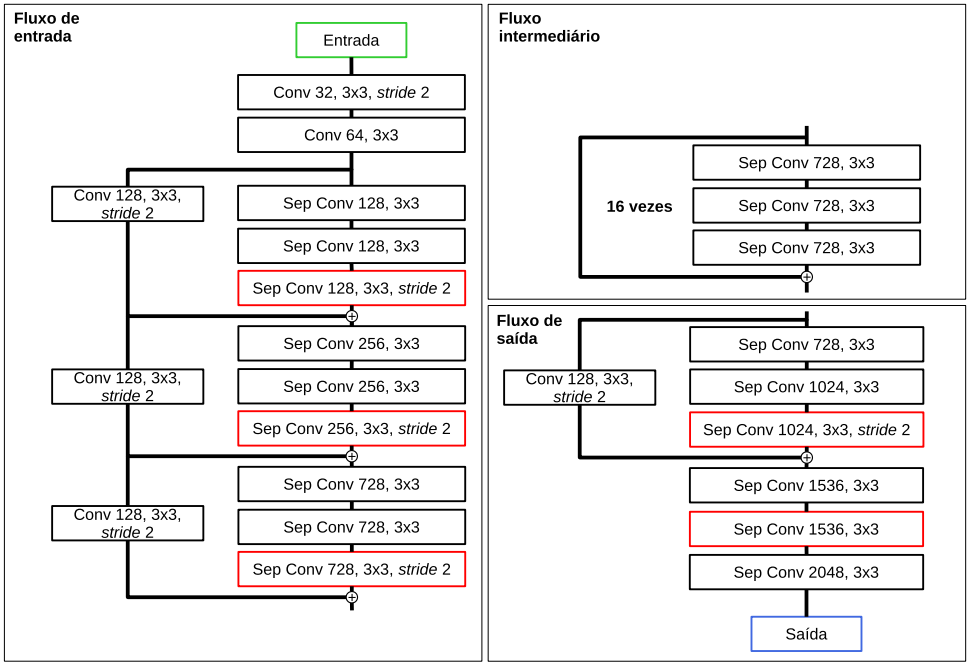
Fonte: Elaborada pela autora (2023).

2.3.5 *Aligned Xception* modificada

O modelo *Xception* (CHOLLET, 2017) foi originalmente proposto para a tarefa de classificação, tendo obtido resultados promissores no conjunto ImageNet (RUSSAKOVSKY et al., 2015). A estrutura deste modelo conta com três fluxos: de entrada, usado para reduzir a dimensão espacial da imagem de entrada; intermediário, que aprende associações e otimiza recursos; e o de saída, responsável por classificar os recursos e retornar um mapa de pontuação.

Qi et al. (2017) modificaram sua arquitetura para uma versão mais profunda (nomeada *Aligned Xception*) que alcançou um bom desempenho na tarefa de detecção de objetos. Inspirados por esta versão, Chen et al. (2018) realizaram novas alterações na arquitetura, conforme ilustra a Figura 8, substituindo todas as operações de *max pooling* por convoluções separáveis em profundidade (em vermelho na Figura 8) com tamanho do passo igual a dois. Além disso, foram adicionadas uma normalização extra do lote e funções de ativação ReLU após cada convolução em profundidade 3×3. Posteriormente, a arquitetura proposta foi pré-treinada no conjunto de imagens ImageNet.

Figura 8 – Arquitetura da rede *Aligned Xception* modificada.



Fonte: Adaptado de Chen et al. (2018).

3 TRABALHOS RELACIONADOS

Este capítulo apresenta trabalhos da literatura relacionados com medição a partir de imagens. No limite do nosso conhecimento, não foram encontrados modelos que estimam a área da superfície de um objeto com base na previsão da área relativa dos *pixels* da imagem.

Nos últimos anos, foram publicados diversos trabalhos para medição da superfície foliar. Nas abordagens convencionais, que utilizam apenas técnicas de processamento de imagens, a identificação da região da imagem pertencente à folha é comumente realizada por meio de uma filtragem por cor (LIANG; KIRK; GREENE, 2018) ou por uma limiarização (POLUNINA; MAIBORODA; SELEZNOV, 2018; TECH et al., 2018; SABOURI et al., 2021; SILVA et al., 2023). Além disso, geralmente os experimentos são realizados com as folhas planificadas e dispostas sobre um fundo que garanta contraste. Desse modo, é comum que os métodos sejam sensíveis à iluminação e aos ruídos que apareçam na imagem. Tu et al. (2021) propuseram um método de medição que combina calibração da cor de fundo, correção de distorção da câmera e correção de margem para extrair os *pixels* da folha. As folhas são fotografadas utilizando um suporte com fontes de luz fixas e há a necessidade de recalibração sempre que a câmera se move. Isto pode ser bastante trabalhoso considerando a análise de amostras com folhas de tamanhos variados. Uma possível solução seria adicionar um padrão de escala conhecido na cena e não apenas uma área de medição delimitada no plano de fundo.

Em alguns métodos recentes, rede neurais de segmentação são utilizadas para mapear a imagem, contornando-se assim o problema com ruídos das abordagens anteriores. Triki et al. (2021) propuseram a rede neural Deep Leaf, baseada na Mask R-CNN (*Mask Region Convolutional Neural Network*), para detecção e dimensionamento de folhas digitalizadas de herbário (coleção de plantas secas prensadas). Após a identificação de cada folha, as dimensões são estimadas utilizando coordenadas de *pixel* e a escala presente na imagem. A proposta do trabalho se limita a um tipo muito específico de imagem, *Digitized Herbarium Specimens* (DHS), se comparado com imagens de folhas realizadas por câmeras ou dispositivos móveis. Somado a isso, os autores puderam validar apenas as medições de largura e comprimento das superfícies (tendo estimado área e perímetro) devido à fragilidade das folhas secas. Richardson et al. (2023) desenvolveram o sistema chamado PhenoBot, baseado em DL, para a estimativa da área foliar em um ambiente comercial. Os autores treinaram um modelo baseado na rede de segmentação U-Net para gerar imagens binárias, onde as folhas estão em branco, e aplicam um segundo método para identificação de cada folha. Por fim, os respectivos números de *pixels* das folhas são correlacionados com um quadrado vermelho de calibração que, por sua vez, é identificado com uma limiarização fixa. Utilizar informação de cor para identificar a escala pode não ser efetivo em algumas condições de iluminação. Nesse sentido, a rede poderia ter sido

usada para segmentar todos os objetos da cena.

Tratando-se da medição não destrutiva da AF, em trabalho recente, Li et al. (2023) propuseram um método para monitoramento do crescimento de folhas de *Brassica napus* utilizando uma rede neural profunda para segmentar as imagens. Os autores testaram três configurações para aquisição das imagens da folha na planta: (1) posicionando a folha entre duas placas acrílicas, (2) com a folha em frente a um fundo quadriculado e (3) fixando a folha com cinco pequenos *clips* sobre um fundo branco. Diferentes modelos de segmentação também foram comparados, como PSPNet, DeepLabv3+ e U-Net. Ao final, propôs-se um modelo com base na U-Net com bloco de atenção para segmentação. A estimava da AF é realizada por um algoritmo que compara o número de *pixels* da folha com os *pixels* de calibração. A proposta requer bastante contato com as folhas, o que poderia gerar prejuízos danificando plantas mais sensíveis.

Para além da medição da superfície foliar, alguns trabalhos para monitoramento de plantas e sua área foliar total têm sido desenvolvidos. Zhang et al. (2020) investigaram o uso de uma CNN com cinco camadas convolucionais para realizar a regressão de parâmetros morfológicos associados ao crescimento de alface, como peso de folhas frescas e secas e a área foliar. A previsão de cada medida foi ligada a uma saída da camada totalmente conectada. A proposta alcançou resultados R^2 acima de 0,89 para todas as estimativas. No entanto, o método exige aquisição controlada das imagens com altura de captura fixa e vista superior da planta. Lüling, Reiser e Griepentrog (2021) apresentaram um método baseado na rede Mask R-CNN para estimativa do volume e área foliar total de repolho. O método conta com imagens RGB e de profundidade, geradas a partir de uma gravação em perspectiva vertical dos vegetais. O modelo de segmentação foi treinado para dividir as imagens coloridas em três classes de interesse. Na sequência, o número de *pixels* de cada região é convertido para centímetros e usado em diferentes equações para estimar o volume e a área foliar total. A proposta depende do ângulo e distância da câmera e a medição das folhas é realizada com base no comprimento médio das folhas visíveis.

Abordagens para medição de outros objetos com base em segmentação e DL também são encontradas na literatura. Por exemplo, Fernandes et al. (2020) utilizaram diferentes modelos para segmentar imagens de tilápia e realizar medições do corpo do peixe (área, largura, comprimento e excentricidade). A partir da segmentação, o tamanho de cada *pixel* é estimado com base em uma régua afixada na mesa. A proposta é limitada por um ambiente de captura controlado, com a câmera fixa a 0,5 m e o peixe colocado sobre um fundo verde. No contexto da construção civil, Dang et al. (2022) desenvolveram um método para medir o comprimento de rachaduras em alvenaria. Os autores treinaram diferentes modelos de DL, dentre eles U-Net, DeepLabv3+ e FPN, para segmentar as fissuras das paredes. A rede DeepLabv3+ superou os demais modelos com IoU (*Intersection over Union*) igual a 0,92. Para identificar as rachaduras, as segmentações foram combinadas com pós-processamento das imagens e um método de esqueletonização. Além disso, uma

rede baseada na Mask R-CNN foi usada para detectar os tijolos da parede e estimar a medida real de milímetro por *pixel*. Por fim, obtém-se o comprimento pela contagem dos *pixels* do esqueleto gerado. Ao utilizar a medida real de uma unidade de tijolo, os autores conseguiram desenvolver um método capaz de contornar cenários complexos.

No contexto da medicina, Brummen et al. (2021) apresentaram um modelo customizado de aprendizado para segmentar faces de pacientes e realizar medições da pálpebra e periorbitais. Para o treinamento da rede neural, dividiram-se as imagens originais na vertical gerando duas imagens, as quais foram escaladas e espelharam-se horizontalmente as imagens com os olhos esquerdos. A partir das máscaras geradas, reconstruídas para o tamanho original, o algoritmo proposto realiza as medições em *pixels* que são convertidas para milímetros com base no diâmetro da íris e a largura média da córnea. Oghli et al. (2021) propuseram um método para medir parâmetros biométricos fetais (*e.g.*, circunferência da cabeça e do abdômen, tamanho do fêmur) a partir de imagens de ultrassom. Os autores propuseram um modelo baseado na MFP-Unet para detectar regiões salientes da imagem. Além disso, um algoritmo de limiarização foi usado para pré-processamento das imagens, destacando as bordas. Posteriormente, um método para ajuste de elipses é usado para extrair esqueletos das regiões e assim estimar as dimensões. Estas propostas se assemelham as demais quanto ao ambiente controlado e aplicações para imagens específicas.

A maioria das propostas descritas neste capítulo realiza algum nível de pós-processamento dos *pixels* para realizar as medições. Pensando-se nisso, propôs-se desenvolver um método capaz de estimar diretamente o tamanho dos *pixels* dos objetos presentes na cena, uma folha saliente e um marcador. Além disso, almejou-se não recorrer a um ambiente fixo de aquisição das imagens, de modo que, a comparação entre as proporções das regiões dos objetos presentes na imagem (folha e marcador) servem como base para estimativa da área da superfície foliar.

4 CONSTRUÇÃO DA BASE DE DADOS

Neste capítulo são apresentadas todas as etapas da construção da base de dados, uma das principais contribuições deste trabalho. A Figura 9 apresenta as etapas principais deste processo, detalhadas nas seções a seguir. A Seção 4.1 aborda informações sobre a espécie escolhida e o processo de cultivo das plantas. Posteriormente, a Seção 4.2 abrange aspectos relacionados ao marcador, o dispositivo de captura e a metodologia de aquisição das imagens. Na Seção 4.3, são descritos os processos manuais de medição das dimensões foliares. A Seção 4.4 apresenta os processos realizados para anotação das imagens. Por fim, a Seção 4.5 apresenta os detalhes do processamento das imagens e a obtenção das máscaras de segmentação e para estimativa de área.

Figura 9 – Principais etapas da construção da base de imagens.

Fonte: Elaborada pela autora (2023).

4.1 CULTIVO DAS PLANTAS

Neste trabalho, propõe-se realizar a estimativa da área de superfícies foliares ainda na planta. Ao projetar uma superfície não-plana como a de uma folha, sem planá-la, pode-se subestimar ou superestimar suas dimensões. Almejando-se não intervir demasiadamente na folha, durante a captura das imagens, uma espécie ideal para a investigação seria aquela com folhas o mais planas possível.

Oliveira et al. (2018) descrevem que o feijoeiro no início de seu desenvolvimento, especificamente na fase vegetativa (V2), apresenta folhas primárias totalmente expandidas na posição horizontal. Esta característica favorece a captura de imagens das superfícies foliares no sentido perpendicular da câmera.

Adicionalmente, salienta-se que no estádio V2 ocorrem muitos ataques de pragas, tanto no solo quanto nas folhas, que reduzem a área foliar. Além disso, há maior vulnerabilidade a estresse hídrico, que reduz o incremento de área foliar ao longo do tempo. A ocorrência destes problemas no início do cultivo pode desencadear danos irreversíveis aos feijoeiros (OLIVEIRA et al., 2018). Portanto, do ponto de vista fisiológico é relevante realizar o monitoramento da cultura ao longo desta fase do desenvolvimento.

Na Figura 10, é possível observar as duas folhas no início deste estádio de desenvolvimento, que possui duração de aproximadamente 1 semana. Ao passo que, o tempo de cultivo, do plantio até a fase V2, leva de 2 a 3 semanas.

Figura 10 – Estádio de desenvolvimento das folhas primárias.

Fonte: Arquivo da autora (2022).

Para este trabalho, foram selecionados grãos de feijão-preto e o plantio foi realizado no final do mês de abril de 2022, na cidade de Ouro Branco, Minas Gerais, Latitude -20.535912, Longitude -43.711031, Brasil. Semearam-se em média 3 sementes em cada cova, feita com auxílio de enxada, ao longo de 9 fileiras de 30 plantas. O terreno utilizado nunca havia sido cultivado e possuía rejeitos de material de construção na lateral. Posteriormente, esse fato contribuiu positivamente para uma variação do fundo da cena conforme a posição da planta. A disposição das plantas no local é mostrada na Figura 11. Foram usadas 306 plantas, 612 folhas no total, para aquisição das imagens conforme descrito na Seção 4.2.

Figura 11 – Disposição das plantas no terreno.

Fonte: Arquivo da autora (2022).

4.2 AQUISIÇÃO DAS IMAGENS

Nesta seção, são fornecidos detalhes sobre a aquisição das imagens: aspectos relacionados ao marcador fiducial adotado (Subseção **4.2.1**), além de informações do dispositivo (Subseção **4.2.2**) e metodologia de captura (Subseção **4.2.3**).

4.2.1 Marcador fiducial

Como referencial de escala e perspectiva foi utilizado um marcador fiducial da biblioteca ArUco. Originalmente desenvolvidos por Garrido-Jurado et al. (2014), estes marcadores são amplamente utilizados em aplicações de realidade virtual e aumentada por apresentarem boa detecção sob condições de iluminação não uniforme (ROMERO-RAMIREZ; MUÑOZ-SALINAS; MEDINA-CARNICER, 2018). Uma das principais vantagens desta biblioteca é a possibilidade de criação de dicionários configuráveis. É possível personalizar o tamanho e o número de bits do marcador conforme a aplicação. Deste modo, a biblioteca trata apenas marcadores específicos, reduzindo o tempo de computação.

Foi definido um dicionário contendo apenas um marcador de tamanho 3×3 composto por bits iguais a [[1,0,1],[1,0,0],[1,1,1]] e largura de borda padrão (equivalente a um bit interno). Buscou-se uma combinação de bits que apresentasse linhas e ângulos bem definidos, a fim que estas características auxiliem no aprendizado da rede neural. O marcador customizado foi impresso com dimensões 5 cm $\times$ 5 cm (25 cm^2) em papel adesivo e, posteriormente, afixado em uma placa de MDF (*Medium Density Fiberboard*) de tamanho similar. Para facilitar o posicionamento do marcador na cena, foi usado um bastão de *selfie* como suporte.

4.2.2 Dispositivo de captura

Antecipando a possibilidade de desenvolvimento de uma aplicação para dispositivos móveis futuramente, optou-se por utilizar um *smartphone* modelo Galaxy M31 (SM-M315F) como dispositivo de captura. Utilizou-se a câmera do dispositivo no Modo Pro, com ajuste do foco manual e o restante das configurações (valor de ISO, velocidade do disparo, exposição, controle de branco e tom da cor) com ajuste automático. Assim, a distância focal foi fixada em 5,23 mm com abertura de f/1,8. A resolução das imagens originais em formato JPEG (*Joint Photographic Experts Group*) obtidas foi de 3468×4624 *pixels* (64-*megapixels*) e 3:4 de razão de aspecto.

Para calibração da câmera, foi usado o algoritmo de Zhang (2000) que obtém parâmetros intrínsecos e extrínsecos da câmera a partir de visualizações de um padrão de dimensões reais conhecidas. Para isso, foram capturadas 15 imagens de um tabuleiro de xadrez com 9×6 quadrados de 15 mm. O padrão foi impresso e afixado sobre uma placa de vidro. Para capturas das imagens, moveu-se o dispositivo em diferentes ângulos, a aproximadamente 20 cm de distância do plano do tabuleiro.

4.2.3 Metodologia de captura

Para a aquisição das imagens, posicionou-se o *smartphone* perpendicularmente ao plano dos objetos de interesse (folha e marcador), a uma distância aproximada de 20 cm. Observou-se os limites do foco da câmera, de modo que apenas a extensão de uma folha

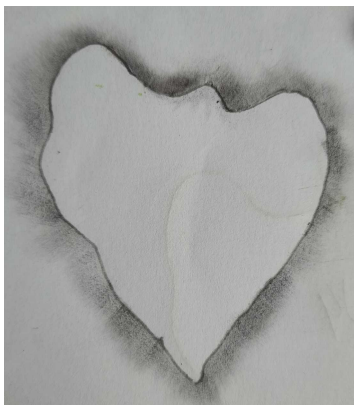
(demais folhas com oclusão) ficasse visível na cena. Além disso, procurou-se assegurar que o marcador planar e a folha estivessem aproximadamente coplanares na cena, considerando o plano médio da folha. Assim, a posição do marcador sofreu alterações, de maneira que pelo menos um de seus lados permanecesse a mesma altura que a folha. Tratando-se de todas as imagens de uma folha, buscou-se equilíbrio entre a quantidade de poses com maior e menor inclinação do marcador. A Figura 13 mostra exemplos da disposição dos objetos na cena. É possível observar que o marcador aparece em diferentes perspectivas ao redor da folha (em baixo, à esquerda e à direita).

A aquisição das imagens aconteceu no decorrer dos dias 5 a 13 de maio de 2022. Não foi utilizado *zoom* ou *flash* no momento da captura. Ao longo do processo, foram capturadas imagens no período da manhã e da tarde, em dias ensolarados com alguns momentos nublados. Além disso, foram fotografadas plantas expostas diretamente sob a luz solar e à sombra. Inicialmente um total de 9645 fotos foram capturadas e, após um processo de filtragem, 7488 permaneceram no conjunto. Foram descartadas imagens com oclusões de partes da folha ou do marcador e imagens borradas ou muito semelhantes entre si.

4.3 AFERIÇÃO DAS DIMENSÕES REAIS

Após serem fotografadas, as duas folhas do feijoeiro receberam numerações de identificação. Posteriormente, cada folha foi destacada e imediatamente contornada sobre uma folha de papel sulfite A4 evitando assim a perda de massa por desidratação. Devido à fragilidade das folhas do feijão, para extração do contorno, primeiramente foi utilizado um pigmento em pó e depois reforçou-se a forma com um lápis grafite (Figura 12). Optou-se por desconsiderar recortes no interior da superfície. Assim, nestes casos especiais, somente o contorno de maior perímetro foi obtido (Figura 13d).

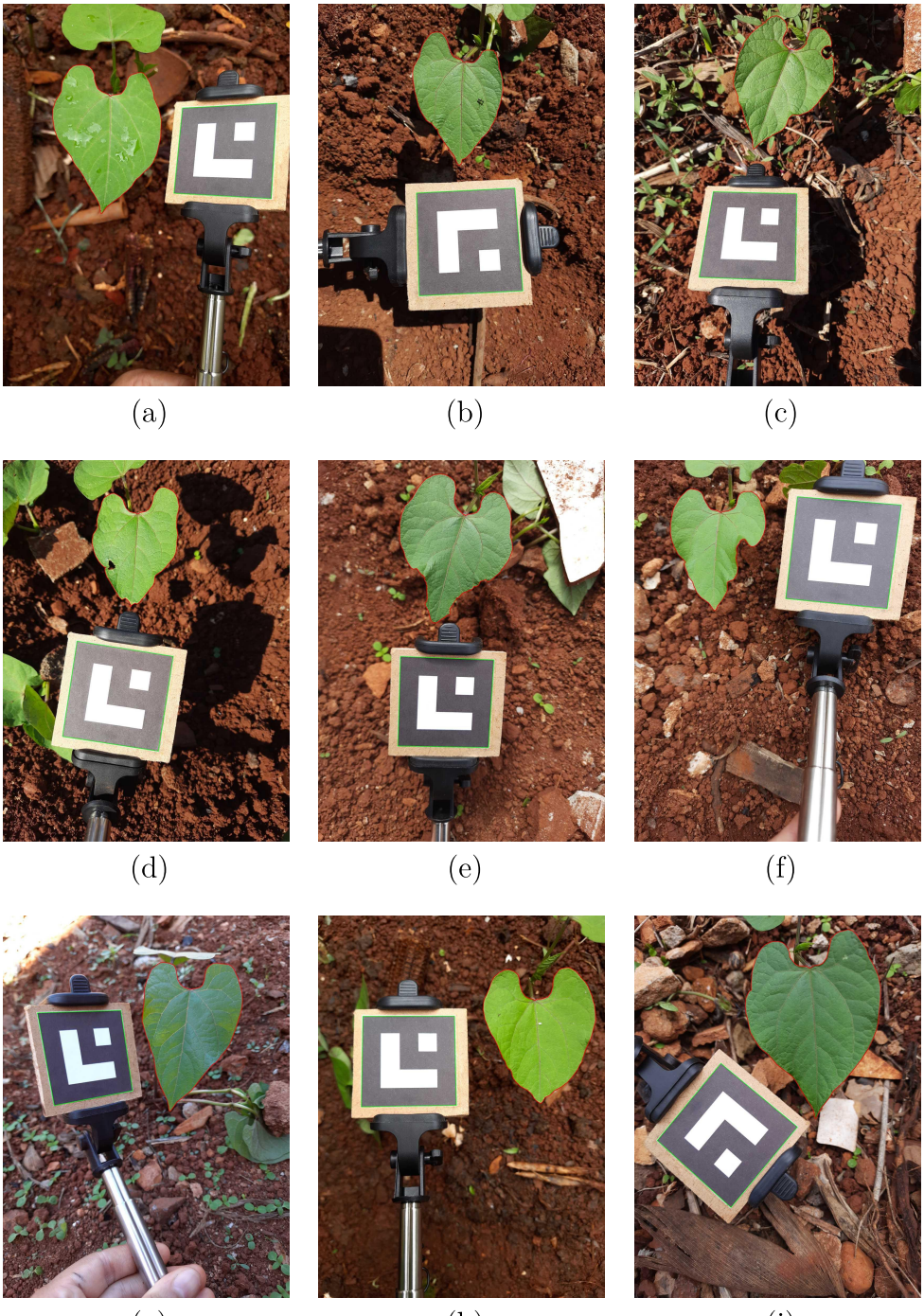
Figura 12 – Extração do contorno da folha.



Fonte: Arquivo da autora (2022).

Cada um dos contornos foi devidamente recortado e utilizado para obtenção das dimensões reais (em centímetros) das folhas. Neste trabalho, usaram-se métodos manuais

Figura 13 – Exemplos de anotação das imagens. Os contornos das folhas estão destacados em vermelho e dos marcadores em verde. As imagens (a) e (b) exibem folhas com elementos na superfície, água e um inseto, respectivamente. Em (c), (d) e (e) as folhas possuem recortes. Enquanto em (f) e (g) as folhas apresentam ondulações. A imagem (h) apresenta uma folha pequena de coloração clara. Por fim, (i) mostra uma folha maior que o marcador sobre um plano de fundo com rejeitos de construção.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fonte: Elaborada pela autora (2023).

de medição. Para a área, foi utilizado um método descrito por Pandey e Singh (2011), baseado no peso do papel. Para isso, utilizou-se um papel A4 cuja gramatura g é conhecida para realizar o contorno da folha. Na sequência, obtém-se o peso deste contorno p por uma balança de precisão. Este procedimento foi realizado pelo técnico especializado, Alex Rodrigues Borges, no laboratório de Ciências da Natureza do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) - Campus Ouro Branco. Por fim, a área a é obtida através da fórmula:

$$a = \frac{p}{g}. \tag{4.1}$$

Para estimativa do perímetro foliar utilizou-se um curvímetro. Este dispositivo permite realizar medições de percursos que contenham curvas em mapas e cartas por meio de uma roda micro dentada conectada a um ponteiro. Para obter maior precisão, foram realizadas sucessivas medições do mesmo contorno até que um mesmo valor de perímetro fosse obtido duas vezes consecutivas. A largura e o comprimento, por sua vez, foram obtidos com o auxílio de uma régua.

4.4 ANOTAÇÃO DAS IMAGENS

É de conhecimento geral que a tarefa de rotulação de dados é custosa e demorada. Para anotação deste conjunto de imagens fez-se necessário a obtenção de duas máscaras, sendo uma para segmentação das imagens e outra para estimativa de área dos objetos. Para isso, contou-se com o apoio de alunos de iniciação científica da UFJF.

Inicialmente, para rotulação dos *pixels*, é preciso identificar a folha e o marcador na imagem. A Figura 13 mostra exemplos dos polígonos gerados para efeito de segmentação. Devido à maior complexidade das folhas, um procedimento específico foi empregado para sua anotação, conforme descrito na Subseção **4.4.1**. A Subseção **4.4.2** apresenta o modo utilizado para extração da região do marcador. Os algoritmos mencionados nesta seção são implementações da biblioteca OpenCV (BRADSKI, 2000).

4.4.1 Identificação da folha

A região da imagem pertencente à folha possui um formato irregular, com ângulos desafiadores. Diante disso, a anotação em forma de polígono foi a mais adequada, por permitir uma delimitação mais precisa e flexível do seu contorno.

Para isso, o primeiro passo foi a criação de uma máscara inicial de segmentação com a ferramenta *Quick Selection Tool* disponível no Adobe Photoshop® 2023 (*Adobe Systems Incorporated*, San Jose, CA, EUA). Este procedimento foi realizado por um aluno de iniciação científica e foi essencial para facilitar a separação da região de interesse (folha)

do restante da cena. Ao final deste processo, obteve-se uma imagem binária I_b, onde a folha está em branco e o fundo em preto (Figura 14).

Figura 14 – Comparativo entre a imagem original (a) e a máscara binária gerada (b).

(a) (b)

Fonte: Arquivo da autora (2022).

Na sequência, para vetorização inicial do contorno da folha, aplicou-se o algoritmo de Suzuki e Abe (1985) na imagem I_b. Este algoritmo extrai um conjunto C:

$$C = \{C_1, C_2, \ldots, C_n\},$$

onde cada C_i é um contorno, representado pelo conjunto de pontos que o compõem:

$$C_i = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\},$$

em que cada (x_j, y_j) é a posição de um ponto (*pixel*) na imagem que forma o contorno. Caso $|C| > 1$, existem contornos que devem ser descartados, por exemplo, recortes internos da folha. Em seguida, realizou-se uma filtragem com base nas áreas em *pixels* dos contornos, calculadas utilizando a fórmula de Green (STEWART, 2009). Dessa forma, obtém-se um conjunto de áreas:

$$A = \{a_1, a_2, \ldots, a_n\},$$

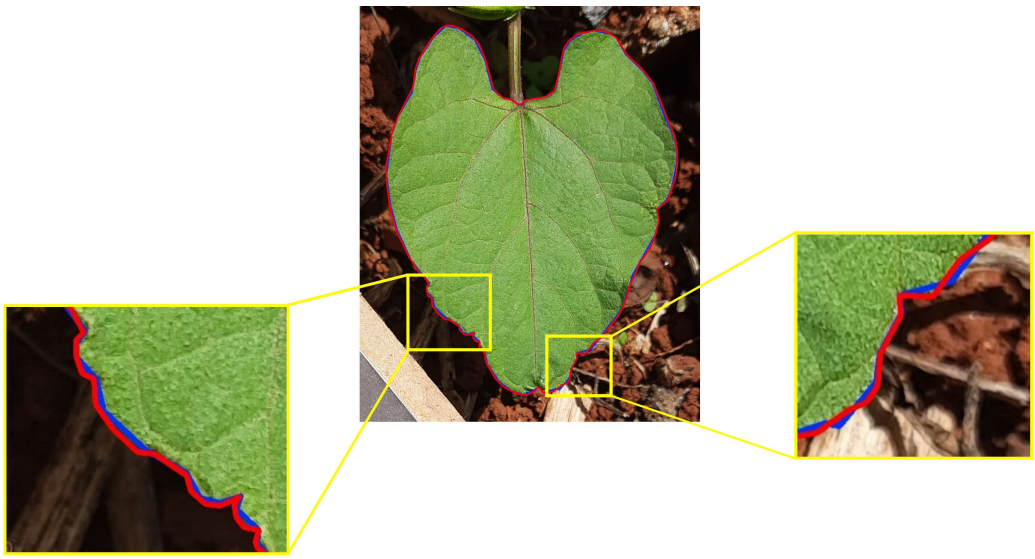
onde cada $a_i \in A$ é a área do respectivo contorno $C_i \in C$.

Considerou-se que o conjunto C_l com maior área em A representa a região da folha. Para minimizar a quantidade de pontos em C_l, foi realizada uma aproximação poligonal usando o algoritmo de Douglas e Peucker (1973), gerando um novo conjunto C_l^p que é um conjunto reduzido de C_l que representa a folha.

Por fim, foi realizado um refinamento manual do polígono utilizando, a ferramenta "*Basic image labeling toolbox*" disponível na plataforma Supervisely (DROZDOV;

KOLOMEICHENKO; BORISOV, 2023). Também foi utilizada uma alternativa a esta ferramenta[1], desenvolvida por um aluno de iniciação cientifica durante o período de anotação das imagens. Os pontos do conjunto C_l^p sofreram ajustes, sendo reposicionados conforme necessário, enquanto novos pontos foram adicionados para cobrir falhas ao longo da borda. Este processo foi realizado em lotes de 15 folhas, com o apoio de alunos de iniciação cientifica. Ao término desta etapa, uma rigorosa revisão da consistência das anotações foi realizada, desenhando-se os polígonos sobre a imagens originais. A Figura 15 ilustra um comparativo entre o contorno extraído da máscara gerada pelo Photoshop (azul) e o contorno revisado (vermelho). Os pontos obtidos foram normalizados pela altura da imagem e armazenados em um arquivo XML (*Extensible Markup Language*).

Figura 15 – Comparativo entre a segmentação inicial da folha (azul) e o contorno revisado (vermelho). Os destaques salientam partes da folha com curvaturas mais irregulares.



Fonte: Elaborada pela autora (2023).

4.4.2 Identificação do marcador

O marcador possui uma geometria plana simples, portanto, extrair os 4 pontos referentes aos seus cantos é suficiente para representá-lo. Inicialmente, foi utilizado o algoritmo de Garrido-Jurado et al. (2014), implementado pelo módulo ArUco, o qual realiza uma detecção dos marcadores presentes em uma imagem com base em um dicionário pré-definido. Dessa forma, obtém-se um conjunto M:

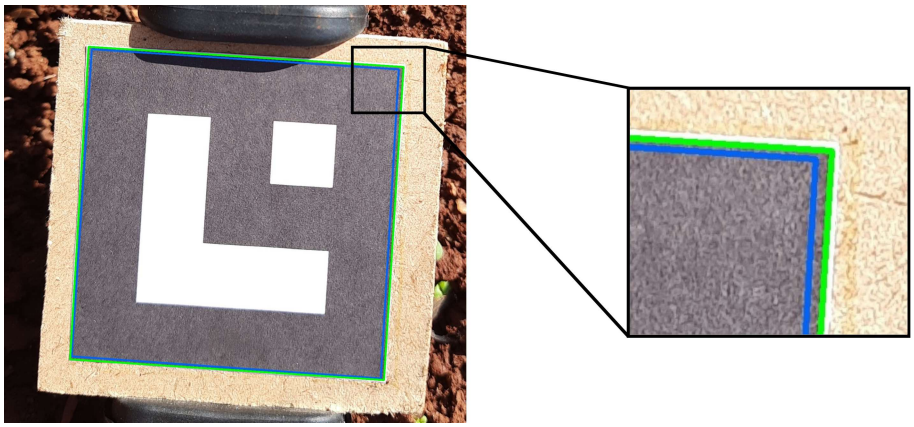
$$M = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\},$$

onde cada (x_j, y_j) é a posição na imagem referente a um dos cantos do marcador. Na sequência, este conjunto foi exportado e revisado utilizando-se as ferramentas de anotação

[1] https://paulo-rozatto.github.io/subvisor/

descritas na Subseção **4.4.1**. Para análise visual, desenhou-se o polígono formado pelos pontos do conjunto M sobre a imagem original. A Figura 16 mostra um comparativo entre o marcador detectado pelo ArUco e o marcador revisado.

Figura 16 – Comparativo entre a detecção inicial do marcador (azul) e o polígono revisado (verde). O destaque mostra que o canto detectado pelo ArUco ficou deslocado do esperado.



Fonte: Elaborada pela autora (2023).

Posteriormente, os vetores de translação $\mathbf{t}$ e rotação $\mathbf{r}$ do marcador foram obtidos com o algoritmo de estimativa de pose de Collins e Bartoli (2014). Este algoritmo utiliza os cantos do marcador juntamente com a matriz de câmera e os coeficientes de distorção obtidos a partir da calibração da câmera, descrita na Subseção **4.2.2**. Feito isso, usou-se a fórmula de Rodrigues para obter a matriz de rotação $\mathbf{R}$ (3×3):

$$\mathbf{R} = \cos(\theta)\mathbf{I} + (1 - \cos(\theta)\mathbf{r}\mathbf{r}^T + \sin(\theta) \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix}, \tag{4.2}$$

onde θ é igual à norma de $\mathbf{r}$ e $\mathbf{I}$ é a matriz identidade. Desta forma, obtém-se a matriz extrínseca homogênea $\mathbf{M}_{ext}$ (4×4):

$$\mathbf{M}_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.3}$$

onde os valores de r_{11} a r_{33} pertencem à matriz $\mathbf{R}$ e t_x, t_y, t_z ao vetor $\mathbf{t}$. A matriz de projeção do modelo $\mathbf{M}_{proj}$ (4×3) é obtida multiplicando-se a matriz da câmera pela matriz homogênea:

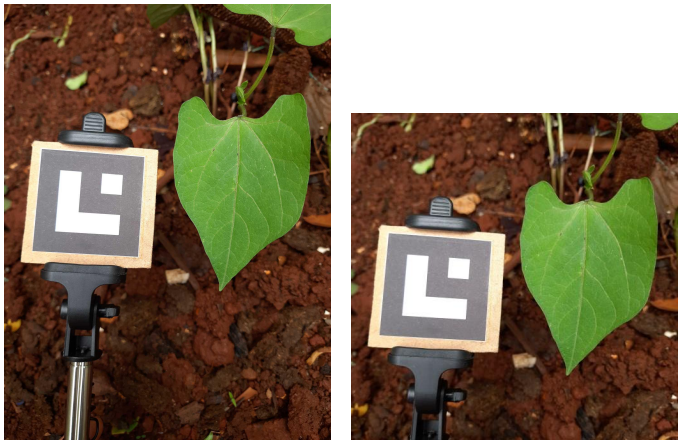
$$\mathbf{M}_{proj} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}. \tag{4.4}$$

Por fim, os cantos revisados do marcador foram normalizados pela altura da imagem e armazenados em um arquivo XML juntamente com a área real do marcador, os coeficientes de distorção, os vetores de $\mathbf{t}$ e $\mathbf{r}$ e as matrizes $\mathbf{M}_{cam}$, $\mathbf{R}$, $\mathbf{M}_{ext}$ e $\mathbf{M}_{proj}$.

4.5 PROCESSAMENTO DAS IMAGENS

As imagens originais foram capturadas em alta resolução, com dimensões iguais a 3468×4624 *pixels*. No entanto, para o treinamento de redes neurais normalmente usam-se entradas menores, devido ao elevado consumo de recursos computacionais e ao tempo de processamento necessário. Somado a isso, os arquivos originais também são consideravelmente grandes (em torno de 4 MB). Portanto, fez-se necessário o redimensionamento das imagens para um tamanho gerenciável. Antes disso, fez-se uma transformação da razão de aspecto original 3:4 em uma razão de 1:1, visando adaptar melhor as imagens às entradas típicas de redes como a DeepLabv3+ e, consequentemente, eliminar informações menos relevantes do fundo da cena. Na Figura 17b, é possível observar que a extensão da segunda folha (canto superior direito) e o bastão (canto inferior esquerdo) foram descartados da imagem reduzida em comparação à imagem original (Figura 17a).

Figura 17 – Comparativo entre a imagem original (a) e a imagem reduzida (b).



(a) (b)

Fonte: Arquivo da autora (2022).

Inicialmente, para auxiliar no deslocamento vertical da origem da imagem original $\mathbf{I}_{orig}$, calculou-se uma caixa envolvente com base nas anotações da folha e do marcador, de modo que os objetos ficassem por completo na nova imagem reduzida $\mathbf{I}_{redu}$. Para isso, a altura da região calculada deveria ser menor ou igual à largura original da imagem. Na sequência, a imagem foi recortada e aplicou-se uma amostragem por área para redução da

resolução, diminuindo o fenômeno de *aliasing*. Desta forma, ao final do processo obteve-se uma nova imagem de 512×512 *pixels*.

A Subseção **4.5.1** descreve a obtenção das máscaras para a tarefa de segmentação e a Subseção **4.5.2** o processo realizado para as máscaras de estimativa da área dos objetos (folha e marcador).

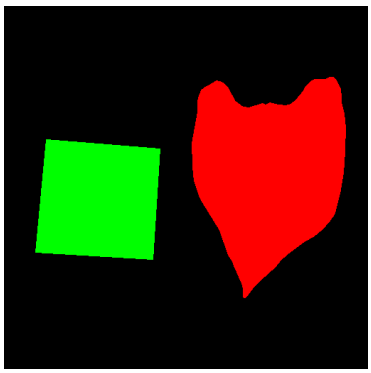
4.5.1 Máscara de segmentação

Para a tarefa de segmentação, os *pixels* das imagens foram rotulados em 3 classes: fundo (0 - preto), folha (1 - vermelho) ou marcador (2 - verde), conforme ilustra a Figura 18. Para isso, foram usados os conjuntos C_l e M anotados para identificação da folha e do marcador, conforme descrito na Seção 4.4. Cada ponto do conjunto foi transportado de $\mathbf{I}_{orig}$ para $\mathbf{I}_{redu}$, por meio de uma transformação de escala:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} r & 0 & 0 \\ 0 & r & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{4.5}$$

onde (x, y) é o ponto ou *pixel* na imagem original, r é a razão entre a resolução original e reduzida, o_y é a altura da nova origem e (x', y') é o *pixel* resultante na imagem reduzida. Por fim, a máscara ou *ground truth* (GT) da segmentação foi obtido desenhando-se polígonos fechados formados pelos pontos de C_l e M, conforme cada um dos rótulos, sobre uma imagem preta.

Figura 18 – Exemplo da máscara de segmentação. Os *pixels* pretos, vermelhos e verdes representam o fundo, a folha e o marcador, respectivamente.



Fonte: Arquivo da autora.

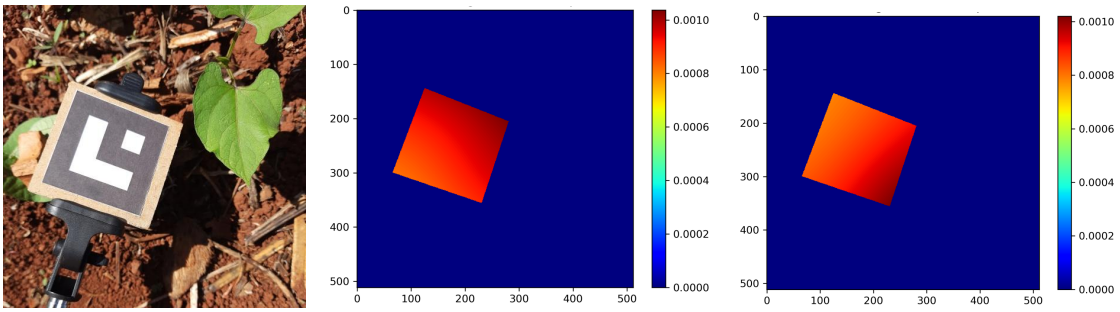
4.5.2 Máscara para estimativa de área

Para a tarefa de estimativa de área, o GT proposto é uma matriz em que cada posição contém o tamanho do *pixel*, estimado com base na área esperada dos objetos (Seção 4.3). Considerando que o fundo não possui objetos com dimensões a serem estimadas, as posições da matriz referentes a ele foram igualadas à zero.

A região da folha também foi rotulada uniformemente, de modo que o tamanho do *pixel* foi definido calculando-se a razão entre a área real (cm^2) da folha e o total de *pixels* daquela região na imagem. Para o marcador, por sua vez, primeiro foi necessário contornar o problema de ambiguidade de pose do ArUco em algumas imagens. Isso ocorre devido à estimativa de pose com uma única imagem de um objeto planar possuir duas possíveis soluções (BENLIGIRAY; TOPAL; AKINLAR, 2019).

Neste trabalho, analisou-se todas as rotações possíveis dos cantos do marcador até encontrar a ordem correta, verificando-se visualmente cada solução via mapa de cores como mostra a Figura 19.

Figura 19 – Comparativo entre a imagem reduzida (a), o mapa de cores da ordem correta dos cantos do marcador (b) e um exemplo de solução incorreta (c).



(a) imagem reduzida (b) mapa de cores correto (c) mapa de cores incorreto

Fonte: Elaborada pela autora (2023).

Após esta etapa, para rotular o marcador, considerou-se a perspectiva da cena, projetando os quatro vértices de cada *pixel* que o compõem no plano associado ao marcador no espaço tridimensional. Dessa forma, dado um *pixel* (x, y) pertencente à região do marcador, as coordenadas de cada vértice podem ser obtidas da seguinte forma:

$$\begin{aligned} \mathbf{v_1} &= (x_1 - 0,5, y_1 - 0,5), \\ \mathbf{v_2} &= (x_2 - 0,5, y_2 + 0,5), \\ \mathbf{v_3} &= (x_3 + 0,5, y_3 + 0,5), \\ \mathbf{v_4} &= (x_4 + 0,5, y_4 - 0,5), \end{aligned} \tag{4.6}$$

onde $\mathbf{v_1}$, $\mathbf{v_2}$, $\mathbf{v_3}$ e $\mathbf{v_4}$ referem-se aos cantos superior-esquerdo, superior-direito, inferior-direito e inferior-esquerdo, respectivamente. Este conjunto V de vértices foi obtido da imagem original com resolução 3468×4624 *pixels*, para uma maior precisão no cálculo. Além disso, a distorção da lente foi removida, obtendo os pontos (x'_i, y'_i) dos cantos em coordenadas de tela (*e.g.*, função *undistortPoints* do módulo ArUco). Na sequência, cada ponto (x'_i, y'_i), $i \in \{1, 2, 3, 4\}$ foi convertido para o SCC multiplicando-o pela inversa da

matriz de câmera (Equação 2.7):

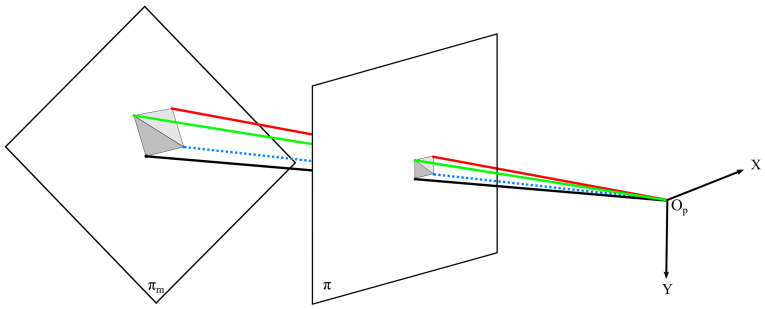
$$\mathbf{d}_i = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_c \\ 0 & f_y & v_c \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}, \tag{4.7}$$

onde $\mathbf{d}_i = (x_i, y_i, 1)$ indica a direção 3D do vértice $\mathbf{v_i}$. O conjunto de valores que compõem a matriz de câmera são descritos na Seção 2.1.

Para encontrar o plano relacionado ao marcador, extraiu-se o vetor normal unitário contido na matriz de rotação e usou-se o vetor de translação como ponto situado no plano. Substituindo esses valores na equação geral do plano, $ax + by + cz + d = 0$, incluindo o cálculo do termo d, obtém-se a equação do plano do marcador π_m. Por fim, tomando-se as interseções em π_m das retas que passam pelo ponto do observador (0, 0, 0) e cada direção $\mathbf{d}_i$ dos vértices, obtém-se o rótulo atribuído no GT calculando-se a soma das áreas dos dois triângulos formados sobre o plano π_m, conforme ilustra a Figura 20. Ao final, obtém-se uma imagem com valores reais que contém as áreas calculadas para todos os *pixels* do marcador.

É importante salientar que a área projetada pode não somar 25 cm^2. Dessa forma, optou-se por utilizar apenas as imagens dos marcadores com uma diferença absoluta entre área projetada e área real menor ou igual a 1 cm^2.

Figura 20 – Representação da projeção dos cantos do *pixel* no plano π_m do marcador passando pelo plano π de formação da imagem.



Fonte: Elaborada pela autora.

5 MÉTODO PROPOSTO: REDE ESTIMADORA DE ÁREA FOLIAR

Este capítulo apresenta a segunda contribuição do presente trabalho, uma nova arquitetura de rede neural profunda para estimativa da área da superfície foliar. Uma rede para prever a área relativa dos *pixels* da imagem foi construída a partir da modificação de uma rede neural de segmentação semântica. O propósito foi a obtenção de um modelo capaz de identificar e comparar as proporções dos objetos de interesse representados por uma imagem. A Seção 5.1 aborda o ajuste fino realizado no modelo base adotado. Na sequência, a Seção 5.2 descreve a proposta para rede estimadora de área foliar. Por fim, a Seção 5.3 apresenta a função de perda proposta para o novo modelo.

5.1 AJUSTE FINO DA DEEPLABV3+

Conforme apresentado no Capítulo 4, este trabalho propõe realizar a estimativa da AF a partir de imagens da folha ainda na planta, acompanhada de um marcador fiducial. Para alcançar este objetivo, é necessário que a rede neural proposta aprenda a isolar os objetos de interesse do restante da cena e obter suas informações espaciais. A função de modelos de segmentação é mapear *pixels* em classes.

A rede proposta neste trabalho se baseia na arquitetura da DeepLabv3+, detalhada na Seção 2.3, a qual tem alcançado resultados significativos para segmentação semântica em diversas aplicações e contextos. O aprendizado da DeepLabv3+ é supervisionado, utilizando máscaras rotuladas para o treinamento. Para isso, foram utilizadas as máscaras rotuladas de acordo com processo descrito na Subseção **4.5.1**.

Propôs-se incluir um novo decodificador na estrutura da rede, o qual será descrito em detalhes na Seção 5.2. Durante o treinamento do modelo completo, realiza-se um ajuste fino nos pesos da rede original, responsáveis pela segmentação. Esse ajuste objetiva treinar o modelo para localizar a folha saliente e o marcador, enquanto o restante da cena é segmentado como fundo (Figura 18).

Como extrator de características utilizou-se a rede neural *Aligned Xception* modificada, descrita na Subseção **2.3.5**, especificamente a versão Xception-65 que conta 65 camadas convolucionais. A escolha da Xception como extrator de características se deu pelo fato de que ela é comumente utilizada na DeepLabv3+, tendo alcançado anteriormente resultados rápidos e acurados na classificação de imagens e detecção de objetos (CHEN et al., 2018).

Para o treinamento, foram reutilizados todos os pesos da DeepLabv3+ pré-treinados no conjunto PASCAL VOC 2012 (EVERINGHAM et al., 2015), exceto os *logits*, visto que o número de classes para estimativa de área foliar é menor (3 classes: fundo, folha e marcador). A função de perda padrão usada no modelo é apresentada na Subseção **2.3.4**. Além disso, optou-se por desabilitar o ajuste de parâmetros da normalização do lote por

limitação na memória da GPU.

5.2 DECODIFICADOR DE ÁREA PROPOSTO

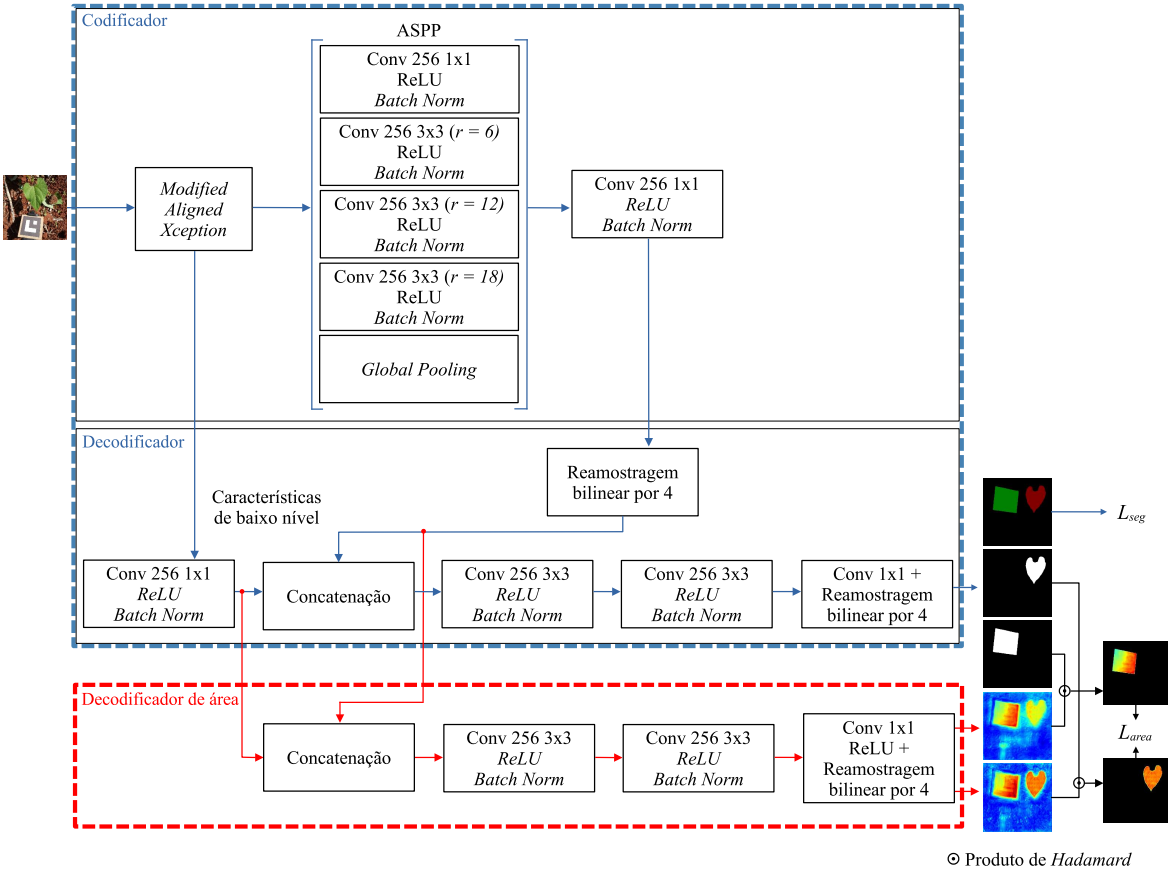
Nesta seção, é apresentado o decodificador proposto para estimativa de área, que é uma das principais contribuições deste trabalho. A proposta do decodificador parte da hipótese principal de que é possível adaptar uma CNN para realizar a regressão da área dos *pixels*. O objetivo é obter uma arquitetura capaz de relacionar informações espaciais entre dois objetos para prover medidas relativas para os elementos de imagens que compõem os objetos. Nesse sentido, assume-se que o modelo seja capaz de considerar perspectiva e proporção das regiões projetadas (folha e marcador). Em especial, assume-se que a superfície planar do marcador com áreas conhecidas para cada pixel sirva de referência para estimar a área foliar.

Para isso, propôs-se estender a arquitetura da DeepLabv3+ com um segundo decodificador, destacado em vermelho na Figura 21. Esse decodificador tem como propósito remapear a representação da imagem aprendida pelo codificador na estimativa do tamanho dos *pixels*. Além disso, para o treinamento supervisionado, propôs-se utilizar máscaras de área para os marcadores usando-se estimativa de pose, conforme descrito na Subseção **4.5.2**. Desta maneira, a perspectiva da cena é aprendida com base no marcador. Na rede original (destacada em azul na Figura 21), o decodificador é um módulo extra para refinar a segmentação, sobretudo nas bordas dos objetos.

O decodificador proposto para estimativa da AF recebe as mesmas características de entrada do decodificador original, que são concatenadas. Sua conexão na DeepLabv3+ adiciona caminhos exclusivos para treinamento do codificador. O intuito é que as características concatenadas no decodificador de área sejam transformadas pela sequência de convoluções 3×3 com 256 filtros, as quais são usadas, nesta ramificação, para inferir a área dos *pixels*. Os recursos extraídos são então combinados por uma convolução 1×1 gerando os *logits* finais. Para essa convolução, optou-se por utilizar a função de ativação ReLU, visto que valores de área não podem ser negativos. Além disso, essa ramificação gera dois canais, um para folha $\mathbf{O}^f$ e outro para o marcador $\mathbf{O}^m$, pois são os únicos objetos na cena que devem ter as áreas estimadas. Ao longo da pesquisa, observou-se que a geração de apenas um canal com todas as áreas tem desempenho inferior. Por fim, realiza-se o aumento bilinear por um fator 4, da mesma forma que o decodificador original. Os dois decodificadores trabalham paralelamente, produzindo duas saídas, uma para segmentação e outra para a estimativa da AF. O intuito é que o modelo associe a localização dos objetos com a informação de área dos *pixels*. Nesse esquema, a segmentação semântica no primeiro decodificador é preservada, sendo possível combinar as duas saídas da rede para determinar a área apenas da região da folha e do marcador.

Neste contexto, a segmentação semântica da DeepLabv3+ define a localização dos

Figura 21 – Fluxograma do método proposto. Os destaques em vermelho salientam as camadas do decodificador de área.



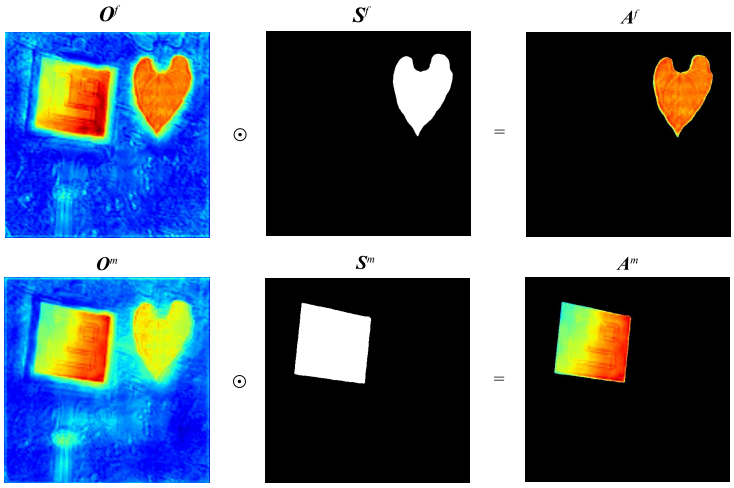
Fonte: Elaborada pela autora (2023).

pixels da folha e do marcador através das máscaras binárias $\mathbf{S}^f$ e $\mathbf{S}^m$, respectivamente. A Figura 22 mostra exemplos destas máscaras, onde os *pixels* que compõem os objetos de interesse são iguais a 1 e os demais 0. O decodificador de área provê duas imagens com estimativas de área: uma para a folha $\mathbf{O}^f$ e uma para o marcador $\mathbf{O}^m$. A área estimada para cada objeto é obtida pelos produtos de Hadamard (HORN, 1990):

$$\begin{aligned} \mathbf{A}^f &= \mathbf{O}^f \odot \mathbf{S}^f, \\ \mathbf{A}^m &= \mathbf{O}^m \odot \mathbf{S}^m, \end{aligned} \tag{5.1}$$

onde as matrizes $\mathbf{A}^f$ e $\mathbf{A}^m$ contêm a estimativa de área apenas para os *pixels* da folha e do marcador, respectivamente. A Figura 22 mostra um exemplo de combinação das saídas dos decodificadores. É possível observar que estimativas para ambos os objetos, folha e marcador, aparecem nos dois canais de saída $\mathbf{O}^f$ e $\mathbf{O}^m$. Nota-se que outros elementos aprendidos pela segmentação, como o suporte do marcador e a folha menos saliente, têm uma pequena resposta do decodificador de área. As estimativas finais $\mathbf{A}^f$ e $\mathbf{A}^m$ contêm apenas áreas estimadas para *pixels* da folha e do marcador. Portanto, a resposta do decodificador de área para o fundo é ignorada nesta proposta, pois não se conhece sua geometria e essa região não é de interesse.

Figura 22 – Representação do produto de Hadamard entre as matrizes extraídas dos *logits* da segmentação $\mathbf{S}^f$ e $\mathbf{S}^m$ pelos canais da estimativa da área $\mathbf{O}^f$ e $\mathbf{O}^m$ da folha e do marcador, respectivamente. As estimativas de área resultantes $\mathbf{A}^f$ e $\mathbf{A}^m$ são usadas na função de perda do decodificador proposto.



Fonte: Elaborada pela autora (2023).

Posteriormente, obtém-se a AF somando-se os valores previstos para a folha. Em síntese, abordou-se a tarefa de estimativa da área como uma inferência fina, na qual, previsões densas inferem as áreas dos *pixels*. É importante ressaltar que uma mesma região (conjunto dos *pixels* de uma classe de objeto) como a do marcador contém *pixels* de diferentes áreas.

5.3 FUNÇÃO DE PERDA

Com o objetivo de prever valores de área para os *pixels*, a função de perda para treinamento do decodificador proposto deve penalizar discrepâncias entre as predições realizadas pelo modelo e os valores de área esperados. Utiliza-se o erro quadrático médio (*Mean Squared Error* – MSE), frequentemente usada para avaliar modelos de regressão e controlar previsões atípicas.

Para o marcador, o MSE é calculado entre a estimativa do modelo $\mathbf{A}^m$ e a máscara de área do marcador $\mathbf{E}^m$ (GT) (Seção **4.5.2**). A máscara de área média da folha $\mathbf{E}^f$ (GT) é calculada distribuindo-se uniformemente a área conhecida da folha para todos os *pixels* que a formam. Embora de fácil aplicação, essa abordagem tem a desvantagem de não considerar a geometria e a perspectiva da folha, informações de difícil obtenção durante a construção da base de dados. Assim, para a folha, o MSE é calculado entre a estimativa do modelo $\mathbf{A}^f$ e essa máscara de área média da folha $\mathbf{E}^f$.

As máscaras de área que compõem o GT são compostas por valores decimais na ordem de 10^{-3}. Para beneficiar a rede, multiplicou-se um escalar β pelas máscaras $\mathbf{E}^f$ e $\mathbf{E}^m$. Definiu-se empiricamente o valor $\beta = 1000$, fazendo com que as previsões deixem de

ser realizadas em centímetros quadrados, conforme sugere a rotulação original das imagens. Dessa forma, a função de perda proposta para o decodificador de área é:

$$L_{area} = \frac{1}{N}\sum_{i=1}^{N}(a_i^f - \beta t_i^f)^2 + \frac{1}{N}\sum_{i=1}^{N}(a_i^m - \beta t_j^m)^2, \tag{5.2}$$

onde N é o número de *pixels* das máscaras, cada valor $t_i^f \in \mathbf{E}^f$ é a área conhecida do i-ésimo *pixel* da folha, e $a_i^f \in \mathbf{A}^f$ é a estimativa da área para este *pixel*. Analogamente, os valores $t_i^m \in \mathbf{E}^m$ e $a_i^m \in \mathbf{A}^m$ são, respectivamente, a área conhecida e a predição da área do i-ésimo *pixel* do marcador. A perda do decodificador de área proposto, por conseguinte, é a soma dos erros da estimativa de área da folha e do marcador.

O método proposto, ilustrado na Figura 21, conta com a supervisão de duas funções de perda: a função L_{seg}, definida pela Equação 2.11 para medir o erro original do DeepLabv3+ para segmentação, e a função de perda L_{area} definida anteriormente pela Equação 5.2 para medir o nível de discrepância da estimativa da área relativa dos *pixels*. Dessa forma, o custo total é dado por:

$$L_{total} = L_{seg} + L_{area}. \tag{5.3}$$

Assim, busca-se minimizar o erro do treinamento do modelo para conseguir a cada passo a segmentação dos objetos que, por usa vez, indicam onde a estimativa de área deve ser considerada. Essa localização dos objetos também é reforçada pela função de perda do decodificador de área.

6 RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados os experimentos realizados para avaliar o modelo de estimativa de área foliar proposto. A Seção 6.1 apresenta a divisão das imagens para o treinamento e avaliação do modelo. Na Seção 6.2, descreve-se a configuração dos hiperparâmetros e o protocolo de treinamento. A Seção 6.3 apresenta a análise quantitativa e a Seção 6.4 a análise qualitativa dos resultados. Por fim, na Seção 6.5, realiza-se uma discussão geral dos experimentos realizados.

6.1 BASE DE DADOS

O conjunto de imagens próprio desta dissertação, apresentado no Capítulo 4, contém imagens de 612 folhas distintas. No entanto, em razão da anotação das imagens não ter sido completamente finalizada em tempo da defesa de mestrado da autora, devido à grande quantidade de trabalho manual necessária, os experimentos foram realizados usando somente as primeiras 300 folhas do conjunto, totalizando 3374 imagens.

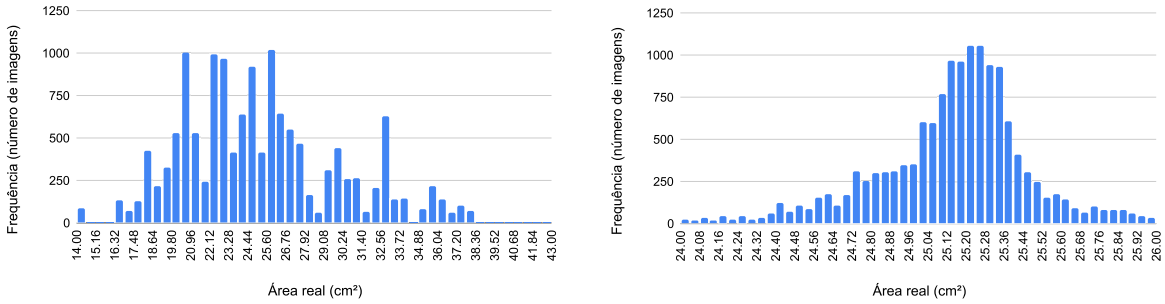
Foram selecionadas aleatoriamente 70% destas folhas para o treinamento e 30% para teste. É importante ressaltar que se mantiveram as imagens de cada folha sempre no mesmo conjunto. No conjunto de treinamento, usou-se a biblioteca Pillow (CLARK et al., 2015) para realizar rotações de 90°, 180° e 270° graus, espelhamento vertical e horizontal nas imagens originais, totalizando 14040 imagens de 210 folhas. Utilizou-se a interpolação bicúbica na aplicação das transformações, visando manter a qualidade das imagens.

A Figura 23 mostra o histograma com a distribuição das áreas das folhas e dos marcadores do conjunto de treinamento. A área foliar média deste subconjunto é de 25,243 cm^2 com desvio padrão de 4,811 cm^2. Enquanto a área projetada média dos marcadores é igual a 25,144 cm^2 com desvio de 0,313 cm^2. A área projetada se difere da área real do marcador devido ao processo descrito na Subseção **4.5.2** para obtenção dos valores esperados de seus *pixels*. O conjunto de teste é consequentemente formado pelas 90 folhas restantes, com 1033 imagens, sendo sua distribuição exibida no histograma da Figura 24. A AF média deste conjunto é de 25,076 cm^2 com desvio 4,655 cm^2. Para os marcadores, a área projetada média é de 25,129 cm^2 com desvio de 0,319 cm^2. Portanto, foi possível alcançar uma distribuição semelhante nos dois conjuntos.

6.2 CONFIGURAÇÃO E PROTOCOLO DE TREINAMENTO

A rede neural proposta foi implementada na plataforma Tensorflow 1.15.0. O uso desta versão é motivada pela implementação original do modelo base desta proposta. Todos os modelos foram executados utilizando uma GPU NVIDIA GeForce RTX 2080 Ti com 12GB de memória. Os modelos foram treinados por 55 mil *steps*, número observado

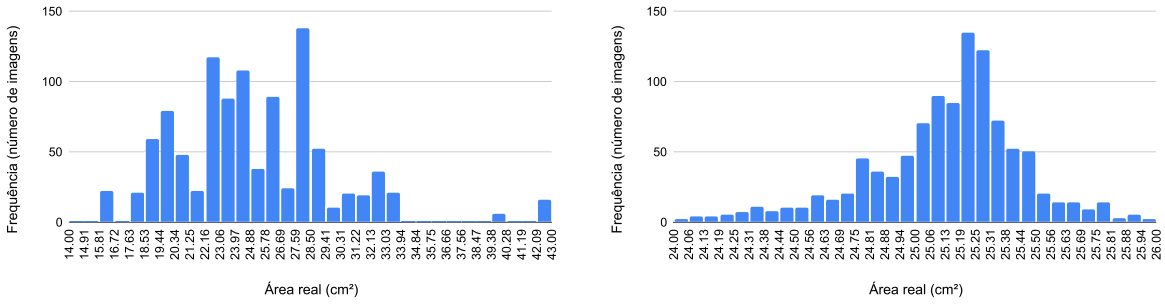
Figura 23 – Histograma da área no subconjunto de treinamento: (a) distribuição da área foliar e (b) área projetada do marcador.



(a) (b)

Fonte: Elaborada pela autora (2023).

Figura 24 – Histograma da área no subconjunto de teste: (a) distribuição da área foliar e (b) área projetada do marcador.



(a) (b)

Fonte: Elaborada pela autora (2023).

ao longo da pesquisa como suficiente para convergência do modelo.

O tamanho de entrada utilizado é de 512×512 com tamanho de lote (*batch size*) fixo em 4. A dimensão da imagem de entrada e o tamanho do lote foram inspirados em artigos que usam a DeepLabv3+ (AYHAN; KWAN, 2020; WU et al., 2021; ZHU et al., 2023). Além disso, o tamanho exíguo da memória na GPU impediu experimentos com lotes maiores.

Dado que o treinamento ocorre por *steps* e não por épocas, para cada experimento foram salvos diferentes modelos para avaliação posterior. Para escolha da taxa de aprendizado inicial mais adequada e seleção do melhor modelo treinado, foi realizado um estudo apresentado na Subseção **6.2.1**.

Inspirando-se nos resultados alcançados no trabalho de Wu et al. (2021) para segmentação de imagens de folhas de alface, decidiu-se utilizar os hiperparâmetros padrões de treinamento da rede DeepLabv3+, listados na Tabela 1. As taxas de dilatação 6, 12, 18 são usadas como padrão nas convoluções do módulo ASPP. Além disso, a regularização

Ridge ou L2 é utilizada no aprendizado, com decaimento do peso de $4 \cdot 10^{-5}$.

A política de aprendizado padrão usada por Chen et al. (2018) é a polinomial, definida como:

$$\eta = \eta_0 \cdot \left(1 - \frac{i}{N_i}\right)^{power}, \tag{6.1}$$

onde η_0 é a taxa de aprendizado inicial, i é o número da iteração, N_i é o total de iterações. O termo *power* (configurado com valor 0,9) controla o decaimento da taxa de aprendizado corrente η.

Como otimizador, é utilizado o *Momentum*, com termo *momentum* $(m) = 0{,}9$. A descida do gradiente (*Gradient Descent* – GD) definida por $\theta \leftarrow \theta - \eta\nabla_\theta J(\theta)$, pode levar muito tempo para convergir. Inspirado-se na física, o termo m e um fator de desconto γ são inseridos no cálculo do GD:

$$\begin{aligned} m &\leftarrow \gamma m + \eta\nabla_\theta J(\theta) \\ \theta &\leftarrow \theta - m, \end{aligned} \tag{6.2}$$

onde γ é usado para adicionar mais importância às últimas atualizações.

Tabela 1 – Hiperparâmetros padrões de treinamento da DeepLabv3+.

Parâmetros	Valores
Taxas de dilatação do ASPP	[6, 12, 18]
Decaimento do peso	0,00004
Política de aprendizado	Polinomial
power	0,9
Otimizador	*Momentum*
momentum (m)	0,9

Fonte: Elaborada pela autora (2023).

O codificador e o decodificador da segmentação foram inicializados com pesos pré-treinados, conforme descreve a Subseção 5.1. As camadas do decodificador proposto foram inicializadas com valores iniciais a partir de funções disponíveis no TensorFlow 1.15.0. As convoluções separáveis são inicializadas com uma distribuição normal truncada, com desvio padrão de 0,33 e média 0,0. Para a convolução tradicional são usados pesos do inicializador Xavier, projetado por Glorot e Bengio (2010) para extrair amostras de uma distribuição uniforme.

6.2.1 Seleção do melhor modelo

Para um bom funcionamento do modelo, alguns hiperparâmetros precisam ser definidos. A taxa de aprendizado η é um deles, responsável por indicar o ritmo no qual os pesos do modelo são atualizados. A sua função principal é controlar o tamanho da etapa na descida do gradiente. Uma taxa muito pequena pode tornar o tempo de aprendizado

da rede proibitivo, enquanto uma alta taxa pode implicar em oscilações no treinamento, impossibilitando a convergência.

Dessa forma, um estudo inicial é necessário para definir a taxa de aprendizado mais adequada para convergência do modelo. Somado a isso, fez-se necessário realizar uma análise das possibilidades de seleção do melhor modelo treinado.

O critério empregado para avaliação quantitativa dos resultados foi a taxa de erro relativa (*Relative Error Rate* – RER), a qual é a porcentagem da diferença absoluta entre a estimativa da área a_{est} realizada pelo método proposto e a área real a_{real} esperada, definida por:

$$\text{RER}(\%) = \frac{|a_{est} - a_{real}|}{a_{real}} \cdot 100. \tag{6.3}$$

Observou-se experimentalmente que a curva de aprendizado do modelo se estabiliza por volta de 50 mil *steps*, então decidiu-se por salvar modelos a cada 50 *steps* partindo-se desta. Posteriormente, calculou-se o RER referente à predição de área da folha e do marcador, para cada uma das 14040 amostras do conjunto de treinamento em cada um dos 101 modelos salvos. Para cada modelo foram calculados a média μ e o desvio padrão σ do RER da folha e do marcador, totalizando $101 \times 14040 \times 2$ operações. Por fim, para determinar o melhor modelo, realizou-se uma análise extensiva dos seguintes critérios de seleção:

1. μ_f = Menor RER médio da folha.
2. $\mu_f + \sigma_f$ = Menor: RER médio da folha + desvio padrão RER da folha.
3. μ_m = Menor RER médio do marcador.
4. $\mu_m + \sigma_m$ = Menor: RER médio do marcador + desvio padrão RER do marcador.
5. $\mu_f + \mu_m$ = Menor: RER médio da folha + RER médio do marcador.
6. $\mu_f + \sigma_f + \mu_m + \sigma_m$ = Menor: RER médio da folha + desvio padrão RER da folha + RER médio do marcador + desvio padrão RER do marcador.

A Tabela 2 mostra os valores de RER obtidos no conjunto de imagens de teste para as taxas de aprendizado iniciais $\eta_0 \in \{0{,}005,\ 0{,}0025,\ 0{,}001,\ 0{,}0005,\ 0{,}0001\}$. O melhor resultado tratando-se do RER médio da folha foi obtido com o modelo treinado usando a taxa de aprendizado inicial $\eta_0 = 0{,}0005$. Sendo a melhor *step* de treinamento a de número 53250, selecionada pelos critérios 3, 4, 5 e 6, com $\mu_f = 5{,}827\%$ e $\sigma_f = 4{,}748\%$, células destacadas em cinza escuro.

Os valores de RER para o marcador nesta *step* foram $\mu_m = 1{,}541\%$ e $\sigma_m = 1{,}154\%$. É possível observar também que este foi o melhor modelo entre os selecionados para taxa

$\eta_0 = 0{,}0005$. O que aponta para uma influência do aprendizado do marcador na acurácia final do modelo, visto que os critérios 1 e 2, relacionados apenas com o erro para folha, selecionaram modelos com erros mais elevados.

O comportamento do treinamento com $\eta_0 = 0{,}005$, apresentado na primeira linha da tabela, corrobora com esta análise ao passo que os critérios 3 e 4 selecionaram a *step* com os menores valores de RER. Adicionalmente, o modelo selecionado apresenta o melhor desempenho na estimativa de área do marcador, com $\mu_m = 0{,}786\%$ e $\sigma_m = 0{,}652\%$, células destacados em cinza claro, o que indica não ser suficiente o superajuste da estimativa da área dos *pixels* do marcador para que o modelo se torne acurado também na predição da área foliar.

O comportamento dos modelos filtrados para as taxas iniciais $\eta_0 = 0{,}0025$ e $\eta_0 = 0{,}001$ foi semelhante. Os critérios 1 e 2 filtraram modelos com uma maior acurácia para a folha, mas que erram um pouco mais na estimativa da área do marcador. Por outro lado, os critérios 3 e 4 filtraram modelos que acertam mais para o marcador, porém afetam a estimativa da área foliar. Em específico, para a taxa de aprendizado $\eta_0 = 0{,}001$, é possível observar que o modelo selecionado pelos critérios 3 e 4 apresenta o RER médio da folha mais elevado da tabela, com $\mu_f = 8{,}322\%$. Isso demonstra que o uso apenas de critérios de ganho para o marcador pode não ser interessante. Por fim, os critérios 5 e 6, que combinam os erros de ambos os objetos, conseguiram selecionar modelos com um desempenho mais equilibrado. Portanto, pode-se dizer que é mais favorável utilizar os critérios da folha ou ainda a combinação destes com erros do marcador, que selecionar um modelo apenas com base no erro para o marcador. A última taxa de aprendizado experimentada, com $\eta_0 = 0{,}0001$, apresentou valores RER médios da folha acima de 6,4% em todos os modelos filtrados. Além disso, apresentou os piores desempenhos para área do marcador da tabela, com RER médio do marcador atingindo valores acima de 3%.

Analisando-se todos os experimentos apresentados nesta seção, tem-se uma evidência positiva de que o modelo foi capaz de aprender a estimar as áreas dos objetos, sobretudo para o marcador. Porém, é notório que o modelo comete alguns erros expressivos para a estimativa da área da superfície foliar, com erros mínimo e máximo em torno de 0,0007% e 24,7189% para o melhor modelo selecionado. Resumidamente, os critérios 3, 4, 5 e 6 tendem a selecionar modelos que favorecem a estimativa de área da folha. O melhor critério efetivamente depende da base de dados. Para o caso médio, porém sugere-se o uso do critério 6 ($\mu_f + \sigma_f + \mu_m + \sigma_m$) que não favorece individualmente nenhum dos objetos.

6.3 RESULTADOS QUANTITATIVOS

Visando compreender melhor a dimensão do problema abordado no presente trabalho, estimar áreas utilizando uma rede neural, foram gerados resultados quantitativos a partir das predições do melhor modelo, selecionado na Seção **6.2.1**. Na Subseção **6.3.1**,

Tabela 2 – Variação da taxa de aprendizado inicial η_0 para o treinamento do modelo por 55 mil *steps*. RER médio (μ) e desvio padrão do RER (σ) obtidos para a folha e o marcador, conforme os critérios de escolha definidos para seleção da melhor modelo. As células em cinza destacam os menores RER obtidos para a folha e o marcador.

			Folha		Marcador	
η_0	Critério(s) que selecionou(aram) o melhor modelo	Melhor *step*	RER (μ)	RER (σ)	RER (μ)	RER (σ)
0,005	3) μ_m , 4) $\mu_m + \sigma_m$	52700	5,910	4,511	0,786	0,652
	1) μ_f , 2) $\mu_f + \sigma_f$, 5) $\mu_f + \mu_m$, 6) $\mu_f + \sigma_f + \mu_m + \sigma_m$	53250	6,041	4,557	0,828	0,728
0,0025	3) μ_m , 4) $\mu_m + \sigma_m$	53050	6,831	5,123	0,959	0,866
	5) $\mu_f + \mu_m$, 6) $\mu_f + \sigma_f + \mu_m + \sigma_m$	53250	6,047	4,470	1,062	0,966
	1) μ_f , 2) $\mu_f + \sigma_f$	53700	5,898	4,425	1,268	1,141
0,001	5) $\mu_f + \mu_m$, 6) $\mu_f + \sigma_f + \mu_m + \sigma_m$	51850	5,862	4,651	1,944	1,269
	3) μ_m , 4) $\mu_m + \sigma_m$	53050	8,322	5,711	1,774	1,197
	1) μ_f , 2) $\mu_f + \sigma_f$	53650	5,878	4,641	2,397	1,396
0,0005	3) μ_m , 4) $\mu_m + \sigma_m$, 5) $\mu_f + \mu_m$, 6) $\mu_f + \sigma_f + \mu_m + \sigma_m$	53250	5,827	4,748	1,541	1,154
	2) $\mu_f + \sigma_f$	53550	5,895	4,748	2,267	1,328
	1) μ_f	53650	5,861	4,698	2,561	1,381
0,0001	3) μ_m , 4) $\mu_m + \sigma_m$, 5) $\mu_f + \mu_m$, 6) $\mu_f + \sigma_f + \mu_m + \sigma_m$	53250	6,685	4,990	1,975	1,652
	1) μ_f	53550	6,444	4,846	3,112	2,099
	2) $\mu_f + \sigma_f$	53950	6,504	4,888	3,209	2,115

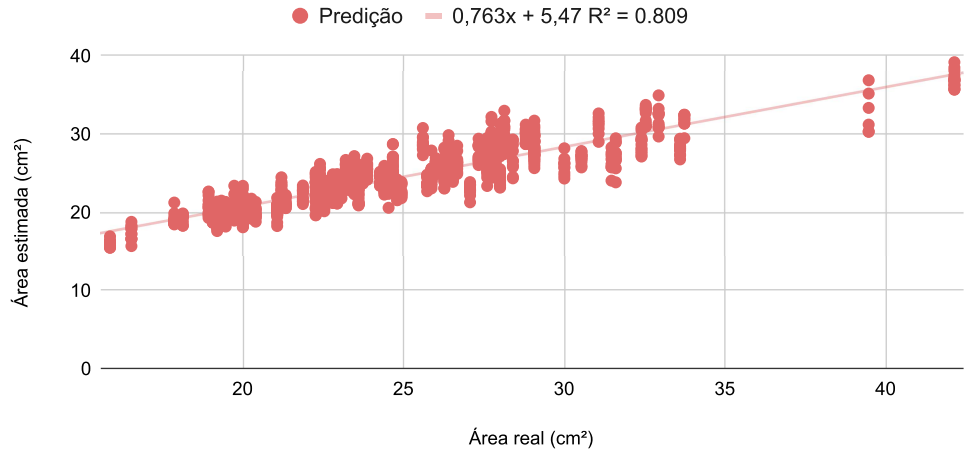
Fonte: Elaborada pela autora (2023).

são apresentados os resultados obtidos considerando-se todas as predições do conjunto de teste. Adicionalmente, a Subseção **6.3.2** apresenta os resultados obtidos com base em uma única estimativa de área por folha. Por fim, na Subseção **6.3.3**, é realizada uma análise complementar do método proposto, em termos da influência do marcador nas estimativas, impacto das modificações propostas na segmentação e custo computacional do modelo final.

6.3.1 Predição por imagem

A fim de avaliar a consistência nas predições, primeiro plotou-se um gráfico de dispersão das áreas foliares estimadas (eixo x) em relação às áreas reais esperadas (eixo y), conforme exibe a Figura 25. As predições para todas as imagens das mesmas folhas são consideradas. O gráfico apresenta a equação de regressão linear e o coeficiente de determinação R^2. É possível observar que os valores de AF estão linearmente relacionados. Além disso, há uma correlação positiva para maioria dos pontos no gráfico, com R^2 igual a 0,809. Os pontos vermelhos na mesma coluna do gráfico representam predições da mesma folha. Por conseguinte, nota-se que para algumas folhas, os valores de predição de área foram bastante discrepantes.

Figura 25 – Gráfico de dispersão das áreas foliares estimadas com o melhor modelo selecionado.



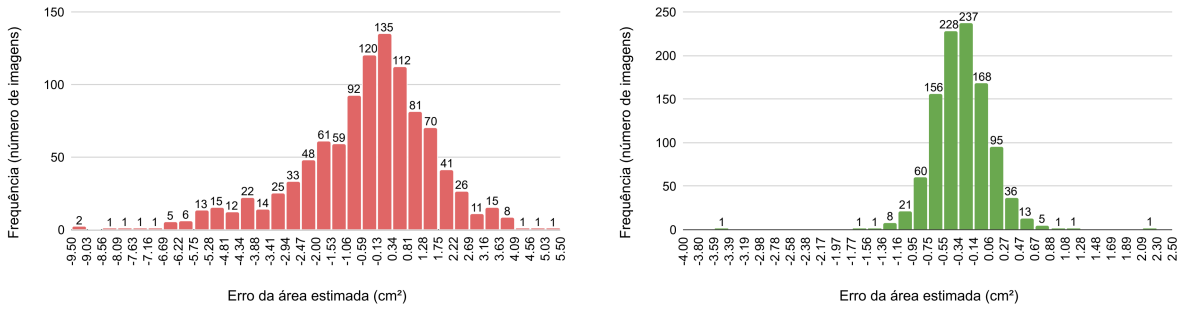
Fonte: Elaborada pela autora (2023).

Com o intuito de observar o impacto destas predições atípicas, foram gerados os gráficos da distribuição de erros para a folha e o marcador, exibidos na Figura 26. Os erros foram calculados fazendo-se a diferença entre a área estimada menos a área real. Conforme mostra a Figura 26a, os erros das predições da folha apresentam uma distribuição assimétrica com tendência a cauda pesada, em que os valores têm uma maior inclinação negativa. Isto aponta para uma tendência do modelo em subestimar a área da superfície foliar na maioria das imagens e/ou situações de captura. Ao analisar as extremidades do histograma, tem-se que seis das estimativas subestimaram a AF em -9,50 a -6,69 cm^2. Em contrapartida, há três predições que mais superestimaram a AF, com 4,09 a 5,50 cm^2 de área excedente. Por fim, tem-se que a maior concentração dos resultados está contida no intervalo de -2,00 a 1,75 cm^2 de erro, sendo que a frequência mais alta do histograma, com uma contagem de 135/1033 imagens, apresenta erros no intervalo de -0,13 a 0,34 cm^2 de área.

Para o marcador, conforme mostra a Figura 26b, a distribuição dos erros também é

assimétrica. Contudo, a maior concentração dos erros das estimativas está no intervalo de -0,75 a 0,27 cm^2 de área. Dessa forma observa-se que, apesar da inclinação do modelo em subestimar a área, as predições para o marcador foram bastante satisfatórias. Além disso, é possível observar que a maior frequência da distribuição, com um total de 237/1033 imagens, apresenta valores de erro que vão de -0,34 a 0,14 cm^2 de área. No entanto, nos extremos das caudas da distribuição estão os valores de erro 2,27 cm^2 de área superestimada e -3,52 cm^2 de subestimada da área da superfície do marcador. Isso corrobora para a conclusão de que há situações de captura mais e menos complexas para o modelo. Em síntese, percebe-se que para a maior parte das imagens de teste, o desempenho do modelo foi satisfatório. Porém, há uma tendência de estimativa de área menor que a esperada para os objetos, com o extremo de cauda mais pesado à esquerda dos histogramas.

Figura 26 – Histograma dos erros das predições no conjunto de teste realizadas pelo melhor modelo. A Figura (a) mostra a distribuição dos erros da área foliar e (b) da área do marcador.



(a) (b)

Fonte: Elaborada pela autora (2023).

6.3.2 Predição por folha

Como visto no gráfico de dispersão da AF, na Figura 25, entre as múltiplas predições de uma única folha existem valores próximos do esperado e outros mais distantes. A fim de contornar essa situação, foram propostas formas de determinar uma única estimativa de área, para cada uma das 90 folhas do conjunto de teste, a partir das estimativas de suas múltiplas imagens. Portanto, nesse refinamento, para cada folha consideram-de todas as suas imagens e respectivas predições de área. Duas áreas são calculadas para cada conjunto de imagens da mesma folha: a média das predições e a mediana das predições.

A Tabela 3 apresenta os resultados obtidos: RER médio e o desvio padrão, e os coeficientes de correlação de Pearson r entre as áreas reais e as áreas preditas a partir do conjunto de imagens de cada folha. É possível observar uma melhora significativa na taxa de erro com ambas as alternativas, obtendo-se μ_f de 4,80% com a predição média e μ_f de 4,83% com a predição mediana das folhas. Contudo, o erro máximo permanece significativo, atingindo 16,80% e 18,31% com a predição média e predição mediana, respectivamente.

Para realizar um comparativo da correlação dos resultados do modelo, ao considerar todas as estimativas obtém-se $r = 89{,}95$. Portanto, tem-se que a correlação das estimativas também foi melhorada, com $r = 92{,}74\%$ com a média e $r = 92{,}43\%$ para a mediana. A correlação tem sido utilizada em trabalhos recentes para avaliar o potencial de uso de métodos de estimativa de área em aplicações para biologia (HAGHSHENAS; EMAM, 2022). Em vista disso, as correlações positivas indicam a possibilidade de implementação e utilização prática do método apresentado nesta dissertação.

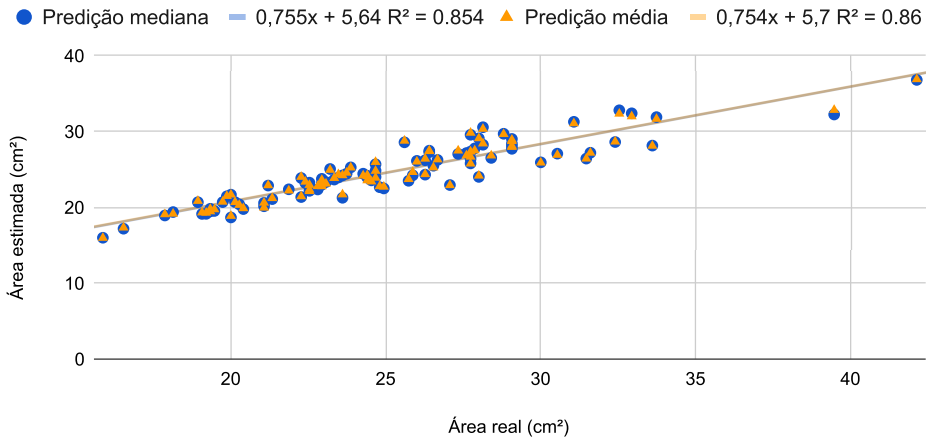
Tabela 3 – Comparativo entre a correlação de Pearson (r), RER médio (μ) e desvio padrão do RER (σ) da área das folhas, considerando-se todas as predições, a predição média e mediana de cada folha, estimadas com as predições do melhor modelo.

	Todas as predições	Predição média	Predição mediana
r	89,95%	92,74%	92,43%
RER (μ)	5,83%	4,80%	4,83%
RER (σ)	4,75%	4,34%	4,42%

Fonte: Elaborada pela autora (2023).

Adicionalmente, na Figura 27 exibe-se o gráfico de dispersão das áreas foliares médias e medianas estimadas (eixo x) em relação às áreas reais esperadas (eixo y). Comparando-se os valores de R^2 e as equações da linha de regressão ajustada, com os do gráfico da Figura 25 nota-se que houve um pequeno ganho no ajuste das estimativas do modelo. Em geral, os resultados alcançados com a predição média foram ligeiramente melhores. Portanto, os resultados quantitativos apresentados fortalecem a evidência da capacidade do modelo em estimar a área dos objetos. Em razão das estimativas destoantes para área da folha e do marcador, na Seção 6.4 são apresentados resultados de um estudo qualitativo de casos significativos.

Figura 27 – Gráficos de dispersão das áreas foliares médias e medianas estimadas com as predições do melhor modelo.



Fonte: Elaborada pela autora (2023).

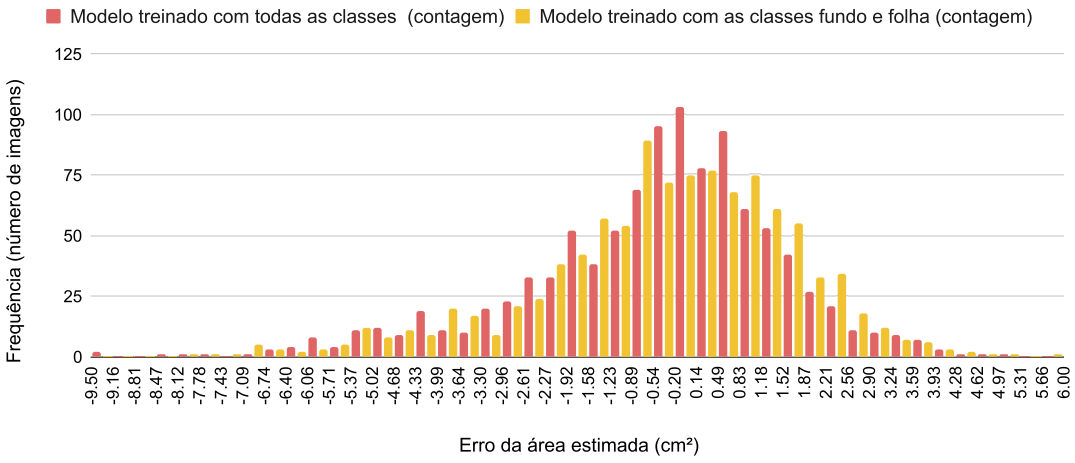
6.3.3 Análise complementar

Em adição às análises apresentadas nas Subseções **6.3.1** e **6.3.2**, realizou-se o treinamento da rede neural proposta somente com as classes fundo e folha, a fim de avaliar a influência do marcador no aprendizado do modelo. Para tanto, utilizaram-se os mesmos hiperparâmetros do melhor modelo. De acordo com os critérios 1 e 2, a melhor *step* do treinamento foi a de número 53650, com RER médio da folha de 6,081% e desvio padrão igual a 4,638%. Em comparação com o resultado do melhor modelo, apresentado na Tabela 2 observa-se um leve aumento do erro.

A Figura 28 mostra um comparativo dos histogramas das distribuições dos erros das estimativas realizadas pelo modelo obtido sem utilizar o marcador (em amarelo) e o melhor modelo (em vermelho) treinado com todas as classes. É possível observar um comportamento semelhante entre os dois, porém, nota-se que houve uma maior frequência de superestimativas.

Somado a isso, tem-se que 570/1033 amostras tiveram sua estimativa de área prejudicada, ou seja, o valor do erro absoluto foi maior no modelo que não utilizou o marcador. Portanto, os resultados indicam que o marcador favorece as predições do método proposto, o que evidencia o propósito estabelecido no Capítulo 5 sobre o método proposto comparar as proporções dos objetos de uma imagem.

Figura 28 – Comparativo dos histogramas dos erros das predições no conjunto de teste realizadas pelo melhor modelo (em vermelho) e o modelo treinado somente com as classes fundo e folha (em amarelo).



Fonte: Elaborada pela autora (2023).

Por fim, a Tabela 4 mostra um comparativo entre o desempenho do modelo proposto e a rede DeepLabv3+, em termos da média da interseção de *pixels* sobre a união (*Mean Intersection over Union* – mIOU), na tarefa de segmentação das imagens do conjunto de teste. Para isso, treinou-se a rede DeepLabv3+ com a mesma configuração de treinamento do melhor modelo da rede proposta. Quantitativamente, os resultados demonstram

que as modificações na arquitetura original não afetaram a qualidade da segmentação, mantendo-se a mIoU geral acima de 99% e acima de 98% para cada uma das três classes.

Tabela 4 – Comparativo do desempenho (mIOU) da rede DeepLabv3+ e do modelo proposto para estimativa de área foliar, na segmentação das amostras de teste.

Modelo	mIOU			
	Fundo	Folha	Marcador	Geral
DeepLabv3+	99,71%	98,72%	99,12%	99,18%
Rede estimadora de área foliar	99,70%	98,69%	99,08%	99,16%

Fonte: Elaborada pela autora (2023).

Em termos do tempo para realizar as predições de todas as imagens do conjunto de teste, a rede DeepLabv3+ leva em média 42 segundos (24 imagens por segundo), enquanto o modelo proposto estimou as duas saídas, segmentação e estimativa de área dos *pixels*, com um tempo médio de 47 segundos (21 imagens por segundo). Os tempos foram calculados pela média de três predições em todo o conjunto de teste. A taxa de amostras preditas por segundo indica a possibilidade do método ser aplicado em tempo real. Adicionalmente, o arquivo salvo contendo os pesos do modelo original da DeepLabv3+ possui 337,7 MB, ao passo que o melhor modelo para estimativa de área possui 340 MB. Portanto, pode-se dizer que a arquitetura proposta não elevou demasiadamente o custo computacional de tempo e espaço necessário da rede original.

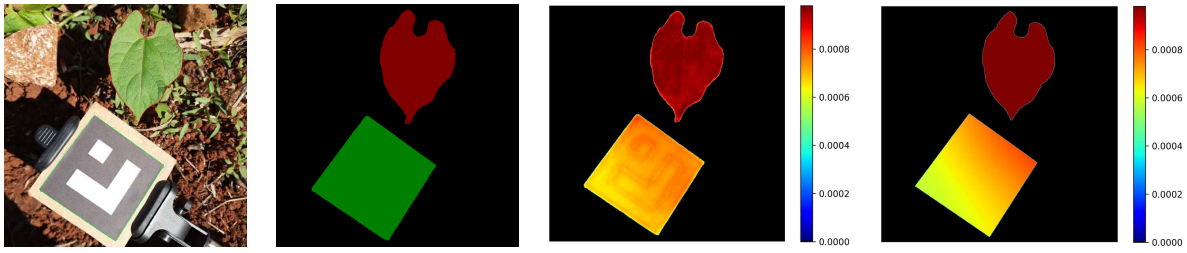
6.4 RESULTADOS QUALITATIVOS

Para se ter percepções visuais da qualidade das predições, realizadas pelo melhor modelo selecionado, propôs-se uma análise qualitativa das amostras do conjunto de teste com estimativas mais e menos corretas. Dessa forma, para cada caso de estudo é apresentada a imagem original com os contornos da segmentação estimada, acompanhada de visualizações da saída da segmentação, do GT da área e da saída da estimativa de área. Para simplificar a análise, uniram-se as duas saídas do decodificador de área em uma imagem.

Primeiro, com foco nos resultados da folha, observou-se que, para algumas das imagens da folha de número 146 com área esperada igual a 22,533 cm^2, foi obtida a estimativa da AF mais próxima do esperado, com RER igual a 0,0007%. A Figura 29 mostra visualizações dos resultados das predições. Esta folha apresenta uma leve curvatura na borda, sem recortes, tendo sido fotografada sob a incidência direta de luz solar. Em relação à segmentação, é possível notar algumas falhas ao longo das bordas dos dois objetos, sobretudo na parte inferior do folíolo, englobando parte do fundo. Tratando-se da visualização da estimativa da AF, observa-se que o aprendizado da perspectiva do marcador foi satisfatório. Os *pixels* mais claros contêm valores de área menores, enquanto os de cores mais escuras representam uma área maior. Além disso, é possível perceber

uma tentativa do modelo em replicar o aprendizado da orientação do marcador na região da folha. O resultado é que a metade superior da folha apresenta colorações mais escuras que a inferior. Interessantemente, o RER do marcador presente na imagem foi igual a 5,0421%, com uma superestimativa de 1,260 cm^2 acima da área esperada que é 25 cm^2.

Figura 29 – Resultados qualitativos da amostra com menor RER da folha. A Figura (a) mostra a imagem de entrada com os contornos extraídos da saída da segmentação (b). Nas Figuras (c) e (d), são exibidos os mapas de cores referentes a área estimada e real dos *pixels* dos objetos, respectivamente.



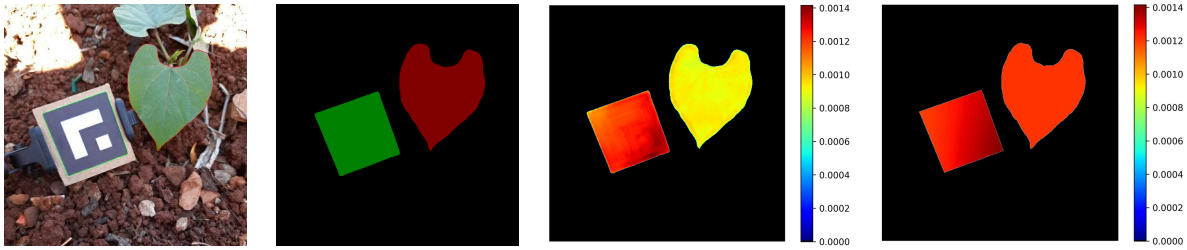
(a) (b) (c) (d)

Fonte: Elaborada pela autora (2023).

Em contrapartida, a amostra com maior RER da folha (24,7189%) é de uma das imagens da folha de número 276, cuja área real é 31,600 cm^2. A estimativa da área subestimou a superfície da folha em 7,811 cm^2. Uma das hipóteses para a discrepância da estimativa pode ser atribuída a menor frequência de folhas com áreas acima de 30 cm^2 no conjunto (Figuras 23 e 24). Conforme exibe a Figura 30 os objetos foram fotografados sob sombra e a folha apresenta topologia próxima de um plano, sem recortes. Outra hipótese é que a vista lateral do feijoeiro pode ter colaborado para a subestimativa da predição. Adicionalmente, para o marcador da imagem, o RER foi igual a 2,8200%, também considerado alto, com uma subestimativa de 0,450 cm^2. Observando-se a visualização da segmentação, é possível notar um arredondamento dos cantos do marcador, além disso, há uma perda de área na extensão do folíolo. Na visualização da AF, como no exemplo anterior, percebe-se uma tentativa do modelo em projetar a perspectiva do marcador na superfície da folha.

Analisando-se os resultados com foco no marcador, descobriu-se que para a amostra exibida na Figura 31, foi obtido o menor RER com 0,0115%. Na Figura 31b, é possível observar uma boa previsão de segmentação. Acredita-se que as condições de captura dos objetos, sob um céu nublado e o fundo da cena sem muitas informações, auxiliaram no aprendizado. No que diz respeito à visualização da estimativa de área, a perspectiva aprendida para o marcador foi bastante satisfatória em comparação com o GT, apresentando coloração mais clara à esquerda e mais escura no lado direito. Em relação à região da folha, nota-se que os *pixels* dos seus veios e do seu entorno receberam valores de área maiores. Comportamento semelhante ocorre no marcador, destacando o contorno da forma geométrica da sua superfície. Este efeito pode ser atribuído às características

Figura 30 – Resultados qualitativos da amostra com maior RER da folha. A Figura (a) mostra a imagem de entrada com os contornos extraídos da saída da segmentação (b). Nas Figuras (c) e (d), são exibidos os mapas de cores referentes a área estimada e real dos *pixels* dos objetos, respectivamente.

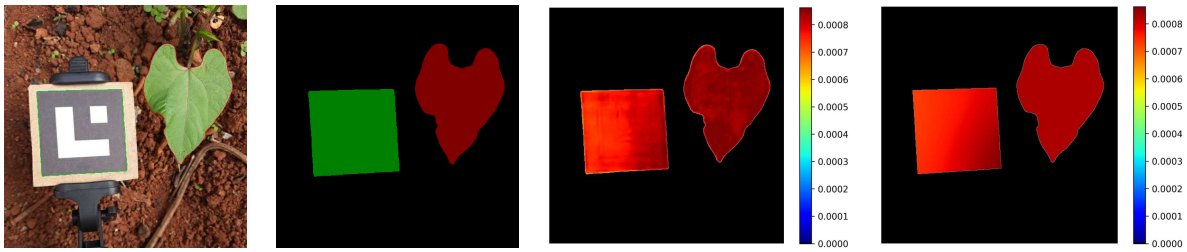


(a) (b) (c) (d)

Fonte: Elaborada pela autora (2023).

compartilhadas entre os decodificadores da rede neural proposta. O RER da folha presente na imagem foi igual a 3,0472%, com uma superestimativa de 0,723 cm^2 da sua área, abaixo da média do conjunto de teste.

Figura 31 – Resultados qualitativos da amostra com menor RER do marcador. A Figura (a) mostra a imagem de entrada com os contornos extraídos da saída da segmentação (b). Nas Figuras (c) e (d), são exibidos os mapas de cores referentes a área estimada e real dos *pixels* dos objetos, respectivamente.



(a) (b) (c) (d)

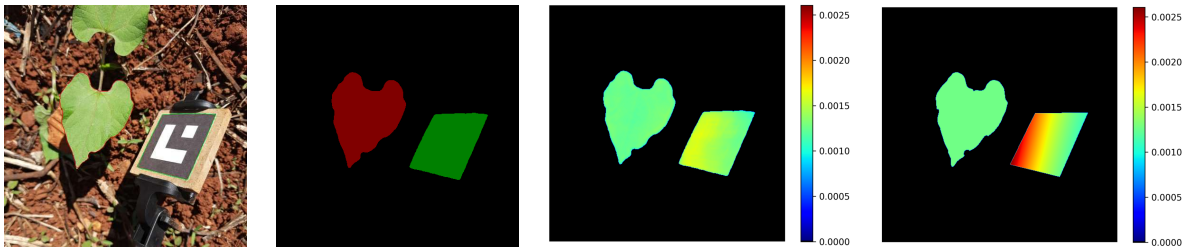
Fonte: Elaborada pela autora (2023).

Com o maior valor de RER do marcador (14,0835%), encontrou-se a amostra da Figura 32. É possível observar que a pose deste marcador não favorece a estimativa, com sua superfície bastante inclinada em direção à folha, contrária a visão da câmera. Além disso, na Figura 32a é possível notar falhas nas bordas dos dois objetos, em que partes de suas superfícies foram segmentadas como fundo. Pela visualização da estimativa de área, nota-se valores de área maiores tanto à esquerda quanto à direita. Por outro lado, apesar do erro alto na área do marcador, para a folha de 24,533 cm^2 presente na imagem, o RER foi igual a 1,1309% com uma subestimativa de apenas 0,277 cm^2. Este resultado indica que, apesar da posição desafiadora do marcador, a rede foi capaz de estimar a AF de forma satisfatória.

Adicionalmente, para análise de um caso médio, selecionou-se a amostra com RER da folha igual a 3,6954% e RER do marcador igual a 2,5603%. Para seleção da amostra

somaram-se os RER dos objetos, ordenou-se os valores e buscou-se pela amostra central do conjunto. A Figura 33 mostra as predições do modelo. A imagem de entrada contém a folha de número 285 de área esperada igual a 23,067 cm^2, que apresenta uma superfície curva e pequenos recortes na borda. Percebe-se que os objetos foram fotografados à sombra. Na Figura 33a, nota-se que alguns *pixels* do fundo foram segmentados como folha, incluindo parte do fundo adicionado à extensão do folíolo. Tratando-se da visualização da estimativa de área, nota-se uma distribuição de cores na região do marcador em conformidade com o esperado. Além disso, percebe-se valores um pouco mais uniformes no interior da região folha, em relação às amostras anteriores.

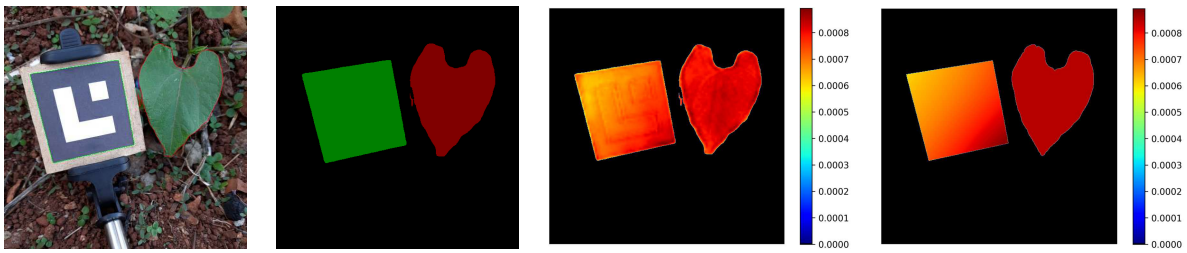
Figura 32 – Resultados qualitativos da amostra com maior RER do marcador. A Figura (a) mostra a imagem de entrada com os contornos extraídos da saída da segmentação (b). Nas Figuras (c) e (d), são exibidos os mapas de cores referentes a área estimada e real dos *pixels* dos objetos, respectivamente.



(a) (b) (c) (d)

Fonte: Elaborada pela autora (2023).

Figura 33 – Resultados qualitativos da amostra com RER médio. A Figura (a) mostra a imagem de entrada com os contornos extraídos da saída da segmentação (b). Nas Figuras (c) e (d) são exibidos os mapas de cores referentes a área estimada e real dos *pixels* dos objetos, respectivamente.



(a) (b) (c) (d)

Fonte: Elaborada pela autora (2023).

Em resumo, analisando-se todas as amostras de teste apresentadas, observa-se que a segmentação apresenta resultados suficientemente adequados, mas pode ser melhorada principalmente nas extremidades dos objetos. Em geral, o aprendizado da perspectiva visto na região do marcador é satisfatório. Em relação à folha, observaram-se evidências de que a rede pode aproximar a área. Entretanto, a ausência da informação de profundidade

da folha produz incerteza na predição dos seus *pixels*, uma das limitações da presente proposta.

6.5 DISCUSSÃO

Com base nas análises apresentadas ao longo do capítulo, conclui-se que o objetivo principal do presente trabalho foi alcançado. Quantitativamente, o erro médio de 5,827% considerando todas as imagens do conjunto de teste mostra uma evidência de que é possível estimar a área relativa da superfície foliar utilizando uma rede neural profunda, tendo apenas uma imagem como entrada. No entanto, conforme visto na análise quantitativa apresentada na Seção 6.3, conclui-se que a abordagem proposta neste trabalho pode obter estimativas de área com menor erro ao se utilizar várias imagens de uma mesma folha. Ao utilizar apenas a predição média ou a predição mediana de cada folha, obteve-se o erro médio de aproximadamente $4,8\%$ na estimativa. Contudo, o desvio na ordem de 4% nos resultados do melhor modelo obtido sinaliza que para algumas folhas com predições nos extremos da distribuição de erros, mesmo a estimativa a partir de várias amostras gera erros altos. Em geral, essas folhas têm geometria com grandes curvaturas. Dessa forma, a abordagem proposta é válida para folhas com aproximadamente a topologia do plano. Considerando o conjunto de dados proposto, observa-se que o marcador deve estar o mais alinhado possível com a direção de captura da câmera, e deve estar o mais coplanar possível com a folha a ser medida.

De acordo com resultados qualitativos apresentados na Seção 6.4, atribuem-se os erros mais elevados às condições de captura dos objetos. A baixa iluminação combinada com uma elevada quantidade de informação no fundo da cena prejudica a segmentação das imagens, principalmente dos cantos do marcador e na parte inferior da folha. Além disso, o ângulo de captura e a pose do marcador são fatores que podem comprometer a estimativa. Além de dificultarem o aprendizado da orientação do marcador, a inclinação da câmera provoca distorções na projeção dos objetos da imagem. Porém, é possível que a baixa representatividade de certos cenas no conjunto de dados impeça que a rede neural aprenda suas peculiaridades. Por exemplo, a baixa frequência de folhas com dimensões superiores a 30 cm^2 no conjunto impactou negativamente a generalização do modelo. Visto que, conforme os resultados obtidos, há uma maior tendência em subestimar as áreas. O não conhecimento da área real dos *pixels* da folha também limita o aprendizado pela rede neural. De fato, a ausência de informações adicionais da geometria e perspectiva da folha impossibilita alcançar um nível maior de comparação entre folha e marcador.

O coeficiente de correlação de Pearson entre as medições manuais e as estimativas de área foliar realizadas pelo modelo, foi acima de 89%. Ao considerar apenas uma predição por folha, o coeficiente obtido foi acima de 92%. Tendo em vista que, este é um indicativo relevante para investigar potenciais aplicações que realizam medições na

biologia, a correlação positiva alcançada pode indicar a viabilidade do emprego do método em uma ferramenta de suporte para especialistas em fisiologia vegetal, futuramente. Ao avaliar este valor juntamente com o RER médio obtido, sobretudo considerando-se apenas uma predição, acredita-se na possibilidade de uso do modelo para monitoramento de plantas e culturas no início da fase fenológica.

7 CONCLUSÃO

Nesta dissertação, buscou-se responder à indagação central: "É possível obter uma rede neural capaz de comparar dois objetos, um de dimensões conhecidas, e outro cuja dimensão será calculada em relação ao primeiro, prevendo assim uma estimativa de área?". Com esse propósito, apresenta-se um método não destrutivo para estimativa de área foliar, desenvolvido e implementado ao longo deste trabalho. O método proposto é baseado em uma nova arquitetura de rede neural profunda, treinada em amostras de um conjunto de imagens construído para esta tarefa. A arquitetura foi construída a partir de uma rede de segmentação semântica e tem como objetivo realizar a comparação entre as proporções da folha saliente e do marcador presente na imagem de entrada. A partir dessa comparação obtém-se a área relativa dos *pixels* de ambos os objetos. Foram realizados experimentos, análises quantitativas e qualitativas para analisar a viabilidade da proposta. Os resultados obtidos demonstraram uma evidência de que o modelo é capaz de estimar a área dos objetos. Observou-se que o alinhamento do marcador com a folha, o ângulo de captura e a baixa iluminação da cena impactam a estimativa da área. Além disso, a falta de informações adicionais da perspectiva da folha limita a comparação entre os objetos. Portanto, este trabalho oferece uma resposta positiva à pergunta de pesquisa.

Adicionalmente, apresenta-se uma nova base de dados com imagens de folhas de feijão-comum acompanhadas por um marcador de realidade virtual e aumentada. Essa base constitui uma das principais contribuições do trabalho, um conjunto inédito de imagens com anotações de máscaras, informações de calibração e medições das folhas, realizadas por método padrão. Acredita-se que a base proposta possui potencial a ser explorado pela comunidade de visão computacional no desenvolvimento e validação de metodologias para medição de superfícies foliares. Adicionalmente, espera-se que as imagens do conjunto sejam utilizadas também em outras aplicações como o treinamento de modelos para identificação de folhas e/ou espécies.

Como trabalhos futuros, a princípio pretende-se usar o restante das imagens da base de dados proposta. Além disso, planeja-se explorar o uso de múltiplas imagens como entradas para a rede. Teoricamente, o uso de uma sequência de quadros de uma mesma folha, pode beneficiar a estimativa da área ao fornecer diferentes poses do marcador para comparação. Além disso, pretende-se treinar o modelo para estimar as demais dimensões anotadas (*e.g.*, comprimento, largura e perímetro). Outra linha a ser seguida consiste na utilização dos dados anotados de calibração de câmera no treinamento do modelo. Estes dados podem ser uma alternativa para aprimorar o aprendizado da perspectiva pelo modelo. Outras funções de perda podem ser incorporadas para reforçar a capacidade de comparação das informações aprendidas. Abordagens que utilizem a reconstrução 3D da folha também podem ser interessantes para obtenção de uma reconstrução com maior acurácia. O objetivo, diante disso, é prover uma forma de contornar a falta da informação

de perspectiva da folha. Não obstante, acredita-se que a rede proposta não está limitada apenas a folhas de feijão. Neste caso, o objetivo é expandir a base dados com outras espécies que apresentem morfologia semelhante. Além disso, considera-se o uso de outras arquiteturas de redes neurais para segmentação como métodos baseados em *transformer* ou modelos de difusão.

REFERÊNCIAS

ACHARYA, Tinku; RAY, Ajoy K. **Image Processing: Principles and Applications**. New Jersey: John Wiley & Sons, Inc., Hoboken, 2005.

ALE, Laha; SHETA, Alaa; LI, Longzhuang; WANG, Ye; ZHANG, Ning. Deep learning based plant disease detection for smart agriculture. In: **IEEE Globecom Workshops (GC Wkshps)**. Waikoloa, HI, USA: IEEE, 2019. p. 1–6.

AYHAN, Bulent; KWAN, Chiman. Tree, shrub, and grass classification using only rgb images. **Remote Sensing**, MDPI, v. 12, n. 8, p. 1333, 2020.

BENLIGIRAY, Burak; TOPAL, Cihan; AKINLAR, Cuneyt. Stag: A stable fiducial marker system. **Image and Vision Computing**, Elsevier, v. 89, p. 158–169, 2019.

BRADSKI, Gary. The opencv library. **Dr. Dobb's Journal: Software Tools for the Professional Programmer**, Miller Freeman Inc., v. 25, n. 11, p. 120–123, 2000.

BRUMMEN, Alexandra Van; OWEN, Julia P; SPAIDE, Theodore; FROINES, Colin; LU, Randy; LACY, Megan; BLAZES, Marian; LI, Emily; LEE, Cecilia S; LEE, Aaron Y et al. Periorbitai: artificial intelligence automation of eyelid and periorbital measurements. **American Journal of Ophthalmology**, Elsevier, v. 230, p. 285–296, 2021.

CARVALHO, Paulo Cezar; VELHO, Luiz; SÁ, Asla; MEDEIROS, Esdras; MONTENGRO, Anselmo Antunes; PEIXOTO, Adelailson; ESCRIBA, Luis Antonio Rivera. **Fotografia 3D**. Rio de Janeiro: Associação Instituto de Matemática Pura e Aplicada, IMPA, 2005.

CHEN, Liang-Chieh; PAPANDREOU, George; KOKKINOS, Iasonas; MURPHY, Kevin; YUILLE, Alan L. Semantic image segmentation with deep convolutional nets and fully connected crfs. **arXiv preprint arXiv:1412.7062**, 2014.

CHEN, Liang-Chieh; PAPANDREOU, George; KOKKINOS, Iasonas; MURPHY, Kevin; YUILLE, Alan L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 40, n. 4, p. 834–848, 2017.

CHEN, Liang-Chieh; PAPANDREOU, George; SCHROFF, Florian; ADAM, Hartwig. Rethinking atrous convolution for semantic image segmentation. **arXiv preprint arXiv:1706.05587**, 2017.

CHEN, Liang-Chieh; ZHU, Yukun; PAPANDREOU, George; SCHROFF, Florian; ADAM, Hartwig. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. Munich, Germany: ECCV, 2018. p. 801–818.

CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, Hawaii: CVPR, 2017. p. 1251–1258.

CLARK, Alex et al. **Pillow (pil fork) documentation**. **readthedocs**, 2015.

COLLINS, Toby; BARTOLI, Adrien. Infinitesimal plane-based pose estimation. **International Journal of Computer Vision**, Springer, v. 109, n. 3, p. 252–286, 2014.

DANG, L Minh; WANG, Hanxiang; LI, Yanfen; NGUYEN, Le Quan; NGUYEN, Tan N; SONG, Hyoung-Kyu; MOON, Hyeonjoon. Deep learning-based masonry crack segmentation and real-life crack length measurement. **Construction and Building Materials**, Elsevier, v. 359, p. 129438, 2022.

DHANYA, VG; SUBEESH, A; KUSHWAHA, NL; VISHWAKARMA, Dinesh Kumar; KUMAR, T Nagesh; RITIKA, G; SINGH, AN. Deep learning based computer vision approaches for smart agricultural applications. **Artificial Intelligence in Agriculture**, Elsevier, 2022.

DOUGLAS, David H; PEUCKER, Thomas K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. **Cartographica: The International Journal for Geographic Information and Geovisualization**, University of Toronto Press, v. 10, n. 2, p. 112–122, 1973.

DROZDOV, D; KOLOMEICHENKO, M; BORISOV, Y. **Supervisely**. 2023. <https://supervisely.com/>. Acesso em: 04 de novembro de 2023.

DYRMANN, Mads; KARSTOFT, Henrik; MIDTIBY, Henrik Skov. Plant species classification using deep convolutional neural network. **Biosystems Engineering**, Elsevier, v. 151, p. 72–80, 2016.

EVERINGHAM, Mark; ESLAMI, SM Ali; GOOL, Luc Van; WILLIAMS, Christopher KI; WINN, John; ZISSERMAN, Andrew. The pascal visual object classes challenge: A retrospective. **International Journal of Computer Vision**, Springer, v. 111, p. 98–136, 2015.

EVERT, Ray F; EICHHORN, Susan E. **Raven Biology of Plants**. New York: W.H. Freeman/Palgrave Macmillan, 2013.

FAN, Xijian; ZHOU, Rui; TJAHJADI, Tardi; CHOUDHURY, Sruti Das; YE, Qiaolin. A segmentation-guided deep learning framework for leaf counting. **Frontiers in Plant Science**, Frontiers, v. 13, p. 844522, 2022.

FANOURAKIS, Dimitrios; KAZAKOS, Filippos; NEKTARIOS, Panayiotis A. Allometric individual leaf area estimation in chrysanthemum. **Agronomy**, MDPI, v. 11, n. 4, p. 795, 2021.

FERNANDES, Arthur FA; TURRA, Eduardo M; ALVARENGA, Erika R de; PASSAFARO, Tiago L; LOPES, Fernando B; ALVES, Gabriel FO; SINGH, Vikas; ROSA, Guilherme JM. Deep learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in nile tilapia. **Computers and Electronics in Agriculture**, Elsevier, v. 170, p. 105274, 2020.

GARRIDO-JURADO, Sergio; MUÑOZ-SALINAS, Rafael; MADRID-CUEVAS, Francisco José; MARÍN-JIMÉNEZ, Manuel Jesús. Automatic generation and detection of highly reliable fiducial markers under occlusion. **Pattern Recognition**, Elsevier, v. 47, n. 6, p. 2280–2292, 2014.

GLOROT, Xavier; BENGIO, Yoshua. Understanding the difficulty of training deep feedforward neural networks. In: **Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics**. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010. (Proceedings of Machine Learning Research), p. 249–256.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA: MIT Press, 2016.

HAGHSHENAS, Abbas; EMAM, Yahya. Accelerating leaf area measurement using a volumetric approach. **Plant Methods**, Springer, v. 18, n. 1, p. 61, 2022.

HARIHARAN, Bharath; ARBELÁEZ, Pablo; GIRSHICK, Ross; MALIK, Jitendra. Hypercolumns for object segmentation and fine-grained localization. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. Boston, Massachusetts: CVPR, 2015. p. 447–456.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Spatial pyramid pooling in deep convolutional networks for visual recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 37, n. 9, p. 1904–1916, 2015.

HORN, Roger A. The hadamard product. In: **Proc. Symposium on Applied Mathematics**. Providence, Rhode Island: American Mathematical Society, 1990. v. 40, p. 87–169.

ISLAM, Md Parvez; NAKANO, Yuka; LEE, Unseok; TOKUDA, Keinichi; KOCHI, Nobuo. Thelnet270v1–a novel deep-network architecture for the automatic classification of thermal images for greenhouse plants. **Frontiers in Plant Science**, Frontiers Media SA, v. 12, p. 630425, 2021.

JADON, Madhu. A novel method for leaf area estimation based on hough transform. **Journal of Multimedia Processing and Technologies**, v. 9, n. 2, p. 33–44, 2018.

JAIN, A. K. **Fundamentals of Digital Image Processing**. New Jersey: Prentice-Hall, Inc., 1989.

JANHÄLL, Sara. Review on urban vegetation and particle air pollution–deposition and dispersion. **Atmospheric Environment**, Elsevier, v. 105, p. 130–137, 2015.

KAUR, Gurjot; DIN, Salam; BRAR, Amandeep Singh. Design and development of software for the implementation of image processing approach for leaf area measurement. **International Journal of Computer Science and Information Technologies**, Citeseer, v. 5, p. 4793–4797, 2014.

KIM, Taehyeon; LEE, Sang-Ho; KIM, Jong-Ok. A novel shape based plant growth prediction algorithm using deep learning and spatial transformation. **IEEE Access**, IEEE, v. 10, p. 37731–37742, 2022.

KOESTER, Robert P; SKONECZKA, Jeffrey A; CARY, Troy R; DIERS, Brian W; AINSWORTH, Elizabeth A. Historical gains in soybean (glycine max merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. **Journal of Experimental Botany**, Oxford University Press UK, v. 65, n. 12, p. 3311–3321, 2014.

KOUBOURIS, Georgios; BOURANIS, Dimitris; VOGIATZIS, Efraim; NEJAD, Abdolhossein Rezaei; GIDAY, Habtamu; TSANIKLIDIS, Georgios; LIGOXIGAKIS, Eleftherios K; BLAZAKIS, Konstantinos; KALAITZIS, Panagiotis; FANOURAKIS, Dimitrios. Leaf area estimation by considering leaf dimensions in olive tree. **Scientia Horticulturae**, Elsevier, v. 240, p. 440–445, 2018.

KRÄHENBÜHL, Philipp; KOLTUN, Vladlen. Efficient inference in fully connected crfs with gaussian edge potentials. **Advances in Neural Information Processing Systems**, v. 24, 2011.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012.

LAUGHLIN, Daniel C. Nitrification is linked to dominant leaf traits rather than functional diversity. **Journal of Ecology**, Wiley Online Library, v. 99, n. 5, p. 1091–1099, 2011.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Yann; BOTTOU, Léon; BENGIO, Yoshua; HAFFNER, Patrick. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LEE, Sue Han; CHAN, Chee Seng; WILKIN, Paul; REMAGNINO, Paolo. Deep-plant: Plant identification with convolutional neural networks. In: IEEE. **IEEE International Conference on Image Processing (ICIP)**. Quebec City, QC, Canada, 2015. p. 452–456.

LI-COR. **LI-3100 area meter instruction manual.** Lincoln: LI-COR, p. 34, 1996.

LI, Mengcheng; LIAO, Yitao; LU, Zhifeng; SUN, Mai; LAI, Hongyu. Non-destructive monitoring method for leaf area of brassica napus based on image processing and deep learning. **Frontiers in Plant Science**, Frontiers Media SA, v. 14, 2023.

LIANG, Wei-zhen; KIRK, Kendall R; GREENE, Jeremy K. Estimation of soybean leaf area, edge, and defoliation using color image analysis. **Computers and Electronics in Agriculture**, Elsevier, v. 150, p. 41–51, 2018.

LIU, Bin; ZHANG, Yun; HE, DongJian; LI, Yuxiang. Identification of apple leaf diseases based on deep convolutional neural networks. **Symmetry**, MDPI, v. 10, n. 1, p. 11, 2017.

LIU, Wei; RABINOVICH, Andrew; BERG, Alexander C. Parsenet: Looking wider to see better. **arXiv preprint arXiv:1506.04579**, 2015.

LU, Jinzhu; EHSANI, Reza; SHI, Yeyin et al. Detection of multi-tomato leaf diseases (late blight, target and bacterial spots) in different stages by using a spectral-based sensor. **Scientific Reports**, Nature Publishing Group, v. 8, n. 1, p. 2793, 2018.

LÜLING, Nils; REISER, David; GRIEPENTROG, Hans W. Volume and leaf area calculation of cabbage with a neural network-based instance segmentation. In: **13th European Conference on Precision Agriculture Conference.** Budapest: ECPA, 2021. p. 2719–2745.

MALLAT, Stéphane. **A Wavelet Tour of Signal Processing**. San Diego, California: Elsevier, 1999.

MAREK, Janaina; AZEVEDO, Dione de; ONO, Elizabeth Orika et al. Photoynthetic and productive increase in tomato plants treated with strobilurins and carboxamides for the control of alternaria solani. **Scientia Horticulturae**, Elsevier, v. 242, p. 76–89, 2018.

OGHLI, Mostafa Ghelich; SHABANZADEH, Ali; MORADI, Shakiba; SIRJANI, Nasim; GERAMI, Reza; GHADERI, Payam; TAHERI, Morteza Sanei; SHIRI, Isaac; ARABI, Hossein; ZAIDI, Habib. Automatic fetal biometry prediction using a novel deep convolutional network architecture. **Physica Medica**, Elsevier, v. 88, p. 127–137, 2021.

OLIVEIRA, LFC De; OLIVEIRA, MG de C; WENDLAND, A; HEINEMANN, AB; GUIMARÃES, CM; FERREIRA, EP de B; QUINTELA, ED; BARBOSA, FR; CARVALHO, M da; JUNIOR, M Lobo et al. **Conhecendo a fenologia do feijoeiro e seus aspectos fitotécnicos.** Brasília, DF: Embrapa, 2018., 2018.

ONAKPOYA, Igho; ALDAAS, Salsabil; TERRY, Rohini; ERNST, Edzard. The efficacy of phaseolus vulgaris as a weight-loss supplement: a systematic review and meta-analysis of randomised clinical trials. **British Journal of Nutrition**, Cambridge University Press, v. 106, n. 2, p. 196–202, 2011.

PANDEY, SK; SINGH, Hema. A simple, cost-effective method for leaf area estimation. **Journal of Botany**, Hindawi Publishing Corporation, v. 2011, n. 2011, p. 1–6, 2011.

POLUNINA, O. V.; MAIBORODA, V. P.; SELEZNOV, A. Y. Evaluation methods of estimation of young apple trees leaf area. **Bulletin of Uman National University of Horticulture**, Uman National University of Horticulture, n. 2, p. 80–82, 2018.

POORTER, Hendrik; NIINEMETS, Ülo; NTAGKAS, Nikolaos et al. A meta-analysis of plant responses to light intensity for 70 traits ranging from molecules to whole plant performance. **New Phytologist**, Wiley Online Library, v. 223, n. 3, p. 1073–1105, 2019.

QI, H.; ZHANG, Z.; XIAO, B.; HU, H.; CHENG, B.; WEI, Y.; DAI, J. **Deformable convolutional networks – coco detection and segmentation challenge 2017 entry**. 2017.

RICHARDSON, Grant A; LOHANI, Harshit K; POTNURU, Chaitanyam; DONEPUDI, Leela Prasad; PANKAJAKSHAN, Praveen. Phenobot: an automated system for leaf area analysis using deep learning. **Planta**, Springer, v. 257, n. 2, p. 36, 2023.

ROMERO-RAMIREZ, Francisco J; MUÑOZ-SALINAS, Rafael; MEDINA-CARNICER, Rafael. Speeded up detection of squared fiducial markers. **Image and Vision Computing**, Elsevier, v. 76, p. 38–47, 2018.

RUSSAKOVSKY, Olga; DENG, Jia; SU, Hao; KRAUSE, Jonathan; SATHEESH, Sanjeev; MA, Sean; HUANG, Zhiheng; KARPATHY, Andrej; KHOSLA, Aditya; BERNSTEIN, Michael et al. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, Springer, v. 115, p. 211–252, 2015.

SABOURI, Hossein; SAJADI, Sayed Javad; JAFARZADEH, Mohammad Reza; REZAEI, Mohsen; GHAFFARI, Sanaz; BAKHTIARI, Samira. Image processing and prediction of leaf area in cereals: A comparison of artificial neural networks, an adaptive neuro-fuzzy inference system, and regression methods. **Crop Science**, Wiley Online Library, v. 61, n. 2, p. 1013–1029, 2021.

SAGAN, Carl. **Pale Blue Dot: A Vision of the Human Future in Space**. New York, NY: USA: Random House, 1994. v. 1st ed.

SAKHRAVI, Ameneh; DEHDARI, Massoud; FAHLIANI, Reza Amiri. Genetic relationships among common bean (phaseolus vulgaris l) genotypes using issr markers. **Gene Reports**, v. 32, p. 101797, 2023. ISSN 2452-0144. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2452014423000596>.

SARKER, Iqbal H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. **SN Computer Science**, Springer, v. 2, n. 6, p. 420, 2021.

SIFRE, Laurent. **Rigid-motion scattering for image classification**. **PhD thesis**, École Polytechnique, France, 2014.

SILVA, Karla Gabriele Florentino da; MOREIRA, Jonas Magalhães; CALIXTO, Gabriel Barreto; MACIEL, Luiz Maurílio da Silva; MIRANDA, Márcio Assis; MORAIS, Leandro Elias. A simple and low-cost method for leaf surface dimension estimation based on digital images. In: **Brazilian Conference on Intelligent Systems**. Cham: Springer Nature Switzerland, 2023. p. 146–161.

SONG, Xiao-Peng; HANSEN, Matthew C; POTAPOV, Peter; ADUSEI, Bernard; PICKERING, Jeffrey; ADAMI, Marcos; LIMA, Andre; ZALLES, Viviana; STEHMAN, Stephen V; BELLA, Carlos M Di et al. Massive soybean expansion in south america since 2000 and implications for conservation. **Nature Sustainability**, Nature Publishing Group UK London, v. 4, n. 9, p. 784–792, 2021.

SRINIVASAN, Venkatraman; KUMAR, Praveen; LONG, Stephen P. Decreasing, not increasing, leaf area will raise crop yields under global atmospheric change. **Global Change Biology**, Wiley Online Library, v. 23, n. 4, p. 1626–1635, 2017.

STEWART, James. **Calculus: Concepts and Contexts**. United States: Cengage Learning, 2009.

SUZUKI, Satoshi; ABE, Keiichi. Topological structural analysis of digitized binary images by border following. **Computer Vision, Graphics, and Image Processing**, v. 30, n. 1, p. 32–46, 1985. ISSN 0734-189X.

TAIZ, L; ZEIGER, E. Auxin: The first discovered plant growth hormone. **Plant Physiology. 5th ed.**, Sinauer Associates, Sunderland, MS, p. 545–582, 2010.

TECH, Adriano Rogério Bruno; SILVA, André Luis Céspedes da; MEIRA, Luiz Antonio et al. Methods of image acquisition and software development for leaf area measurements in pastures. **Computers and Electronics in Agriculture**, Elsevier, v. 153, p. 278–284, 2018.

TRIKI, Abdelaziz; BOUAZIZ, Bassem; GAIKWAD, Jitendra; MAHDI, Walid. Deep leaf: Mask r-cnn based leaf detection and segmentation from digitized herbarium specimen images. **Pattern Recognition Letters**, Elsevier, v. 150, p. 76–83, 2021.

TU, Li-fen; PENG, Qi; LI, Chun-sheng; ZHANG, Aiqun. 2d in situ method for measuring plant leaf area with camera correction and background color calibration. **Scientific Programming**, Hindawi Limited, v. 2021, p. 1–11, 2021.

WELLSTEIN, Camilla; POSCHLOD, Peter; GOHLKE, Andreas; CHELLI, Stefano; CAMPETELLA, Giandiego; ROSBAKH, Sergey; CANULLO, Roberto; KREYLING, Jürgen; JENTSCH, Anke; BEIERKUHNLEIN, Carl. Effects of extreme drought on specific leaf area of grassland species: A meta-analysis of experimental studies in temperate and sub-mediterranean systems. **Global Change Biology**, Wiley Online Library, v. 23, n. 6, p. 2473–2481, 2017.

WRIGHT, Ian J; REICH, Peter B; WESTOBY, Mark et al. The worldwide leaf economics spectrum. **Nature**, Nature Publishing Group, v. 428, n. 6985, p. 821–827, 2004.

WU, Zhenchao; YANG, Ruizhe; GAO, Fangfang; WANG, Wenqi; FU, Longsheng; LI, Rui. Segmentation of abnormal leaves of hydroponic lettuce based on deeplabv3+ for robotic sorting. **Computers and Electronics in Agriculture**, Elsevier, v. 190, p. 106443, 2021.

YANG, Ming-Der; TSENG, Hsin-Hung; HSU, Yu-Chun; TSAI, Hui Ping. Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date uav visible images. **Remote Sensing**, MDPI, v. 12, n. 4, p. 633, 2020.

ZHANG, Lingxian; XU, Zanyu; XU, Dan; MA, Juncheng; CHEN, Yingyi; FU, Zetian. Growth monitoring of greenhouse lettuce based on a convolutional neural network. **Horticulture Research**, Oxford Academic, v. 7, 2020.

ZHANG, Zhengyou. A flexible new technique for camera calibration. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 22, n. 11, p. 1330–1334, 2000.

ZHU, Shisong; MA, Wanli; LU, Jiangwen; REN, Bo; WANG, Chunyang; WANG, Jianlong. A novel approach for apple leaf disease image segmentation in complex scenes based on two-stage deeplabv3+ with adaptive loss. **Computers and Electronics in Agriculture**, Elsevier, v. 204, p. 107539, 2023.