BRUNO CONCEIÇÃO DO NASCIMENTO

CLASSIFICAÇÃO DE PLÂNTULAS DE SOJA COM RELAÇÃO AO GENÓTIPO E CONDIÇÕES DO SOLO COM REDES NEURAIS CONVOLUCIONAIS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Michel Melo da Silva

VIÇOSA - MINAS GERAIS
2022

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
N244c
2022

Nascimento, Bruno Conceição do, 1992-
Classificação de plântulas de soja com relação ao genótipo e condições do solo com Redes Neurais Convolucionais / Bruno Conceição do Nascimento. – Vicosa, MG, 2022.
1 dissertação eletrônica (83 f.): il. (algumas color.).

Orientador: Michel Melo da Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Informática, 2022.
Referências bibliográficas: f. 79-83.
DOI: https://doi.org/10.47328/ufvbbt.2023.115
Modo de acesso: World Wide Web.

1. Redes neurais (Computação). 2. Aprendizado do computador. 3. Solos - Compactação. 4. Soja - Genética. I. Silva, Michel Melo da, 1990-. II. Universidade Federal de Viçosa. Departamento de Informática. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDD 23. ed. 006.32

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

BRUNO CONCEIÇÃO DO NASCIMENTO

CLASSIFICAÇÃO DE PLÂNTULAS DE SOJA COM RELAÇÃO AO GENÓTIPO E CONDIÇÕES DO SOLO COM REDES NEURAIS CONVOLUCIONAIS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 14 de dezembro de 2022.

Assentimento:

Bruno C. do Nascimento

Bruno Conceição do Nascimento
Autor

M. de Melo

Michel Melo da Silva
Orientador

Agradecimentos

Agradeço primeiramente a minha família, meus pais Nelson e Leonice e meu irmão Leonardo, por todo apoio e amor que tive durante toda a minha vida.

A minha namorada, Raquel, por todo o carinho e incentivo durante todo esse período, sendo essencial na minha vida.

Ao meu orientador, Prof. Michel Melo da Silva e ao meu Coorientador, Prof. Marcos Henrique Fonseca Ribeiro, por todos os ensinamentos, pelo apoio e compreensão.

Aos meus amigos de Ponte Nova e Viçosa.

Ao professor Laércio Junio da Silva por todas as reuniões e apoio.

A Nayara Capobiango, pela gentileza e esclarecimentos de dúvidas.

À Universidade Federal de Viçosa, ao Centro de Ciências Exatas, ao Departamento de Informática da UFV.

À empresa GDM Seeds.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

NASCIMENTO, Bruno Conceição do, M.Sc., Universidade Federal de Viçosa, dezembro de 2022. **Classificação de plântulas de soja com relação a condições de solo e genótipo com Redes Neurais Convolucionais**. Orientador: Michel Melo da Silva.

A compactação do solo é um dos fatores que geram prejuízos no plantio de várias culturas. O aumento do tamanho e peso das máquinas agrícolas, associado ao manejo intensivo das áreas cultivadas, tem grande potencial para intensificar a compactação do solo. Sementes cultivadas em solos com esse impedimento físico ao germinarem apresentam um sistema radicular com características que podem influenciar no desenvolvimento da plântula, gerando raízes curtas e grossas sendo uma resposta da plântula para tentar se adequar ao solo, mas com esses características podem ter dificuldade para captar nutrientes e água do solo, essenciais para o seu crescimento. Alguns estudos visam identificar genótipos de soja que conseguem se desenvolver em condições adversas no solo sem alterações que afetem o sistema radicular de forma prejudicial. A análise de imagens de plântulas de soja através de Redes Neurais Convolucionais (CNNs) para identificação do genótipo e condição de solo onde foram cultivadas, podem auxiliar na identificação de qual genótipo melhor se desenvolve em uma determinada condição de solo. Neste trabalho, é gerado um conjunto de dados de plântulas de soja cultivadas em solo compactado e não compactado, com duas versões, uma com a imagem completa da plântula e outra com a plântula sem raiz. O principal objetivo deste trabalho é o treinamento de CNNs para a classificação da condição do solo, e a abordagem de uma Rede Convolucional Multitarefa desenvolvida para a classificação de genótipos. Para a classificação do genótipo todos os modelos treinados obtiveram boas taxas de precisão, sendo a VGG-16 alcançando os maiores valores com uma acurácia média na de 91,5% e 85,6% na fase de teste para conjunto de dados com plântula completa e sem raiz respetivamente. No problema de classificação de genótipo a rede Multitarefa obteve os melhores resultados com 52,3% para o conjunto com plântula completa e 47,1% com plântula sem raiz ambos na fase de teste.

Palavras-chave: Redes Convolucionais Multitarefa. Redes Neurais Convolucionais. Plântula de Soja. Identificação de Condição do Solo. Classificação de Genótipo.

Abstract

NASCIMENTO, Bruno Conceição do, M.Sc., Universidade Federal de Viçosa, December 2022. **Classificação de plântulas de soja com relação a condições de solo e genótipo utilizando Redes Neurais Convolucionais**. Advisor: Michel Melo da Silva.

Soil compaction is one of the factors that generate losses in the planting of several crops. Increase in the size and weight of agricultural machinery, associated with the intensive management of cultivated areas, has great potential to intensify soil compaction. Seeds grown in soils with this physical barrier during the germination phase develop a root system with characteristics that can negatively impact seedling growth. The development of short and thick roots may be a seedling's response to trying to adapt to the soil. These physical characteristics can hinder the uptake of nutrients and water from the soil, essential for its growth. Some studies aim to identify soybean genotypes that manage to develop in adverse conditions in the soil without changes that affect the root system in a harmful way. The analysis of images of soybean seedlings through Convolutional Neural Networks (CNNs) to identify the genotype and soil condition where they were grown, can increase the productivity rate by identifying which genotype best developed in a given soil condition. In this work, a soybean seedlings dataset grown in compacted and non-compacted soil is generated, with two versions, one with the complete seedling and the other with the rootless seedlings. The main objective of this work is the training of CNNs for soil condition classification, and the approach of a Multitasking Convolutional Network developed for genotype classification. For genotype classification, all trained models obtained good accuracy rates, with VGG-16 achieving the highest values with an average accuracy of 91.5% and 85.6% in the test phase for a dataset with complete seedling and rootless respectively. In the genotype classification problem, the Multitasking network obtained the best results with 52.3% for the dataset with complete seedling and 47.1% with seedling rootless, both in the test phase.

Keywords: Multitasking Convolutional Networks. Convolutional Neural Networks. Soybean Seedling. Soil Condition Identification. Genotype Classification.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

A implantação de técnicas de visão computacional com finalidade de correlacionar características são utilizadas para auxiliar especialistas na tomada de decisão em diversas áreas no mundo real como: na saúde, indústria e na agricultura de precisão. Esse impulso teórico e conceitual abrange a aplicação de métodos convolucionais profundos. Podem ser citadas o emprego de técnicas de classificação de imagens e localização de objetos em tarefas como análise de imagens para detecção de doenças em humanos (MAHDY et al., 2020; SUN et al., 2016), classificação de animais marinhos (CAO et al., 2015) e componentes de linhas de alta tensão (SILVA et al., 2021). Assim como problemas denominados classificação de imagens *Fine-Grained*, onde a tarefa de classificação concentra-se na diferenciação entre classes de objetos difíceis de distinguir, por exemplo tem-se a distinção entre modelos de carros (YANG et al., 2015), diferenciação de modelos de avião (MAJI et al., 2013) e classificação de espécies de pássaros (HORN; PERONA, 2017).

No contexto da agricultura de precisão, existem na literatura problemas com várias abordagens e aplicações com arquiteturas de CNNs para identificação de espécies vegetais e doenças através de imagens de folhas, sementes, raízes, flores, mudas e frutos (KUMAR; CHAUDHARY; CHANDRA, 2021; SHRIVASTAVA; PRADHAN, 2021; LIANG; WANG; LING, 2021). Além dessas implicações, as culturas agrícolas sofrem com problemas de condições adversas do clima e do solo, como por exemplo a compactação do solo. Entre essas culturas, encontra-se a soja.

O estudo do cultivo de sementes de soja em solo compactado é um problema relevante devido aos impactos que a compactação pode exercer sobre produtividade. Nas próximas seções, o problema da compactação do solo será exposto, indicando a sua importância e justificativa. Logo, a hipótese e os objetivos deste trabalho serão apresentadas. Por último, será mostrada a organização deste trabalho.

1.1 O problema e sua importância

O cultivo da soja gera interesse comercial de vários países no mundo principalmente pela versatilidade do grão podendo ser utilizado como alimento e como matéria prima para a produção de cosméticos e medicamentos. O Brasil é o maior produtor dessa cultura, atingindo esse patamar através do investimento em pesquisa e desenvolvimento de novas tecnologias para o aumento da produtividade. A produção brasileira na safra de 2021/2022 foi de 123,8 bilhões de toneladas empregando para o cultivo 40,9 bilhões de hectares, alcançando uma produtividade de 3.026 kg/ha. O valor da produção foi cerca de 341,7 bilhões de reais (EMBRAPA, 2022; IBGE, 2022).

Para manter a produção da soja em alta são necessários investimentos no desenvolvimento de tecnologias capazes de controlar pragas, doenças e problemas com o solo como a compactação para evitar a baixa produtividade e perda do valor de produção. A soja como diversas culturas agrícolas, é sensível à compactação do solo (BEUTLER et al., 2006), essa condição afeta negativamente o crescimento das plântulas de soja devido ao aumento da dificuldade de penetração das raízes no solo, modificando o desenvolvimento do sistema radicular gerando raízes curtas e grossas (BATEY, 2009). Como a captação de nutrientes e água do solo ficam comprometidos devido a má formação das raízes, a plântula pode não se desenvolver de forma adequada, acarretando no crescimento desigual das plantas e prejudicando a escolha de qual época é apropriada para a dessecação e colheita. Atualmente, com o aumento do tamanho e peso das máquinas agrícolas, associado ao manejo intensivo das áreas cultivadas, existe um grande potencial para intensificar a compactação do solo. Empresas de melhoramento genético têm realizado estudos para avaliar o impacto da condição do solo sobre o genótipo da soja, visando encontrar genótipos menos afetados pela compactação do solo. Um problema em aberto é identificar a condição do solo em que a soja foi cultivada por meio da análise de imagens de plântulas.

Observando esse cenário, este trabalho propõe através de imagens de plântulas de soja a identificação automática do genótipo e da condição do solo em que a plântula foi cultivada, com a finalidade de auxiliar especialistas na escolha do melhor genótipo para uma determinada região.

1.2 Justificativa

Os trabalhos encontrados na literatura abordam o estudo dos atributos das plântulas de soja cultivadas em solo compactado, não existindo uma proposta através das imagens de plântulas. Em (CAPOBIANGO et al., 2022) por exemplo, foram realizadas medições e análise das características físicas e químicas das plântulas de soja objetivando-se em identificar genótipos que melhor se desenvolvem em solo compactados. Para o estudo foi necessário o cultivo e extração das plântulas, sendo que esta última atividade mata a plântula.

Devido a importância da classificação automática de plântulas de soja com associação ao genótipo e a condição do solo e pela não existência dessa abordagem na literatura, este trabalho se torna relevante ao propor uma metodologia de classificação através de imagens.

1.3 Hipótese

Com o aumento do volume de dados representados por imagens e geração de diferentes conjuntos de dados compostos por imagens, surgiram várias abordagens de

aplicações das Redes Neurais Convolucionais para problemas de classificação, segmentação e identificação de objetos em imagens devido a boas taxas de precisão obtido por essas redes observado em trabalhos presentes na literatura.

Portanto, para o problema classificação de plântulas de soja em imagens pelo genótipo e condição do solo, pode ser proposto o uso de Redes Neurais Convolucionais.

1.4 Objetivos

O objetivo principal deste trabalho é através da análise de imagens identificar o genótipo e a condição de solo em que plântulas de soja foram cultivadas. Para alcançar esse objetivo, são propostos os seguintes objetivos específicos:

1. Criar um conjunto de dados composto por imagens de plântulas de soja individuais, variando o genótipos da plântula e a condições de compactação de solo em que a mesma foi cultivada. De forma que o conjunto garanta a representabilidade dos dados, tenha balanceamento entre as classe e não possua enviesamento.
2. Utilizar técnicas avançadas de treinamento e propor arquiteturas de Redes Neurais Convolucionas que não seja simplesmente a aplicação de um *backbone* clássico, para a classificação das imagens de plântulas de soja com relação a condição do solo e genótipo.
3. Analisar a interpretabilidade dos modelos treinados com relação as características das entradas para a classificação da condição do solo.

1.5 Organização desta dissertação

O restante deste texto está estruturado como detalhado abaixo:

- Capitulo 2 [Fundamentação Teórica]: exibe os fundamentos teóricos para do Processamento Digital de Imagens, assim como Redes Neurais, Aprendizado Profundo, *Autoencoders*, e Aprendizado Multitarefa.
- Capitulo 3 [Trabalhos Relacionados]: são apresentados trabalhos de criação e treinamento de CNNs, assim como diferentes abordagens de *Autoencoders* e redes Multitarefas.
- Capitulo 4 [Criação do Conjunto de Dados de Plântulas de Soja]: a abordagem do problema de classificação da condição do solo e identificação de genótipos de plântulas de soja com modelos CNNs precisa de um conjunto de dados balanceado e não enviesado, possibilitando o treinamento de modelos que possam ser capazes

de atingir boas taxas de acerto na classificação. Com essa finalidade, o capítulo apresenta o processo de criação da base de dados de plântulas de soja, abordando o cultivo das sementes de soja, coleta das plântulas germinadas, forma de aquisição das imagens e todo o processamento para o corte das plântulas contidas nas imagens para a formação do conjunto de dados com imagens individuais de plântula.

- Capitulo 5 [Rede de Aprendizagem Multitarefa]: o capítulo aborda a Rede Multitarefa desenvolvida para reconstrução de imagens e classificação de genótipos. A rede é dividida em Braço Codificador (Compartilhado); Braço de Reconstrução e Braço Classificador.

- Capitulo 6 [Material e Métodos]: são descritos os métodos propostos, implementados e testados para geração do conjunto de dados e classificação das imagens de plântulas de soja na condição de solo compactado e não compactado e 30 genótipos distintos.

- Capitulo 7 [Resultados e Discussão]: expõe os resultados obtidos nos treinamentos dos modelos.

- Capitulo 8 [Conclusões]: apresenta as considerações finais e possíveis trabalhos futuros.

2 Fundamentação Teórica

Neste capítulo, serão exibidos os fundamentos teóricos para do Processamento Digital de Imagens, Redes Neurais Artificiais, Aprendizado Profundo, *Autoencoders*, e Aprendizado Multitarefa.

2.1 Processamento Digital de Imagens

Processamento de imagem digital é a utilização de um computador através de algoritmos para processar imagens digitais. Dentre os algoritmos, podem ser realizadas operações para remoção de ruídos, segmentações, e identificação de contornos de objetos (CHAKRAVORTY, 2018).

2.1.1 Conversão de imagens RGB para escala de cinza

Existem vários métodos para a conversão de imagens RGB (do inglês *Red Green Blue*) em uma imagem em tons de cinza, como por exemplo: o método da média e o método ponderado.

- **Método da média:** calcula o valor médio de R, G e B para obter o valor correspondente em escala de cinza, através da fórmula: $(R + G + B)/3$;
- **Método ponderado** O Método Ponderado adiciona uma ponderação referente a cada canal. Como exemplo tem-se a fórmula: $0{,}299R + 0{,}587G + 0{,}114B$. (SARAVANAN, 2010)

Exemplos da aplicação dos dois métodos são apresentados na Figura 1. A imagem produzida pelo método da média apresenta uma tonalidade mais escura em comparação com a gerada pelo método ponderado. Essa característica é devido ao uso da intensidade resultante para a conversão da imagem, que é a média dos valores de intensidade de cor de cada pixel. Podendo levar a uma imagem mais escura, sendo que a média dos valores de intensidade de cor geralmente é menor do que os valores máximos de cada canal.

Figura 1 – Exemplos da aplicação dos métodos para conversão em escala de cinza. Da esquerda para a direita, imagem original, método da média e ponderado.

2.1.2 Binarização

A Binarização ou Limiarização é um processo de segmentação das imagens cuja finalidade é separar as regiões de interesse de uma imagem. Um dos métodos é o *Threshold* que utiliza um valor de corte para a limiarização separando regiões da imagem em pixels pretos e brancos. Para determinadas situações, um limiar fixo não retorna bons resultados, sendo necessário técnicas de limiarização variáveis (GONZALEZ; WOODS, 2006). A limiarização multinível é um processo que determina mais de um limiar para uma imagem em escala de cinza, segmentando-a em várias regiões distintas que podem corresponder ao fundo ou a outros objetos (ARORA et al., 2008). Já o método de limiarização local seleciona diferentes valores de limiar para cada pixel na imagem com base na análise dos pixels vizinhos, obtendo bons resultados para imagens com diferentes níveis de contraste (SAUVOLA; PIETIKÄINEN, 2000). O algoritmo procura iterativamente o limite que minimiza a variância dentro da classe, definida como uma soma ponderada das variâncias das duas classes (*background* e *foreground*). As cores em tons de cinza geralmente estão entre 0-255 (0-1 no caso de *float*). Portanto, se escolhermos um limite de 100, todos os pixels com valores menores que 100 se tornarão o plano de fundo e todos os pixels com valores maiores ou iguais a 100 se tornarão o primeiro plano da imagem.

Uma das técnicas de binarização mais comuns é o método de Otsu. Para aplicação do método inicialmente é obtido o histograma da imagem, este se refere a distribuição dos pixels na imagem. O algoritmo processa o histograma e busca iterativamente o limiar que minimiza a variância dentro da classe, definida como uma soma ponderada das variâncias das duas classes: fundo e objeto, sendo expressa por:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t), \tag{2.1}$$

sendo $\sigma_w^2(t)$ representa a variância das classes, $w_1(t), w_2(t)$ as probabilidades das duas classes divididas por um limiar t.

A Figura 2, exibe um exemplo de binarização realizada com o método de Otsu.

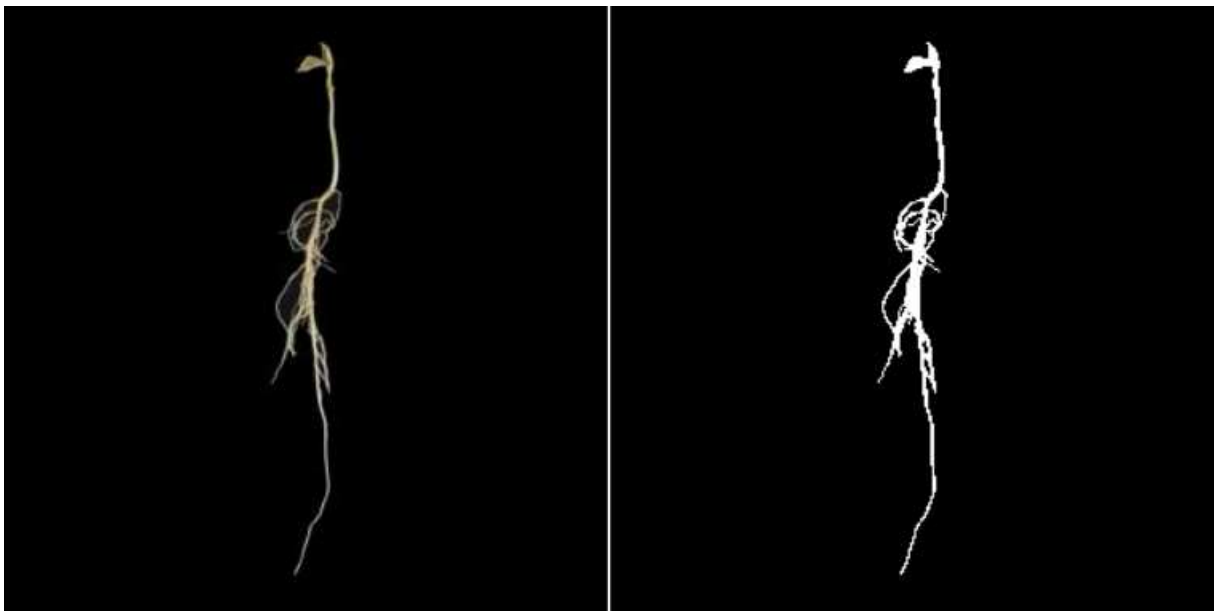


Figura 2 – Exemplo de imagem binarizada pelo método de Otsu. Imagem esquerda original, imagem direita binarizada.

2.1.3 Operações Morfológicas

A morfologia é um conjunto de operações de processamento de imagens. As operações morfológicas aplicam um elemento estruturante a uma imagem de entrada resultando em outra com mesmo tamanho. A dilatação e erosão são as operações morfológicas mais simples. A dilatação adiciona pixels nas fronteiras dos objetos em uma imagem, enquanto a erosão remove os pixels das fronteiras do objeto.

2.1.3.1 Erosão

A erosão encolhe a imagem de acordo com o elemento estruturante passado para a operação. Semelhante ao processo de dilatação, o elemento estruturante desloca-se da esquerda para a direita e de cima para baixo.

Formalizando o processo de erosão, sendo A e B pertencentes ao conjunto $\mathbb{Z}^2$, a erosão de A com B denotada por $A \ominus B$, é definida:

$$A \ominus B = \{z|(B)_z \subseteq A\}. \tag{2.2}$$

Essa equação indica que a erosão de A por B é o conjunto de todos os pontos z tais que B, transladado por z, está contido em A, sendo o conjunto B o elemento estruturante e A o conjunto que passará por erosão. Como a afirmação de que B deve estar contido em A é equivalente a B não compartilhar nenhum elemento comum com o fundo, podemos expressar a erosão da seguinte forma equivalente:

$$A \ominus B = \{z|(B)_z \cap A^C = \oslash\}, \tag{2.3}$$

onde, A^C é o complemento de A e $\oslash$ é o conjunto vazio (GONZALEZ; WOODS, 2006).

Na Figura 3-(b) é exibido o resultado da operação de erosão.

2.1.3.2 Dilatação

A dilatação é um processo onde a imagem binária é expandida de sua forma original, dependendo do elemento estruturante determinado. O processo de dilatação é semelhante a uma convolução, sendo o elemento estruturante refletido e deslocado da esquerda para a direita e de cima para baixo, a cada deslocamento.

Matematicamente, sendo A e B pertencentes ao conjunto $\mathbb{Z}^2$, a dilatação de A com B denotada por $A \oplus B$, é definida:

$$A \oplus B = \{z|(\hat{B})_z \cap A \neq \oslash\}. \tag{2.4}$$

A equação se baseia em refletir B sobre sua origem, e deslocar esta reflexão por z. A dilatação A por B é o conjunto de todos os deslocamentos, z, tais que $\hat{B}$ e A se sobrepõem em pelo menos um elemento. Com essa interpretação, temos a equação equivalente:

$$A \oplus B = \{z| \left[(\hat{B})_z, \cap A\right] \subseteq A\}, \tag{2.5}$$

assumindo para ambas as equações que B é o elemento estruturante e A o conjunto a ser dilatado (GONZALEZ; WOODS, 2006).

Na Figura 3-(c) é exibido o resultado da operação de dilatação.

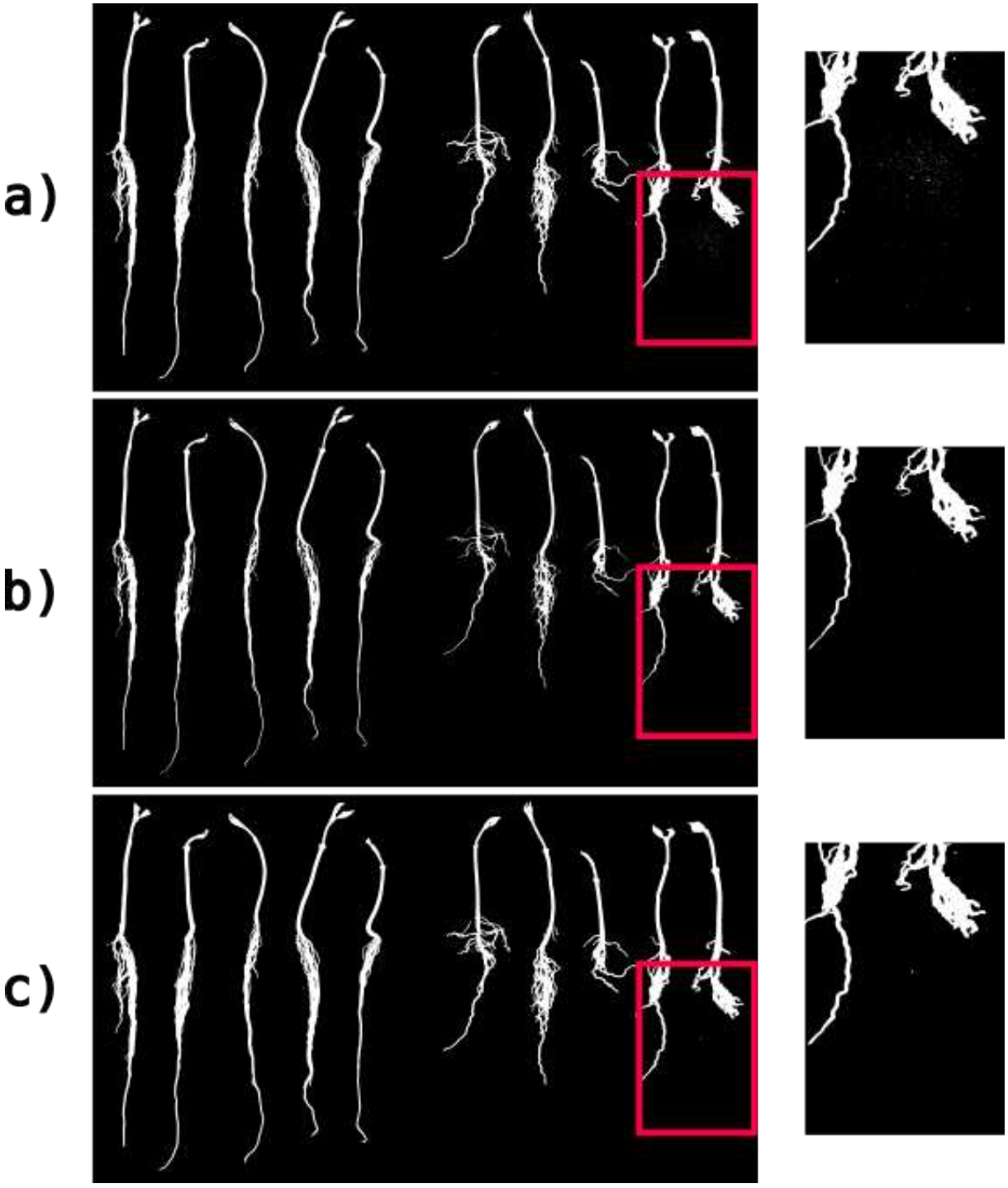


Figura 3 – Exemplo de operações morfológicas em uma imagem binarizada. (a)Imagem binarizada. (b) Imagem após operação de erosão. (c) Imagem após operação de dilatação.

2.2 Redes Neurais Artificiais

O estudo das reações do cérebro de animais (principalmente de seres humanos) proporcionou conhecimento do funcionamento de Redes Neurais Biológicas, sendo vistas como redes complexas formadas por neurônios que se comunicam por ligações sinápticas.

Na década de 1950, essas redes motivaram o surgimento das Redes Neurais Artificiais (RNA) como alternativas para a solução de problemas computacionais difíceis (ROSENBLATT, 1961).

As RNAs são formadas por unidades de processamento simples, chamados neurônios, tendo o potencial para adquirir e transmitir conhecimento (RUSSELL; NORVIG, 2009). Durante o processo de treinamento, os ajustes dos parâmetros (pesos atrelados aos neurônios) permitem a aprendizagem das redes, sendo a forma do aprendizado definido de acordo com as alterações dos parâmetros (HAYKIN, 2007).

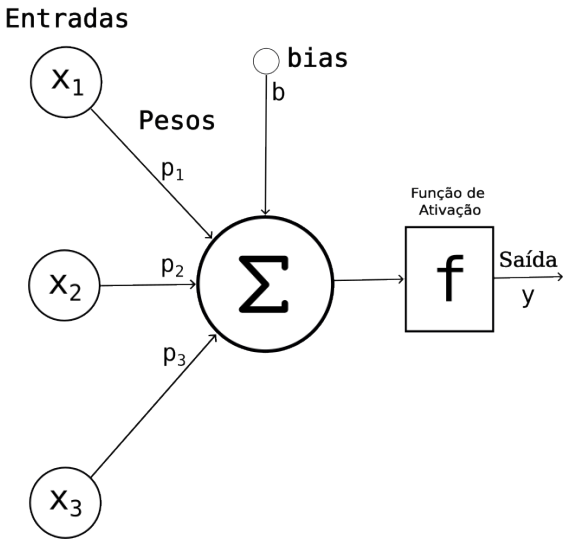


Figura 4 – Modelo Perceptron

A Equação define a modelagem matemática do neurônio artificial:

$$yk = f\left(bk + \sum_{i=1}^{n} x_i \cdot w_{ki}\right), \tag{2.6}$$

onde yk é a saída do gerada pelo neurônio k. O vetor de entrada no neurônio xi e os pesos atrelados wi são multiplicados, formando a saída do combinador linear posteriormente somado ao bias b. O bias influência na entrada da função de ativação f, sendo esta, utilizada para criar a não linearidade na formulação (TAN; STEINBACH; KUMAR, 2005). A Figura 4, mostra o modelo conhecido como modelo perceptron.

A extensão do perceptron é denominada *Multilayer Perceptron (MLP)*. A MLP é unidirecional composta por camadas formadas por neurônios que enviam informações para as camadas seguintes (RUSSELL; NORVIG, 2009). As camadas internas ou ocultas, ficam entre a camada de entrada e a de saída, a finalidade das camadas internas é aumentar o aprendizado da rede atuando como um extrator de características complexas (HAYKIN, 2007).

A Figura 5 ilustra uma MLP, com 1 camada oculta, 3 atributos de entrada e dois estados de saída.

O algoritmo de correção de erros amplamente utilizado nas RNAs é o *backpropagation*, definido pela propagação dos sinais e a retropropação do erro. Na propagação, os neurônios produzem sinais de acordo com a entrada recebida, os sinais são disparados e propagados para os neurônios das camadas seguintes até a última camada gerando a saída da rede. O erro é calculado pela distância do resultado disparado pela da rede com relação à saída esperada, a cada camada, o erro é retro-propagado realizando ajustes nos valores dos pesos das redes para diminuir o erro na próxima saída (HAYKIN, 2007).

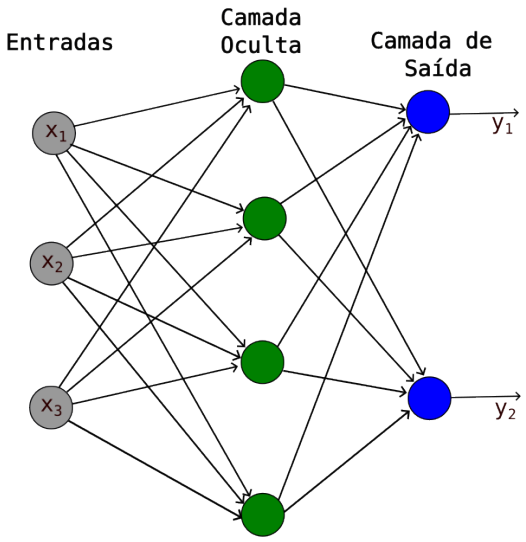


Figura 5 – Exemplo de uma rede MLP

A qualidade do treinamento de uma RNA dá-se devido ao balanceamento dos dados, arquitetura utilizada, escolha da função de loss, inicialização dos pesos, número de épocas necessárias e configuração dos hiperparâmetros visando superar os problemas de *underfitting* e *overfitting*. *Underfitting* ocorre quando o erro no conjunto de treinamento é alto, não sendo capaz de aprender e correlacionar as características de entrada com as saídas; o *overfitting* ocorre quando o erro no conjunto de treinamento é baixo mas é alto para o conjunto de teste, indicando pouca capacidade de generalização do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.3 Redes Convolucionais

O amplo acesso a conjunto de dados com um grande volume de imagens como: ImageNet (DENG et al., 2009), PASCAL VOC (EVERINGHAM et al., 2010), e MS COCO (LIN et al., 2014), possibilitou o avanço das CNNs (PONTI et al., 2017). Por outro lado, é relevante o acréscimo da capacidade de processamento paralelo das GPUs,

proporcionando a elaboração de modelos profundos com alto volume de parâmetros (GOODFELLOW; BENGIO; COURVILLE, 2016).

As Redes Neurais Convolucionais (do inglês *Convolutional Neural Networks - CNNs*) são uma das classes de métodos de *Deep Learning*, sendo constituídas essencialmente por camadas convolucionais para o tratamento de informações visuais (PONTI et al., 2017). As CNNs são redes neurais que usam convolução no lugar da multiplicação geral de matrizes em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016), elas podem ser divididas em: a) camadas convolucionais; b) camadas de pooling; e c) camadas totalmente conectadas. A organização das dá-se por intercalação das camadas podendo ter camada de pooling e uma totalmente conectada. A seguir são retratadas as particularidades de cada camada.

2.3.1 Camada Convolucional

Nas camadas convolucionais ocorrem o processo de convolução nos mapas de características resultantes da camada antecedente, onde são utilizados um conjunto de filtros com o objetivo de identificar atributos. Sendo l o índice da camada convolucional formada por um grupo de K filtros, com W_k^l e bias b_k^l, para $k \in [1, \ldots, K]$. O volume, M^{l-1}, é composto por F mapas de características gerados da camada precedente $l-1$. Para constituir o volume de saída M^l, são convolucionados cada parte $f \in [1, ..., F]$ do volume de saída M^{l-1} para cada $k \in [1, ..., K]$ filtros W_k^l. A Equação 2.7 exibe a formulação de um componente do M_k^l.

$$M_k^l = \sum_{f=1}^{F} M_f^{l-1} * W_f^l + b_k^l : \tag{2.7}$$

Como exemplo de convolução, dada uma entrada com as dimensões 3×3 e f o filtro de tamanho f $\times$ f (não precisar ser necessariamente quadrado), onde f = 3, 4, 5 e assim por diante desde que o tamanho do filtro seja menor que o da entrada. A máscara de filtro desliza sobre toda a imagem de entrada, calculando o produto escalar entre os pesos dos filtros com os valores da entrada, o que resulta na produção de um mapa de ativação. A Figura 6, exibe o resultado de uma convolução.

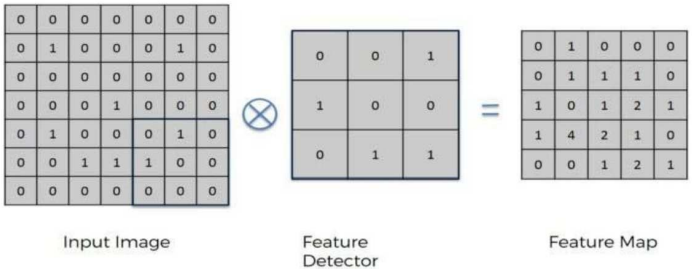


Figura 6 – Resultado de uma convolução. **Fonte**: (LI et al., 2019)

2.3.2 Camada de *Pooling*

As camadas de *pooling* geralmente aparecem em sequência a uma de convolução, tendo como finalidade reduzir o tamanho dos mapas de ativação. A operação de *pooling* envolve o deslizamento de um filtro bidimensional sobre cada canal do mapa de características. Os principais tipos de *pooling* são:

Average-pooling: calcula a média dos elementos presentes na região do mapa de características coberta pelo filtro, retornando como saída um mapa de características contendo a média do valores de cada região, como exibido na Figura 7-(a).

Max-pooling: seleciona o elemento máximo da região do mapa de características sobreposta pelo pelo filtro, gerando como saída um mapa de características contendo somente os maiores valores de cada região, pode ser visto na Figura 7-(b).

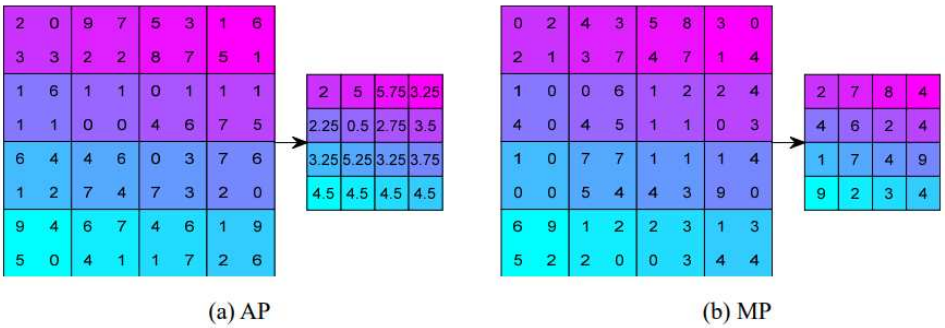


Figura 7 – *Average-pooling* e *Max-pooling*. *Average-pooling* (a) calcula a médias dos valores; *Max-pooling* (b) retorna o maior valor. **Fonte**: (ALBAWI; MOHAMMED; AL-ZAWI, 2017)

2.3.3 Camada Totalmente Conectada

As Camadas Totalmente Conectadas são camadas da rede que possuem todos os seus neurônios ligados a todos os elementos de entrada da camada. Essas camadas são computacionalmente mais custosas, dependendo da arquitetura essas camadas podem possuir grande parte dos parâmetros de uma rede convolucional.

2.3.4 Blocos Residuais

Os blocos residuais tem a finalidade de solucionar o problema do *Vanishing Gradient*, que ocorre quando a representação de uma imagem é transformada em várias camadas e acaba perdendo características relevantes ao longo do caminho. Os blocos residuais permitem que a informação original seja mantida através da adição de uma conexão residual entre a entrada e a saída da camada. A rede deixa de aprender a transformar a entrada completamente, para aprender apenas a diferença entre a entrada e a saída desejada. Resultando em um gradiente mais robusto, permitindo que o modelo aprenda de forma mais eficiente (HE et al., 2016). Na Figura 8, é exibido a formulação de um bloco residual.

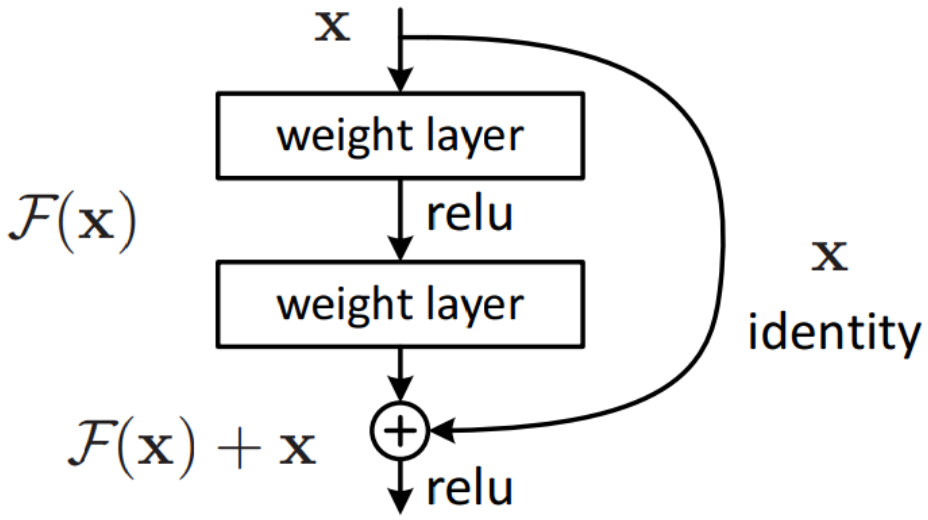


Figura 8 – Formulação de um Bloco Residual. **Fonte**: (HE et al., 2016)

2.4 Análise visual baseados na retropropagação de gradiente

Durante o treinamento de modelos é relevante reconhecer os padrões que estão sendo aprendidos, sobretudo se estão levando a um resultado correto. As CNNs funcionam como uma espécie de "caixa preta", sendo difícil identificar quais características a rede está correlacionando para um disparar um resultado final. Análises visuais foram desenvolvidas para distinguir esses padrões aprendidos pelas redes. Alguns métodos baseados na visualização da retropropagação dos gradientes permitem identificar regiões de interesse e pixels da imagem que foram responsáveis pelo resultado do modelo.

2.4.1 *Guided Backpropagation*

O *Guided Backpropagation* (GBP) é uma técnica que detecta o gradiente em relação às imagens durante o processo de retropropagação. Através da função de ativação ReLU, ocorre o fluxo apenas dos gradientes positivos, alterando os valores dos negativos para zero, possibilitando visualizar as características da imagem que ativam os neurônios da rede. Seja f o mapa de características de uma camada l, logo, a passagem a frente é $f_i^{l+1} = Relu(f_i^l, 0)$, sendo o retorno $R_i^l = (f_i^l > 0) \cdot (R_i^{l+1} > 0) \cdot (R_i^{l+1})$, onde R é um resultado intermediário no cálculo da retropropagação para a camada l. A saída final do GBP é uma imagem com as mesmas dimensões da entrada, exibindo os atributos existentes na imagem que aumentaram a ativação dos mapas de características (SPRINGENBERG et al., 2015). Um exemplo pode ser visto na Figura 9.

Figura 9 – *Guided Backpropagation* para uma imagem de plântula de soja

2.4.2 *Gradient-weighted Class Activation Mapping*

O *Gradient-weighted Class Activation Mapping* (Grad-CAM) utiliza gradientes de uma determinada classe alvo que fluem para a para a última camada convolucional, destacando regiões de interesse na imagem de entrada através de um *Heat-map*. Para obter o $L^c_{Grad-CAM} \in \mathbb{R}^{u \times v}$ referindo-se como mapa de localização do Grad-CAM, sendo c a classe alvo, u largura e v altura de A, inicialmente é calculado o gradiente de y^c (pontuação para cada classe c) em relação aos mapas de características A de uma camada convolucional, ou seja, $\frac{\partial y^c}{\partial A^k_{ij}}$ (SELVARAJU et al., 2017). Posteriormente esses gradientes passam por *Global-average-pooled* para obter pesos α^c_k:

$$\alpha^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}}. \tag{2.8}$$

Este peso α^c_k, representa uma linearização parcial da rede saindo de A obtendo a 'importância' do mapa de características de índice k para uma classe de destino c. Uma combinação ponderada dos mapas de características aplicadas a ReLU geram o *Heat-map* do Grad-CAM:

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right). \tag{2.9}$$

A Figura 10, mostra o resultado da execução do Grad-CAM em uma imagem dada como entrada.

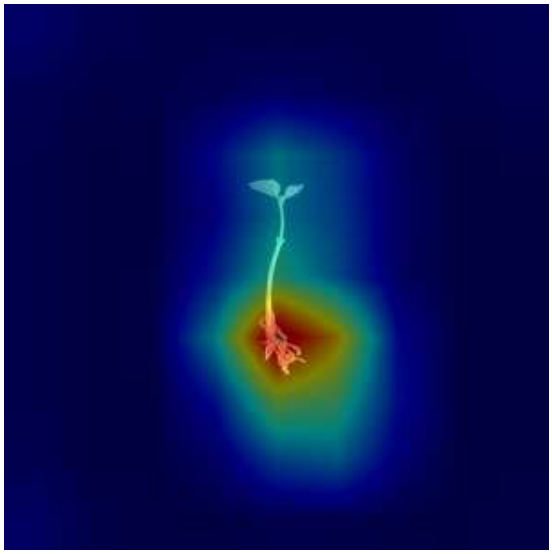


Figura 10 – *Grad-CAM* para uma imagem de plântula de soja

2.4.3 *Guided Grad-CAM*

O *Guided Grad-CAM* é uma combinação do Grad-CAM com o *Guided Backpropagation*, visando juntar os métodos, obtendo a discriminação de classes do Grad-CAM e alta resolução da *Guided Backpropagation*. Para produzir o *Guided Grad-CAM* é realizado uma multiplicação elemento a elemento das imagens resultados dos dois métodos, na Figura 11, vemos o resultado da junção.

Figura 11 – *Guided Grad-CAM* para uma imagem de plântula de soja

A combinação da discriminação de classes do Grad-CAM e alta resolução da *Guided Backpropagation* é realizada com a multiplicação elemento a elemento com a finalidade de se obter o *Guided Grad-CAM*. A combinação é exibida na Figura (SELVARAJU et al., 2017).

2.5 Autoencoders

Um *Autoencoders* (AE) é uma rede neural treinada com finalidade de gerar uma representação intermediária de menor dimensionalidade, de forma a não perder informações relevantes. Contendo uma ou mais camadas ocultas que descrevem um código representativo da entrada. A rede pode ser separada em duas partes: uma função codificadora $h = f(x)$ e um decodificador que produz uma reconstrução $r = g(h)$. Como o modelo é forçado a priorizar quais aspectos da entrada devem ser replicados, ele frequentemente aprende propriedades úteis dos dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Os AEs modernos generalizaram a ideia de um codificador e um decodificador, compartilhando um vetor denominado Espaço Latente, este é constituído pelas principais características da entrada, pode ser visto na Figura 12. Tradicionalmente, AEs eram usados para redução de dimensionalidade ou aprendizado de características, atualmente, conexões teóricas com modelos de variáveis latentes aumentaram os estudos dos AEs para a modelagem generativa como: Shen et al. (2020), Voynov e Babenko (2020) e Mukherjee et al. (2019).

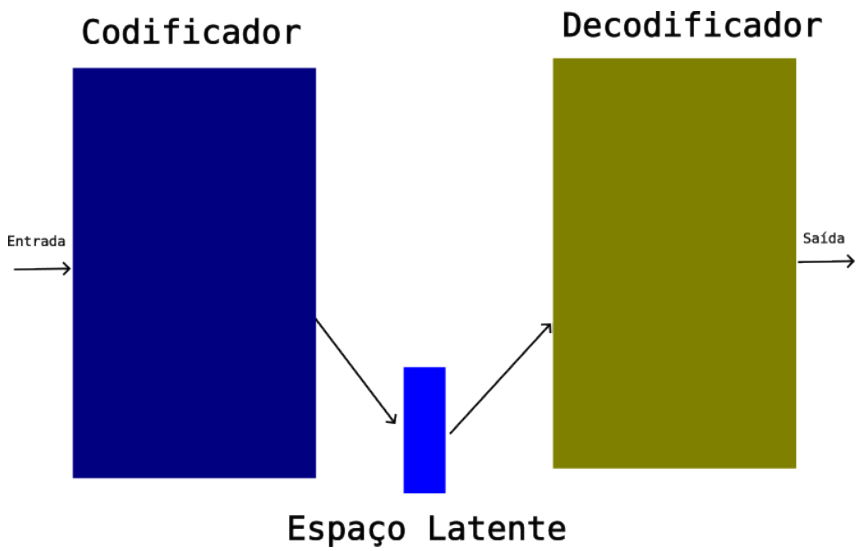


Figura 12 – Modelo de um *Autoencoder*

2.6 Rede Multitarefa

Em *Machine Learning*, geralmente objetiva-se em otimizar para uma métrica específica, como uma pontuação em um determinado *benchmark*. Para isso, normalmente são treinados um único modelo ou um grupo de modelos com base em um conjunto de dados para executar uma determinada tarefa, que pode ser uma regressão, classificação, reconstrução e segmentação de imagens, reconhecimento de discurso entre outras existentes. Os modelos são ajustados até que seu desempenho não aumente, alcançando um resultado aceitável, mas ao manter a concentração em uma única tarefa, são ignoradas as informa-

ções que podem ajudar a melhorar a métrica utilizada. Em específico, os dados de entrada podem ter anotações para mais de uma única tarefa, sendo possível o compartilhamento das características da entrada para todas as tarefas durante o processo de treinamento. Ao compartilhar representações entre as tarefas, é aumentada a capacidade de generalização do modelo para a tarefa original. Essa abordagem é denominada de *Multi-Task Learning* (RUDER, 2017).

O aprendizado multitarefa vem sendo empregado em várias abordagens de aprendizado do máquina como: processamento de linguagem natural (COLLOBERT; WESTON, 2008); reconhecimento de discurso (DENG; HINTON; KINGSBURY, 2013) para visão computacional (GIRSHICK, 2015) e descoberta de medicamentos (RAMSUNDAR et al., 2015).

As Redes Neurais Multitarefas recebem como entrada instâncias de treinamento referentes a todas as tarefas. As saídas de cada camada oculta podem ser vistas como a representação de características comuns aprendida por todas as tarefas. A transformação da representação original para a aprendida depende dos pesos que conectam a entrada e as camadas ocultas, assim como a função de ativação adotada. A arquitetura de redes multitarefas possui várias saídas de acordo com o número de tarefas existente, enquanto nas redes neurais para uma única tarefa existe uma única saída. Na Figura 13, é mostrada um exemplo de rede multitarefa, possuindo as tarefas A e B.

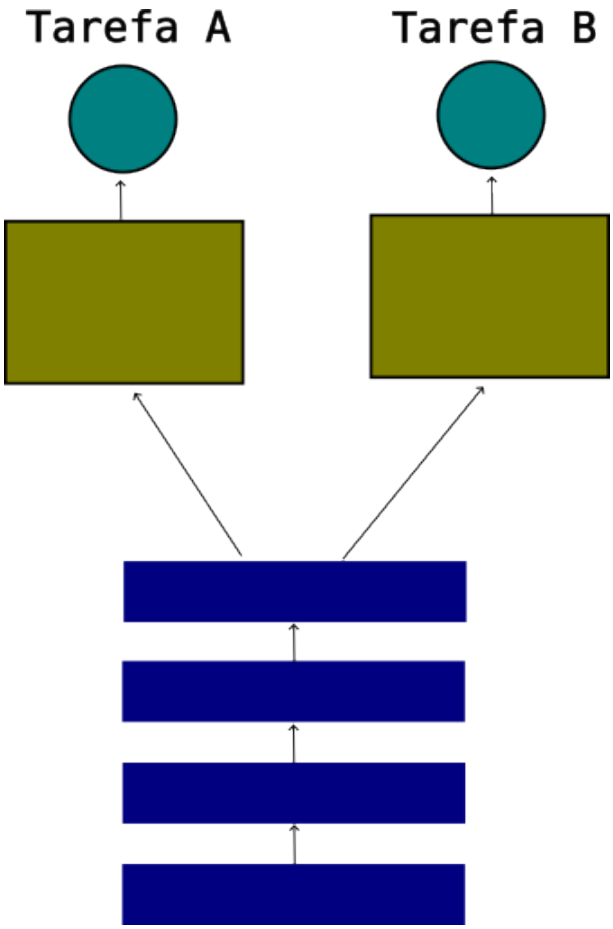


Figura 13 – Modelo de Rede Multitarefa

3 Trabalhos Relacionados

Neste capítulo, serão apresentados trabalhos de criação e treinamento de CNNs, assim como diferentes abordagens de *Autoencoders* e Redes Multitarefas. O capítulo é dividido em três seções. Inicialmente serão destacados a importância de ter conjuntos de dados balanceados para aplicações de aprendizado profundo em diferentes áreas. A segunda seção aborda um trabalho com aplicação de um *Autoencoder* com Blocos Residuais e a terceira seção a apresentação da utilização de rede multitarefas.

3.1 Criação de Conjunto de Dados e Treinamento de CNNs

Nesta seção, serão destacados trabalhos que abordam a construção de conjunto de dados para treinamento de redes convolucionais. A construção de um conjunto de dados tem a sua importância principalmente pela necessidade da qualidade dos dados, sendo que esses afetam diretamente a precisão e o desempenho dos modelos. Outro fator importante é representatividade do conjunto de dados, se o conjunto de dados de treinamento não é representativo, o modelo pode ter dificuldades em generalizar e fazer previsões precisas.

Os trabalhos seguintes visam a criação de conjuntos de dados através da captação de instâncias de forma automática, possibilitando a geração de uma base com um grande volume de imagens.

No trabalho de Laroca et al. (2018) é abordado o problema de Reconhecimento Automático de Placas de Veículos (do inglês *Automatic License Plate Recognition*), sendo divididos em três estágios: Detecção veículos e placas, segmentação e reconhecimento de caracteres. No trabalho de Laroca et al. (2018) são apresentados um sistema fundamentado no modelo detector de objetos YOLO (REDMON et al., 2016). No primeiro estágio, foram treinadas duas CNNs, uma para detecção de veículo na imagem de entrada e outra para detecção da placa no veículo detectado utilizando as arquiteturas Fast-YOLO (SHAFIEE et al., 2017) e YOLOv2 (DONG et al., 2018). Após a detecção da placa foi empregada a arquitetura CR-NET (MONTAZZOLLI; JUNG, 2017) para segmentação e reconhecimento de caracteres, sendo usado uma rede para segmentar os caracteres e depois outras duas para reconhecê-los. Removeu-se as quatro primeiras camadas da CNN para o reconhecimento de dígitos, sendo após a realização de testes obteve-se resultados semelhantes com a CNN completa porém com um custo computacional menor. Entretanto, para reconhecimento de letras (mais classes e menos exemplos), foram mantidas todas as camadas da arquitetura. Para treinamento e testes em um cenário mais realístico, o trabalho aborda a criação do conjunto de dados UFPR-ALPR, contendo 4.500 imagens tiradas de dentro de um veículo em trânsito regular em ambiente urbano. As imagens fo-

ram obtidas de 150 vídeos com duração de 1 segundo e taxa de quadros de 30 quadros por segundo. Assim, o conjunto de dados é dividido em 150 veículos, cada um com 30 imagens com apenas uma placa visível em primeiro plano. Existindo pequenas variações na posição da câmera devido a montagens repetidas e também para simular uma condição real, onde a câmera nem sempre é colocada exatamente na mesma posição. Por fim, foram coletadas 1.500 imagens com cada câmera, assim divididas: 900 de carros com placa cinza, 300 de carros com placa vermelha e 300 de motocicletas com placa cinza. Como no Brasil, as placas possuem variações de tamanho e cor conforme o tipo do veículo e sua categoria, as placas dos carros têm um tamanho de 40 cm×13 cm, enquanto as placas das motocicletas têm 20 cm×17 cm. Veículos particulares têm placas cinzas, enquanto ônibus, táxis e outros veículos de transporte têm placas vermelhas. Essas diferenças de layouts e posições das placas são um dos desafios existentes no conjunto de dados. O modelo foi testado em dois conjuntos de dados distintos, o primeiro, *SegPlate Database* (GONÇALVES et al., 2016) formado por 2.000 quadros de 101 vídeos de veículos, alcançando uma acurácia de 93,53% a 47 quadros por segundo superando o sistema *Sighthound* (MASOOD et al., 2017) que teve acurácia de 89,80%. Para o conjunto de dados UFPR-ALPR, as versões de teste de sistemas comerciais alcançaram taxas de reconhecimento abaixo de 70%. Por outro lado, o modelo proposto no trabalho obteve melhor desempenho, com taxa de reconhecimento de 78,33%.

O trabalho de Laroca et al. (2018) emprega o uso das redes Fast-YOLO e YOLOv2 para identificação de carros e segmentação de placas, o trabalho desta diferencia-se pela abordagem de um modelo identificador e classificador de objetos o *Single Shot Detector* (LIU et al., 2016), com a finalidade de identificar nas imagens regiões que compõem as plântulas de soja, sendo essas regiões divididas em dois objetos rotulados de parte aérea e sistema radicular.

A classificação de modelos de carros pertence ao conjunto de problemas *Finegrained Visual Classification Tasks*, é abordado no trabalho de Kuhn e Moreira (2021). Devido ao avanço das CNNs em tarefas de classificação de imagens, aumentou-se os estudos de classificação de imagens de domínio específico como a distinção entre espécie de animais, plantas e modelos de veículos, sendo esses problemas denominados de *Finegrained Visual Classification Tasks*. No trabalho de Kuhn e Moreira (2021), é criado o conjunto de dados BRCars, constituído com imagens de modelos de carros brasileiros. Para a construção do conjunto de dados, foram coletadas imagens de anúncios de carros de um dos maiores sites de anúncios de carros do Brasil. No total, foi coletado um conjunto de 2.808.846 imagens, que estão distribuídas entre 1.005 modelos diferentes de carros. As imagens pertencem a um conjunto de 336.660 anúncios de carros. A coleta de imagens gerou um alto desequilíbrio de classe sendo alguns modelos de carros contendo um número maior de imagens do que outros, com objetivo de equilibrar o balanceamento foram gerados dois conjuntos para avaliar a tarefa FGVC, sendo denominados de BRCars-196 e BRCars-427.

O BRCars-196 é constituído por 196 classes, com cada classe relacionada a um modelo de carro. Na construção desse conjunto foram selecionados modelos com pelo menos 200 imagens, sendo escolhidos para cada modelo 200 instâncias de forma aleatória. Por fim, o BRCars-196 foi formado com 212.609 imagens, das quais 170.151 são separadas para a fase de treinamento, enquanto 42.458 são destinadas ao teste. BRCars-427 junta o BRCars-196 e 231 classes adicionais referentes a modelos com menos de 200 instâncias de carros. Classes com menos de 20 instâncias foram removidas para evitar um desbalanceamento capaz de enviesar o conjunto de dados. A adição das 231 classes teve a finalidade de replicar o desafio de lidar com modelos mais raros. Por fim, o BRCars-427 é composto por 300.325 imagens, das quais 239.668 são destinadas ao treinamento, e 60.657 são destinadas ao teste. Para experimentos com os conjuntos de dados, foram utilizadas as arquiteturas de CNN, InceptionV3 e ResNet50, e versões siamesas dessas redes. De acordo com os resultados obtidos, notou-se que o BRCars-427 possui baixos valores em todas as métricas, sendo um resultado esperado, devido ao desbalanceamento desse conjunto de dados.

Problemas visando o aumento de segurança de trabalhadores também são estudados e CNNs utilizadas em conjuntos de dados para esse fim. Veículos aéreos não tripulados *Unmanned Aerial Vehicle* estão sendo utilizados para para inspeção de linhas de energia captando imagens em regiões de risco. O trabalho de Silva et al. (2021) abordam a aplicação de CNNs em imagens de linhas de transmissão geradas por drones, os autores criam o conjunto de dados *STN Power Line Assets* contendo imagens reais e de alta resolução de vários componentes de linhas de energia de alta tensão. Posteriormente a criação do conjunto de dados resultados de métodos de detecção e classificação de objetos são comparados com uma abordagem proposta o *Multi-Size Power line Asset Detection*. As imagens do conjunto de dados foram geradas por um drone, mantendo sempre uma distância semelhante com relação à torre de transmissão, os ângulos de visão do drone foram variados com objetivo de garantir uma maior variabilidade da posição dos objetos nas imagens e igualmente adquirir características de condições diurnas, climáticas e iluminação. Com esse processo de coleta, obteve-se uma média de 18,1 instâncias por imagem capturada, com uma área média de pelo menos $2{,}89 \times 10^3$ pixels. Para anotar os 2.409 objetos em todas as 133 imagens capturadas foi utilizado um programa de computador, onde os autores marcavam cada objeto com uma caixa delimitadora. Segundo os autores o número relativamente baixo de imagens pertencentes ao conjunto de dados, é compensado devido o STN PLAD ter consideravelmente mais informações do que imagens regulares de conjuntos de dados comuns, como ImageNet (DENG et al., 2009) e MS COCO (LIN et al., 2014). Para a detecção e classificação de objetos, foram utilizadas *Single Shot MultiBox Detector* e Faster R-CNN (GIRSHICK, 2015), analisando os resultados obtidos, os autores propuseram uma simples modificação do pipeline para melhorar desempenho geral da detecção de ativos de linha de energia denominada *Multi-Size Power line Asset Detection*. A imagem de entrada é redimensionada para a primeira rede. Duas redes independentes

foram treinadas separadamente, enquanto uma usa o pipeline original que redimensiona sua entrada, sendo treinada sem a classe *Stockbridge damper* por essa ter um tamanho muito menor, dificultando a identificação após o redimensionamento da imagem. A segunda rede é responsável por detectar pequenos objetos, incluindo a classe *Stockbridge damper*. A entrada dessa rede é dividida em uma grade, gerando 16 outras imagens menores, processo que aumenta a taxa de identificação de objetos menores. Esta modificação simples permitiu uma melhoria de *Mean Average Precision* de 75,5% para 89,2%.

Como visto, em alguns trabalhos é essencial a criação do conjunto de dados para a realização de treinamento de modelos. Kuhn e Moreira (2021) expõem a criação conjunto de dados de modelos de carros brasileiros e Silva et al. (2021) a criação da base de dados com imagens de linhas de transmissão. Este trabalho aborda a criação de um conjunto de dados com imagens de plântulas de soja, diferenciando dos trabalhos citados pelas metodologias de aquisição de imagens e pré-processamento destas com técnicas de Processamento Digital de Imagens.

No contexto de trabalhos agronômicos, são encontrados na literatura muitas abordagens do uso de CNN com conjunto de dados constituído por imagens para a identificação e classificação de espécies. Como a forma de obtenção das instâncias que formam a base de dados geralmente são obtidas de forma manual, as bases geradas possuem um menor número de instâncias mas os autores visam manter uma boa representatividade de balanceamento dos conjuntos de dados.

A rotação de milho e soja é comumente utilizada pelos seus benefícios agronômicos e econômicos. Um incidente usual no ciclo do milho é o encontro de grãos e espigas que são deixados no campo por vários fatores, como acamamento e perdas na colheita. As sementes de milho que suportam ao inverno podem germinar junto com a soja plantada, originando uma população indesejável de milho voluntário *Volunteer Corn*. O milho voluntário age como erva daninha na cultura da soja e pode diminuir a qualidade da soja, além de reduzir o rendimento ao competir por água, nutrientes e luz. Como a inspeção manual é trabalhosa Flores et al. (2021) propuseram uma solução automática para a diferenciação da soja com o VC. Inicialmente foi proposto um conjunto de dados com 191 imagens de plântulas de soja e 220 de plântulas de milho para diferenciar as espécies em estágios iniciais. Para aquisição das imagens, as sementes de soja foram cultivadas em vasos retangulares ($30 \times 10 \times 10$ cm) com espaçamento de 9 cm, enquanto as sementes de milho foram plantadas em posições aleatórias, simulando os padrões reais de cultivo. A soja e o milho voluntário começaram a germinar entre quatro a seis dias após o plantio. A aquisição das imagens foi iniciada após sete dias de plantio através de uma câmera fotográfica e uma câmera infravermelha colorida, obtendo o conjunto de dados com imagens RGB e *Color-Infrared*. Além das versões obtidas os autores geraram uma nova através da fusão de imagens RGB e *Color-Infrared*. Para a fusão foram utilizadas as informações das imagens RGB e sua correspondente em *Color-Infrared* para gerar uma nova imagem

composta única, sendo utilizado a transformação *Wavelet*. Após a fusão das imagens, foram obtidos seis conjuntos de dados: RGB - milho voluntário e soja, *Color-Infrared*: milho voluntário e soja e fundido: milho voluntário e soja, com cada conjunto integrando 220 imagens milho voluntário e 191 imagens de soja. Foram treinados quatro diferentes classificadores de aprendizado de máquina e quatro modelos CNNs. Tanto os classificados de aprendizado de máquina e CNNs alcançaram maiores taxas de precisam com o conjunto de dados com imagens fundidas, sendo o GoogLeNet (SZEGEDY et al., 2015) selecionado por sua alta precisão de 99,9%, tempo de computação razoável (0,02 s por planta).

No trabalho de Flores et al. (2021) são propostas três versões do conjunto de dados para diferenciar plântulas de milho de milho voluntario através de modelos classificadores CNNs, diferentemente, na atual dissertação são treinados modelos de CNNs para a classificação da condição do solo onde a plântula de soja se desenvolveu e distinção entre genótipos.

O trabalho de Yang et al. (2022), objetiva-se na identificação e segmentação de vagens de soja em galhos para adquirir características fenotípicas morfológicas de plantas de soja. Especialmente, a contagem de vagens de soja é realizada manualmente. No entanto, as vagens de soja são pequenas possuindo várias formas e com número incerto de vagens por planta, bem como a ocorrência da sobreposição por galhos e outras vagens. No entanto, a contagem manual é cara, propensa a erros, e trabalhosa. Os autores propõem a utilização de dados sintéticos para a segmentação de vagens de soja, o uso de dados sintéticos visa a redução custos da anotação manual. Mesmo que os dados sintéticos não sejam fiéis em comparação com imagens reais, os autores afirmam que características críticas do conjunto de dados sintéticos podem ser obtidas automaticamente, sendo essa abordagem capaz de criar uma quantidade quase ilimitada de conjuntos de dados rotulados. No trabalho são criados dois conjuntos de dados para o problema de segmentação de plantas de soja: (*i*) Vagens de Soja *In-vitro* Sintéticas; e (*ii*)Planta de soja madura. Para a geração do conjunto de dados sintéticos, vagens de soja colhidas manualmente foram espalhadas sobre uma flanela de cor preta, posteriormente através de uma câmera posicionada acima das vagens, foram obtidas as imagens. Através de um software, as viagens foram cortadas e posicionadas em um fundo preto. O conjunto de dados de planta de soja madura do mundo real foi composto pela seleção aleatória de 60 espécimes de plantas, passando pelo o mesmo processo das imagens sintéticas, as espécimes foram posicionadas. Posteriormente a segmentação das imagens, foi realizada através da transferência de aprendizado em duas etapas. Na primeira etapa, uma rede de segmentação de instância pré-treinada no conjunto de dados MS COCO (LIN et al., 2014) foi treinada tendo como entrada o conjunto de dados de vagens de soja sintéticas. Na segunda etapa, realizou-se *finetunning* com o modelo treinado anteriormente treinando o novo modelo com amostras de plantas de soja do mundo real. Através dos resultados obtidos, os autores concluíram que o modelo treinado no conjunto dados planta de soja madura sem a transferência de

aprendizagem adquire apenas conhecimento grosseiro das regiões das vagens de soja, mas o modelo proposto onde se utiliza a transferência de aprendizado do conjunto de dados de vagem de soja sintética na primeira etapa e, em seguida, ajusta o conjunto de dados da planta de soja do mundo real, gera resultados melhores e mais satisfatórios que o primeiro modelo.

No trabalho de Yang et al. (2022) são utilizadas estratégias de *finetunning* para alcançar melhores taxas de acerto na segmentação de vagens. A dissertação diferencia do trabalho por não focar na tarefa de segmentação de imagens. São empregadas a transferência de aprendizagem em modelos pré-treinados com o conjunto de dados ImageNet e para a rede Multitarefa visando melhorar as taxas de acurácia.

3.2 *Autoencoder* com Blocos Residuais

A seção expõe o trabalho que apresenta redes *Autoencoder* com blocos residuais. O *Autoencoder* é uma técnica de aprendizado profundo que visa codificar e decodificar informações onde um modelo é treinado para codificar informações de entrada em um espaço de codificação de dimensão reduzida, e posteriormente decodificar esse espaço de codificação para recuperar a representação aproximada dos dados de entrada.

O trabalho de Wickramasinghe, Marino e Manic (2021) aborda o problema da degradação de desempenho do aprendizado não supervisionado profundo. Redes profundas como o *Autoencoder* podem sofrer pela deterioração das informações aprendidas devido ao desaparecimento do gradiente quando possuem um grande número de camadas ocultas. O estudo apresenta o *Residual Autoencoder* e sua versão convolucional *Convolutional Autoencoder* para aprendizado profundo não supervisionado, essas arquiteturas tem adição de conexões residuais para o aumento da capacidade da rede sem incidir na degradação da aprendizagem profunda em comparação com os *Autoencoders* padrão. As *Residual Autoencoders* e o *Convolutional Autoencoder* possuem as mesmas características das *Autoencoders*, como extração automatizada de recursos não lineares e aprendizado não supervisionado, eles possibilitam a projeção de redes maiores sem efeitos adversos no desempenho da aprendizagem. Os autores realizaram treinamentos para avaliar a *Residual Autoencoder* e a *Convolutional Autoencoders* com relação aos *Autoencoders* nos conjuntos de dados MNIST (DENG, 2012), Fashion MNIST (XIAO; RASUL; VOLLGRAF, 2017) e CIFAR10 (KRIZHEVSKY; HINTON et al., 2009). O acréscimo do número de camadas do *Convolutional Autoencoder* superou o *Autoencoder*, mostrando uma menor degradação da acurácia de classificação, maior acurácia média e baixa variância na acurácia em comparação com o *Autoencoder*.

Nesta dissertação, será utilizada a proposta de adição de blocos residuais em uma CNN, a proposta de rede CNN que será vista no próximos capítulos dessa dissertação

diferencia do trabalho de Wickramasinghe, Marino e Manic (2021) por ser uma aplicação com rede Multitarefa, sendo utilizada para classificação e reconstrução de imagens de plântulas de soja.

3.3 Abordagem com Rede Multitarefas

Para o contexto de redes Multitarefas, encontram-se trabalhos que abordam redes que objetivam-se em diversas tarefas como: classificação, reconstrução e segmentação de imagens.

O trabalho de Nguyen et al. (2019), destaca a análise de imagens e vídeos digitais manipulados por *deepfake*, sendo um problema que tem ganhado destaque pelo compartilhamento em redes sociais de imagens e vídeos alterados. A identificação de partes de uma imagem que foram alteradas pode ser realizada ao segmentar toda a imagem de entrada e executar uma classificação binária através de uma janela deslizante. O método proposto pelos autores adota essa abordagem de identificação com uma diferença, sendo apenas as áreas faciais consideradas em vez de toda a imagem, essa modificação visa diminuir custo computacional ao lidar com grandes entradas. Logo, os autores propõem uma rede neural convolucional utilizando aprendizado multitarefa para detecção simultânea de imagens e vídeos manipulados, também identificando as regiões alteradas. Os dados obtidos pela execução de uma tarefa são compartilhadas com uma outra, melhorando o desempenho de ambas as tarefas. O método proposto retorna a probabilidade de uma entrada ser falsa e os mapas de segmentação das regiões manipuladas em cada quadro da entrada. A rede é composta por um *Autoencoder* em forma de Y, com a finalidade de retornar a versão reconstruída da imagem de entrada, a probabilidade da imagem ter sido falsificada e o mapa de segmentação correspondente a essa imagem. O particionamento de recursos latentes e o design do decodificador em forma de Y permitem que o *Autoencoder* compartilhe informações importantes entre as tarefas de classificação, segmentação e reconstrução, melhorando o desempenho geral. O treinamento foi realizado com os conjuntos de dados *FaceForensics* e *FaceForensics++* (ROSSLER et al., 2019). Cada conjunto de dados foi dividido em 704 vídeos para treinamento, 150 para validação e 150 para teste, com a a base de dados fornecendo as máscaras de segmentação correspondentes aos vídeos manipulados. Os resultados foram comparados com a reimplementação do modelos *ForensicsTransfer*, a rede proposta atingiu 54,07% de acurácia para classificação e 84,67% para segmentação. Com objetivo de alcançar melhores resultados os autores utilizaram os vídeos separados anteriormente para validação do conjunto de dados para realizar *fine-tuning* em todos métodos. O conjunto de dados foi dividido em duas partes: 100 vídeos treinamento e 40 avaliação. Foram utilizadas 50 épocas para o treinamento, os melhores modelos foram selecionados e posteriormente foram treinados com os vídeos separados para o treino. O modelo melhorou consideravelmente a precisão para as duas

tarefas, atingindo 83,71% para classificação e 93,01% para segmentação.

A dissertação difere com relação ao trabalho de Nguyen et al. (2019) na aplicação de rede multitarefas, ao possuir em sua estrutura uma arquitetura de *Autoencoder* que retorna somente a reconstrução da entrada, proposta que será abordada no Capítulo 4. A rede multitarefa proposta nesta dissertação objetiva-se principalmente na classificação, a reconstrução da imagem tem a finalidade de ajustar o Espaço Latente para ter uma representação mais simplificada da entrada destacando os atributos mais relevantes que podem ser utilizados para a classificação.

4 Criação do Conjunto de Dados de Plântulas de Soja

Para treinamentos de modelos baseados em CNNs são necessárias bases com uma grande quantidade de instâncias com a finalidade de alcançar altas taxas de precisão e generalização em tarefas como classificação. A abordagem do problema de classificação da condição do solo e identificação de genótipos de plântulas de soja com modelos CNNs precisa de um conjunto de dados balanceado e não enviesado, possibilitando o treinamento de modelos que possam ser capazes de atingir boas taxas de acerto na classificação. Com essa finalidade, neste capítulo será apresentado o processo de criação da base de dados de plântulas de soja, abordando o cultivo das sementes de soja, coleta das plântulas germinadas, forma de aquisição das imagens e todo o processamento para o corte das plântulas contidas nas imagens para a formação do conjunto de dados com imagens individuais de plântula.

4.1 Cultivo das plântulas de soja

Para o cultivo das sementes de soja, foi utilizado um sistema com objetivo de simular os efeitos de um solo compactado, este foi desenvolvido por Capobiango et al. (2022).

A montagem do sistema consiste na utilização de um recipiente (bandeja) com dimensão 515 × 300 × 95 mm (comprimento, largura e altura, respectivamente). No interior da bandeja, introduziu-se na região central um tubo de PVC com 200 mm de diâmetro e 200 mm de altura, posicionado verticalmente com relação a bandeja. A Figura 14 mostra o desenho do sistema. No fundo do cano de PVC inseriu-se um tecido telado com a finalidade de impedir o vazamento do substrato existente dentro da coluna para a bandeja. Tanto a bandeja, quanto o cano de PVC, foram preenchidos com substrato. Para o umedecimento do substrato, inseriu-se água na bandeja até o enchimento deste recipiente, por capilaridade a água subiu para o substrato contida no interior do cano de PVC e consequentemente é absorvida pelas raízes das plântulas. Sobre a superfície do cano posicionou-se um disco metálico de 190 mm de diâmetro, perfurado, contendo 25 furos (20 mm de diâmetro cada furo). O substrato foi simulado utilizando areia.

O cultivo foi realizado posteriormente a adição de água na bandeja, as sementes foram tratadas com fungicida Derosal® de acordo com a recomendação do fabricante. Cada semente foi semeada alinhada ao centro de cada furo do disco de metal a 2 mm de profundidade. Para cada sistema foram cultivadas 25 sementes (5 genótipos distintos

sendo 5 sementes por genótipo). Após a semeadura, um objeto foi colocado no topo do disco de metal com a finalidade de exercer força e simular a compactação do solo. Para solo compactado, utilizou-se um objeto com peso de 26 Kg, e para solo não compactado não foi disposto um objeto sobre o disco. Após sete dias, desmontou-se o sistema e as plântulas foram retiradas do solo simulado tentando manter as raízes intactas. Foram realizadas 6 repetições deste sistema para contemplar todos os 30 genótipos, o processo utilizando o sistema para o cultivo das sementes e colheita das plântulas durou 2 semanas, sendo 4 repetições para cada genótipo. Por fim, cultivou-se 600 plântulas de soja em condição de solo compactado e 600 em condição de solo não compactado.

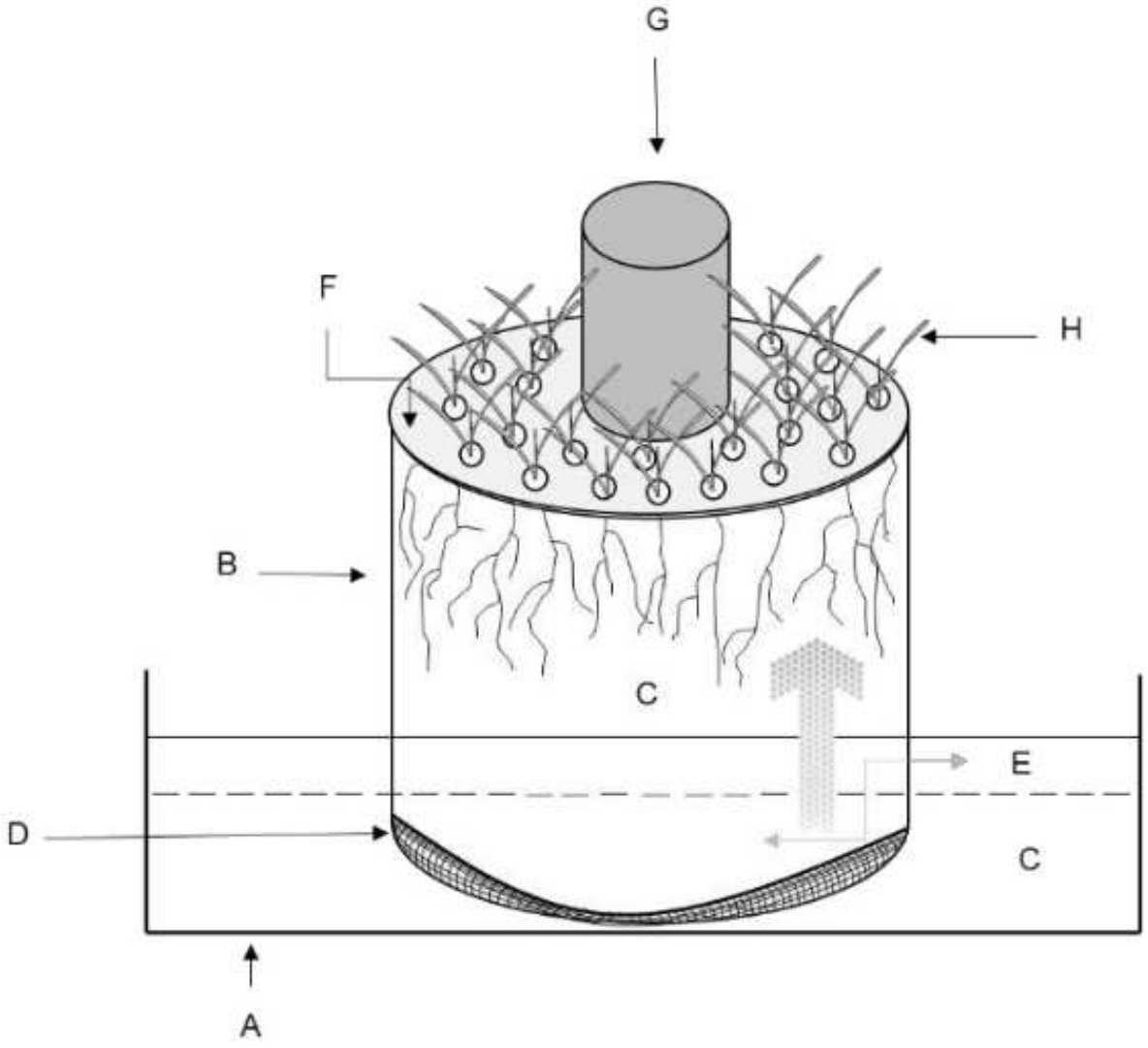


Figura 14 – Sistema para cultivo de plântulas em condição de compactação do solo: (A) bandeja de plástico; (B) cano de PVC; (C) areia; (D) tecido telado; (E) água; (F) disco de metal perfurado; (G) objeto de peso conhecido e (H) sementes/-plântulas. **Fonte**: (CAPOBIANGO et al., 2022)

4.2 Aquisição das imagens das plântulas de soja

Com a obtenção das plântulas de soja para cada execução do sistema relatado na seção anterior, inicialmente as plântulas passaram por procedimento de lavagem em água corrente. A posteriori considerando um genótipo específico de soja, foram selecionadas 5 plântulas cultivadas em solo compactado e 5 cultivadas em solo não compactado, totalizando 10 plântulas. Posteriormente, as 10 plântulas foram posicionadas horizontalmente em cima de uma flanela preta, evitando-se a sobreposição entre elas, de modo que uma folha ou outra parte de uma plântula não cubra a região de uma outra. A ordem da disposição das plântulas seguiu o mesmo padrão para todos os genótipos e repetições, sendo as cinco primeiras plântulas da esquerda para direita as que foram cultivadas em solo não compactado e as outras cinco restantes em solo compactado. As imagens foram captadas por uma câmera com resolução de 13 Megapixels. Para todos os genótipos e repetições, a câmera foi posicionada em um mesmo ângulo de 90 graus com relação ao chão a 200 mm de distância das plântulas, visando impedir sombras ou diferença de iluminação entre as imagens que poderiam gerar um enviesamento da base de dados. O processo de aquisição de imagens resultou em uma imagem por genótipo para cada repetição, como o cultivo teve 4 repetições para cada um dos 30 genótipos, esse processo resultou em 120 imagens em um total de 1.200 plântulas. A Figura 15 exibe uma das imagens captadas de plântulas de um mesmo genótipo.

Figura 15 – Imagem de plântulas de soja obtida após o processo de cultivo e posicionamento sobre um tecido preto. As cinco primeiras das esquerda para direita foram cultivadas em solo não compactado e as cinco últimas em solo compactado

4.3 Criação da base de dados

Nessa seção, será abordada a criação da base de dados com imagens individuais de plântulas de soja. Foram aplicadas técnicas de Processamento Digital de Imagens abordadas no Capítulo 2, como a remoção de ruídos, detecção das bordas e extração individual das plântulas na imagem.

A princípio foram geradas versões em tons de cinza das imagens através do método ponderado para que fosse realizado o processo de binarização, revisitar a Subseção 2.1.1. Com as versões binarizadas obtidas através do método de Otsu, facilitou-se a detecção e remoção de ruídos e partículas indesejadas do fundo das imagens, rever a Subseção 2.1.2. O resultado pode ser visto na Figura 16-(b). Para a remoção de ruídos, foram realizadas as operações morfológicas de Erosão e Dilatação, nesta ordem, para detalhes das operações, visitar a Subseção 2.1.3. A Erosão foi aplicada para remover pequenos pixels indesejados que não pertencem as plântulas na imagem, estes podem ser fragmentos de solo, pequenas folhas caídas, manchas no tecido utilizado como fundo onde as plântulas foram posicionadas ou alguma região que teve o brilho mais acentuado. Dado que a Erosão afeta pixels de interesse, como as bordas da plântula, algumas partes das raízes e folhas desaparecem ou ficam desmembradas podendo dividir uma plântula em dois ou mais objetos na imagem, com o objetivo de restaurar esses pixels removidos pela Erosão e evitar a perda de partes da plântula foi aplicada a Dilatação com elemento estruturante quadrado de tamanho 5 Figura 16-(c). Dado que o objetivo era fazer um corte preciso das plântulas, utilizou-se o algoritmo de detecção de borda. Mesmo após a Erosão e Dilatação, alguns pequenos elementos indesejados também tiveram suas bordas detectadas. Para filtrá-los, analisou-se o número de pixels dentro das bordas selecionadas. Como as plântulas possuem uma área maior que os ruídos, consequentemente o número de pixels da região relacionada com a plântula será maior. Utilizou-se o método de componentes conectados *Spaghetti Labeling* (BOLELLI et al., 2019) para obter a localização de cada componente e seus números de pixels. Os dez maiores componentes foram selecionados para manter as dez plântulas e os demais componentes referentes a ruídos foram descartados. A Figura 16-(d) mostra as bordas das plântulas detectadas destacadas em vermelho e colocadas no topo da imagem, apenas para verificação visual.

Após a detecção das bordas das plântulas de soja, foram recortadas as regiões de interesse no interior das bordas para ser obtida uma plântula individual. Cada representação de uma plântula foi inserida separadamente sobre uma imagem de fundo uniforme de cor preta, com a dimensão de 300 x 300 pixels. Mesmo as plântulas sendo objetos verticais, as imagens têm dimensões quadradas para serem utilizadas como entrada em redes convolucionais tradicionais, assim as plântulas ocupam boa parte do espaço vertical das imagens mas horizontalmente o fundo é predominante. Todo o processo foi realizado nas 120 imagens com 10 plântulas cada, gerando um total de 1.200 imagens de plântulas

individuais com fundo preto. A Figura 17 exibe algumas imagens geradas.

Analisando todas as imagens da base de dados, observou-se uma necessidade de filtragem, pelo motivo de algumas plântulas não terem crescido adequadamente no intervalo de 7 dias. Imagens referentes a um mesmo genótipo e condição de solo de cultivo apresentam plântulas com características físicas muito distintas, como pode ser visto na Figura 18, essas diferenças podem impactar no processo de treinamento de modelos, sendo as plântulas não germinadas consideradas ruídos na base de dados. Para filtrar a base de dados, foram retiradas imagens de plântulas que não se desenvolveram de acordo com as demais amostras do mesmo genótipo e condição de solo, considerando o tamanho médio do sistema radicular, folha e parte aérea (região sem raiz). Após esse processo de filtragem, a base de dados ficou com 1.026 imagens.

O número total, média e desvio padrão de imagens de plântulas cultivadas em solo compactado e não compactado para cada genótipo é mostrado na Figura 19. A remoção de plântulas não germinadas ou com mau desenvolvimento gerou uma diferença no número de imagens entre as classes relacionadas a condição de solo e genótipo, sendo que o número de imagens de plântulas cultivadas em solo não compactado é relativamente maior do que em solo compactado, essa desigualdade impactou na média por genótipos em solo não compactado que é um pouco maior com relação ao solo compactado Figura 19-(b). Como a diferença entre o número de instâncias de cada classe é pequena, a base de dados possui uma boa distribuição para as duas classes. A Figura 19-(a), mostra um adequado balanceamento do conjunto de dados, não existindo um genótipo com uma quantidade de imagens consideravelmente menor que os demais, com todos possuindo um número de imagens maior que 25. Portanto, a base gerada mantém um bom balanceamento, fator desejável principalmente para a generalização de modelos, evitando que sejam treinados com um número muito maior de amostras de uma determinada classe com relação a outra, resultando em altos erros para classes com poucas instâncias nas fases de validação e testes.

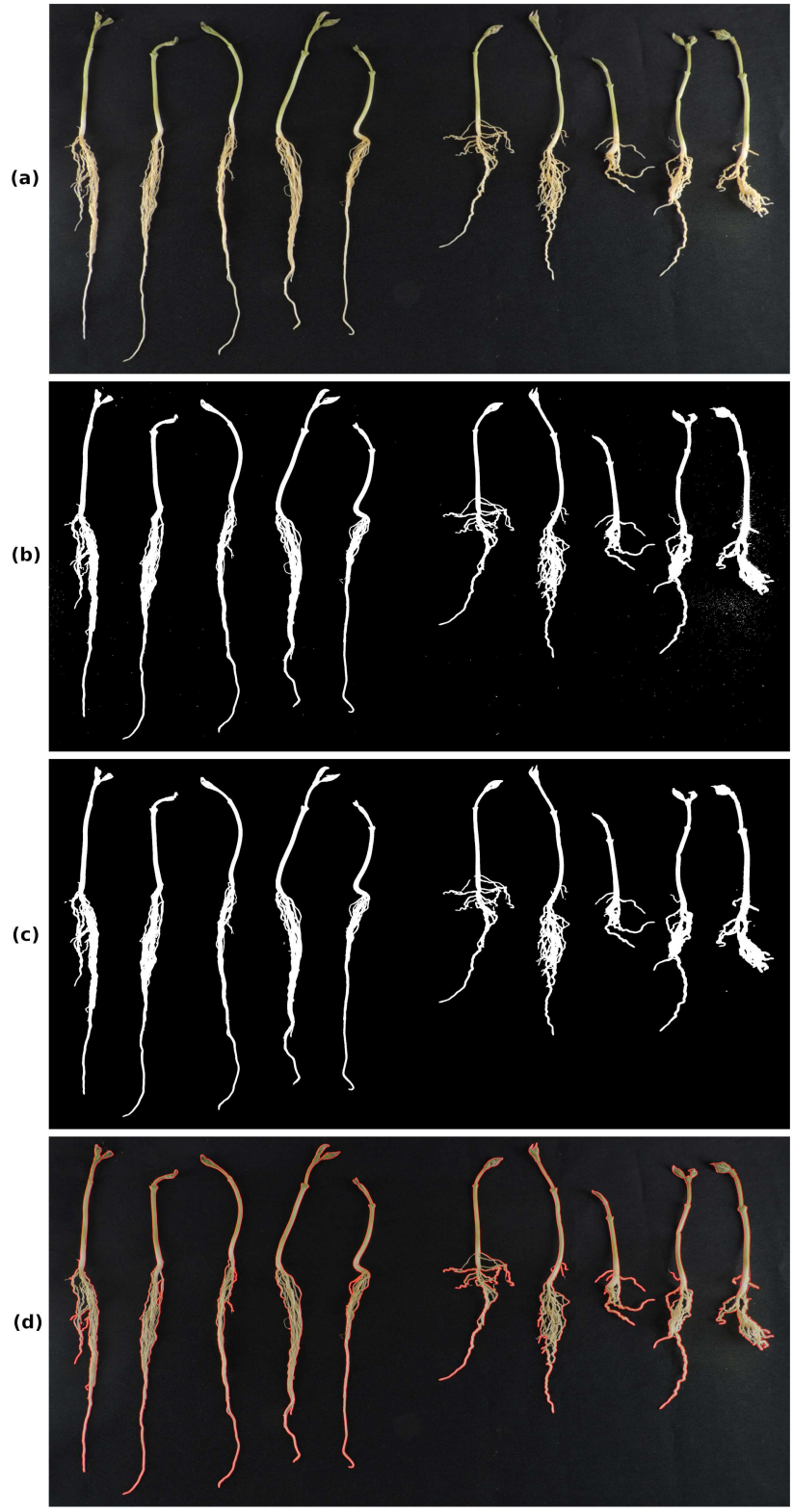


Figura 16 – Processamento da imagem para geração da base de dados com plântulas individuais. (a) imagem coletada com 5 plântulas cultivadas em solo compactado e 5 em solo não compactado para um genótipo específico. (b) imagem binária; (c) imagem binária após operações morfológicas de erosão e dilatação; (d) Sobreposição das bordas detectadas com a imagem original com propósitos estritamente visual.

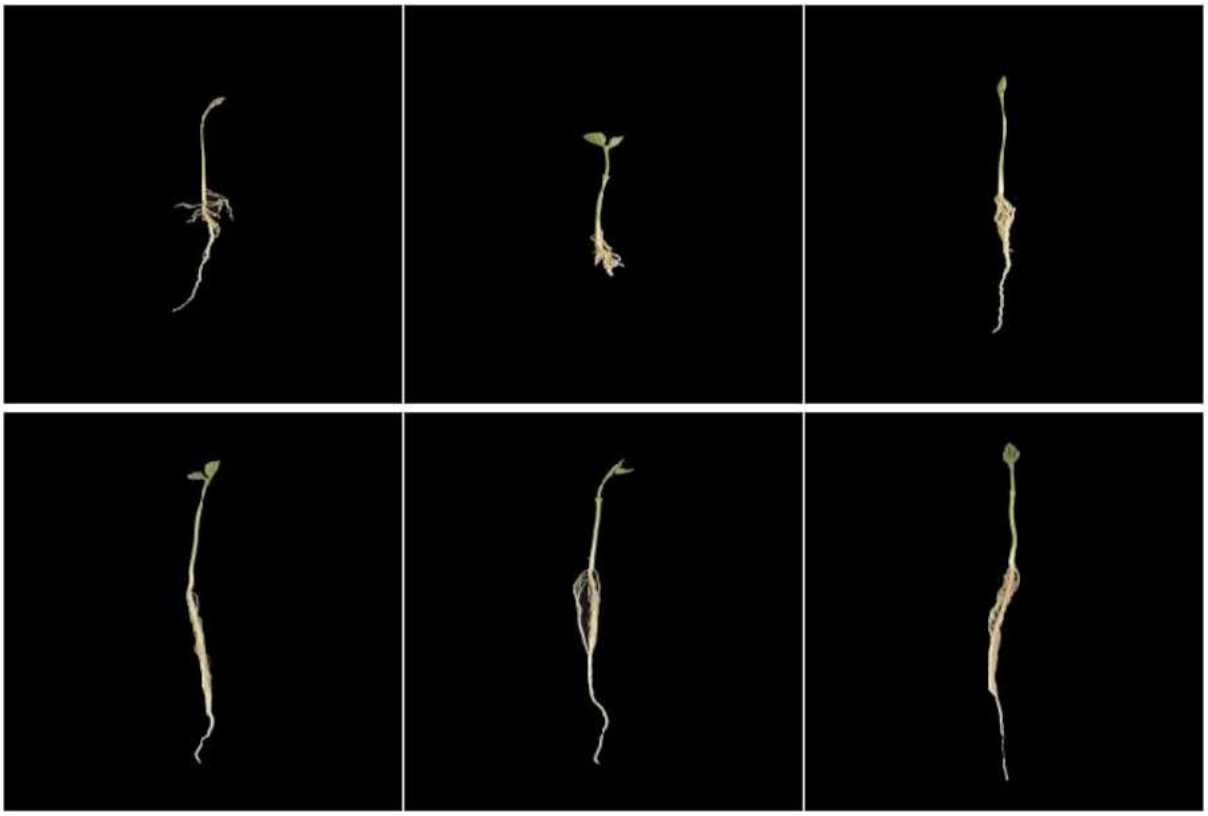
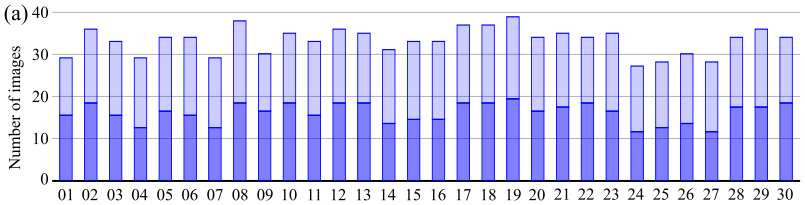


Figura 17 – Imagens da base de dados gerada. Amostras de imagens de plântulas de soja cultivadas em solo compactado (linha superior) e em solo não compactado (linha inferior). As plântulas individuais foram inseridas em uma imagem de fundo uniforme preto com a dimensão de 300×300 pixels.

Figura 18 – Imagem com plântulas de soja mal germinadas. A sexta, sétima e última plântula da esquerda para a direita, foram removidas da base de dados por não terem se desenvolvido adequadamente com a finalidade de reduzir o ruído da base de dados.



(b)

	Compacted	Non-Compacted
Total images	498	528
Mean	16.6	17.6
Standard Deviation	2.4	1.5

Figura 19 – Estatísticas da base de dados. (a) Número de plântulas cultivadas em solo compactado e não compactado por genótipo. Cada genótipo possui quarenta imagens, divididas igualmente em solo compactado e não compactado. No entanto, plântulas fora do padrão comum foram descartadas, gerando uma pequena diferença no número de imagens de um genótipo para outro. (b) Tabela com número total por classe e média e desvio padrão em relação ao número de imagens por genótipo.

4.4 Criação da versão sem raiz

Para o processo de classificação do solo onde as plântulas foram cultivadas, possuir informações suficientes da plântula para a tarefa impacta diretamente no resultado final dos treinamentos de modelos. Imagens de plântula completa (com as folhas, parte aérea e raiz) possuem uma quantidade adequada de características físicas que um modelo pode aprender e correlacionar com possíveis classes para uma classificação. O processo para adquirir plântulas com raiz visível consiste em retirar a plântula do solo, o que pode matar

e interromper o seu ciclo de desenvolvimento. Portanto mantê-la no solo por mais tempo aumenta as possibilidades de estudos e análises do seu crescimento, logo, é importante realizar as classificações de condição do solo e genótipo apenas com imagens da parte aérea das plântulas, evitando a necessidade de retirada da plântula do solo.

Com esse propósito, foi criado uma versão do conjunto de dados com plântulas sem raízes. Para a remoção das raízes foi necessário a identificação da região da plântula que faz parte do sistema radicular (região com raiz). Para tal finalidade, foi treinado um modelo CNN para detectar e classificar objetos, sendo cada imagem possuindo dois objetos, um é o sistema radicular e o outro é a parte aérea da plântula.

Para gerar caixas delimitadoras que possuem a informação das coordenadas de cada objeto existente na imagem, foi realizada uma marcação de forma manual através de um pixel de referência na fronteira entre o sistema radicular e a parte aérea de uma imagem de plântula, com o objetivo de assinalar os limites de separação das duas regiões e facilitar a geração das caixas delimitadoras. A marcação de forma manual é suscetível a erro podendo agrupar pixels da região pertencente ao sistema radicular a parte aérea da plântula. De forma a ter mais precisão na separação das regiões da plântula foi aplicada a rede de detecção e classificação de objetos, o *Single Shot Detector* (SSD) (LIU et al., 2016).

Com as caixas delimitadoras iniciais, as imagens das plântulas, juntamente com as informações de coordenadas de cada objeto existente nelas, foram passadas como entrada para um modelo que realiza duas tarefas, detectar objetos nas imagens e rotulá-los quanto a suas classes. O modelo foi treinado para detectar os objetos e classificá-los como sistema radicular ou parte aérea da plântula.

Para gerar a versão sem raiz do conjunto de dados, o objeto identificado e classificado como sistema radicular em cada imagem foi sobreposto por um retângulo uniforme de cor preta e com a mesma dimensão do objeto. Dessa forma, a região passa a ser considerada como fundo e deixando visível apenas a parte aérea da plântula conforme ilustrado na Figura 20.

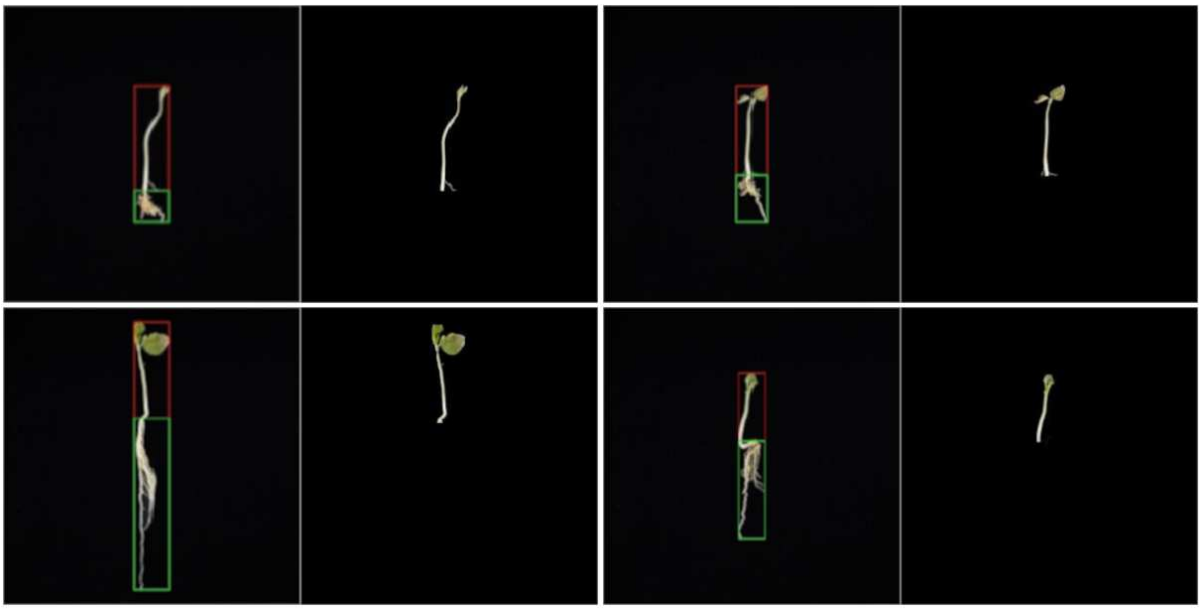


Figura 20 – Detecção do sistema radicular e parte aérea das plântulas nas imagens através da rede SSD. O método de detecção de objetos identificou os objetos retornando as caixas delimitadoras para parte aérea (retângulo vermelho) e raiz (retângulo verde). Através da detecção do sistema radicular, a região correspondente a este foi sobreposta por um retângulo de cor preta originando a versão da base de dados sem raíz. Linha superior: plântulas de soja cultivadas em solo compactado; Linha inferior: plântulas cultivadas em solo não compactado.

Nesta seção, foi demonstrada a construção da base de dados de plântulas de soja cultivadas em condições de solo compactado e não compactado com 30 genótipos distintos. Para a construção da base de dados, foram aplicadas técnicas de Processamento Digital de Imagens para segmentar, extrair e filtrar imagens de plântulas. Uma versão sem raízes da base de dados, apenas com a parte aérea, também foi criada, visando possibilitar a classificação das plântulas sem extraí-las do solo aumentando o seu tempo de desenvolvimento. Nos próximos capítulos, serão abordadas novas questões de pesquisa sobre a base de dados, incluindo abordagens de transferência de aprendizado carregando pesos pré-treinados com outros modelos e iniciando um novo treinamento com esses valores pré-definidos, modelos capazes de codificar em um vetor as características intrínsecas das imagens e através delas realizar a reconstrução das imagens. E também abordando o problema da classificação das imagens por genótipos, sendo este último mais complexo devido a grande diversificação de genótipos com relação ao total de imagens existentes na base de dados.

5 Rede de Aprendizagem Multitarefa

Nesta seção, é abordada a Rede Multitarefa desenvolvida para reconstrução de imagens e classificação de genótipos.

O problema de classificação das imagens de plântulas por genótipo pode ser considerado mais complexo, principalmente por não existir características visuais distintas tanto na parte aérea quanto no sistema radicular, essas características poderiam facilitar o trabalho de profissionais da área na tarefa de diferenciar uma da outra. Além disso, a existência de um maior número de classes para o problema impacta em um menor número de imagens por classe, se comparado com o número de instâncias por classe no problema de condição do solo. Esse aspecto pode influenciar os modelos treinados, aumentando as chances de serem incapazes de generalizar devido ao baixo número de amostras obtidas durante a fase de treinamento. De acordo com essas complexidades, visando obter um resultado superior a arquiteturas presentes na literatura, é proposto neste trabalho uma arquitetura Multitarefa visando o treinamento de dois problemas, a reconstrução das imagens de plântulas de soja e classificação de genótipos.

As Redes Multitarefas apresentam uma arquitetura composta por um braço compartilhado que distribui as informações aprendidas para outros braços específicos. O objetivo principal dessas redes é permitir que diferentes tarefas sejam realizadas de forma eficiente, compartilhando informações e características importantes. A Figura 21 apresenta a estrutura da rede, na qual o braço compartilhado funciona como um codificador da entrada, gerando um código que é compartilhado com o braço de reconstrução e o braço classificador.

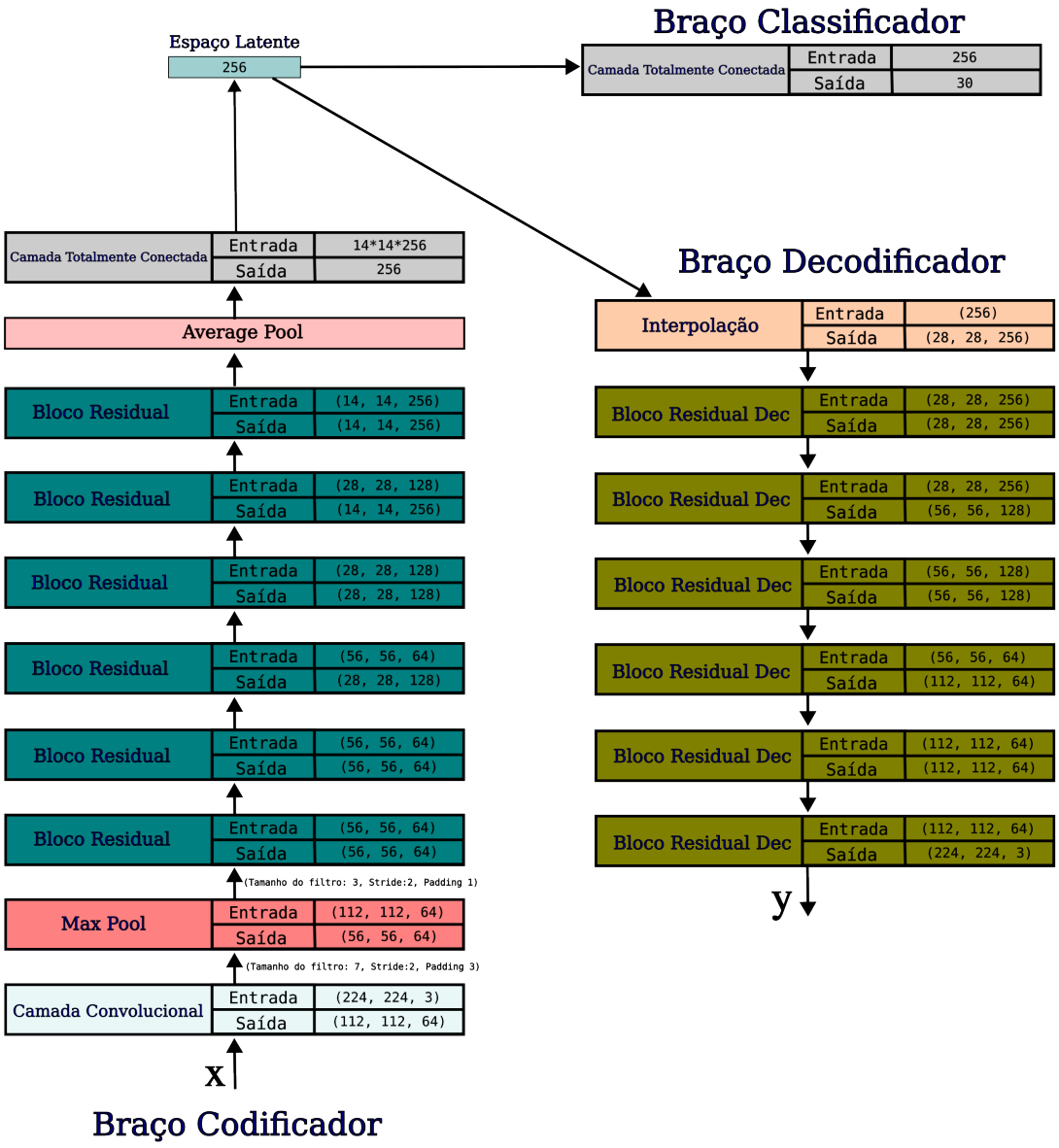


Figura 21 – Rede Convolucional Multitarefa. Composta por três braços: Braço compartilhado; braço de reconstrução e braço classificador

5.1 Bloco Residual

A rede possui blocos residuais para contornar o problema do *Vanishing Gradient* ao permitirem que as informações fluam através da rede passando por menos camadas intermediárias, visitar a Subseção 2.3.4. A Figura 22 mostra o esquema de bloco residual utilizado nos braços da rede multitarefa, a Figura 22-(b) destaca a presença de camadas de redimensionamento no braço decodificador, que realiza o aumento das dimensões dos mapas de características através de uma técnica de interpolação. Essa abordagem permite que a rede mantenha alta precisão mesmo quando a profundidade é aumentada.

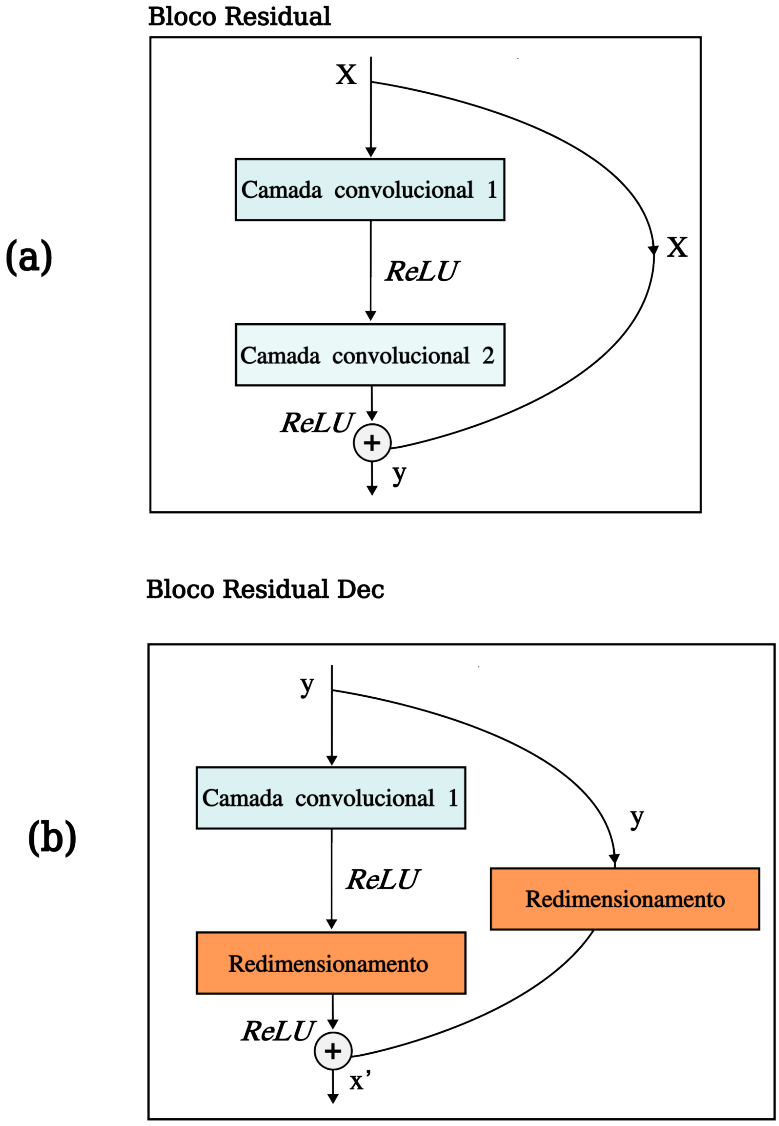


Figura 22 – Bloco Residual e Bloco Residual Decodificador

Nas próximas seções serão abordados os braços que constituem a rede e a função de perda empregada para o erro de reconstrução.

5.2 Braço Codificador (Compartilhado)

O *Hard Parameter Sharing* é uma abordagem de Aprendizagem Multitarefa que visa compartilhar os parâmetros das camadas ocultas entre várias tarefas, diminuindo consideravelmente o risco de *overfitting*. Quanto mais tarefas são aprendidas simultaneamente, maior a necessidade de encontrar uma representação comum que abranja todas elas, prevenindo o superajuste a uma tarefa específica (CARUANA, 1993). O Braço Compartilhado realiza essa função, compartilhando informações durante o treinamento com os demais braços.

O Braço Compartilhado da rede proposta foi inspirado na arquitetura Resnet-18, principalmente pela existência de blocos residuais, sendo visto na Figura 22. Os blocos residuais têm a finalidade de evitar o desaparecimento do gradiente e saturação de redes profundas (HE et al., 2016). As camadas ocultas da Resnet-18 extraem características abstratas, os mapas de características associados a essas camadas contêm uma grande quantidade de informação semântica sendo empregadas nas classificações. Uma alteração é realizada na arquitetura removendo os dois últimos blocos residuais existentes na Resnet-18 com intuito de decrementar o número de parâmetros da arquitetura possibilitando um treinamento menos custoso. Ao fim do Braço Compartilhado é gerado um vetor z com 256 posições denominado Espaço Latente, que é compartilhado com o Braço Classificador e Braço de Reconstrução. A finalidade do Braço Compartilhado é encontrar valores representativos da entrada e codificá-los no Espaço Latente descartando informações da entrada que não são relevantes para a reconstrução ou classificação de uma imagem, como por exemplo regiões que representam fundo.

5.3 Braço de Reconstrução

Como o Braço Compartilhado da rede é responsável por codificar a entrada x em um código de dimensão reduzida (Espaço Latente) z. O Braço Decodificador tem a finalidade de reconstruir a entrada de acordo com os valores que Espaço Latente possui. Para essa reconstrução é necessário a utilização de métodos que restaurem as informações da imagem, assim como as suas dimensões. Para esse fim é utilizado o processo inverso da Convolução chamado de Decovolução. O problema dessa técnica é a geração de artefatos durante as reconstruções, durante o processo de decovolução ocorrendo uma sobreposição desigual quando o tamanho do *kernel* não é divisível pelo *stride* (SUGAWARA; SHIOTA; KIYA, 2019). Como forma de contornar esse problema, foi utilizado no Braço de Reconstrução a Interpolação do Vizinho mais Próximo para as restaurações. A interpolação é utilizada na primeira camada do Decodificador e nos blocos residuais para aumentar as dimensões dos mapas de características.

A combinação destes dois braços permite que o *Autoencoder* formado por eles, aprenda uma representação codificada dos dados de entrada. Esta representação pode ser utilizada para outras tarefas, como a classificação, se o espaço latente for corretamente ajustado para refletir as relações entre as características da entrada e os rótulos de classificação. Pode ser visto em detalhes na Figura 21 os braços da rede e a região do *Autoencoder*.

5.4 Braço Classificador

As camadas do Braço Compartilhado identificam contornos, formas e partes de objetos da entrada, por fim, resultando em um valor codificado z que é definido como recursos extraídos ou como a representação simplificada das informações de entrada. Após a transformação do codificador das amostras de entrada (x) em uma representação z, a classificação ou agrupamento pode ser realizada neste espaço latente, que é uma representação simplificada e condensada das informações de entrada. Neste caso, o Braço Classificador pode ser integrado como uma camada totalmente conectada e usar um algoritmo de classificação supervisionada. A classificação através da representação codificada em um espaço latente pode ter um menor grau de complexidade do que classificar as amostras de entrada brutas, sendo que essas possuem informações não relevantes para uma classificação. Neste trabalho, o braço classificador é composto por uma camada totalmente conectada.

5.5 Funções de Perda

A função de perda calcula o quão perto uma saída está do resultado alvo. No caso da reconstrução de imagens, a função retorna o valor representativo da distância da imagem reconstruída com relação a entrada.

A rede proposta realiza durante o treinamento ajustes nos parâmetros dos três braços que a constituem. Como o Reconstrutor e o Classificador realizam tarefas distintas, cada um utiliza uma função de perda distinta, com o Classificador utilizando a Entropia Cruzada e o Decodificador uma função customizada apresentada na Equação 5.1. Essa função é semelhante ao Erro quadrático médio diferenciando apenas pelo fator $(\epsilon + x_i)$, como as imagens do conjunto de dados são constituídas em grandes proporções por fundo (pixels pretos), o fator inserido na função customizada tem objetivo de diminuir os valores de erro para pixels que correspondem ao fundo mantendo valore maiores para as regiões de interesse, no caso, pixels que correspondem a partes da plântula. Durante o treinamento os valores retornados pelas funções de perda são somados para gerar um único valor que considere os erros das duas tarefas, com base nesse erro que os gradientes são calculados.

A função de perda é representada por $J\theta$, a função é calculada usando a diferença entre a entrada(x) e a saída(y), ou seja, o erro:

$$J\theta = \frac{1}{T}\sum_{i=1}^{T}(y_i - x_i)^2 \cdot (\epsilon + x_i), \tag{5.1}$$

,

onde x_i é a i-ésimo pixel da imagem, y_i é a saída para a i-ésima amostra de entrada, θ denota o conjunto de parâmetros do *Autoencoder* (pesos e bias). A Equação 5.1, mostra a modificação com relação ao Erro Quadrático Médio, sendo acrescentado o fator $(\epsilon +$

x_i). Este tem a finalidade de diminuir o produto da multiplicação caso x_i tenha valores próximos de zero, concentrando a importância da reconstrução em valores acima de zero. O ϵ é uma constante que impossibilita zerar a função de perda.

A função de perda utilizada para o calculo de erro do classificador foi a Entropia Cruzada (SHANMUGAMANI, 2018).

6 Material e Métodos

Neste capítulo, são descritos os parâmetros utilizados nos métodos adotados para a geração do conjunto de dados e classificação das imagens de plântulas de soja na condição de solo compactado e não compactado e predição de genótipos. Os experimentos foram executadas nas GPUs: NVIDIA GeForce Tesla K40; NVIDIA GeForce GTX 680; 1070 e TITAN X, também foram utilizadas GPUs disponibilizadas pelo Google Colaboratory. Os códigos foram implementados com a linguagem de programação Python na versão 3.10.6 com o framework Pytorch 1.11 sob a versão 11.4 do CUDA, as técnicas de PDI foram executados com a bibiloteca OpenCV 4.5.5. Utilizando o sistema operacional Ubuntu 22.04 LTS.

6.1 Detalhes de Implementação para Construção do Conjunto de Dados

As imagens de plântulas de soja foram obtidas com precauções para evitar a geração de um conjunto de dados enviesado, logo, foram controlados a luminosidade, posição da câmera e limpeza do tecido onde as plântulas foram posicionadas de forma a evitar o aparecimento de objetos indesejados. Mesmo seguindo esses cuidados é necessário realizar um pré-processamento nos dados para a remoção de ruídos e artefatos. Um dos primeiros processos realizados foi a limiarização das imagens. A limiarização ou binarização objetiva-se em facilitar a análise dos objetos pertencentes a imagem, separando em conjuntos as regiões em comum, assim, pixels com características próximas (valor correspondente a cor) passam a ter o mesmo valor referente a cor, dessa forma, são destacadas as regiões de interesse (regiões com pixels pertencentes as plântulas). Nas imagens trabalhadas, a finalidade é manter os pixels pertencentes a plântula com a cor branca e os demais com a cor preta, sendo determinandos como fundo.

Foram avaliados alguns métodos para a binarização das imagens, como Limiarização Global Simples, Limiarização Adaptativa Gaussiana e Limiarização de Otsu (OTSU, 1979). A Limiarização Adaptativa Gaussiana foi utilizada com o limiar cujo o valor era 127, este método binarizou a imagem com uma grande quantidade de ruídos destacando com a cor branca regiões pertencentes ao fundo da imagem, em algumas plântulas os ruídos ficaram próximos de suas bordas sendo difícil distinguir se os pixels na vizinhança das folhas ou raízes realmente pertenciam a plântula, resultado exibido na Figura 23-(b). O método da Limiarização Adaptativa Gaussiana não levou a um resultado satisfatório principalmente por manter o fundo com pixels de cor branca e na parte interna das folhas encontrava-se pixels com a cor preta, indicando de forma indesejada que aquela área era

fundo, como pode ser visto na Figura 23-(a). O método de Otsu produziu o melhor resultado, a imagem binarizada possui uma quantidade de ruídos consideravelmente menor que outros métodos e nenhuma região interna da plântula tinha pixels pretos. A Figura 23-(c) mostra a imagem binarizada com o método de Otsu. Com o pequeno número de ruídos produzidos pela técnica, foi possível removê-los com a operações morfológica denominada Abertura, sendo constituída por uma operação de Erosão seguida de uma Dilatação. Para os operadores morfológicos, utilizamos um elemento estruturante quadrado com tamanho de 5 pixels.

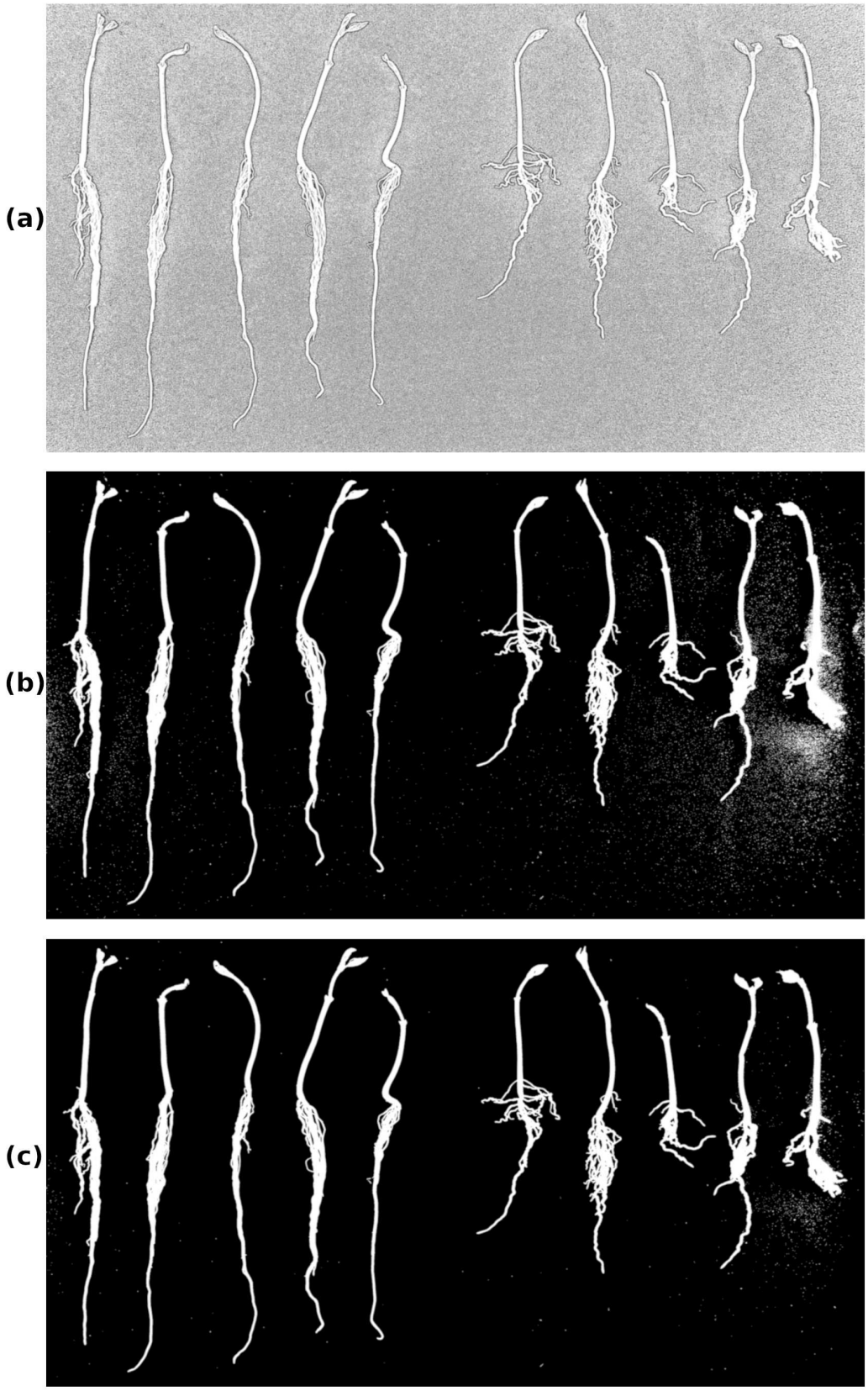


Figura 23 – Processo de binarização das imagens. (a) Limiarização Adaptativa Gaussiana; (b) Limiarização Global Simples; (c) Limiarização de Otsu

Após remover os ruídos das imagens binárias, foi utilizado o procedimento de

detecção de bordas das plântulas, com propósito de analisar a identificação destas de forma que o corte das plântulas para a formação das imagens individuais não exclua regiões da plântula como folhas ou raízes. Para esse procedimento de detecção de bordas utilizou-se o método de Canny (CANNY, 1986).

Posteriormente a fase de pré-processamento das imagens e com o conjunto de dados constituído por imagens individuais de plântulas de soja, foi gerado uma nova versão do conjunto removendo as raízes das plântulas e mantendo somente a parte aérea. Para esse processo, inicialmente as imagens foram marcadas manualmente com um pixel de cor vermelha atráves do software Gimp 2.10, com a finalidade de identificar a região que separa o sistema radicular com a parte aérea. Foi realizado o treinamento da rede SSD para identificar objetos e classificá-los como sistema radicular ou parte aérea da plântula prossegui com os pesos inicializados usando a Normalização Xavier, com taxa de aprendizado 10^{-3}, decaimento de peso de 5.10^{-4} e adotando como otimizador o SGD. Foram executadas 150 épocas para treino e validação. Utilizamos o conjunto de dados com plântulas completas determinando as caixas delimitadoras dos objetos juntamente com as suas respectivas classes (parte aérea ou raiz). A divisão da base consistiu em 656 imagens para treinamento, 164 para validação e 206 para teste.

Nas duas seções seguintes, serão abordados as aplicações para a classificação da condição do solo e genótipos da plântula.

6.2 Classificação da Condição do Solo

Para a classificação da condição do solo foram treinados modelos clássicos de CNNs como AlexNet, VGG-16, GoogLeNet e ResNet-18 e ResNet-50 para as duas versões do conjunto de dados, com imagens de plântulas completas e com plântulas sem raiz. Para cada CNN foram treinados dois modelos, um inicializava os pesos pré-treinados na base de dados ImageNet e outro treinado a partir do zero com a Normalização Xavier para inicialização dos pesos. Todos os modelos passaram por 150 épocas, com Cross Entropy como função de perda e Adam como otimizador. A taxa de aprendizado inicial era de 10^{-5} , tendo a aplicação do decaimento exponencial da taxa de aprendizado da época 50 em diante, esta teve o valor reduzido em 2% a cada época. As duas versões do conjunto de dados com imagens de plântulas completas e sem raiz foram divididos em 872 imagens para treinamento e validação com o 5-Fold, e as 154 imagens restantes separadas para teste. A divisão do 5-Fold foi de 80% para treino e 20% para validação. Como CNNs necessitam de um grande volume de imagens para o treinamento para evitar principalmente o overfitting adicionou-se transformações com o propósito de realizar um aumento dos dados. Foram aplicadas nas imagens utilizadas na fase de treinamento: rotação aleatória, flip horizontal aleatório; flip vertical aleatório, cada uma das transformações eram selecionadas de forma aleatória, podendo ou não ser aplicadas com uma probabilidade de 50%.

6.3 Classificação de Genótipos

Para a realização da classificação do genótipo, foram utilizados os modelos AlexNet, VGG-16, GoogLeNet e ResNet-18 e ResNet-50, como conjunto de dados de plântulas de soja é composto por plântulas de 30 genótipos distintos, a saída da última camada dos modelos foram alteradas de 1.000 para 30. Treinou-se dois modelos para cada arquitetura CNN, sendo um modelo para cada versão do conjunto de dados com plântulas completas e outro para sem raíz e assim como realizado na classificação da compactação do solo, um modelo é pré-treinados na base de dados ImageNet e outro começa o treinamento do zero, tendo a inicialização dos pesos realizada com a Normalização Xavier. A Entropia Cruzada foi utilizada como função de perda e o Adam como otimizador. Foram testadas várias taxas de aprendizagem juntamente com o decaimento delas de 2% após a época 20 para obter o melhor resultado para cada modelo. O conjunto de dados teve a mesmo divisão realizada para a classificação da condição do solo (872 imagens para treinamento e validação e 154 imagens para teste). A aplicação da validação cruzada com 5-Fold foi de 80% para treino e 20% para validação. Para o aumento de dados utilizou-se a rotação aleatória, flip horizontal aleatório; flip vertical aleatório, selecionadas de forma aleatória.

6.4 Treinamento do Modelo Classificador Multitarefa

Na atual seção, serão apresentadas as configurações utilizadas durante os treinamentos.

Foram treinados dois modelos, uma para a versão do conjunto de dados com plântulas completas e outro para plântulas sem raiz. Em ambos utilizou-se, durante o treinamento, a taxa de aprendizagem com o valor 10^{-3} com decaimento desse valor após a época 40 com um taxa de diminuição de 2% a cada época, usando o Adam como otimizador com 10^{-5} para o decaimento dos pesos, estes que foram inicializados pela Normalização Xavier. O corte do conjunto de dados e aumento de dados foi o mesmo realizado no treinamento de modelos para condição do solo.

O segundo treinamento foi realizado considerando os valores de erro para a reconstrução e classificação, como o objetivo desse treinamento é obter os melhores resultados para a tarefa de classificação, o erro resultante da soma dos erros de reconstrução e classificação é composto por 20% do valor correspondente ao erro de reconstrução e o restante pelo erro retornado na classificação. O segundo treinamento também foi compostor por dois modelos um para cada versão do conjunto de dados, sendo utilizado a taxa de aprendizagem com o valor 10^{-3} para a versão com plântulas completas e 10^{-3} para plântulas sem raiz, ambos com decaimento desse valor uma taxa de diminuição de 2% após a época 40 para imagens de plântulas completas e 100 para plântulas sem raiz. O Adam foi adotado como otimizador com 10^{-5} para o decaimento dos pesos, o braço classificador teve os

pesos inicializados pela Normalização Xavier. Os métodos de aumentos de dados e divisão do conjunto de dados foram os mesmos utilizados para o problema da condição do solo.

No próximo capítulo serão apresentados os resultados e as discussões do treinamento dos modelos expostos neste capítulo, além das análises visuais identificados como importantes pelos modelos para classificação.

7 Resultados e Discussão

Neste capítulo serão apresentados os resultados deste trabalho, atingidos por aplicação dos métodos evidenciados no capítulo anterior. Os resultados mostram a importância do sistema radicular para a classificação da condição do solo, onde modelos de CNNs relacionam as características dessa região com plântulas cultivadas em solo compactado. Para a classificação de genótipos os modelos atingiram uma precisão mais baixa devida a complexidade do problema, sendo que a rede Multitarefa proposta no atual trabalho, atingiu os melhores resultados para a classificação.

7.1 Identificação e Classificação das Regiões da Plântula com a Rede SSD

A geração da versão sem raiz do conjunto de dados foi inicialmente realizada através da marcação manual da região que delimita a parte aérea da plântula e a região de raiz. A marcação manual é suscetível a erro, podendo afetar na divisão da plântula onde um pixel representa uma porção pertencente a parte aérea a plântula pode ser considerado como pertencente a raiz, afetando a representabilidade do conjunto de dados e consequentemente os resultados dos modelos treinados. Para ter resultados satisfatórias na marcação, foi utilizado a rede *Single Shot Detector* para detectar e classificar os objetos da imagem (cada imagem continha dois objetos, um com as coordenadas da região da imagem correspondente a parte aérea e outra o sistema radicular). A *Single Shot Detector* segundo a literatura é uma arquitetura que onde é possível ter bons resultados com baixo tempo de treinamento, isso é conseguido usando uma combinação de camadas de convolução e camadas de detecção para encontrar objetos em imagens e vídeos.

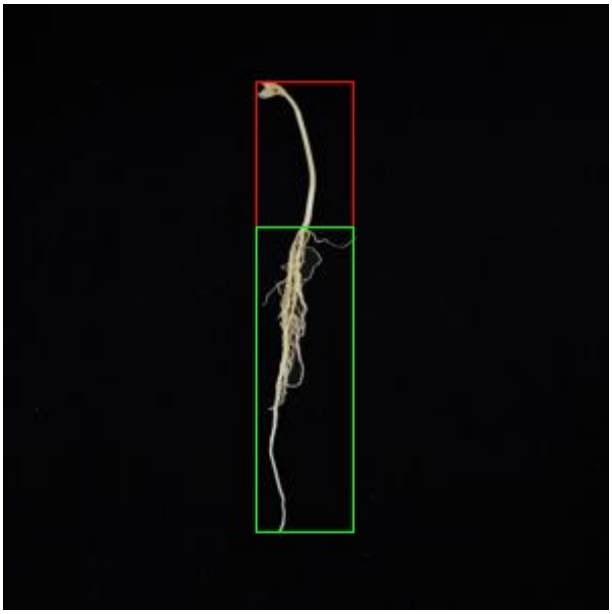


Figura 24 – Imagem de plântula de soja após a detecção da parte aérea e sistema radicular realizado pela rede SSD

O treinamento consistiu em 150 épocas com 656 imagens, posteriormente o modelo treinado alcançou na fase de teste *Mean Average Precision* de 74.4% e 76.1%, *Average Precision* de 76.1% e 78.8% para a parte aérea e sistema radicular respectivamente, como o *Single Shot Detector* é treinado para identificar objetos e classificá-los, as taxas obtidas nas métricas indicam um resultado com uma boa taxa de acerto para a identificação e classificação das raízes. Na Figura 24 é exibido o resultado da demarcação dos objetos das imagens, dividindo a plântula em parte aérea e sistema radicular.

7.2 Critérios de Avaliação

Os critérios de avaliação considerados para classificar as imagens de plântulas de soja pela condição do solo ou pelo genótipo foram com base nas taxas obtidas em cada *fold* da validação cruzada, sendo empregadas as seguintes métricas: a melhor acurácia entre todos os *folds*; a acurácia média de dos 5-*Folds* e desvio padrão das acurácias.

7.3 Problema Compactação do Solo

A remoção do acesso visual ao sistema raiz torna o problema mais difícil devido a diminuição de informação contida nas imagens e pela perda de características físicas relevantes existentes no sistema radicular. Esse aumento de complexidade impactou nas saídas dos modelos, o que é evidenciado na Tabela 1. A diferença nos resultados obtidos para as duas versões do conjunto de dados, com a acurácia alcançada pelos modelos tendo um valor menor quando se comparam a versão plântulas completas com a sem raiz, como

também observado na Figura 25. Mesmo existindo diferenças na parte aérea das plântulas em condição de solo compactado e não compactado devido ao desenvolvimento do sistema radicular, os modelos correlacionam melhor as plântulas com as suas respectivas classes quando a informação das raízes está disponível nas imagens indicando a relevância das informações contidas na raiz, como forma, espessura e comprimento.

Os resultados alcançados por todos os modelos indicam que o conjunto de dados gerado tem boa representatividade e o problema para classificação da condição do solo. Os modelos da arquitetura VGG-16 se destacaram por retornar médias de acurácia superiores aos demais nas duas versões do conjunto de dados. A AlexNet atingiu média de acurácia inferior a VGG-16, mas superior aos demais modelos mesmo sendo uma arquitetura mais simples. Uma explicação pode ser pelo maior número de camadas e parâmetros que as outras arquiteturas utilizam, aumentando a probabilidade de saturação da acurácia durante o treinamento, considerando que a complexidade do problema não é alta, logo, arquiteturas mais simples (com menos camadas e parâmetros, caso da AlexNet) podem alcançar altas taxas de acerto.

Tabela 1 – Acurácia de modelos de 5-fold para a classificação da condição do solo. Os valores estão em %. pt (pré-treinados), Melhor, Média e DP, significam melhor resultado entre os folds, média e desvio padrão entre os folds de 5.

Modelo	Plântulas completas			Plântulas sem raíz		
	Melhor	**Média**	**DP**	**Melhor**	**Média**	**DP**
AlexNet	91.9	90.2	1.42	87.3	83.2	3.00
AlexNet pt	91.9	89.4	1.81	87.9	83.2	3.07
VGG-16	94.2	92.8	1.60	91.3	88.9	3.05
VGG-16 pt	93.1	92.0	0.81	90.1	84.9	4.01
GoogLeNet	87.3	86.8	0.55	87.9	83.2	3.29
GoogLeNet pt	89.0	86.3	1.95	86.0	82.9	2.55
ResNet-18	91.3	86.0	3.10	86.2	83.9	2.93
ResNet-18 pt	87.9	84.1	2.28	85.0	81.4	2.62
ResNet-50	81.7	80.1	1.10	85.0	81.7	1.92
ResNet-50 pt	86.2	81.6	3.39	83.9	79.3	3.46
Média	*89.0*	*86.9*	*1.9*	*85.8*	*82.7*	*2.7*

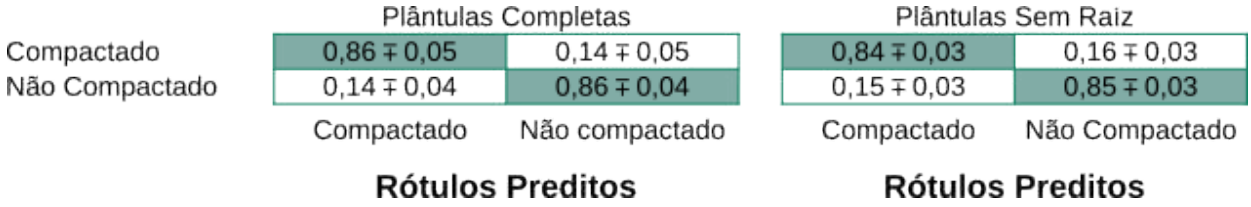


Figura 25 – Matriz de confusão com média e desvio padrão entre todos os modelos treinados.

Embora os resultados da classificação tenham atingindo boa taxa de acerto, é relevante identificar se existem diferenças relativas dessas taxas ao considerar os genótipos

de forma individual. Nas Figuras 26 e 27, são exibidas as matrizes de confusão para o treinamento da condição de solo com relação a cada genótipo. Essas matrizes facilitam a identificação dos genótipos onde as plântulas possuem características que foram melhor correlacionadas pelos modelos para a classificação da condição do solo. Os genótipos denominados como 28 e 1 atingiram as maiores taxas de acerto médio para condição de solo compactado e não compactado respectivamente no conjunto de dados com plântulas completas, já o genótipo 30 alcançou a menor taxa média para as duas condições de solo na mesma versão do conjunto de dados. Para a versão do conjunto de dados sem raiz, os genótipos 15 e 3 tiveram as maiores taxa de acerto médio para plântulas compactadas e o genótipo 3 para plântulas não compactadas. Os genótipos 9 e 8 atingiram as menores taxas médias de acerto para as condições de solo compactado e não compactado respectivamente.

A análise das Figuras 26 e 27 possibilita identificar que os modelos correlacionar melhor as características de alguns genótipos para a classificação, fato que pode ser explicado pelo diferente desenvolvimento das plântulas nas duas condições de solo. Na Figura 28 são exibidas plântulas dos genótipos 28 e 1, onde os modelos classificaram com as maiores taxas de acerto, sendo visível a diferença das plântulas cultivadas em solo compactado (5 últimas) e não compactado (5 primeiras). Esse aspecto dos genótipos justifica uma maior taxa de acerto, pela evidente diferença de características físicas das plântulas cultivadas em diferentes condições de solo. O inverso ocorre com as plântulas do genótipo 30, como pode ser visto na Figura 29, algumas plântulas cultivadas em solo não compactado têm características físicas semelhantes a plântulas cultivadas em solo compactado, aumentando a dificuldade na tarefa de classificação das plântulas e justificando a menor taxa de acerto atingido pelos modelos treinados nesse genótipo.

Plântulas Completas

		CP	NCP
1	CP	0,91 ∓ 0,03	0,04 ∓ 0,03
	NCP	0,08 ∓ 0,01	0,94 ∓ 0,01
2	CP	0,89 ∓ 0,01	0,11 ∓ 0,01
	NCP	0,1 ∓ 0,01	0,9 ∓ 0,01
3	CP	0,93 ∓ 0,02	0,07 ∓ 0,02
	NCP	0,09 ∓ 0,01	0,91 ∓ 0,01
4	CP	0,92 ∓ 0,02	0,08 ∓ 0,02
	NCP	0,07 ∓ 0,02	0,93 ∓ 0,02
5	CP	0,94 ∓ 0,01	0,06 ∓ 0,01
	NCP	0,12 ∓ 0,01	0,88 ∓ 0,01
6	CP	0,84 ∓ 0,01	0,16 ∓ 0,01
	NCP	0,11 ∓ 0,02	0,89 ∓ 0,02
7	CP	0,8 ∓ 0,01	0,2 ∓ 0,01
	NCP	0,18 ∓ 0,01	0,82 ∓ 0,01
8	CP	0,93 ∓ 0,01	0,07 ∓ 0,01
	NCP	0,18 ∓ 0,01	0,82 ∓ 0,01
9	CP	0,95 ∓ 0,01	0,05 ∓ 0,01
	NCP	0,08 ∓ 0,03	0,92 ∓ 0,03
10	CP	0,81 ∓ 0,01	0,19 ∓ 0,01
	NCP	0,23 ∓ 0,01	0,77 ∓ 0,01
11	CP	0,89 ∓ 0,01	0,11 ∓ 0,01
	NCP	0,19 ∓ 0,01	0,81 ∓ 0,01
12	CP	0,79 ∓ 0,01	21 ∓ 0,01
	NCP	0,14 ∓ 0,01	0,86 ∓ 0,01
13	CP	0,85 ∓ 0,01	0,15 ∓ 0,01
	NCP	0,17 ∓ 0,01	0,83 ∓ 0,01
14	CP	0,88 ∓ 0,01	0,12 ∓ 0,01
	NCP	0,2 ∓ 0,01	0,8 ∓ 0,01
15	CP	0,93 ∓ 0,02	0,07 ∓ 0,02
	NCP	0,09 ∓ 0,01	0,91 ∓ 0,01
16	CP	0,83 ∓ 0,01	0,17 ∓ 0,01
	NCP	0,21 ∓ 0,01	0,79 ∓ 0,01
17	CP	0,78 ∓ 0,01	0,22 ∓ 0,01
	NCP	0,24 ∓ 0,01	0,76 ∓ 0,01
18	CP	0,95 ∓ 0,02	0,05 ∓ 0,02
	NCP	0,11 ∓ 0,02	0,89 ∓ 0,02
19	CP	0,92 ∓ 0,01	0,08 ∓ 0,01
	NCP	0,08 ∓ 0,01	0,92 ∓ 0,01
20	CP	0,88 ∓ 0,02	0,12 ∓ 0,01
	NCP	0,1 ∓ 0,01	0,9 ∓ 0,01
21	CP	0,88 ∓ 0,02	0,12 ∓ 0,02
	NCP	0,09 ∓ 0,01	0,91 ∓ 0,01
22	CP	0,94 ∓ 0,01	0,06 ∓ 0,01
	NCP	0,15 ∓ 0,01	0,85 ∓ 0,01
23	CP	0,9 ∓ 0,01	0,1 ∓ 0,01
	NCP	0,12 ∓ 0,01	0,88 ∓ 0,01
24	CP	0,76 ∓ 0,01	0,24 ∓ 0,01
	NCP	0,2 ∓ 0,01	0,8 ∓ 0,01
25	CP	0,93 ∓ 0,01	0,07 ∓ 0,01
	NCP	0,09 ∓ 0,01	0,91 ∓ 0,01
26	CP	0,82 ∓ 0,01	0,18 ∓ 0,01
	NCP	0,17 ∓ 0,01	0,83 ∓ 0,01
27	CP	0,9 ∓ 0,02	0,1 ∓ 0,02
	NCP	0,14 ∓ 0,02	0,86 ∓ 0,02
28	CP	0,96 ∓ 0,03	0,04 ∓ 0,03
	NCP	0,08 ∓ 0,01	0,92 ∓ 0,01
29	CP	0,85 ∓ 0,01	0,15 ∓ 0,01
	NCP	0,1 ∓ 0,01	0,9 ∓ 0,01
30	CP	0,71 ∓ 0,02	0,29 ∓ 0,02
	NCP	0,28 ∓ 0,03	0,72 ∓ 0,03

Figura 26 – Matrizes de confusão com relação aos genótipos 30 da base com plântulas completas, com média e desvio padrão entre todos os modelos treinados. CP - Plântulas Compactadas, NCP - Plântulas Não Compactadas

Plântulas Sem Raiz

		CP	NCP
1	CP	0,92 ∓ 0,01	0,08 ∓ 0,01
	NCP	0,1 ∓ 0,02	0,9 ∓ 0,02

		CP	NCP
2	CP	0,91 ∓ 0,01	0,09 ∓ 0,01
	NCP	0,14 ∓ 0,01	0,86 ∓ 0,01

		CP	NCP
3	CP	0,88 ∓ 0,01	0,12 ∓ 0,01
	NCP	0,05 ∓ 0,02	0,95 ∓ 0,02

		CP	NCP
4	CP	0,73 ∓ 0,02	0,27 ∓ 0,02
	NCP	0,12 ∓ 0,01	0,88 ∓ 0,01

		CP	NCP
5	CP	0,84 ∓ 0,02	0,16 ∓ 0,02
	NCP	0,16 ∓ 0,01	0,84 ∓ 0,01

		CP	NCP
6	CP	0,93 ∓ 0,01	0,07 ∓ 0,01
	NCP	0,08 ∓ 0,03	0,92 ∓ 0,03

		CP	NCP
7	CP	0,89 ∓ 0,01	0,11 ∓ 0,01
	NCP	0,15 ∓ 0,01	0,85 ∓ 0,01

		CP	NCP
8	CP	0,79 ∓ 0,01	0,21 ∓ 0,01
	NCP	0,24 ∓ 0,03	0,76 ∓ 0,03

		CP	NCP
9	CP	0,72 ∓ 0,01	0,28 ∓ 0,01
	NCP	0,21 ∓ 0,01	0,79 ∓ 0,01

		CP	NCP
10	CP	0,77 ∓ 0,02	0,23 ∓ 0,02
	NCP	0,2 ∓ 0,02	0,8 ∓ 0,02

		CP	NCP
11	CP	0,85 ∓ 0,01	0,15 ∓ 0,01
	NCP	0,17 ∓ 0,01	0,83 ∓ 0,01

		CP	NCP
12	CP	0,8 ∓ 0,03	0,2 ∓ 0,03
	NCP	0,18 ∓ 0,01	0,82 ∓ 0,01

		CP	NCP
13	CP	0,81 ∓ 0,01	0,19 ∓ 0,01
	NCP	0,2 ∓ 0,03	0,8 ∓ 0,03

		CP	NCP
14	CP	0,86 ∓ 0,01	0,14 ∓ 0,01
	NCP	0,06 ∓ 0,01	0,94 ∓ 0,01

		CP	NCP
15	CP	0,97 ∓ 0,01	0,03 ∓ 0,01
	NCP	0,07 ∓ 0,01	0,93 ∓ 0,01

		CP	NCP
16	CP	0,8 ∓ 0,01	0,2 ∓ 0,01
	NCP	0,15 ∓ 0,01	0,85 ∓ 0,01

		CP	NCP
17	CP	0,83 ∓ 0,01	0,17 ∓ 0,01
	NCP	0,21 ∓ 0,02	0,79 ∓ 0,02

		CP	NCP
18	CP	0,89 ∓ 0,01	0,11 ∓ 0,01
	NCP	0,19 ∓ 0,01	0,81 ∓ 0,01

		CP	NCP
19	CP	0,89 ∓ 0,01	0,11 ∓ 0,01
	NCP	0,13 ∓ 0,01	0,87 ∓ 0,01

		CP	NCP
20	CP	0,9 ∓ 0,01	0,1 ∓ 0,01
	NCP	0,09 ∓ 0,01	0,91 ∓ 0,01

		CP	NCP
21	CP	0,84 ∓ 0,02	0,16 ∓ 0,02
	NCP	0,14 ∓ 0,01	0,86 ∓ 0,01

		CP	NCP
22	CP	0,96 ∓ 0,03	0,04 ∓ 0,03
	NCP	0,08 ∓ 0,02	0,92 ∓ 0,02

		CP	NCP
23	CP	0,78 ∓ 0,01	0,22 ∓ 0,01
	NCP	0,18 ∓ 0,01	0,82 ∓ 0,01

		CP	NCP
24	CP	0,79 ∓ 0,01	0,21 ∓ 0,01
	NCP	0,14 ∓ 0,02	0,86 ∓ 0,02

		CP	NCP
25	CP	0,8 ∓ 0,01	0,2 ∓ 0,01
	NCP	0,18 ∓ 0,01	0,82 ∓ 0,01

		CP	NCP
26	CP	0,86 ∓ 0,03	0,14 ∓ 0,03
	NCP	0,17 ∓ 0,01	0,83 ∓ 0,01

		CP	NCP
27	CP	0,92 ∓ 0,01	0,08 ∓ 0,01
	NCP	0,17 ∓ 0,01	0,83 ∓ 0,01

		CP	NCP
28	CP	0,82 ∓ 0,02	0,18 ∓ 0,02
	NCP	0,11 ∓ 0,01	0,89 ∓ 0,01

		CP	NCP
29	CP	0,76 ∓ 0,01	0,24 ∓ 0,01
	NCP	0,22 ∓ 0,02	0,78 ∓ 0,01

		CP	NCP
30	CP	0,77 ∓ 0,01	0,23 ∓ 0,01
	NCP	0,15 ∓ 0,02	0,85 ∓ 0,02

Figura 27 – Matrizes de confusão com relação aos genótipos 30 da base com plântulas sem raiz, com média e desvio padrão entre todos os modelos treinados. CP - Plântulas Compactadas, NCP - Plântulas Não Compactadas

Figura 28 – Plântulas dos genótipos 28 e 1, onde a condição de solo foi predita com as maiores taxas de acerto. Genótipo 28 (cima) e genótipo 1 (baixo).

Figura 29 – Plântulas do genótipo 30 onde a condição de solo foi predita com a menor taxa de acerto na classificação da condição do solo em compactado e não compactado.

Na próxima subseção será explorada as análises visuais das imagens onde a arquitetura VGG-16 identificou como importantes para a classificação.

7.3.1 Análise da aprendizagem

Durante o processo de treinamento, os modelos podem identificar características consideráveis para uma classificação que não são perceptíveis por um especialista. Analisar regiões da imagem que orientaram o processo de classificação é uma forma de verificar o aprendizado e reconhecer os motivos que o levaram a uma classificação correta ou incorreta. Um modelo de classificação pode atingir alta precisão, mas pelos motivos errados, devido a um viés nos dados que o levam a correlacionar atributos de forma falha a uma classe. Um exemplo seria a classificação de uma mulher vestindo jaleco branco como enfermeira, ao invés de médica, apenas por causa de seu gênero.

Para avaliar o aprendizado do modelo, portanto, o equilíbrio e o viés do conjunto de dados e identificar as regiões de interesse para a classificação de uma imagem, foram executados em todos os modelos da arquitetura VGG-16 três métodos de análise visual baseados na retropropagação de gradiente: (*i*) *Guided Backpropagation* (SPRINGENBERG et al., 2015); (*ii*) *Gradient-weighted Class Activation Mapping* (Grad-CAM) (SELVARAJU et al., 2017); e (*iii*) *Guided Backpropagation* + Grad-CAM (SELVARAJU et al., 2017).

O Grad-CAM explora as informações espaciais mantidas pelas camadas convolucionais para destacar as partes da imagem de entrada que foram relevantes para a classificação. Ele calcula os gradientes da pontuação de classificação com relação aos mapas de características convolucionais, e esses gradientes retornam inferindo a importância do neurônio para cada mapa de característica. O *Guided Backpropagation* visualiza o gradiente em relação às imagens ao retropropagar através da função de ativação ReLU (SPRINGENBERG et al., 2015), permitindo o fluxo apenas dos gradientes positivos, alterando os valores de gradiente negativo para zero. Dessa forma, possibilita a visualização das características da imagem que ativam os neurônios através do destaque dos pixels determinantes para a classificação do modelo dada uma classe de destino. *Guided Grad-CAM* combina o melhor das duas técnicas mencionadas realizando uma combinação linear com os resultados obtidos por elas e gerando uma imagem que realça os pixels mais relevantes das regiões mais importantes da imagem para a classificação do modelo dada uma classe alvo.

Em seguida são discutidos os resultados da execução dessas técnicas de retropropagação baseadas em gradiente, passando como argumento o modelo treinado para classificar a condição do solo e as 206 imagens separadas para teste. São exibidos na Figura 30-(a). os resultados obtidos pelas técnicas aplicando como entrada uma imagem de uma plântula de soja cultivada em solo compactado. O Grad-CAM gera um mapa de calor no qual a região mais representativa (mais ativada pelo modelo) para a classificação da imagem é colorida com vermelho, enquanto a região menos representativa tem a cor azul. Observa-se pelo mapa de calor, que a parte central do sistema radicular é foco de interesse possuindo características importantes aprendidas pelo modelo para a classificação, em contrapartida, a parte aérea juntamente com as folhas da plântula não são regiões que afetam de forma significativa a classificação. O *Guided Backpropagation* ativa os pixels das bordas e do interior da plântula, sendo a ativação mais perceptível nas folhas e raízes. Combinando os dois resultados, o *Guided Grad-CAM* ofusca os pixels da folha por não ser um região de interesse identificado pelo Grad-CAM e destaca aqueles pertencentes à raiz, de modo que o resultado mostra visualmente que o modelo identificou na raiz um padrão para classificar a imagem como compactada. O resultado mostra que o modelo está classificando com precisão a imagem, e também que a decisão está fundamentada nos elementos visuais corretos.

Para a situação de solo não compactado, o Grad-CAM não indica a importância do sistema radicular para a classificação, como pode ser visto na Figura 30-(b). O mapa de calor assinala a parte aérea e a terminação da raiz principal como relevantes. O *Guided Backpropagation* ativa os pixels pertencentes a plântula, mas a parte central do sistema radicular é visivelmente mais obscurecida, diferentemente das outras regiões. O *Guided Grad-CAM* exibe uma mais acentuada ofuscação dos pixels pertencentes à raiz por meio do resultado da combinação dos outros dois métodos, indicando que o modelo discerniu

padrões relevantes nos outros locais da plântula não pertencentes ao centro das raízes.

Como observado na Figura 30-(a), o sistema radicular possui informações que são relacionadas pelo modelo para a classificação da imagem como plântula cultivada em solo compactado, mas na Figura 30-(b), indica a não relevância dessa região para solo não compactado. Como a raiz interfere nas predições é importante o estudo com a versão de plântula sem raiz para um melhor entendimento dos locais das imagens onde o modelo focaliza e reconhece como significativo. Na Figura 30-(c), os resultados são do modelo treinado na versão da base de dados com imagens com plântulas sem raízes. A principal diferença destacada é a mudança na região de importância indicada pelo Grad-CAM. Como o modelo possui somente informações da parte aérea, esta se torna a região de interesse com um pouco de ênfase nas folhas que podem ser vistas como um leve borrão de seus pixels no resultado do *Guided Grad-CAM*. O mesmo comportamento pode ser observado para este modelo tendo como entrada uma plântula sem raiz cultivada em solo não compactado. Figura 30-(d) região de interesse é afetada drasticamente, por não possuir a raiz, o centro da parte aérea tem uma coloração mais avermelhada assinalando uma mudança do foco do modelo na imagem para a classificação, e neste caso, pouca ou nenhuma atenção é dada para as folhas.

Uma vez que os resultados reforçam a importância do sistema radicular neste problema de classificação, podemos argumentar que a base de dados é representativa e sem viés. Além disso, ao tornar o problema mais árduo na retirada das raízes, o modelo é forçado a identificar novos padrões na parte aérea das plântulas que podem ser imperceptíveis ou desconhecidos por um especialista da área.

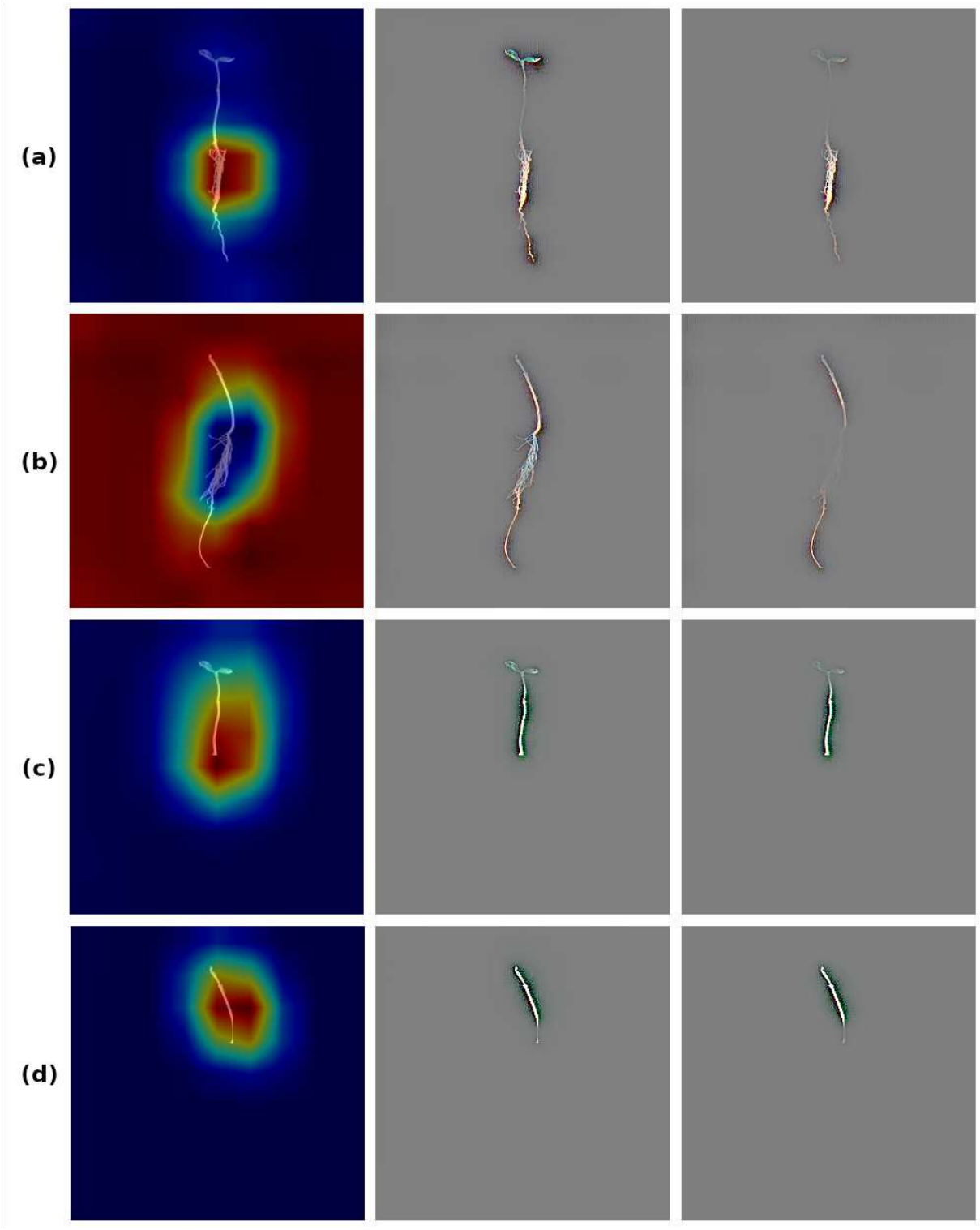


Figura 30 – Resultado para técnicas de retropropagação baseadas em gradiente. Da esquerda para a direita, o resultado do Grad-CAM, *Guided Backpropagation*, e *Guided Backpropagation* + Grad-CAM. (a) plântulas cultivadas em solo compactado e (b) em solo não compactado, para modelo treinado com plântulas completas. (c) plântulas cultivadas em solo compactado e (d) em solo não compactado, para modelo treinado com plântula sem raízes.

7.4 Problema de Classificação de Genótipos

Nesta seção será abordada o problema da classificação por genótipos, assim como os resultados dos treinamento de modelos das arquiteturas existentes na literatura e da Rede Convolucional Multitarefa proposta nesta dissertação.

A Rede Convolucional Multitarefa foi treinada em duas fases, sendo a primeira com os Braços Codificador e Decodificador, o objetivo era alcançar uma boa reconstrução da saída e utilizar os pesos do modelo treinado na segunda fase de treinamento da rede. Os resultados de reconstrução das entradas na primeira fase de treinamento após a execução de 1.000 épocas são mostrados nas Figuras 31 e 32. As reconstruções possuem plântulas com um formato semelhante das encontradas nas imagens de entrada, assim como a coloração das folhas e raízes. Alguns artefatos encontram-se presentes no fundo, principalmente nas extremidades da imagem e as plântulas tiveram uma reconstrução levemente borrada mas são detalhes que não atrapalharam a representatividade do Espaço Latente gerado, como será observado nos resultados obtidos no próximo parágrafo.

tem muitos parâmetros, o que a torna propensa ao *overfitting*, especialmente se o conjunto de dados de treinamento é pequeno

Os resultados de acurácia dos modelos treinados são exibidos na Tabela 2. É notório o aumento da complexidade do problema, sendo as taxas de acerto estando distantes de níveis mais altos. Um resultado que se destaca os demais é o da VGG-16 por estar longe dos alcançados pelos demais modelos, a arquitetura possui uma grande quantidade de parâmetros, característica que aumenta a dificuldade no treinamento e a generalização em conjunto de dados com poucas instâncias, evidenciando que esta arquitetura não possui a configuração de camadas e parâmetros capazes de alcançar um resultado aceitável. Por outro lado, a rede Multitarefa proposta atingiu os maiores valores de acurácia nas duas versões do conjunto de dados, assinalando que a abordagem de treinamento para obtenção de uma representação menor da entrada e posteriormente utilizar esse valor para classificar a entrada de acordo com o genótipo da plântula se sobressai principalmente ao realizar o treinamento em duas fases.

Tabela 2 – Acurácia dos modelos 5-*folds* para a classificação do genótipo. Os valores estão em %. pt (pré-treinados), Melhor, Média, DP e CV, significam, melhor resultado, média, desvio padrão e coeficiente de variação entre os 5*folds*.

Modelo	Plântulas completas				Plântulas sem raiz			
	Melhor	**Média**	**DP**	**CV**	**Melhor**	**Média**	**DP**	**CV**
AlexNet pt	43.0	41.3	**1.20**	0.03	36.5	33.0	2.67	0.08
VGG-16 pt	9.7	7.4	2.10	0.28	8.0	7.2	0.78	0.10
GoogLeNet pt	52.0	49.2	1.69	0.03	46.8	46.2	**0.69**	0.01
ResNet-18 pt	52.0	49.8	1.48	0.03	46.3	44.6	1.49	0.03
ResNet-50 pt	49.0	47.4	1.67	0.03	46.5	44.3	1.95	0.04
Rede Multitarefa	**54.0**	**52.3**	1.30	**0.02**	**47.9**	**47.1**	0.7	**0.01**

Como o foco do treinamento da rede Multitarefa é ter um bom resultado na tarefa de classificação, a reconstrução das imagens tem uma qualidade inferior quando comparado com as reconstruções realizadas no primeiro treinamento sendo somente o Codificador e Decodificador eram ajustados. A piora na qualidade de reconstrução pode ser explicado sobretudo pela forma de cálculo do erro no segundo treinamento, sendo este constituído por apenas 20% do erro de reconstrução. A Figura 33 mostra a imagem reconstruída durante o treinamento dos braços de reconstrução e classificação.

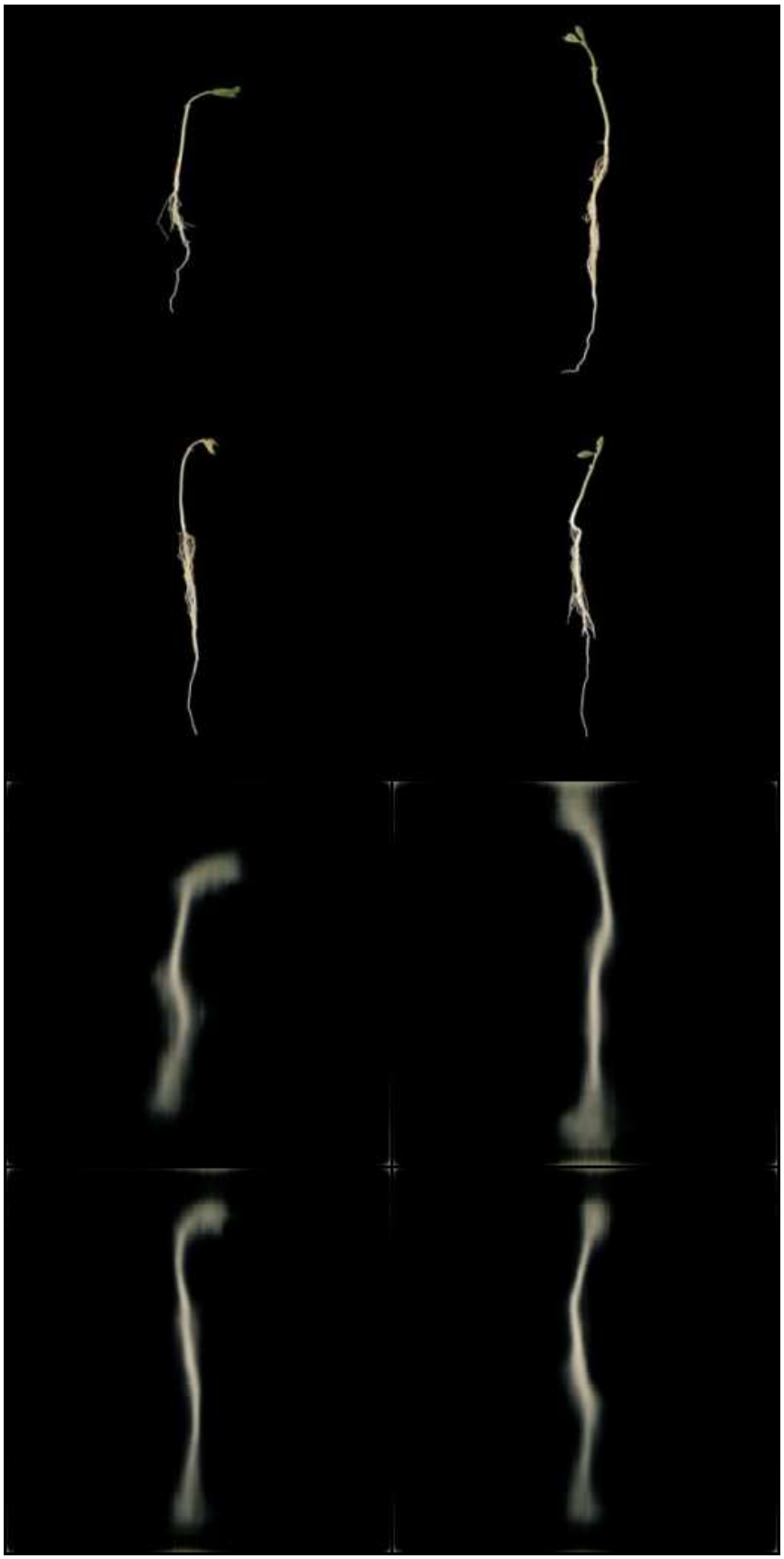


Figura 31 – Resultado da reconstrução das imagens de plântulas completas durante a primeira fase de treinamento da Rede Multitarefa. As quatro imagens superiores foram utilizadas na entrada e as quatro abaixo foram as reconstruções obtidas pela rede.

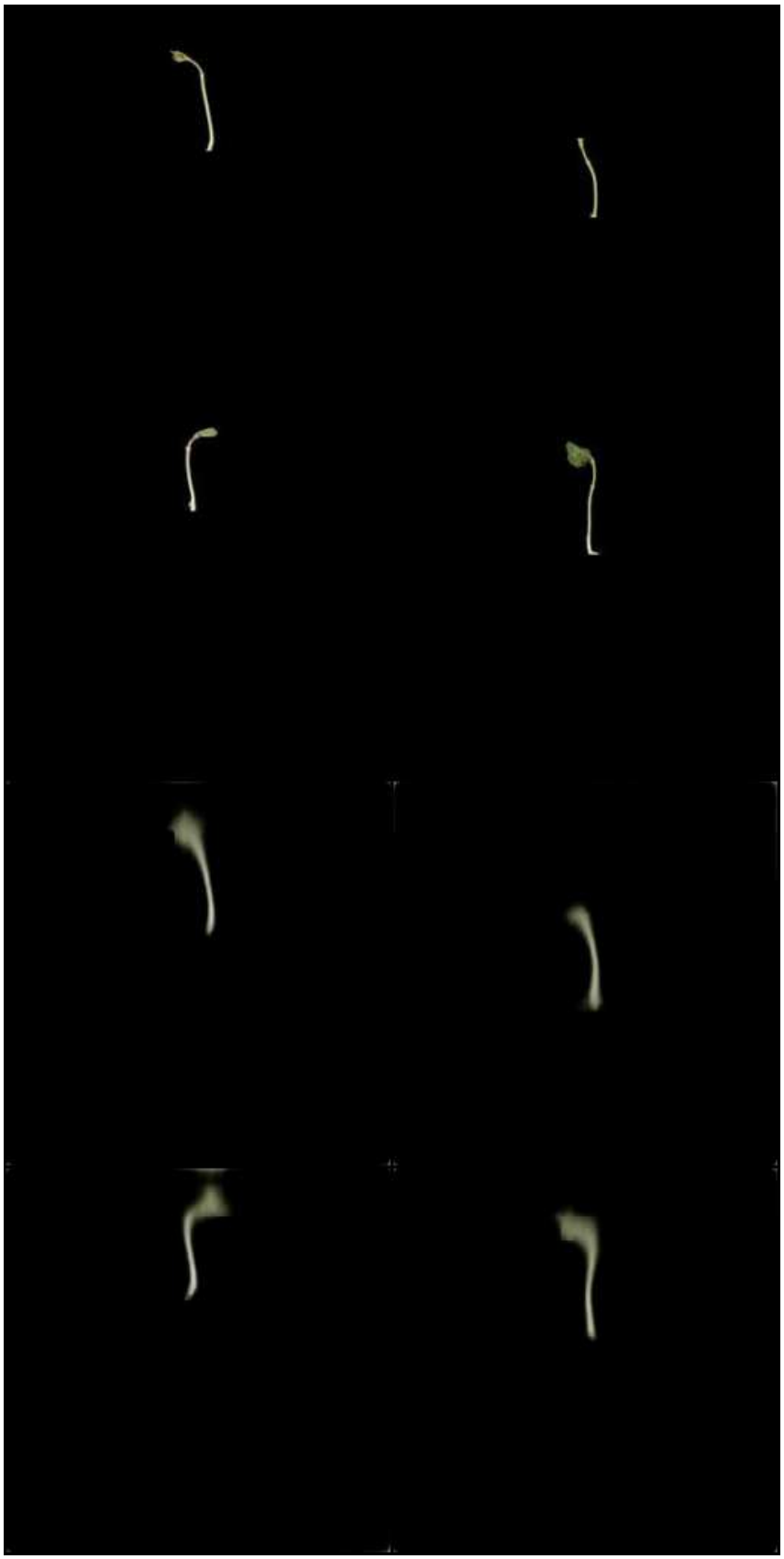


Figura 32 – Resultado da reconstrução das imagens de plântulas sem raiz durante a primeira fase de treinamento da Rede Multitarefa. As quatro imagens superiores foram utilizadas na entrada e as quatro abaixo foram as reconstruções obtidas pela rede.

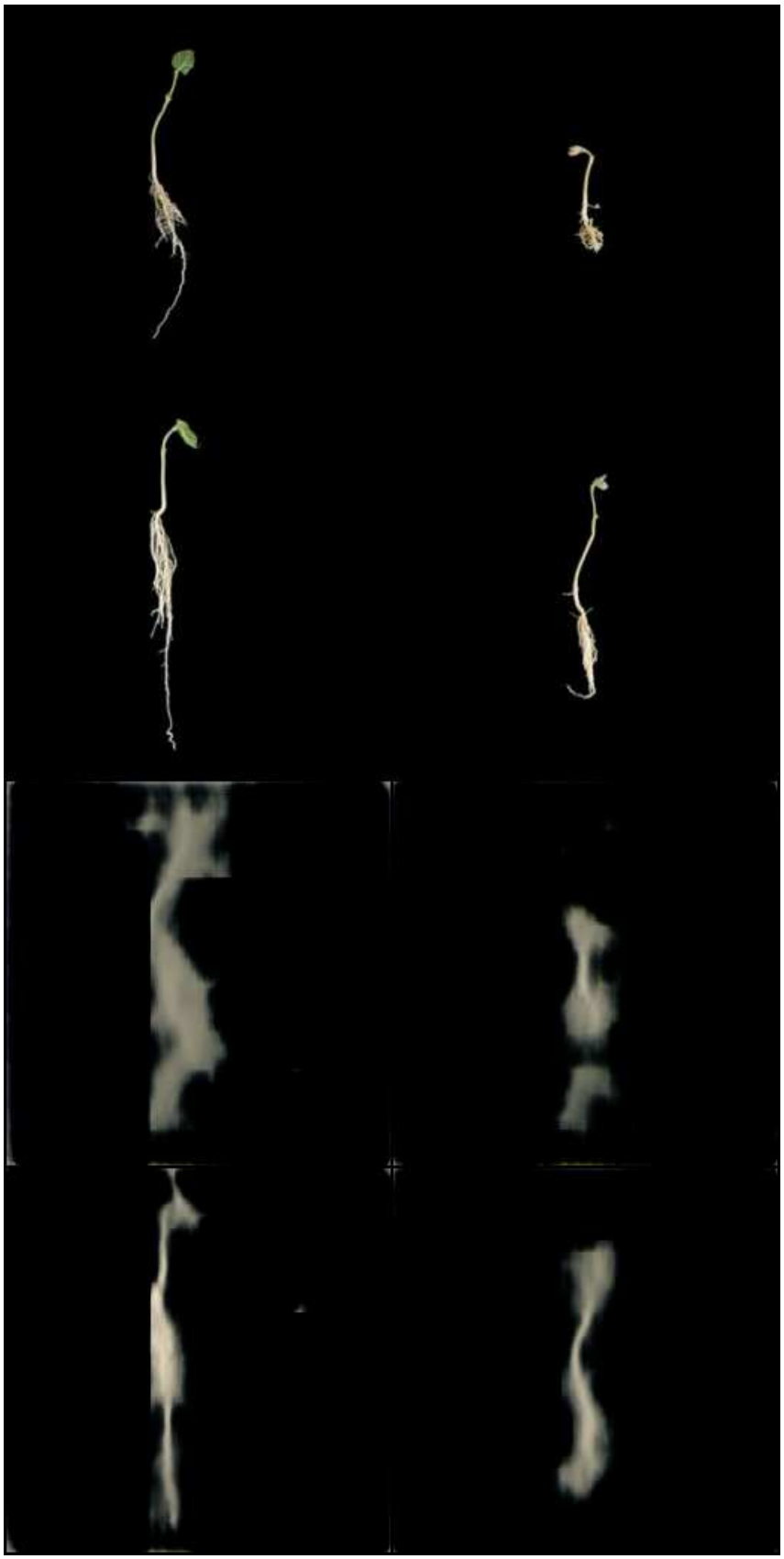


Figura 33 – Resultado da reconstrução das imagens de plântulas durante o treinamento das tarefas de classificação e reconstrução. As quatro imagens superiores foram utilizadas na entrada e as quatro abaixo foram as reconstruções obtidas pela rede.

8 Conclusões

Na atualidade, a compactação do solo é intensificada com o aumento do tamanho e peso das máquinas agrícolas. Este fator, associado ao manejo intensivo das áreas cultivadas, impacta o desenvolvimento das plantas devido a dificuldade de penetração das raízes no solo (BATEY, 2009). Essa combinação de fatores diminui a produtividade do cultivo de soja, gerando perdas para os produtores e para toda a cadeia econômica. Visando abordar este problema, o trabalho propôs uma análise de imagens de plântulas de soja para a classificação do genótipo e da condição de solo em que as plântulas foram cultivadas.

A primeira contribuição do trabalho foi a criação de um conjunto de dados composto por imagens de plântulas de soja individuais com 30 genótipos distintos e cultivadas em solo compactado e não compactado. Inicialmente foram cultivadas sementes de soja com 30 genótipos distintos em condição de solo compactado e não compactado através do sistema proposto por (CAPOBIANGO et al., 2022), após sete dias as plântulas foram retiradas do sistema e posicionadas sobre um pano preto para a geração de imagens, de forma a possuir dez plântulas de um mesmo genótipo em cada imagem obtida, sendo cinco cultivadas em solo não compactado e cinco em solo compactado. Através de técnicas de Processamento Digital de Imagens, as imagens obtidas passaram por pré-processamento para remoção de ruídos, posteriormente as regiões das imagens que representam as plântulas foram cortadas e inseridas em um fundo preto, gerando um conjunto de dados com imagens individuais de plântulas. As imagens de plântulas que não se desenvolveram adequadamente foram descartadas com a finalidade de evitar o enviesamento mas mantendo o balanceamento entre as classes. Uma versão do conjunto de dados foi gerado com imagens de plântulas de soja sem raiz. Para gerar a segunda versão foi treinando o modelo *Single Shot Detector* com o objetivo de identificar e classificar os objetos existentes na imagem, sendo a parte aérea e sistema radicular da plântula. Modelos baseados em arquiteturas presentes na literatura foram treinados para o problema de classificação da condição do solo, as taxas de acurácia na fase de teste foram satisfatórias, reforçando que o conjunto de dados é balanceado e possui boa representabilidade dos dados.

Com o conjunto de dados formado, foram realizados treinamentos de modelos com base em Redes Neurais Convolucionais. Para o problema da classificação de condição de compactação do solo, observou-se que a versão sem raiz torna a classificação mais complexa ao diminuir as informações contidas nas imagens que são referentes as plântulas e pelo sistema radicular possuir características relevantes para a distinção da condição do solo. Os modelos atingiram boas taxas de acerto com destaque para a VGG-16 que atingiu as acurácias médias de 92,8% e 88,9% para a versão do conjunto de dados com plântu-

las completas e sem raiz respectivamente. Para o problema da classificação de genótipo através da imagem da plântula, foi proposta uma Rede Convolucional Multitarefa. Esta sendo constituída por três braços: Braço Codificador; Braços de Reconstrução e Braço Classificador. O Braço Codificador extrai as características e a semântica das entradas com a finalidade de reproduzir essas informações no Espaço Latente. O Braço Codificador juntamente com o Braço de Reconstrução formam uma rede *Autoencoder*, estes foram treinados inicialmente para obter boas reconstruções da entrada e ter um Espaço Latente representativo. Posteriormente, foi realizado o segundo treinamento carregando os pesos obtidos no primeiro treinamento e considerando o Braço Classificador para realizar a classificação de genótipos. Os resultados de acurácia atingidos superaram aos resultados obtidos no treinamento de outros modelos baseados em arquiteturas clássicas, mostrando a vantagem em utilizar representações menores das entradas para a classificação, possibilitando o descarte de informações não relevantes como valores de pixel referentes ao fundo da imagem.

Para a análise da interpretabilidade dos modelos treinados com relação as características das entradas para a classificação da condição do solo, foi utilizado o modelo VGG-16. O modelo foi avaliado através dos métodos de análise visual baseados na visualização da retropropagação de gradiente. Um dos métodos foi o Grad-CAM, este apontou regiões de interesse nas imagens através do mapa de calor, onde as informações contidas nessas regiões foram empregados pelo modelo para determinar uma saída de classificação. O *Guided Backpropagation* gerou uma imagem com realce nos pixels ativados durante a classificação, as saídas dos dois métodos foram combinadas para gerar uma nova imagem através do *Guided Grad-CAM*, mantendo o realce dos pixels somente das regiões de interesse identificados pelo Grad-CAM. Com base nas imagens obtidas pelos três métodos, foi possível constatar a importância do sistema radicular neste problema de classificação e que a base de dados é representativa e sem viés. Além disso, ao tornar o problema mais complexo com a retirada da informação das raízes nas imagens, o modelo foi forçado a identificar novos padrões na parte aérea das plântulas que podem ser imperceptíveis ou desconhecidos por um especialista da área.

De acordo com os resultados apresentados neste trabalho, para o problema classificação de plântulas de soja em imagens pelo genótipo e condição do solo, Redes Neurais Convolucionais podem ser utilizadas, auxiliando especialistas na identificação de padrões existentes nas plântulas que se correlacionam a genótipos que melhor se adaptam a condição de solo compactado.

8.1 Contribuições

O trabalho gerou a publicação Nascimento et al. (2022), sendo composta pelas contribuições:

- Criação do conjunto de dados composto por imagens de plântulas de soja individuais com 30 genótipos distintos e cultivadas em solo compactado e não compactado.
- Comparação dos resultados obtidos no treinamento de modelos baseados em Redes Neurais Convolucionas para a classificação das imagens de plântulas de soja com relação a condição do solo.
- Análise visual das regiões das imagens identificadas como relevantes por um modelo treinado para a classificação da condição do solo, através de métodos baseados na visualização da retropropagação de gradiente.

8.2 Trabalhos Futuros

As contribuições do trabalho possibilitam novos estudos em trabalhos futuros como:

- ***Obter imagens de plântulas de soja cultivadas em lavouras:*** captar imagens do mundo real inclusive com fundo não uniforme aumentando o número de instâncias por classe e a diversidade do conjunto de dados. Essa ampliação dos dados pode, consequentemente, influenciar no aumento da taxa de acerto dos modelos treinados.
- ***Obter imagens de plântulas de soja artificialmente:*** como o registro de imagens de plântulas de soja é complicado e trabalhoso podem ser adotadas as Redes Adversárias e Generativas (do inglês *Generative Adversarial Networks* - GANs) para gerar artificialmente novas imagens a partir das existentes no conjunto, podendo contornar o problema do baixo número de instâncias.
- ***Novas arquiteturas para classificações:*** A abordagem deste trabalho para redes Multitarefas mostrou-se relevante para a classificação de genótipos abrindo a possibilidade para uso de arquiteturas já existentes na literatura e novas propostas de CNNs com a finalidade de melhorar os resultados obtidos.

Referências

ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 international conference on engineering and technology (ICET)**. [S.l.], 2017. p. 1–6.

ARORA, S. et al. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. **Pattern Recognition Letters**, Elsevier, v. 29, n. 2, p. 119–125, 2008.

BATEY, T. Soil compaction and soil management–a review. **Soil Use and Management**, Wiley Online Library, v. 25, n. 4, p. 335–345, 2009.

BEUTLER, A. N. et al. Efeito da compactação na produtividade de cultivares de soja em latossolo vermelho. **Revista Brasileira de Ciência do Solo**, SciELO Brasil, v. 30, p. 787–794, 2006.

BOLELLI, F. et al. Spaghetti labeling: Directed acyclic graphs for block-based connected components labeling. **Trans. on Image Processing**, IEEE, v. 29, p. 1999–2012, 2019.

CANNY, J. A computational approach to edge detection. **Trans. on Pattern Analysis and Machine Intelligence**, IEEE, n. 6, p. 679–698, 1986.

CAO, Z. et al. Marine animal classification using combined cnn and hand-designed image features. In: IEEE. **OCEANS**. [S.l.], 2015. p. 1–6.

CAPOBIANGO, N. P. et al. A proposal for evaluation of seedling emergence and growth under mechanical impedance. **Communications in Soil Science and Plant Analysis**, Taylor & Francis, p. 1–14, 2022.

CARUANA, R. Multitask learning: A knowledge-based source of inductive bias1. In: CITESEER. **Proceedings of the Tenth International Conference on Machine Learning**. [S.l.], 1993. p. 41–48.

CHAKRAVORTY, P. What is a signal?[lecture notes]. **IEEE Signal Processing Magazine**, IEEE, v. 35, n. 5, p. 175–177, 2018.

COLLOBERT, R.; WESTON, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: **Proceedings of the 25th international conference on Machine learning**. [S.l.: s.n.], 2008. p. 160–167.

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. **Conf. on Computer Vision and Pattern Recognition**. [S.l.], 2009. p. 248–255.

DENG, L. The mnist database of handwritten digit images for machine learning research. **Signal Processing Magazine**, IEEE, v. 29, n. 6, p. 141–142, 2012.

DENG, L.; HINTON, G.; KINGSBURY, B. New types of deep neural network learning for speech recognition and related applications: An overview. In: IEEE. **2013 IEEE international conference on acoustics, speech and signal processing**. [S.l.], 2013. p. 8599–8603.

DONG, E. et al. An improved convolution neural network for object detection using yolov2. In: IEEE. **2018 IEEE International Conference on Mechatronics and Automation (ICMA)**. [S.l.], 2018. p. 1184–1188.

EMBRAPA. **Brasil maior produtor mundial de soja**. 2022. Disponível em: <https://www.embrapa.br/soja/cultivos/soja1/>.

EVERINGHAM, M. et al. The pascal visual object classes (voc) challenge. **Inter. Journal of Computer Vision**, Springer, v. 88, n. 2, p. 303–338, 2010.

FLORES, P. et al. Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning. **Industrial Crops and Products**, Elsevier, v. 161, p. 113223, 2021.

GIRSHICK, R. Fast r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2015. p. 1440–1448.

GONÇALVES, G. R. et al. Benchmark for license plate character segmentation. **Journal of Electronic Imaging**, SPIE, v. 25, n. 5, p. 053034, 2016.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation (3rd Edition)**. USA: Prentice-Hall, Inc., 2007. ISBN 0131471392.

HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778.

HORN, G. V.; PERONA, P. The devil is in the tails: Fine-grained classification in the wild. **ArXiv e-prints**, 2017.

IBGE. **Produção Agropecuária**. 2022. Disponível em: <https://www.ibge.gov.br/explica/producao-agropecuaria/>.

KRIZHEVSKY, A.; HINTON, G. et al. Learning multiple layers of features from tiny images. Toronto, ON, Canada, 2009.

KUHN, D. M.; MOREIRA, V. P. Brcars: a dataset for fine-grained classification of car images. In: IEEE. **Conf. on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.], 2021. p. 231–238.

KUMAR, S.; CHAUDHARY, V.; CHANDRA, S. K. M. Plant disease detection using cnn. **Turkish Journal of Computer and Mathematics Education**, v. 12, n. 12, p. 2106–2112, 2021.

LAROCA, R. et al. A robust real-time automatic license plate recognition based on the yolo detector. In: IEEE. **Inter. Joint Conf. on Neural Networks**. [S.l.], 2018. p. 1–10.

LI, Z. et al. Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. **International Journal of Imaging Systems and Technology**, Wiley Online Library, v. 29, n. 4, p. 577–583, 2019.

LIANG, J.; WANG, D.; LING, X. Image classification for soybean and weeds based on vit. In: IOP PUBLISHING. **Journal of Physics: Conf. Series**. [S.l.], 2021. v. 2002, n. 1, p. 012068.

LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: **European Conf. on Computer Vision**. [S.l.: s.n.], 2014. p. 740–755.

LIU, W. et al. Ssd: Single shot multibox detector. In: SPRINGER. **European Conf. on Computer Vision**. [S.l.], 2016. p. 21–37.

MAHDY, L. N. et al. Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine. **MedRxiv e-prints**, Cold Spring Harbor Laboratory Press, 2020.

MAJI, S. et al. Fine-grained visual classification of aircraft. **ArXiv e-prints**, 2013.

MASOOD, S. Z. et al. License plate detection and recognition using deeply learned convolutional neural networks. **arXiv preprint arXiv:1703.07330**, 2017.

MONTAZZOLLI, S.; JUNG, C. Real-time brazilian license plate detection and recognition using deep convolutional neural networks. In: IEEE. **2017 30th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)**. [S.l.], 2017. p. 55–62.

MUKHERJEE, S. et al. Clustergan: Latent space clustering in generative adversarial networks. In: **Proceedings of the AAAI conference on artificial intelligence**. [S.l.: s.n.], 2019. v. 33, n. 01, p. 4610–4617.

NASCIMENTO, B. C. d. et al. A soybean seedlings dataset for soil condition and genotype classification. In: **Conf. on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.: s.n.], 2022.

NGUYEN, H. H. et al. Multi-task learning for detecting and segmenting manipulated facial images and videos. In: IEEE. **2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)**. [S.l.], 2019. p. 1–8.

OTSU, N. A threshold selection method from gray-level histograms. **IEEE Trans. on Systems, Man, and Cybernetics**, IEEE, v. 9, n. 1, p. 62–66, 1979.

PONTI, M. A. et al. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: IEEE. **2017 30th SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)**. [S.l.], 2017. p. 17–41.

RAMSUNDAR, B. et al. Massively multitask networks for drug discovery. **arXiv preprint arXiv:1502.02072**, 2015.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 779–788.

ROSENBLATT, F. **Principles of neurodynamics. perceptrons and the theory of brain mechanisms**. [S.l.], 1961.

ROSSLER, A. et al. Faceforensics++: Learning to detect manipulated facial images. In: **Proceedings of the IEEE/CVF international conference on computer vision**. [S.l.: s.n.], 2019. p. 1–11.

RUDER, S. An overview of multi-task learning in deep neural networks. **arXiv preprint arXiv:1706.05098**, 2017.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Upper Saddle River, NJ, USA: [s.n.], 2009.

SARAVANAN, C. Color image to grayscale image conversion. In: IEEE. **2010 Second International Conference on Computer Engineering and Applications**. [S.l.], 2010. v. 2, p. 196–199.

SAUVOLA, J.; PIETIKÄINEN, M. Adaptive document image binarization. **Pattern recognition**, Elsevier, v. 33, n. 2, p. 225–236, 2000.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Inter. Conf. on Computer Vision**. [S.l.: s.n.], 2017. p. 618–626.

SHAFIEE, M. J. et al. Fast yolo: A fast you only look once system for real-time embedded object detection in video. **arXiv preprint arXiv:1709.05943**, 2017.

SHANMUGAMANI, R. **Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras**. [S.l.]: Packt Publishing Ltd, 2018.

SHEN, Y. et al. Interpreting the latent space of gans for semantic face editing. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2020. p. 9243–9252.

SHRIVASTAVA, V. K.; PRADHAN, M. K. Rice plant disease classification using color features: a machine learning paradigm. **Journal of Plant Pathology**, Springer, v. 103, n. 1, p. 17–26, 2021.

SILVA, A. L. B. Vieira-e et al. Stn plad: A dataset for multi-size power line assets detection in high-resolution uav images. In: **Conf. on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.: s.n.], 2021. p. 215–222.

SPRINGENBERG, J. T. et al. Striving for simplicity: The all convolutional net. **ArXiv e-prints**, 2015.

SUGAWARA, Y.; SHIOTA, S.; KIYA, H. Checkerboard artifacts free convolutional neural networks. **APSIPA Transactions on Signal and Information Processing**, Cambridge University Press, v. 8, 2019.

SUN, X. et al. A benchmark for automatic visual classification of clinical skin disease images. In: SPRINGER. **European Conf. on Computer Vision**. [S.l.], 2016. p. 206–222.

SZEGEDY, C. et al. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 1–9.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. Introduction to data mining. ed. **Addison-Wesley Longman Publishing Co., Inc.**, 2005.

VOYNOV, A.; BABENKO, A. Unsupervised discovery of interpretable directions in the gan latent space. In: PMLR. **International conference on machine learning**. [S.l.], 2020. p. 9786–9796.

WICKRAMASINGHE, C. S.; MARINO, D. L.; MANIC, M. Resnet autoencoders for unsupervised feature learning from high-dimensional data: Deep models resistant to performance degradation. **IEEE Access**, IEEE, v. 9, p. 40511–40520, 2021.

XIAO, H.; RASUL, K.; VOLLGRAF, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. **arXiv preprint arXiv:1708.07747**, 2017.

YANG, L. et al. A large-scale car dataset for fine-grained categorization and verification. In: **Conf. on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2015. p. 3973–3981.

YANG, S. et al. Transfer learning from synthetic in-vitro soybean pods dataset for in-situ segmentation of on-branch soybean pods. In: **Conf. on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2022. p. 1666–1675.