



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE RUSSAS
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JANNE KELLY OLIVEIRA PEREIRA

**USO DE VISÃO COMPUTACIONAL PARA RECONHECIMENTO DE IMAGENS DE
FRUTAS EM IMAGENS RGB**

RUSSAS

2021

JANNE KELLY OLIVEIRA PEREIRA

USO DE VISÃO COMPUTACIONAL PARA RECONHECIMENTO DE IMAGENS DE
FRUTAS EM IMAGENS RGB

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciência da Computação
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Ciência da Computação.

Orientadora: Prof. Dra. Tatiane Fernan-
des Figueiredo

RUSSAS

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- P492u Pereira, Janne Kelly Oliveira.
Uso de visão computacional para reconhecimento de imagens de frutas em imagens RGB / Janne Kelly Oliveira Pereira. – 2021.
33 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, 3, Russas, 2021.
Orientação: Profa. Dra. Tatiane Fernandes Figueiredo.
1. Redes neurais convolucionais. 2. Classificação de frutas em imagens. 3. Visão computacional. 4. Aprendizagem profunda. I. Título.

CDD

JANNE KELLY OLIVEIRA PEREIRA

USO DE VISÃO COMPUTACIONAL PARA RECONHECIMENTO DE IMAGENS DE
FRUTAS EM IMAGENS RGB

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciência da Computação
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Ciência da Computação.

Aprovada em: 20 de Agosto de 2021

BANCA EXAMINADORA

Prof. Dra. Tatiane Fernandes
Figueiredo (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Ms. Eurinaldo Rodrigues Costa
Universidade Federal do Ceará (UFC)

Prof. Dr. Bofim Amaro Júnior
Universidade Federal do Ceará (UFC)

À minha família, por acreditar e investir em mim.
Mãe, seu cuidado e dedicação foram o que me deram , em alguns momentos, a esperança para seguir, sua presença significou segurança e certeza de que não estou sozinha nessa caminhada.

AGRADECIMENTOS

A minha mãe Elizangela, irmãos Francisjunior, Mayk, Diego e Cauã e todos meus amigos, que nos momentos de minha ausência dedicados ao estudo superior, sempre fizeram entender que o futuro é feito a partir da constante dedicação no presente!

A Prof. Dra. Tatiane Fernandes Figueiredo por me orientar em meu trabalho de conclusão.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender.

“Escreva algo que valha a pena ler ou faça algo
que valha a pena escrever.”

(Benjamin Franklin)

RESUMO

Dentre as propostas de disponibilização de informação nutricional com o objetivo de promover hábitos alimentares saudáveis, destaca-se o uso da tecnologia com uso de técnicas de Processamento Digital de Imagens e Visão Computacional para identificar alimentos em imagens digitais e fornecer informações precisas baseadas em tabelas de composição alimentar de referência. Neste trabalho, apresentamos modelos inteligentes para detecção e classificação de frutas de bananas, maçãs e laranjas em imagens *Red Green Blue* (RGB) utilizando Redes Neurais Convolucionais. Os modelos propostos foram executados por meio de dois experimentos. Os experimentos foram realizados com o *dataset* de (KUANG, 2015), onde as imagens dos frutos possuem fundos variados. O primeiro experimento obteve uma acurácia de 94,33%. O segundo experimento, utilizando uma ampliação da base de dados através de uma função de transformações aleatória, obteve uma acurácia de 98,7%. Espera-se em trabalhos futuros, utilizar os resultados obtidos pelo modelo inteligente para criação de um aplicativo, onde os usuários possam obter informações nutricionais, promovendo assim um canal de educação alimentar para a população.

Palavras-chave: Redes neurais convolucionais. Classificação de frutas em imagens. Visão computacional. Aprendizagem profunda.

ABSTRACT

Among the proposals to provide nutritional information to promote healthy eating habits, the use of Digital Image Processing and Computer Vision techniques to identify foods in digital images and provide accurate information based on food composition tables stands out. of reference. In this work, we present a model for detecting and classifying bananas, apples and oranges in Red Green Blue (RGB) images using Convolutional Neural Networks. The proposed model was executed through two experiments. The experiments were carried out with the *dataset* of (KUANG, 2015), where the images of the fruits have varied backgrounds. The first experiment had an accuracy of 94.33%. The second experiment, using a database expansion through a random transformation function, obtained an accuracy of 98.7%. It is expected that in future works, use the results obtained by the model to create an application, where users can obtain nutritional information, thus promoting a food education channel for the population.

Keywords: Convolutional neural networks. Classification of fruits in images. Computer vision. Deep learning.

LISTA DE FIGURAS

Figura 1 – Hierarquia de processamento	17
Figura 2 – Diferenças entre um modelo de classificação e um modelo de regressão. . .	18
Figura 3 – Exemplo de aplicação de um algoritmo de aprendizagem não-supervisionada para clusterização de dados.	19
Figura 4 – Exemplo de uma metodologia simplificada utilizada por um algoritmo de aprendizagem por reforço.	20
Figura 5 – Estrutura de uma Rede Neural Convolutacional	21
Figura 6 – Exemplo de bananas no <i>dataset</i>	25
Figura 7 – Exemplo de laranjas no <i>dataset</i>	25
Figura 8 – Exemplo de maçãs no <i>dataset</i>	26
Figura 9 – Cenário completo do <i>loss</i> por época para o experimento 1	29
Figura 10 – Cenário completo da acurácia por época para o experimento 1	30
Figura 11 – Cenário completo da acurácia por época para o experimento 2	31
Figura 12 – Cenário completo do <i>loss</i> por época para o experimento 2	31

LISTA DE TABELAS

Tabela 1 – Comparação entre esta monografia e o estado da arte.	23
Tabela 2 – Divisão <i>Dataset</i>	25
Tabela 3 – Configurações dos cenários distintos.	28

LISTA DE ABREVIATURAS E SIGLAS

CNN Rede Neural Convolucional

PDI Processamento Digital de Imagens

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.1.1	<i>Objetivos Gerais</i>	15
1.1.2	<i>Objetivos Específicos</i>	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Visão Computacional	16
2.1.1	<i>Processamento de Imagens Digitais</i>	16
2.2	Aprendizado de Máquina	17
2.2.1	<i>Aprendizagem Supervisionada</i>	18
2.2.2	<i>Aprendizagem Não-Supervisionada</i>	18
2.2.3	<i>Aprendizagem Por Reforço</i>	19
2.2.4	<i>Aprendizagem Profunda</i>	20
2.2.5	<i>Redes Neurais Convolucionais</i>	20
3	TRABALHOS RELACIONADOS	22
3.1	Classificação de imagens de maçãs, bananas e outras frutas utilizando aprendizado de máquina	22
3.2	Classificação de laranjas por meio de processamento digital de imagens	22
3.3	Implementação de um modelo <i>Bag Of Features</i> para classificação de ma- ças, bananas e outras frutas	23
3.4	Comparação entre esta monografia e o estado da arte	23
4	METODOLOGIA	24
4.1	Revisão Bibliográfica	24
4.2	Seleção do algoritmo para gerar o modelo de classificação dos frutos de bananas, laranjas e maçãs	24
4.3	<i>Dataset</i>	24
4.4	Pré-processamentos	26
4.5	Geração do modelo	27
5	RESULTADOS	29
5.1	Primeiro Experimento	29
5.2	Segundo Experimento	30

6	CONCLUSÃO	32
6.1	Considerações Finais	32
6.2	Trabalhos Futuros	32
	REFERÊNCIAS	33

1 INTRODUÇÃO

A busca pela qualidade de vida e bem-estar tem se tornado cada vez mais presente na sociedade. A Organização Mundial de Saúde define a qualidade de vida como sendo “a percepção do indivíduo de sua inserção na vida, no contexto da cultura e sistemas de valores, nos quais ele vive e em relação aos seus objetivos, expectativas, padrões e preocupações” (OMS, 1997).

Considerando a alimentação como um dos pilares necessários ao bem-estar humano, uma alimentação inadequada pode acarretar em sobrepeso ou obesidade, o que contribui para o surgimento de outros tipos de doenças como a diabetes, hipertensão, doenças cardiovasculares, colesterol elevado, desenvolvimento de câncer, entre outras, que são as chamadas Doenças Crônicas Não Transmissíveis. O percentual de pessoas obesas em idade adulta no país mais do que dobrou em 17 anos, indo de 12,2%, entre 2002 e 2003, para 26,8%, em 2019. No mesmo período, a proporção da população adulta com excesso de peso passou de 43,3% para 61,7%, representando quase dois terços dos brasileiros. Os dados são do segundo volume da Pesquisa Nacional de Saúde (PNS) 2019, divulgada em 20 de outubro de 2020 pelo IBGE, realizada através de uma parceria com o Ministério da Saúde.

Na tentativa de contornar esse cenário a literatura apresenta diversos estudos com o objetivo de utilizar a tecnologia para auxiliar a população a manter uma alimentação saudável. A telemedicina, por exemplo, apresenta um canal eficiente para levar informações nutricionais para a população. Intitulada de "telesaúde", a ideia surgiu para oferecer serviços de atendimento nutricional, principalmente em situações em que o acesso a médicos e demais profissionais de saúde é prejudicado. Outro exemplo de plataforma, que busca disseminar informes nutricionais entre a população, é a *PlayKids*, uma plataforma de lazer para a criançada entre 2 e 8 anos, com mais de 5 milhões de usuários e 4 mil atividades diferentes. Sendo algumas específicas sobre alimentação saudável, apresentando explicações sobre a origem dos vegetais e sua importância para a saúde (DIALOGANDO, 2018).

Dentre as propostas de disponibilização de informação nutricional, com o objetivo de promover hábitos alimentares saudáveis e, assim, melhorar a qualidade de vida dos consumidores, destaca-se o uso da tecnologia com técnicas de Processamento Digital de Imagens e Visão Computacional para identificar alimentos em imagens digitais com intuito de fornecer informações precisas baseadas em tabelas de composição alimentar de referência (ALVES, 2016). Buscando contribuir cientificamente com a criação de metodologias para detecção de frutas em

imagens RBG. Neste trabalho apresentamos um modelo inteligente utilizando Redes Neurais Convolucionais.

A organização deste trabalho encontra-se da seguinte forma. Na Seção 2 expõe os tópicos chaves da pesquisa, enquanto a Seção 3 apresenta os trabalhos que se assemelham em algum aspecto com este trabalho. A Seção 4 descreve os procedimentos utilizados para realizar a pesquisa e assim como seu desenvolvimento. Na Seção 5 são apresentados os resultados alcançados. Por fim, na Seção 6 são mostradas as conclusões e os trabalhos futuros.

1.1 Objetivos

1.1.1 Objetivos Gerais

Uso de visão computacional e aprendizado de máquinas para detecção e classificação de imagens de frutas em imagens RGB.

1.1.2 Objetivos Específicos

- Criar uma base de dados com imagens de bananas, laranjas e maçãs.
- Criar um modelo inteligente que possa detectar e classificar frutas em imagens RGB;
- Analisar o desempenho do modelo gerado em relação a quantidade de acertos e erros na realização da tarefa de detecção das frutas.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são discutidos conceitos fundamentais para o entendimento da solução proposta neste trabalho. Nas seções 2.1 e 2.2 encontra-se uma breve explicação das áreas de conhecimento de Visão Computacional e Aprendizado de Máquina, respectivamente.

2.1 Visão Computacional

Ballard e Brown (1982), definiram Visão Computacional (VC) como a ciência que estuda e desenvolve tecnologias que permitam que máquinas enxerguem e extraiam características de imagens digitais capturadas por diferentes tipos de sensores e dispositivos. Essas informações extraídas permitem reconhecer, manipular e processar dados sobre os objetos que compõem a imagem capturada. A ideia geral da VC é oferecer um aprendizado parecido com o humano na obtenção de dados sobre uma imagem. Para isso, é necessário a utilização de técnicas de Processamento Digital de Imagem (PDI), assim como algoritmos de Aprendizado de Máquina (AM).

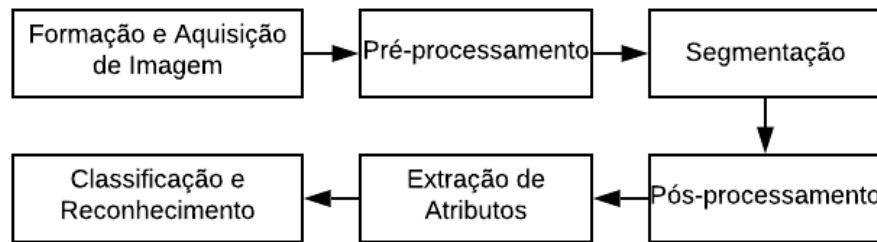
2.1.1 *Processamento de Imagens Digitais*

Uma imagem pode ser definida como uma função bidimensional, $f(x,y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x,y) é chamada de intensidade ou nível de cinza da imagem nesse ponto. Quando x , y e os valores de intensidade de f são quantidades finitas e discretas, chamamos de imagem digital. O campo do processamento digital de imagens se refere ao processamento de imagens digitais por um computador digital. Observe que uma imagem digital é composta de um número finito de elementos, cada um com localização e valor específicos. Esses elementos são chamados de elementos pictóricos, elementos de imagem, pels e pixels. Pixel é o termo mais utilizado para representar os elementos de uma imagem digital (GONZALEZ, 2009).

A Figura 1, apresenta as fases da metodologia comumente utilizada na literatura para a geração de algoritmos de PDI.

O primeiro passo da metodologia é a aquisição do banco de dados de imagens com a formação e digitalização das imagens. Após a obtenção das imagens digitais, o próximo passo consiste no pré-processamento da imagem da etapa anterior. O pré-processamento serve para melhorar a imagem através do uso de filtros para redução de ruídos e realçar contornos ou bordas de objetos contidos na imagem para que as próximas etapas sejam bem-sucedidas. Após o pré-processamento, temos a etapa de segmentação. O papel básico da segmentação é

Figura 1 – Hierarquia de processamento



Fonte: Autoria Própria

separar a imagem em partes ou objetos de interesse. Muitas vezes o resultado da segmentação não é adequado AUGUSTO (2013). Para corrigir esses defeitos, realiza-se o estágio de pós-processamento, com o auxílio de filtros morfológicos. Como exemplo, podemos citar: filtros de dilatação e erosão que são usados para corrigir imperfeições deixadas pela segmentação (ESQUEF *et al.*, 2003). Desta forma, a imagem de saída do pós-processamento está pronta para fornecer informações através da etapa de extração de atributos. Nesta etapa, algumas características dos objetos nas imagens são extraídas, resultando em algum tipo de discriminação entre objetos, sendo estes divididos em classes de acordo com suas discriminações. Por fim, o último estágio envolve o reconhecimento de padrões e classificação. O reconhecimento atribui uma descrição a um objeto, baseado na informação fornecida pelo seu descritor. A classificação confere um significado para um conjunto de objetos reconhecidos AUGUSTO (2013).

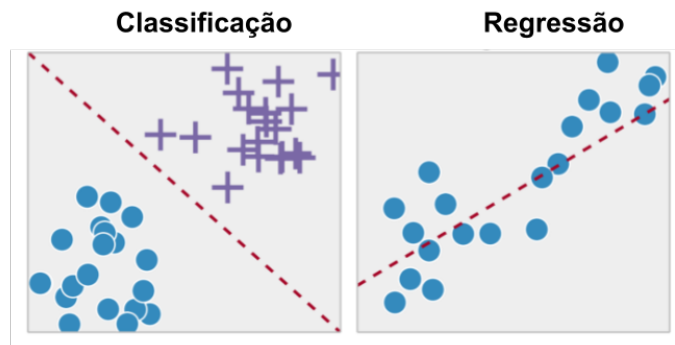
2.2 Aprendizado de Máquina

O Aprendizado de Máquina (AM), do inglês *Machine Learning*, é um ramo da Inteligência Artificial que explora o estudo e elaboração de algoritmos que podem aprender a partir de uma base dados. De acordo com Samuel (1959), o aprendizado de máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados. Além disso, AM explora a construção de algoritmos que podem aprender com seus erros e fazer previsões sobre dados a partir de quatro abordagens de aprendizagem: supervisionada, não supervisionada, por reforço e profunda.

2.2.1 Aprendizagem Supervisionada

Os algoritmos de aprendizagem supervisionada buscam adquirir conhecimentos a partir de dados de treinamento rotulados. O objetivo é encontrar os parâmetros ótimos que ajustem um modelo de forma a prever/classificar rótulos desconhecidos em novas bases de dados, também denominado como conjunto de teste. Se o rótulo é um número real, esta tarefa é denominada como regressão. Se o rótulo vem de um conjunto finito e não ordenado, então esta tarefa é denominada de classificação. A Figura 2 representa um exemplo dos dois tipos de tarefas que podem ser realizadas por algoritmos supervisionados de forma generalizada (KORBUT, 2017).

Figura 2 – Diferenças entre um modelo de classificação e um modelo de regressão.

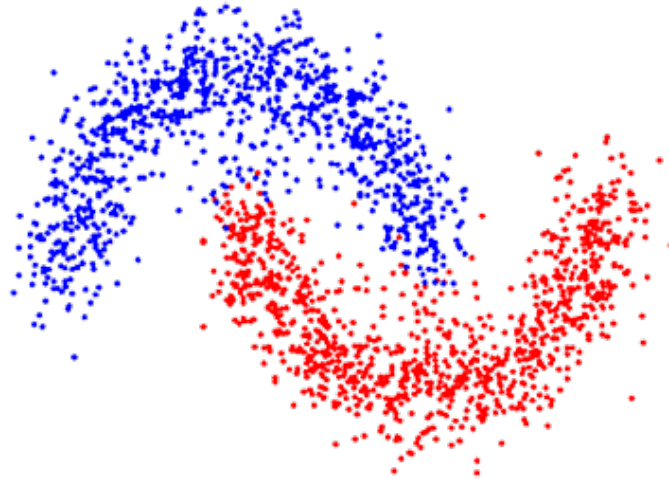


Fonte: Adaptado de (KORBUT, 2017)

2.2.2 Aprendizagem Não-Supervisionada

Os algoritmos de aprendizagem não-supervisionada são utilizados quando há menos informação sobre os dados em estudo, mais especificamente, o conjunto de dados de treinamento não é rotulado. Neste contexto, o objetivo destes algoritmos é observar algumas similaridades entre os dados de estudo, agrupando-os de forma apropriada. Em alguns casos, alguns dados em estudo podem diferir largamente dos demais, não se enquadrando em nenhum grupo similar, em caso como estes, os dados podem ser considerados anomalias ou *outliers* e portanto podem ser removidos da base. A Figura 3 apresenta um exemplo de clusterização (agrupamento) de uma base de dados de forma generalizada (KORBUT, 2017).

Figura 3 – Exemplo de aplicação de um algoritmo de aprendizagem não-supervisionada para clusterização de dados.



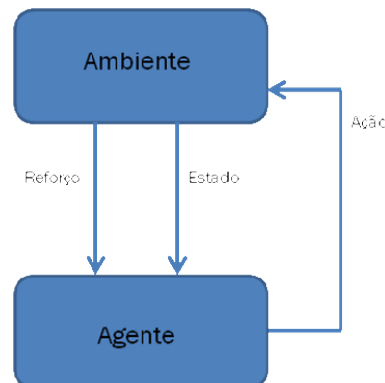
Fonte: (KORBUT, 2017)

2.2.3 Aprendizagem Por Reforço

Os algoritmos de aprendizagem por reforço buscam aprender e gerar conhecimentos de forma bastante diferente das classes de algoritmos anteriores, isto se deve a não utilização explícita de um conjunto de treinamento. Desta forma, os algoritmos de aprendizagem por reforço investigam e determinam ações a serem realizadas em determinados ambientes de modo a maximizar recompensas e minimizar penalidades. Imagine que você é um robô em um lugar desconhecido. Você pode realizar atividades e receber recompensas do ambiente pelas atividades realizadas de forma correta, podendo também receber penalidades quando a tarefa não for realizada, ou for realizada de forma errônea. Após cada ação executada, o comportamento do robô ficará mais inteligente. A Figura 4 apresenta um exemplo de metodologia, de forma simplificada, de um algoritmo de aprendizagem por reforço.

A cada passo, o agente recebe como entrada, uma indicação do estado atual do ambiente. O agente escolhe, uma ação a tomar, e gera sua saída. A ação altera o estado do ambiente, e uma medida dessa mudança de estado é informada ao agente através de um valor de sinal de reforço (KORBUT, 2017).

Figura 4 – Exemplo de uma metodologia simplificada utilizada por um algoritmo de aprendizagem por reforço.



Fonte: (DORÇA, 2012)

2.2.4 *Aprendizagem Profunda*

Aprendizagem profunda, do inglês *Deep Learning*, é uma forma de aprendizado de máquina, realizada através do uso de redes neurais artificiais, que permite que os computadores aprendam com a experiência e entendam o mundo em termos de uma hierarquia de conceitos. Com a hierarquia de conceitos é possível que o computador aprenda conceitos complicados, construindo-os a partir de outros mais simples. Desta forma um gráfico dessas hierarquias possui várias camadas de profundidade (GOODFELLOW *et al.*, 2016).

Em cada canal há redes neurais que buscam simular o cérebro humano e o funcionamento dos seus neurônios, sendo capaz de reconhecer padrões de forma eficiente e obter conhecimentos por meio de experiência. Através de entidades de processamento, os neurônios são interligados, podendo desta forma receber e enviar sinais. Realizando assim, a simulação dos neurônios presentes no cérebro, recebendo e retransmitindo informações (GOODFELLOW *et al.*, 2016).

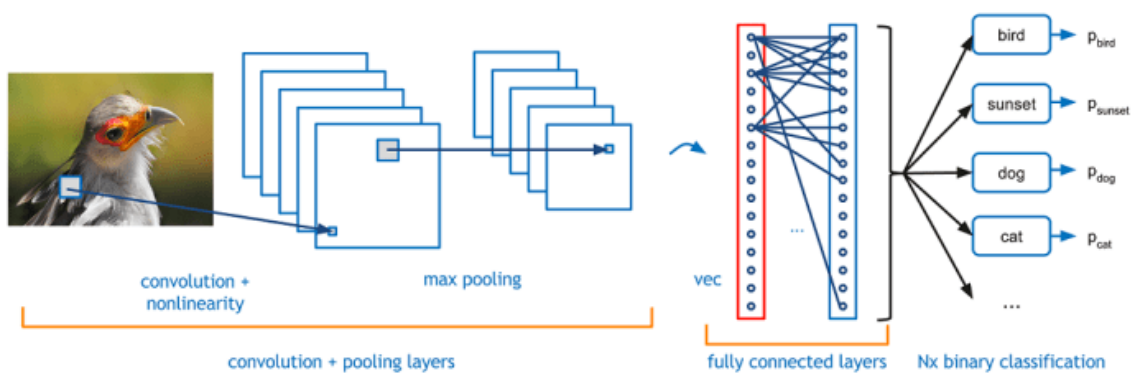
2.2.5 *Redes Neurais Convolucionais*

Dentre os algoritmos existentes na literatura mais utilizados para reconhecimento e classificação de imagens, destaca-se o uso das Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Network*), que será o algoritmo utilizado neste trabalho. Pertencentes a subclasse das Redes Neurais Profundas, pode-se dizer que a principal diferença entre uma Rede Neural Convolucional e uma Rede Neural Regular é o uso de no mínimo uma camada convolucional. Uma CNN reconhece um objeto procurando primeiro por características de baixo nível, como bordas, linhas e curvas, e então segue construindo características mais

abstratas através de uma série de camadas convolucionais (ROSA, 2019). Desta forma, as CNNs conseguem reconhecer padrões bastante complexos, sendo adaptáveis a distorções e variações das imagens.

Na Figura 5 representa a estrutura de uma CNN, que recebe uma imagem de entrada e aplica a convolução em várias camadas, logo após é passada para camadas totalmente conectadas e por fim como saída é dada uma classificação binária que define quais objetos foram detectados na imagem.

Figura 5 – Estrutura de uma Rede Neural Convolucional



Fonte: (BOOK, 2017)

A arquitetura de uma CNN é definida de acordo com a Figura 5:

- **Camada de Entrada (*Input*):** responsável pela leitura da imagem.
- **Camada Convolucional (*Convolution*):** cada camada extrai recursos de alto nível dos dados de entrada e os repassa para a próxima camada na forma de mapas de recursos.
- **Camada de Pooling (*Max pooling*):** utilizada para reduzir o tamanho da dimensão da imagem.
- **Camada Totalmente Conectada (*Fully connected layer*):** seu propósito é classificar a imagem de acordo com classes predeterminadas utilizando as características extraídas das camadas de convolução e de *pooling*.
- **Camada de Saída (*Nx binary classification*):** saída com os resultados da classificação.

3 TRABALHOS RELACIONADOS

Neste capítulo são abordados alguns trabalhos encontrados na literatura que se assemelham em alguns requisitos ao trabalho proposto nesta monografia. Na seção 3.1 é apresentado o trabalho do ROSA (2019) que utilizou aprendizado de máquina para fazer a classificação de imagens de várias frutas, dentre elas maçãs vermelhas, banana, melão, manga e mamão. Na seção 3.2 é apresentado o trabalho dos autores SANTIAGO *et al.* (2019) que tem como foco a classificação de imagens de laranjas. Na seção 3.3 é explicado o trabalho dos autores CHRISTMANN *et al.* (2018) onde os mesmos exploraram uma técnica clássica de visão computacional e fazem uma comparação com o uso de aprendizagem profunda para classificar imagens de maçãs, bananas e outras frutas.

3.1 Classificação de imagens de maçãs, bananas e outras frutas utilizando aprendizado de máquina

Neste trabalho, o autor ROSA (2019) teve como propósito principal utilizar algoritmos de aprendizado de máquina para criação de modelos inteligentes para classificação de várias frutas, dentre elas maçãs vermelhas, banana, melão, manga e mamão. Para tal, foram utilizados os algoritmos de Redes Neurais Convolucionais denominados por VGG, InceptionV3, Xception e MobileNetV2. Os modelos foram testados a partir da base de dados *Google Imagens*, sendo utilizado o conjunto de dados *Fruit Dataset* que pode ser encontrado em Kuang (2015), composto por 20 classes de objetos. Como resultado o melhor modelo gerado atingiu uma acurácia (proximidade entre o valor real esperado e o valor obtido no experimento) superior a 90% para o conjunto de teste determinado.

3.2 Classificação de laranjas por meio de processamento digital de imagens

Neste trabalho, os autores SANTIAGO *et al.* (2019) buscam facilitar o processo de seleção de laranjas através de técnicas de Processamento de Imagens Digitais e Aprendizagem de Máquina. Buscando desenvolver e testar modelos inteligentes para executar esta tarefa, os autores utilizaram algoritmos de aprendizado supervisionado para classificar frutas em uma base de dados com 204 imagens de laranjas divididas em três classes, denominadas de pequena, média e grande. Os modelos testados obtiveram uma acurácia de 82%. Na etapa de treinamento e teste dos classificadores o conceito de calibre, que informa a área circular da secção equatorial da fruta,

foi adicionado a base dados em alguns casos. Como resultado final, os autores demonstram que o modelo classificador de laranjas utilizando calibre obteve melhor desempenho em comparação com o modelo classificador sem utilizar este dado. Já o segundo classificador que foi utilizado, k-vizinhos mais próximos (K-NN), obteve 72,13% de desempenho tanto para o algoritmo utilizando dados sobre o calibre como sem.

3.3 Implementação de um modelo *Bag Of Features* para classificação de maçãs, bananas e outras frutas

Neste trabalho, os autores CHRISTMANN *et al.* (2018) tiveram como objetivo explorar uma técnica clássica de visão computacional, conhecida como *Bag Of Features*, com o intuito de detectar e classificar frutas e vegetais em imagens digitais. O principal objetivo deste trabalho é apresentar o desempenho de modelos de aprendizagem de máquina simples na realização deste tipo de tarefa. Para realização dos testes, foram definidos dois tipos de cenários. Para o primeiro cenário, foi criado um ambiente controlado, havendo apenas diferenças de iluminação e posicionamento dos objetos, enquanto o segundo cenário foi construído considerando uma maior presença de ruídos no fundo das imagens. Como resultado mostrou-se que apesar da tendência atual em se usar modelos mais complexos, o método de *Bag Of Features* continuam sendo eficiente quando utilizado em *datasets* considerados pequenos e com baixo custo computacional de treinamento.

3.4 Comparação entre esta monografia e o estado da arte

A Tabela 1 apresenta as semelhanças e diferenças entre o trabalho proposto e os trabalhos apresentados nesta seção.

Tabela 1 – Comparação entre esta monografia e o estado da arte.

Lista de Atividades	Trabalhos Comparados			
	Rosa (2019)	Santiago (2019)	Christmann (2018)	TCC Janne
Reconhecimento de bananas	X		X	X
Reconhecimento de laranjas		X		X
Reconhecimento de maçãs	X		X	X

Fonte: Elaborado pelo Autor (2020).

4 METODOLOGIA

Este capítulo descreve as etapas necessárias na realização esta pesquisa. O desenvolvimento desta pesquisa foi dividido em cinco etapas apresentadas nas seguintes seções. A Seção 4.1 apresenta a etapa de revisão das tecnologias mais recentes relacionadas à detecção de frutas, enquanto a Seção 4.2 define os algoritmos utilizados para geração do modelo inteligente usado nesta monografia. A Seção 4.3 detalha a estrutura do banco de dados utilizado e a Seção 4.4 descreve como processar os dados. Por fim, a Seção 4.5 apresenta o processo de implementação do modelo.

4.1 Revisão Bibliográfica

Na etapa inicial desta pesquisa, realizou-se uma verificação da literatura, com o intuito de identificar as tecnologias que obtiveram melhores resultados para resolver problemas semelhantes ao problema foco desta monografia.

4.2 Seleção do algoritmo para gerar o modelo de classificação dos frutos de bananas, laranjas e maçãs

Devido a existência de uma grande diversidade de algoritmos para classificação de objetos em imagens RGB, tornou-se fundamental a realização de uma filtragem, dando uma maior atenção para algoritmos que apresentam uma boa *acurácia*, isto significa, um bom percentual de acertos na classificação dos frutos de bananas, laranjas e maçãs. O objetivo dessa etapa é usar os conhecimentos adquiridos no estudo exploratório para o desenvolvimento de um modelo que alcance um resultado satisfatório e competitivo, quando comparado com os trabalhos existentes na literatura. Após o estudo exploratório, foi determinado o uso das Redes Neurais Convolucionais (RNC) para geração do modelo inteligente.

4.3 Dataset

Uma parte importante do trabalho da geração de modelos para detecção e classificação de objetos em imagens, é a seleção ou construção correta de conjuntos de dados utilizado como entrada dos modelos. Neste trabalho foi usado o *dataset* encontrado em Kuang (2015), onde foi selecionado imagens de frutas mais conhecidas na região como os frutos de maçã,

banana e laranja para realizar os treinamentos e testes necessários no modelo criado. A Tabela 2 apresenta a divisão das imagens para o conjunto de treinamento, teste e validação.

Tabela 2 – Divisão *Dataset*

Frutos	Treinamento	Teste	Validação
Banana	354	50	40
Laranja	432	50	40
Maçã	378	50	40

Fonte: Elaborado pelo Autor(2021).

As imagens de cada conjunto estão dividida entre imagens com fundo sem muitas variações e imagens de fundos variados, podendo ser variáveis nas cores ou objetos. Na Figura 6 são apresentados alguns exemplos de imagens pertencentes ao conjunto de bananas. A Figura 7 apresenta as imagens do conjunto de laranjas. A Figura 8 apresenta as imagens do conjunto de maçãs.

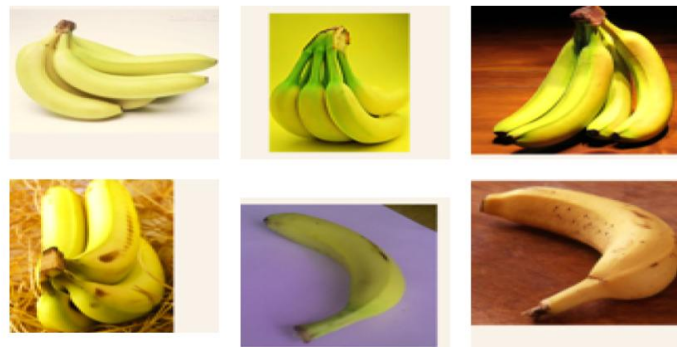


Figura 6 – Exemplo de bananas no *dataset*

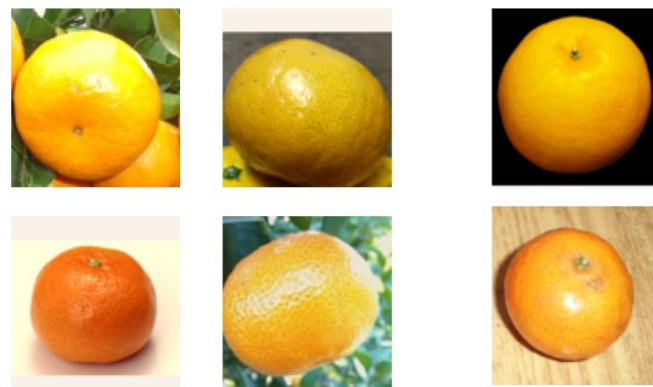


Figura 7 – Exemplo de laranjas no *dataset*



Figura 8 – Exemplo de maçãs no *dataset*

O principal intuito da inclusão de imagens com fundos variados no *dataset* é fazer com que modelo inteligente pudessem lidar com desafios durante a classificação das frutas. Desta forma, o modelo proposto poderia ser utilizado em aplicações inteligentes e disponibilizada a usuários finais, podendo ser utilizado para classificação de frutas em imagens retiradas de ambientes reais.

4.4 Pré-processamentos

Nessa fase realizou-se alguns pré-processamentos nas imagens do *dataset* para que as mesmas ficassem de acordo com os padrões necessários para aplicação do algoritmo RNC. Para isso, utilizou-se a biblioteca *OpenCV* para redimensionamentos das imagens para o tamanho padronizado de 150 x 150 pixels. Essa padronização é necessária para a entrada na rede, trazendo um ganho de performance ao modelo, reduzindo o número de operações matemáticas durante o processamento.

Para aumentar o número de imagens para a etapa de treinamento do modelo, utilizou-se a função *data_augmentation*. Para cada imagem existente na base, foram criadas dez novas imagens variadas. A mesma foi utilizada para melhorar o desempenho do modelo.

4.5 Geração do modelo

O modelo gerado neste trabalho foi implementado na linguagem *Python*¹, com o suporte das bibliotecas, *openCV*², *Keras*³, *Numpy*⁴, *Matplotlib*⁵ e *Pillow/PIL*⁶ no ambiente *Google Colaboratory*⁷, onde é permitido criar códigos em *Python* através do navegador sem a necessidade de configurações, pois o mesmo disponibiliza o uso gratuito a GPU's. Todas as etapas para geração do modelo foram criadas utilizando um Notebook com processador AMD Ryzen 5 3500U e 8GB de memória RAM, placa integrada AMD Radeon RX Vega 8 e o sistema operacional Windows 10.

A estrutura da RNC aplicada foi elaborada com diferentes tipos de camadas:

- Três camadas de convoluções usadas para extrair dados importantes das imagens, onde na primeira camada foi utilizada uma função de ativação *Relu*;
- Três camadas *Pool* usadas para diminuir a proporção das imagens;
- Uma camada *Dropout* para eliminar aleatoriamente neurônios durante o treinamento, com o objetivo de evitar um *overfitting*;
- Uma camada *Flatten* que transforma a matriz da imagem de 2D para 1D;
- Duas camadas densas, uma com a função de ativação *Relu* e outra com a *Softmax*, que indica se a imagem pertence ou não a uma classe.

A rede foi compilada com alguns hiperparâmetros importantes: a indicação de uma função de perda, que tem como propósito calcular o tamanho da perda que o modelo deve tentar minimizar durante a fase de treinamento (KERAS, 2021b); um método de descida gradiente estocástico, baseado em previsão adaptativa de momentos de primeira e segunda ordem (KERAS, 2021a). Nesta monografia foi usado o gradiente estocástico *Sparse_Categorical_Crossentropy* e *loss* como função de perda.

A métrica utilizada para avaliação da qualidade da classificação foi a acurácia, que é calculada através da proporção de casos que foram corretamente previstos, sejam eles verdadeiro positivo ou verdadeiro negativo. Por exemplo: de 100 imagens ele obteve 60 previsões corretas, logo a acurácia seria $60/100=0,6$, ou seja, 60%.

¹ <https://www.python.org>

² <https://opencv.org>

³ <https://keras.io>

⁴ <https://numpy.org/>

⁵ <https://matplotlib.org/>

⁶ <https://pypi.org/project/Pillow/>

⁷ <https://colab.research.google.com>

Com a estrutura da rede criada, inicia-se o processo de treinamento. Na etapa de treinamento foi realizada utilizando experimentos distintos. Os principais configurações para cada experimento realizado foram:

- O conjunto de treinamento e validação, onde o conjunto de validação foi 20% do *dataset*, neste trabalho foi definido um conjunto separado para teste;
- O número de épocas, ou seja, o número de vezes que o modelo é exposto ao conjunto de dados de treinamento.
- *batch_size* Número de amostras por atualização de gradiente. Quando não especificado, o padrão é 32. A utilização de mini lotes para treinar a rede é importante para reduzir o consumo da memória RAM.

Na tabela 3 é mostrado os valores das configurações de cada um dos cenários.

Tabela 3 – Configurações dos cenários distintos.

Parâmetros	Experimento 1	Experimento 2
Épocas	50	150
<i>Batch size</i>	256	512
Imagens	1.244	29.603

Fonte: Elaborado pelo Autor (2021).

5 RESULTADOS

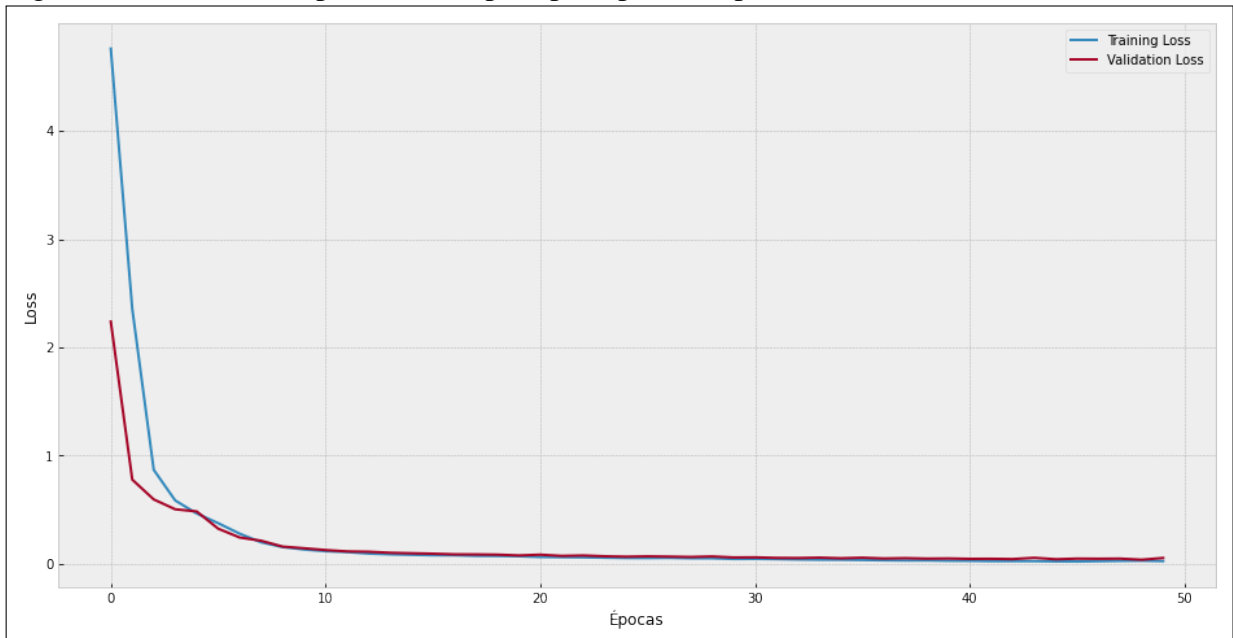
Neste capítulo são apresentados os resultados atingidos pelo modelo inteligente definido neste trabalho. Um conjunto de experimentos foi realizado com propósito de avaliar a metodologia proposta. As seções 5.1 e 5.2 mostram detalhadamente cada classificação, bem como os resultados encontrados em cada cenário apresentado.

5.1 Primeiro Experimento

No primeiro cenário foi considerado o *dataset* original apresentado anteriormente na Seção 4.3. O modelo foi treinado com 50 épocas, usando 1.244 imagens, com 256 amostras por *batch size*.

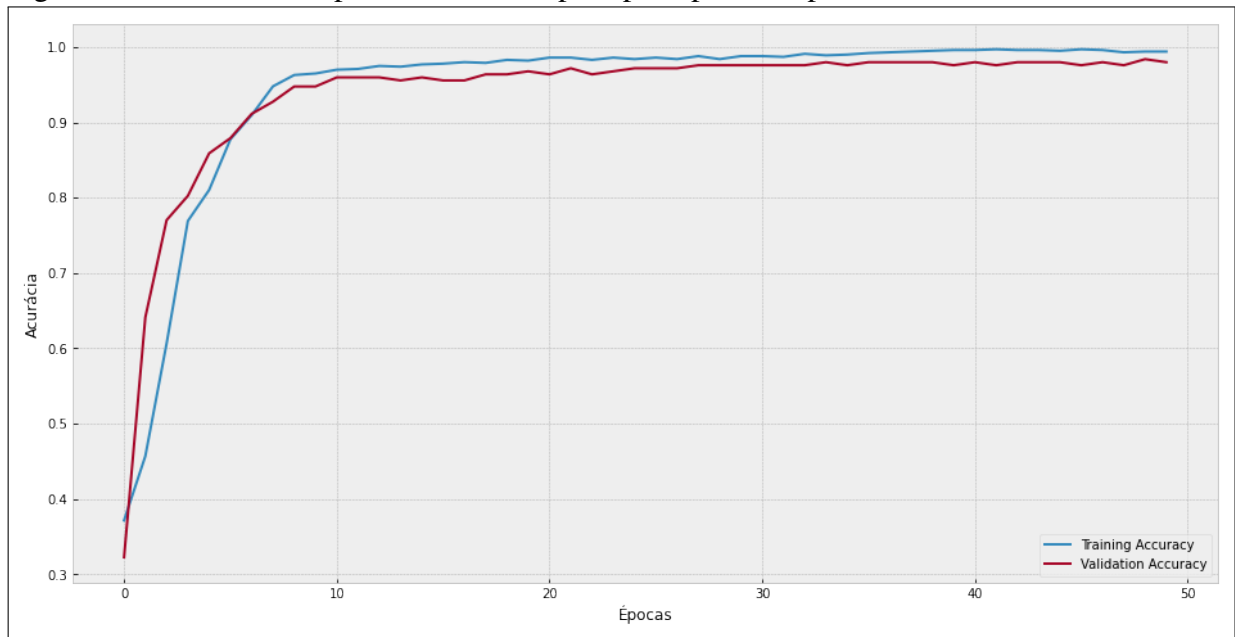
Como mencionado nas seções anteriores, neste trabalho foi escolhida duas métricas para avaliar a rede. A primeira foi o *loss*, onde quanto menor seja o valor do *loss*, melhor é o desempenho da rede. Acurácia também foi analisada, pois a mesma mostra o desempenho do modelo no treinamento. As Figuras 9 e 10 mostram mutuamente as taxas do *loss* e da acurácia, no decorrer do treinamento. A acurácia obtida para o primeiro experimento foi 96,66%.

Figura 9 – Cenário completo do *loss* por época para o experimento 1



Fonte: Resultado do modelo

Figura 10 – Cenário completo da acurácia por época para o experimento 1



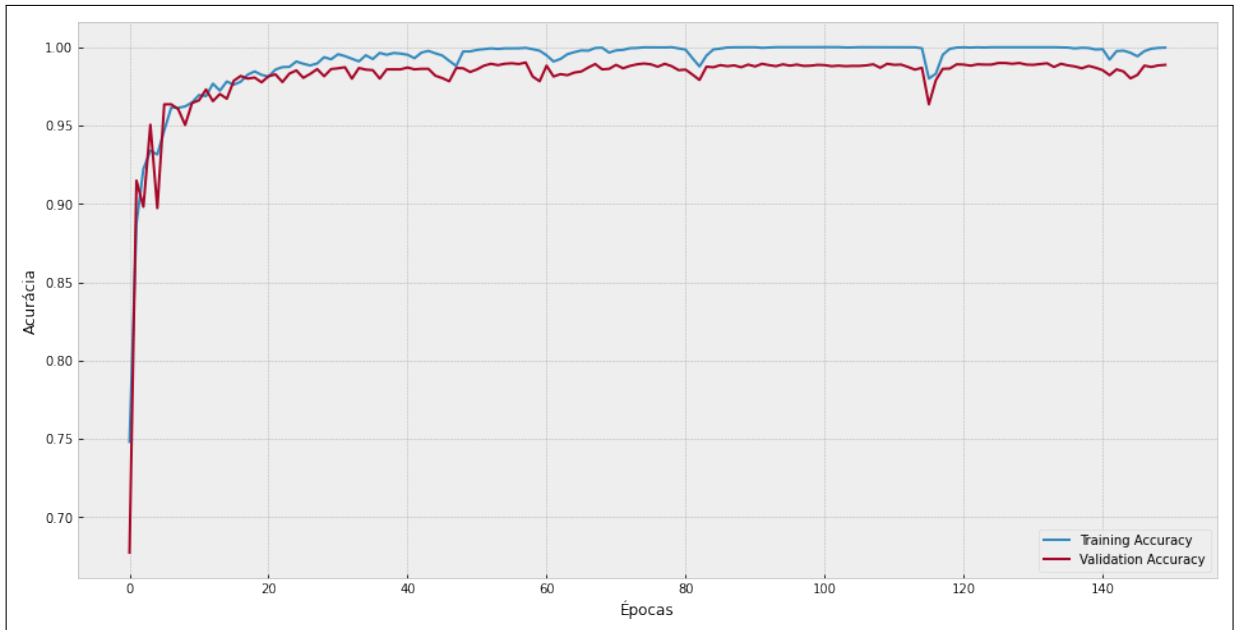
Fonte: Resultado do modelo

5.2 Segundo Experimento

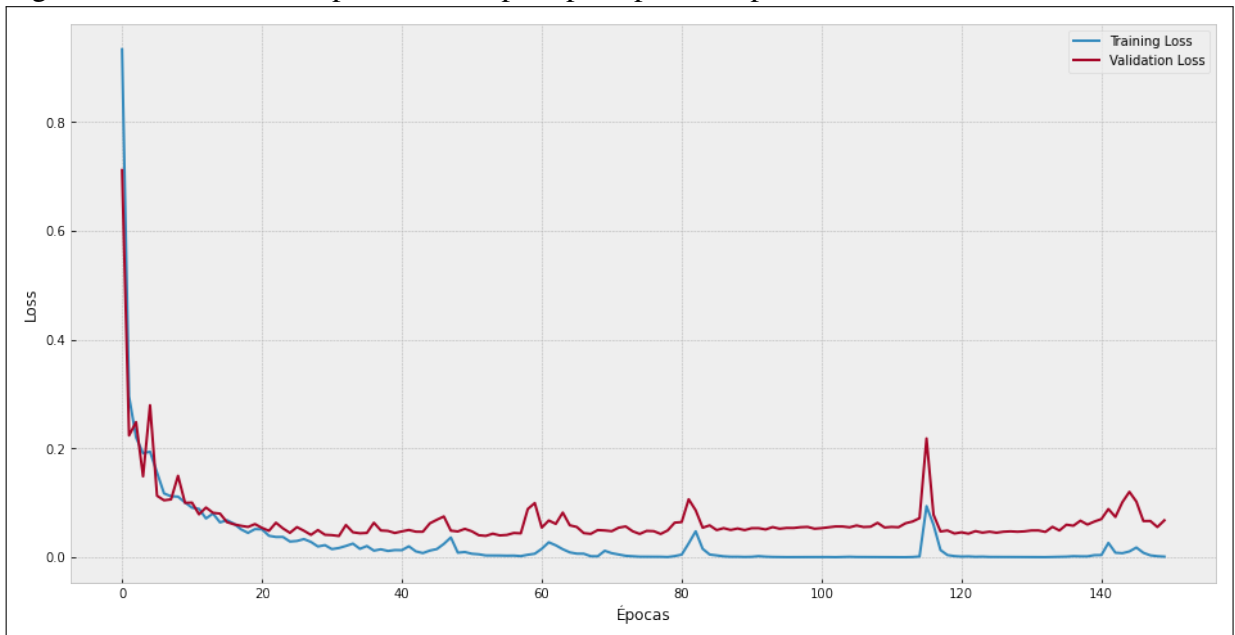
No segundo cenário foi utilizada a função *augmentation* para aumentar o *dataset* apresentado na seção 4.3, a função faz transformações aleatórias na imagem, como rotação, para este treinamento para cada imagem a função fez 10 transformações aleatórias. A rede foi treinada com 150 épocas, utilizando 29.603 imagens e o *batch size* foi definido como 512, para um melhor desempenho no treinamento, foi adicionada uma função de ativação *relu* na primeira camada de convolução.

Nas figuras 11 e 12 é mostrado o cenário completo das duas métricas analisadas neste trabalho, acurácia e *loss*, respectivamente. Neste cenário o modelo apresentou uma acurácia de 98,06%, mostrando assim a importância de possuir um *dataset* com uma boa quantidade de imagens.

Figura 11 – Cenário completo da acurácia por época para o experimento 2



Fonte: Resultado do modelo

Figura 12 – Cenário completo do *loss* por época para o experimento 2

Fonte: Resultado do modelo

6 CONCLUSÃO

6.1 Considerações Finais

Esta monografia apresenta um modelo inteligente para a classificação dos frutos de: banana, maçã e laranja em imagens RGB. Inicialmente, foi escolhido um *dataset* considerado razoável, em número de imagens, para realização das etapas de treinamento e os testes do modelo. Para o treinamento da rede foi usado dois cenários diferente, no primeiro foi utilizado o *dataset* de (KUANG, 2015) sem modificações, já no segundo cenário a rede foi treinada com o mesmo *dataset* ampliado.

Com os resultados obtidos neste trabalho pode-se concluir que o modelo implementado utilizando a estrutura RNC, com algumas modificações foi capaz de classificar satisfatoriamente os frutos de banana, maçã e laranja, mostrando também que ter uma boa quantidade de imagens em um *dataset* pode fazer a diferença em um treinamento.

6.2 Trabalhos Futuros

Um aprimoramento da base de dados para a classificação de outras frutas e também de vegetais, seria uma das ideias de trabalhos futuros. Com o intuito de criar uma aplicação *mobile* que venha a ajudar de alguma forma a reeducação alimentar das pessoas.

REFERÊNCIAS

- ALVES, M. C. M. Modelo de estimativa de valores nutricionais, calóricos e glicêmicos por meio de reconhecimento de alimentos em imagens digitais. 2016.
- AUGUSTO, K. S. Identificação automática do grau de maturação de pelotas de minério de ferro. 2013.
- BALLARD, D. H.; BROWN, C. M. **Computer Vision**. New York: Prentice Hall Professional Technical Reference, 1982. ISBN 978-0-13-165316-0.
- BOOK, D. L. **As 10 Principais Arquiteturas de Redes Neurais**. 2017. Disponível em: <<http://deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais/>>. Acesso em: 31 set. 2020.
- CHRISTMANN, G. H. G.; GRANDO, R. B.; MONTENEGRO, F. J. C.; GUERRA, R. S. Implementação de um modelo bag of features para classificação de frutas e vegetais. 2018.
- DIALOGANDO. **Alimentação saudável: tecnologia auxilia mães e filhos**. 2018. Disponível em: <<https://www.dialogando.com.br/inovacao/alimentacao-saudavel-tecnologia>>. Acesso em: 21 set. 2020.
- DORÇA, F. A. Uma abordagem estocástica baseada em aprendizagem por reforço para modelagem automática e dinâmica de estilos de aprendizagem de estudantes em sistemas adaptativos e inteligentes para educação a distância. 2012.
- ESQUEF, I. A.; ALBUQUERQUE, M. P. de; ALBUQUERQUE, M. P. de. Processamento digital de imagens. 2003.
- GONZALEZ, R. C. **Processamento Digital de Imagens**. 3rd. ed. Pearson, 2009. ISBN 978-8576054016. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=bad9f7e77103e98987a2af8a668c24a4>>.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Deep Learning Book: MIT Press, 2016. <<http://www.deeplearningbook.org>>.
- KERAS. **Adam**. 2021. Disponível em: <<https://keras.io/api/optimizers/adam/>>.
- KERAS. **Losses**. 2021. Disponível em: <<https://keras.io/api/losses/>>.
- KORBUT, D. **Machine Learning Algorithms: Which One to Choose for Your Problem**. 2017. Disponível em: <<https://blog.statsbot.co/machine-learning-algorithms-183cc73197c>>. Acesso em: 17 set. 2020.
- KUANG, H. **Fruitdataset**. 2015. Disponível em: <https://www.researchgate.net/publication/283087342_Fruitdataset>.
- OMS, W. H. O. Measuring quality of life. 1997.
- ROSA, A. L. d. Classificação de imagens de frutas utilizando aprendizado de máquina. 2019.
- Samuel, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210–229, 1959.
- SANTIAGO, W. E.; LOPES, G. D.; FACETO, A. D.; ROCHA, A. L. J.; ; CALDEIRA, R. F. Classificação de laranjas por meio de processamento digital de imagens. 2019.