

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**ENZO SALVADORI**

**IDENTIFICAÇÃO DE PLANTAS DE MILHO UTILIZANDO IMAGENS  
AÉREAS OBTIDAS POR VANTS**

**TRABALHO DE CONCLUSÃO DE CURSO**

**PONTA GROSSA**

**2020**

**ENZO SALVADORI**

**IDENTIFICAÇÃO DE PLANTAS DE MILHO UTILIZANDO IMAGENS  
AÉREAS OBTIDAS POR VANTS**

Trabalho de Conclusão de Curso apresentada como requisito parcial à obtenção do título de Bacharel em Ciência da Computação, do Departamento Acadêmico de Informática, da Universidade Tecnológica Federal do Paraná.

Orientador: Prof.<sup>a</sup> Dra. Simone Bello Kaminski Aires

Coorientador: Prof. Dr. Erikson Freitas de Moraes

**PONTA GROSSA**

**2020**



Ministério da Educação  
**Universidade Tecnológica Federal do Paraná**  
Câmpus Ponta Grossa

Diretoria de Graduação e Educação Profissional  
Departamento Acadêmico de Informática  
Bacharelado em Ciência da Computação



---

## TERMO DE APROVAÇÃO

### IDENTIFICAÇÃO DE PLANTAS DE MILHO UTILIZANDO IMAGENS AÉREAS OBTIDAS POR VANTS

por

ENZO SALVADORI

Este Trabalho de Conclusão de Curso (TCC) foi apresentado em 9 de dezembro de 2020 como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

-----  
Prof.<sup>a</sup> Dra. Simone Bello Kaminski Aires  
Orientador(a)

-----  
Prof. Dr. Erikson Freitas de Moraes  
Coorientador(a)

-----  
Prof.<sup>a</sup> Dra. Mauren Louise Sguario Coelho De Andrade  
Membro titular

-----  
Prof. MSc. Geraldo Ranthum  
Membro titular

-----  
Prof. MSc. Geraldo Ranthum  
Responsável pelo Trabalho de Conclusão de  
Curso

-----  
Prof.<sup>a</sup> Dra. Mauren Louise Sguario Coelho De Andrade  
Coordenadora do curso

## RESUMO

Agricultura de precisão é a área da agronomia que trabalha com recursos tecnológicos de precisão no manejo da lavoura que são utilizados para a gestão dos processos do cultivo e o seu georreferenciamento, com intuito de melhorar a produtividade geral da plantação, o que possibilita ao agricultor um monitoramento de sua produção de forma mais precisa e constante. Este trabalho se justifica uma vez que determinando o número de plantas de um cultivo, obtém-se uma informação que influencia diretamente na produtividade geral de uma plantação. Esta informação pode ser utilizada pelo agricultor de diversas maneiras como contornar a perda, caso o resultado seja extremamente prejudicial para o cultivo ou em casos não extremos identificar os fatores externos que podem ter afetado a produtividade do plantio e assim monitorar tal condição para reduzir a perda de produtividade em próximas plantações, além de poder calcular uma estimativa financeira do valor final da produção. Este trabalho apresenta um método para identificação e contagem de plantas de milho, que se consiste em utilizar técnicas de processamento de imagens e aprendizagem de máquina para treinar uma rede neural para identificar plantas de milho. A rede neural escolhida para os experimentos foi a *RetinaNet*, e utilizou no seu treinamento uma base de imagens gerada a partir de imagens aéreas que foram obtidas por um VANT. Essa base possui 3129 exemplos de rótulos que foram utilizados para o treinamento e validação, e 757 exemplos na base de teste. Os resultados obtidos no cálculo da métrica de precisão apresentaram bons valores comparados com a literatura, onde a análise foi feita por meio da base de teste, que conseguiu alcançar uma precisão de 99,0% e revocação de 97,6%.

## **ABSTRACT**

Precision agriculture is the area of agronomy that works with precision technological resources in the management of crops that are used for the management of cultivation processes and their georeferencing, in order to improve the overall productivity of the plantation, which allows the farmer a monitoring your production more accurately and constantly. This work is justified since determining the number of plants in a crop, information is obtained that directly influences the overall productivity of a plantation. This information can be used by the farmer in several ways, such as circumventing the loss, if the result is extremely harmful to the crop or in non-extreme cases, to identify the external factors that may have affected the productivity of the plantation and thus monitor such condition to reduce the loss of productivity in upcoming plantations, in addition to being able to calculate a financial estimate of the final value of production. This work presents a method for the identification and counting of corn plants, which consists of using image processing techniques and machine learning to train a neural network to identify corn plants. The neural network chosen for the experiments was RetinaNet, and used in its training a base of images generated from aerial images that were obtained by a UAV. This base has 3129 examples of labels that were used for training and validation, and 757 examples in the test base. The results obtained in the calculation of the precision metric showed good values compared to the literature, where the analysis was made through the test base, which managed to achieve an accuracy of 99.0% and recall of 97.6%.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Imagem ilustrando falhas detectadas em um determinado cultivo.....	9
Figura 2 - Metodologia do trabalho de OLIVEIRA et al. ....	12
Figura 3 - Detecção de plantas de eucalipto em uma imagem aérea. ....	13
Figura 4 – Detecção de carros em uma imagem área urbana. ....	14
Figura 5 - Aplicação da rede RetinaNet original e da versão melhorada para detectar pessoas. ....	15
Figura 6 - Detecção de palmeiras feitas pela rede RetinaNet. ....	15
Figura 7 - Imagem aérea de uma produção de cana de açúcar.....	18
Figura 8 - Imagem aérea de uma produção de milho.....	19
Figura 9 - Imagem aérea de uma produção de soja.....	19
Figura 10 - Matriz de pixels de uma imagem digital. ....	20
Figura 11 - Sistema tridimensional de mistura de cores do sistema RGB.....	21
Figura 12 - Espaço de cores HSV.....	21
Figura 13 - Binarização de uma imagem.....	23
Figura 14 - Binarização da cor verde pelo canal HSV.....	24
Figura 15 - Rotulação de componentes conexos.....	25
Figura 16 - Estrutura de uma Rede Neural com múltiplas camadas. ....	28
Figura 17 - Exemplos de âncoras posicionadas em torno do centro da imagem. ....	29
Figura 18 - Exemplo da estrutura de uma CNN de dois estágios.....	30
Figura 19 - Perda Focal para diferentes valores de $\gamma$ . ....	31
Figura 20 - Componentes principais de uma rede RetinaNet.....	32
Figura 21 - VANT utilizado no mapeamento.....	35
Figura 22 - Foto tirada da plantação com o VANT a 55 metros de altura.....	36
Figura 23 - Etapas de metodologia para detecção com uso técnicas de processamento de imagem. ....	37
Figura 24 - Binarização das plantas. ....	38
Figura 25 - Regiões detectadas nas imagens binárias.....	38
Figura 26 - Etapas de metodologia para detecção com uso de CNN.....	39
Figura 27 - Ferramenta Labellmg.....	40
Figura 28 - Formatação dos rótulos salvos no formato Pascal VOC.....	41
Figura 29 - Aplicação da rede RetinaNet no experimento 1.....	43
Figura 30 - Gráfico da função de mAP do treinamento da rede RetinaNet no experimento 2.....	44
Figura 31 - Gráfico da função de perda do treinamento da rede no experimento 3. ....	45
Figura 32 – Imagens utilizadas na base de teste. ....	46
Figura 33 - Avaliação da detecção feita pela rede RetinaNet com o melhor treinamento. ....	50

## LISTA DE TABELAS

Tabela 1 - Avaliação do método utilizando técnicas de processamento de imagem.	47
Tabela 2 - Avaliação da rede RetinaNet treinada por 10 épocas com a base de 1613 rótulos. ....	48
Tabela 3 - Avaliação da rede RetinaNet treinada com auto a avaliação por mAP. ....	49
Tabela 4 - Avaliação do método utilizando a rede RetinaNet com o melhor treinamento. ....	49
Tabela 5 – Comparação de treinamentos. ....	51

## LISTA DE SIGLAS

VANT	<i>Veículo aéreo não tripulado</i>
RBG	<i>Red Blue Green</i>
HSV	<i>Hue Saturation Value</i>
CNN	<i>Convolutional Neural Networks</i>
GNSS	<i>Global Navigation Satellite System</i>
XML	<i>Xtensible Markup Language</i>
mAP	<i>Mean Average Precision</i>



## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>8</b>
1.1 OBJETIVO GERAL	10
1.2 OBJETIVOS ESPECÍFICOS	10
1.3 JUSTIFICATIVA	10
<b>2 TRABALHOS RELACIONADOS</b>	<b>12</b>
<b>3 FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
3.1 AGRICULTURA DE PRECISÃO	16
3.2 VEÍCULOS AÉREOS NÃO TRIPULADOS	17
3.3 IMAGENS DIGITAIS	20
3.3.1 Binarização	22
3.3.2 Componentes Conexos	25
3.4 INTELIGÊNCIA ARTIFICIAL	26
3.4.1 Aprendizagem de Máquina	26
3.4.2 Redes Neurais Convolucionais	27
3.4.3 Redes de Dois e Um Estágio	29
3.4.4 Loss	30
3.4.5 RetinaNet	30
3.5 MÉTRICAS DE AVALIAÇÃO	33
3.5.1 Mean Average Precision	33
3.5.2 Contagem de Indivíduos	33
<b>4 METODOLOGIA</b>	<b>35</b>
4.1 OBTENÇÃO DAS IMAGENS	35
4.2 ABORDAGEM UTILIZANDO TÉCNICAS DE PROCESSAMENTO DE IMAGEM	36
4.2.1 Transformação para HSV e Binarização	37
4.2.2 Detecção de Regiões	38
4.3 ABORDAGEM COM USO DE CNN	39
4.3.1 Criação da Base de Imagens	39
4.3.2 Seleção da CNN	41
4.3.3 Treinamento da Rede	42
4.3.4 Configuração e Entradas	43
4.3.4.1 Experimento 1	43
4.3.4.2 Experimento 2	43
4.3.4.3 Experimento 3	44
4.3 ANÁLISE DOS RESULTADOS	45
4.4 RESULTADO TÉCNICAS DE PROCESSAMENTO DE IMAGEM	47
4.5 RESULTADO RETINANET – EXPERIMENTO 1	48
4.6 RESULTADO RETINANET – EXPERIMENTO 2	48
4.7 RESULTADO RETINANET – EXPERIMENTO 3	49

<b>5 CONCLUSÕES .....</b>	<b>52</b>
<b>REFERÊNCIAS.....</b>	<b>53</b>

## 1 INTRODUÇÃO

A agricultura é um campo de estudos que está em desenvolvimento constante desde os primórdios da humanidade. Teve início com a descoberta da utilização da terra para produção de alimentos, e com o passar dos anos vem desenvolvendo técnicas para o aperfeiçoamento da produção alimentícia (Ministério da Agricultura, Pecuária e Abastecimento, 2011).

Devido ao grande crescimento populacional e a falta de espaço para produção, é importante que se tenha um cultivo onde se possa extrair o máximo de recursos da terra. Nas últimas décadas foram desenvolvidas tecnologias com intuito de aumentar a produtividade agrícola, como exemplo: a utilização de implementos agrícolas, desenvolvimento de novos tipos de defensivos agrícolas, criação de sementes transgênicas e aplicação de sistemas de monitoramento computacional (Ministério da Agricultura, Pecuária e Abastecimento, 2011). Todas essas inovações proporcionam uma maior eficiência na produtividade, e ampliam a capacidade de produção.

A chamada agricultura de precisão tem chamado a atenção de pesquisadores, e fortalece-se desde meados da década de 1990 (SOARES et al., 2017). Essa área da agronomia trabalha com recursos tecnológicos de precisão no manejo da lavoura utilizados para a gestão dos processos do cultivo e para o seu georreferenciamento, com intuito de melhorar a produtividade geral da plantação (SOARES et al., 2017). Com os avanços nesse campo, foi possibilitado ao agricultor um monitoramento de sua produção de forma mais precisa e constante.

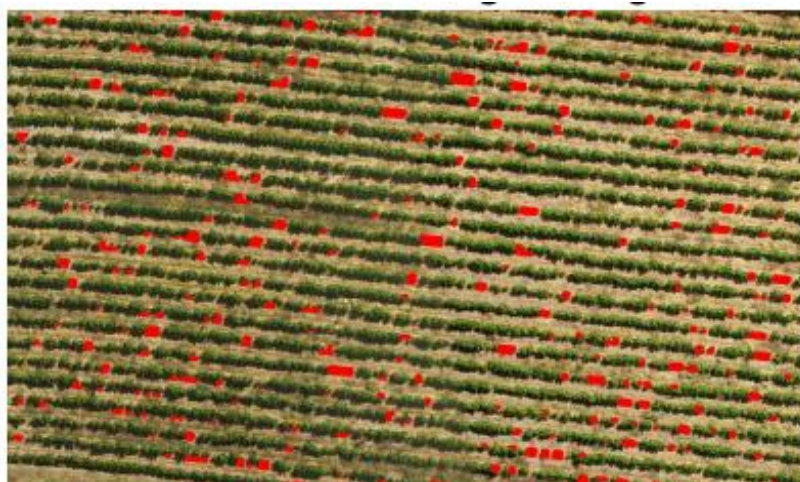
A utilização de veículos aéreos não tripulados popularmente conhecidos como drones surgiu como uma tecnologia ao auxílio de produtores para monitoramento das áreas de plantio por meio de imagens computacionais e tem se difundido devido ao seu baixo custo e a fácil acessibilidade (JORGE; INAMASU, 2014).

Os estudos encontrados referentes a imagens de VANTs na agricultura de precisão procuram soluções relacionadas às áreas de doenças das culturas, estresse hídrico e desequilíbrio de nutrientes do solo, todas essas trabalham com variáveis que interferem diretamente na saúde da planta (OLIVEIRA et al., 2018). No entanto estimar a produção de um cultivo por meio de imagens obtidas por VANTs é uma área de pesquisa a ser explorada, uma vez que esta informação é de grande utilidade, pois com a correção das mesmas pode se evitar uma perda substancial no cenário geral da produção (SOARES et al., 2017).

Uma análise apresentada pelo Jornal Eletrônico da Embrapa Milho e Sorgo (EMBRAPA, 2014), mostra um exemplo de como as falhas nas áreas de plantio podem gerar perdas significativas para os produtores. Considerando que uma cultura de milho que apresentou falhas em 28,5% da sua área de plantio de 1.000 hectares, e que era planejado produzir 70.000 plantas por cada hectare teremos uma perda de produção de 20.000 plantas por hectare. Considerando que cada planta gera uma espiga e que uma espiga pesa em média 0,15 kg teremos uma perda de 3.000 kg por hectare, e uma perda total da produção de 3.000.000 kg. Cada saco de milho pesa 60 kg o que totaliza 50.000 sacos de milho que seriam vendidos a R\$ 23,66 cada, tendo uma perda econômica bruta total de aproximadamente R\$ 1.183.000,00.

A Figura 1 apresenta uma imagem já processada obtida por um VANT, onde pode se observar que as linhas de plantio apresentam áreas pintadas em vermelho que demarcam os possíveis locais onde ocorreram falhas no nascimento das plantas.

**Figura 1 - Imagem ilustrando falhas detectadas em um determinado cultivo.**



Fonte (OLIVEIRA et al., 2018)

Neste trabalho, será desenvolvida uma base de dados de imagens em uma produção de milho utilizando VANTs. Posteriormente serão utilizados métodos computacionais que, a partir das imagens digitais fornecidas pelos VANTs, realizem de forma automática a detecção e contagem de plantas em um cultivo. Para tanto serão empregadas técnicas de processamento digital de imagens e aprendizagem de máquina por meio de um treinamento de uma rede neural convolucional, para gerar os resultados dos experimentos.

## 1.1 OBJETIVO GERAL

O objetivo desse trabalho é automatizar a detecção e quantificação de plantas em cultivos de milho por meio de uma rede neural convolucional utilizando imagens de plantações obtidas por um VANT.

## 1.2 OBJETIVOS ESPECÍFICOS

Como objetivos específicos têm-se:

- Realizar uma revisão bibliográfica sobre os assuntos abordados neste trabalho.
- Realizar a coleta de imagens aéreas de plantações pelo uso de um VANT.
- Construir a base de dados com as imagens obtidas pelo VANT.
- Compreender técnicas de processamento de imagem, inteligência artificial e aprendizagem de máquina.
- Aplicar técnicas de processamento de imagem, inteligência artificial e aprendizagem de máquina.
- Treinar uma Rede Neural Convolucional para detecção de plantas em uma produção.
- Desenvolver um algoritmo de contagem de plantas em uma produção.
- Realizar experimento na base de imagens obtida pelo VANT.
- Analisar os resultados obtidos pelos experimentos.
- Quantificar percentual de produtividade em uma plantação.

## 1.3 JUSTIFICATIVA

Este trabalho se justifica uma vez que determinado o número de plantas de um cultivo, obtemos uma informação que influencia diretamente a produtividade geral de uma plantação.

A tarefa de detecção de plantas em lavouras, utilizando imagens aéreas de VANTs para automatizar o processo, gera uma informação diretamente ligada a

produtividade geral do agricultor e pode ser feita uma estimativa dos lucros finais da sua produção (EMBRAPA, 2014).

Essa tecnologia implantada em culturas de grandes áreas trará resultados com uma praticidade e precisão maior em se obter o percentual de produtividade geral da produção, que anteriormente seria adquirida por métodos estimativos (OLIVEIRA et al., 2018).

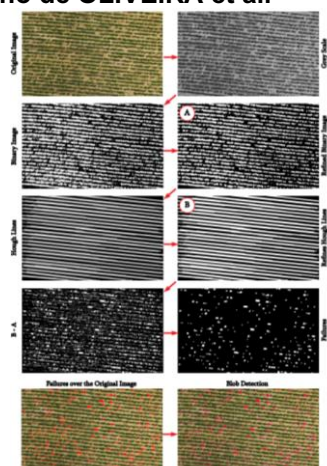
## 2 TRABALHOS RELACIONADOS

A fim de gerar estimativas sobre a produtividade do cultivo de plantas algumas abordagens têm sido propostas por pesquisadores. Dentre essas abordagens algumas estão relacionadas com este trabalho e outras são descritas a seguir com o objetivo de explicar diferentes abordagens que utilizam imagens obtidas por VANTs.

O artigo proposto por (OLIVEIRA et al., 2018) utiliza de métodos de processamento de imagem para identificar falhas dentro das linhas de produção em cultivos de cana de açúcar. O método proposto no trabalho para encontrar as falhas de plantas consiste em 8 etapas: (I) transformar a imagem colorida em escala de cinza; (II) binarizar a imagem para separar o que for planta representado pela cor branca e o fundo pela cor preta; (III) aplicar uma função de abertura na imagem binarizada; (IV) encontrar as linhas da imagem aberta por meio da técnica de Transformada de Hough; (V) aplicar uma função de fechamento na imagem das linhas; (VI) fazer a subtração da imagem gerada na etapa (III) pela da etapa (V); (VII) aplicar novamente uma função de fechamento na nova imagem gerada pela subtração; (VIII) aplicar um detector de componentes para encontrar as áreas finais que representam todas as áreas onde se encontram as falhas. Os resultados do trabalho apresentam completude aproximando-se de 97% nas áreas de teste e possibilitam identificar quantos metros de falhas são encontrados dentro das linhas.

A Figura 2 demonstra todas essas etapas da metodologia.

**Figura 2 - Metodologia do trabalho de OLIVEIRA et al.**

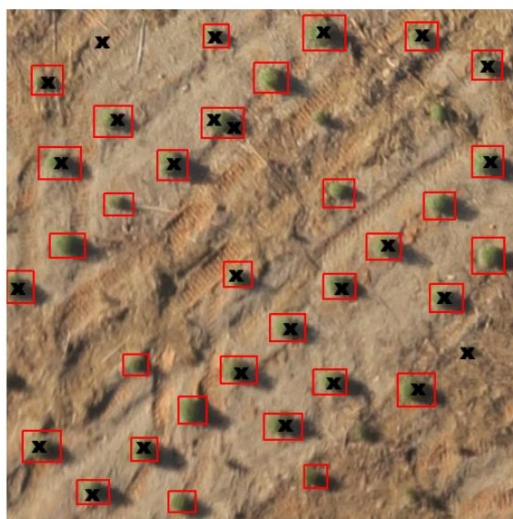


Fonte: Adaptado de (OLIVEIRA et al., 2018).

O trabalho de (SILVA et al., 2018) utiliza de métodos de processamento de imagens juntamente com aprendizagem de máquina para detecção e contagem de plantas em produções de eucalipto. Foram utilizados neste trabalho algumas redes neurais convolucionais, onde o treinamento das mesmas foi realizado com uma base de dados obtida por meio do sobrevoo de uma região de plantação de eucaliptos com um VANT. As diversas redes apresentaram resultados distintos para a detecção das árvores, sendo elas Resnet-101 com 95% de taxa de acerto, SSD Inception V2 com 93% de taxa de acerto e R-CNN Resnet Inception V2 com 95% de taxa de acerto. Os resultados obtidos pelas redes foram promissores com algumas das redes chegando no objetivo esperado da contagem de 95% das plantas. Cada rede apresentou resultados considerados como bons pelo autor, mas diferentes uns dos outros variando em seus tempos de execução e taxas de assertividade, analisando assim cada uma pode ter uma aplicação diferente dependendo da necessidade sobre velocidade ou precisão.

A Figura 3 demonstra um exemplo dessa detecção das plantas de eucalipto.

**Figura 3 - Detecção de plantas de eucalipto em uma imagem aérea.**



Fonte: (SILVA et al., 2018).

A pesquisa de (DOUULLARD, 2018) utiliza um treinamento de uma rede neural convolucional RetinaNet para identificar carros em imagens aéreas de cidades obtidas por satélites. Os resultados obtidos pela pesquisa são mensurados por meio de uma medida denominada de *f1-score* que é a combinação de duas métricas conhecidas *recall* e *precision*. Por meio desta métrica o resultado pode analisar tanto a sua



precisão como as taxas de erros dos falsos positivos e falsos negativos. Para essa métrica de *f1-score* na grande maioria os resultados da detecção dos carros pela pesquisa apresentaram resultados em torno de 95% para imagens de áreas urbanas. Tendo como seu maior problema a detecção de falsos positivos em poços de ventilação no topo de edifícios.

A Figura 4 demonstra uma imagem com pontuação de *f1-score* de 91%, onde a cor verde representa verdadeiros positivos, a cor amarela representa falsos positivos e a cor vermelho representa falsos negativos.

**Figura 4 – Detecção de carros em uma imagem área urbana.**

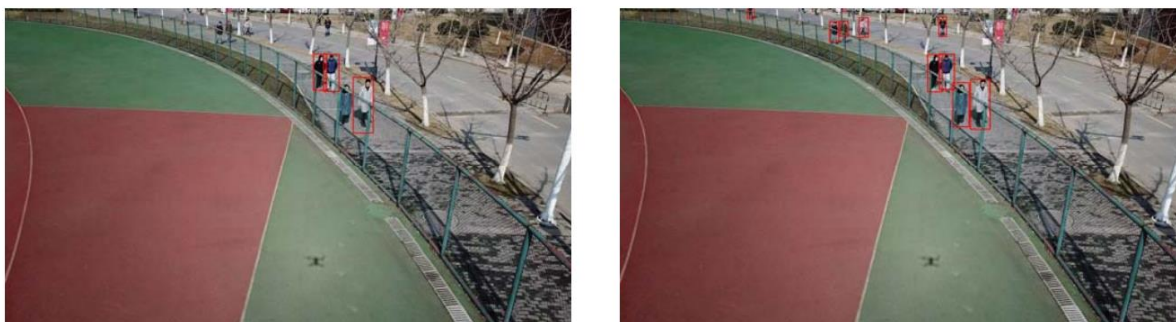


Fonte: (DOUILLARD, 2018).

O trabalho de (XIANG et al., 2019) utiliza do treinamento de redes neurais convolucionais para identificar pedestres em imagens aéreas obtidas por um VANT em ruas de cidades. A rede neural que obteve o melhor resultado nesse trabalho é uma variação da rede RetinaNet que foi modificada pelo próprio autor para se adaptar melhor ao ambiente de estudo. O resultado final da sua rede RetinaNet adaptada chegou a alcançar um resultado de 84,05% na sua média de precisão em seus experimentos de avaliação.

A Figura 5 demonstra a comparação entre a rede RetinaNet Figura 5-a comum e a adaptação feita no trabalho Figura 5-b.

**Figura 5 - Aplicação da rede RetinaNet original e da versão melhorada para detectar pessoas.**



Fonte: (XIANG et al., 2019).

(a)

(b)

O trabalho de (CULMAN; DELALIEUX; VAN TRICHT, 2020) utiliza de um treinamento da rede RetinaNet por meio de uma base de imagens de palmeiras obtidas por satélites, para localizar e identificar palmeiras em florestas, cidades e locais de cultivo. O objetivo da detecção das palmeiras está diretamente ligado ao monitoramento da diversidade vegetativa e a qualidade de saúde das mesmas. Os resultados obtidos pela aplicação da RetinaNet alcançaram bons resultados chegando a obter uma média de precisão de 89,3% no seu conjunto de validação e 76,9% no seu conjunto de avaliação.

A Figura 6 apresenta uma demonstração das detecções das palmeiras.

**Figura 6 - Detecção de palmeiras feitas pela rede RetinaNet.**



Fonte: (CULMAN; DELALIEUX; VAN TRICHT, 2020).

### 3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados conceitos necessários para o desenvolvimento deste estudo, abrangendo os temas de agricultura de precisão, utilização de VANTs, métodos de processamentos de imagens, técnicas de inteligência artificial e aprendizagem de máquina.

#### 3.1 AGRICULTURA DE PRECISÃO

Os processos de análise de características em plantações agrícolas começaram a ser descritos e estudados com maior abrangência a partir da década de 1980 (CAPELLI, 1999). Esses conceitos de analisar e identificar características de solo, histórico de cultivo, clima e outras variáveis do sistema de produção, em diferentes locais dentro do campo foram nomeados de agricultura de precisão (CAPELLI, 1999).

Os métodos mais simples para entender e analisar um cultivo são aqueles que adotam a métrica de tratar as regiões de produção de maneira homogênea. Na prática estes métodos aplicam as quantidades de insumos distribuídas de maneira igualitária em todas as áreas, não analisando necessidades específicas de algumas regiões e gerando como resultado uma lavoura com produtividade não uniforme (CAPELLI, 1999).

A agricultura de precisão vem como uma tentativa de amenizar esse problema permitindo uma melhor categorização das áreas e das suas necessidades específicas. Trabalhando com uso de gerenciamento de informações exatas, fazendo o gerenciamento do campo produtivo de metro a metro, obtendo assim as informações que cada região da fazenda possui propriedades diferentes entre si (ROZA, 2000).

Para Coelho (2005) a agricultura de precisão utiliza muitas tecnologias, porém as mais importantes estão ligadas aos computadores e programas, os computadores têm capacidade de manipular os dados com maior velocidade e organização além de possibilitar a visualização dos mesmos de forma gráfica.

A utilização de computadores aplicados em problemas na agricultura possibilita atividades como, mapeamento de áreas, levantamento de dados de

culturas e amostragem sistematizada de solos. Por meio desses dados coletados, pode-se aplicar soluções computacionais para dispor nos mapas as informações relevantes sobre determinadas áreas, possibilitando ao agricultor um melhor entendimento da sua área, melhorando a sua tomada de decisão (COELHO, 2005).

Segundo Coelho (2005) sistemas computacionais utilizados na agricultura de precisão precisam de aquisição, processamento e análise de uma quantidade de dados muito grande e que variam durante o espaço e o tempo, pois geralmente trabalham com áreas grandes que necessitam dessa demanda computacional para a representação das informações e dados.

Devido às condições do campo onde na maioria dos casos não podemos encontrar conexões a internet e nem fontes de energia elétrica, existe uma necessidade de utilizar sistemas de computação móvel, esses sistemas necessitam alta capacidades de processamento, memória expansiva e capacidade de armazenamento de dados massivos, além de na grande maioria dos casos estarem equipados com sensores para monitorização das culturas e um GNSS, Sistema Global de Navegação por Satélite ou em inglês *Global Navigation Satellite System*, para determinação de coordenadas geográficas (COELHO, 2005) (EMMI et al., 2013).

Juntamente aos avanços nas áreas de tecnologia computacional, desenvolvimento de software, sistemas globais de navegação, sensores de monitoramento e a miniaturização de equipamentos o interesse nos VANTs vem crescendo ao redor do mundo e estão sendo aplicados na agricultura de diversas formas (JORGE; INAMASU, 2014).

### 3.2 VEÍCULOS AÉREOS NÃO TRIPULADOS

Os Veículos Aéreos Não Tripulados chamados popularmente por VANTs ou “drones”, são miniaturas derivadas de aviões (SHIRATSUCHI, 2014). O início de sua produção deu-se na indústria bélica, onde eram utilizados como aeronaves de espionagem norte americanas na guerra fria. Posteriormente na década de 1990 começaram a carregar armas de fogo para serem utilizados em combate (SHIRATSUCHI, 2014).

Inicialmente os VANTs foram criados com intuito militar, mas nos dias atuais ganharam espaço para atuar em diferentes atividades da sociedade, a exemplo do

auxílio no aumento da oferta de alimentos mundial e dos demais produtos provenientes do campo, onde esses aparelhos são utilizados em culturas agrícolas para fazer um monitoramento por registros de imagens dos períodos de plantio, cultivo e colheita de uma produção. Os principais cultivos a utilizar-se dessa prática são as produções de eucalipto e cana-de-açúcar, mas existe uma grande expansão para utilização também nas culturas de soja, milho e algodão (SHIRATSUCHI, 2014) (JORGE; INAMASU, 2014).

A utilização de VANTs na agricultura de precisão qualifica o Brasil como um dos pioneiros, muitos profissionais relacionados à área estão se especializando na utilização desses aparelhos. Em suas aplicações mais simples um VANT possibilita que o seu usuário faça avaliações visuais da evolução do cultivo, verificação de falhas no plantio ou na adubação, identificação de reboleiras, contagem de plantas entre outras atividades. Já utilizando de recursos mais avançados um VANT pode substituir aviões ou satélites na produção de fotos georreferenciadas, o que possibilita uma dimensionalidade das áreas (SHIRATSUCHI, 2014).

As Figuras 7, 8 e 9 representam imagens aéreas de diversos cultivos.

A Figura 7 apresenta um cultivo de cana de açúcar, as plantas são de um porte maior e ficam ligeiramente próximas umas das outras, formando linhas bem delimitadas na imagem o que dificulta a análise das plantas uma a uma.

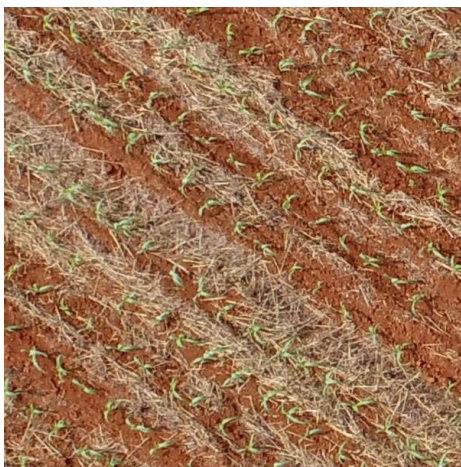
**Figura 7 - Imagem aérea de uma produção de cana de açúcar.**



**Fonte: (SOARES et al., 2017)**

A Figura 8 apresenta um cultivo de milho, as plantas são de um porte menor pois estão em uma fase muito inicial do seu crescimento, devido a isso as plantas podem ser facilmente analisadas uma a uma.

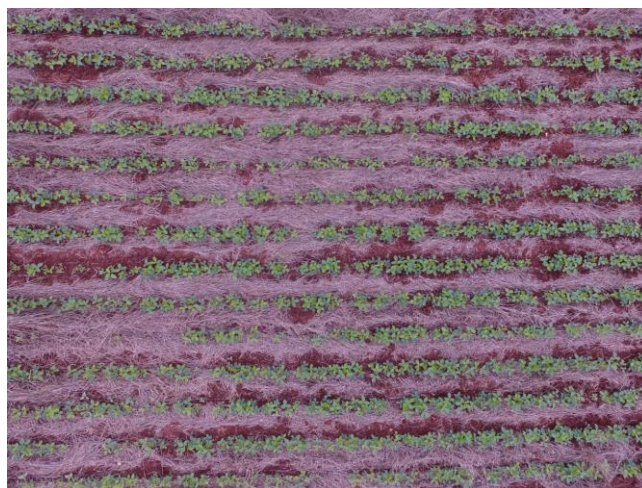
**Figura 8 - Imagem aérea de uma produção de milho.**



**Fonte: autoria própria**

A Figura 9 apresenta um cultivo de soja, as plantas são de um porte médio pois estão em uma fase intermediária do seu crescimento, devido a isso elas já estão mais próximas umas das outras o que dificulta a análise delas uma a uma.

**Figura 9 - Imagem aérea de uma produção de soja.**



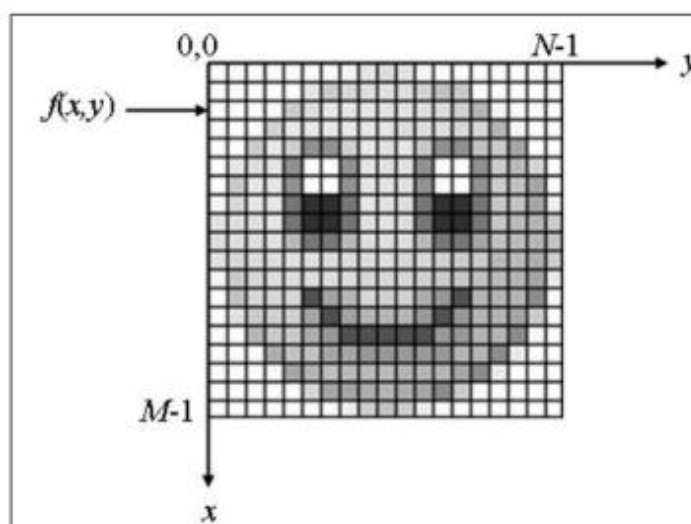
**Fonte: autoria própria**

Este estudo está focado em plantações de milho, desta forma a base de imagens digitais foi construída por um VANT modelo PHANTOM 4 ADVANCED.

### 3.3 IMAGENS DIGITAIS

Uma imagem digital é representada por uma matriz bidimensional  $M \times N$  onde cada célula dessa matriz é denominada de pixel. Na matriz,  $M$  representa o número de linhas e  $N$  o número de colunas. Para representar uma imagem nessa matriz é necessária uma função bidimensional  $f(x, y)$  onde  $x$  e  $y$  são coordenadas espaciais e  $f$  é o valor da amplitude do pixel para essas coordenadas, que pode ser representada de diversas maneiras (GONZALEZ; WOODS, 2000). Essa representação é demonstrada na Figura 10.

**Figura 10 - Matriz de pixels de uma imagem digital.**



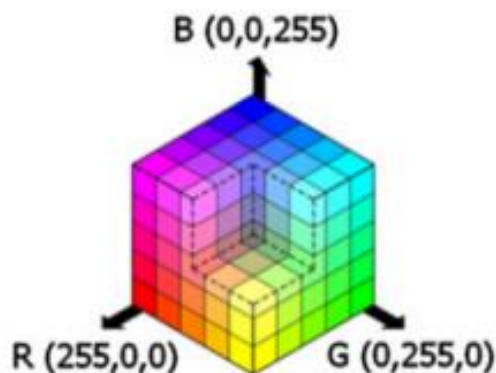
Fonte: (SANCHES, 2009)

Trabalhando com a amplitude dos pixels por escala de cinza podemos representar tons de cinza que geralmente são compostos por 8 bits gerando assim 256 níveis diferentes de amplitude, possibilitando valores no intervalo de 0 cor preta, 255 cor branca e os demais valores intermediários tons de cinza (GONZALEZ; WOODS, 2000).

Trabalhando em imagens com escalas de cores, o padrão mais comum a ser utilizado é o RGB, que consiste na mistura das três cores básicas vermelho (R), verde (G) e azul (B), onde cada uma dessas cores é um elemento de uma matriz tridimensional (KAKUMANU; MAKROGIANNIS, 2007). Esses elementos normalmente são compostos por 8 bits possibilitando uma variação de intensidade de 256 níveis,

variando de 0 ausência total da cor a 255 presença total da cor e a combinação dos 3 elementos resulta em todas as mais de 16 milhões de cores possíveis, como representado na Figura 11 (FEITOSA, 2015).

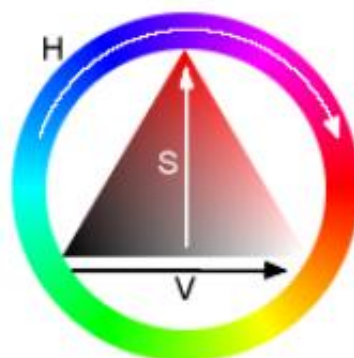
Figura 11 - Sistema tridimensional de mistura de cores do sistema RGB.



Fonte: (FEITOSA, 2015)

Além do padrão RGB outro padrão utilizado para representar imagens coloridas é o HSV, que trabalha com um elemento para cor dominante *Hue* que varia em um espectro angular de 0 a 360 possibilitando a representação de todas as cores do espectro, um elemento para saturação *Saturation* que determina a pureza da cor que varia de 0% a 100% onde quanto menor é a porcentagem, mais acinzentada é a cor e quanto maior mais forte e pura é a cor e um elemento *Value* que também varia de 0% a 100% que determina o brilho daquela cor (FEITOSA, 2015). Esse padrão de cor pode ser representado por meio de um círculo, possuindo as 360 opções de ângulos do canal Hue como demonstrado na Figura 12.

Figura 12 - Espaço de cores HSV.



Fonte: (FEITOSA, 2015)



Processamento de imagem é o processo de analisar essas imagens digitais, para possibilitar tarefas como auxiliar a visão humana melhorando as informações presentes nas imagens e analisar cenas por um computador de forma automática (MARQUES FILHO; NETO, 1999).

As aplicações desses processos nas imagens podem ser utilizados de diversas maneiras, variando desde uma inspeção de qualidade de produtos em uma linha de produção até no entendimento de comportamentos humanos em imagens de um vídeo (BALLARD; BROWN, 1982).

Dentro das aplicações existem alguns campos específicos onde se pode trabalhar de maneiras diferentes com imagens tais como: reconhecimentos, análise de movimento, reconstrução 3D, restauração de imagens (SILVA et al., 2018).

Nesse trabalho a aplicação principal será a área de reconhecimento, que trabalha no âmbito de encontrar padrões dentro das imagens por meio de cores, formatos de objetos e outras informações (SILVA et al., 2018). Esses padrões podem gerar informações relevantes como no caso deste trabalho, encontrar os padrões do que é uma planta e identificá-la em uma imagem, pode ser utilizado para fazer a contagem das mesmas dentro de uma imagem com várias plantas.

Nas seções a seguir serão apresentados alguns métodos e técnicas de processamento de imagem que serão utilizados no desenvolvimento deste trabalho.

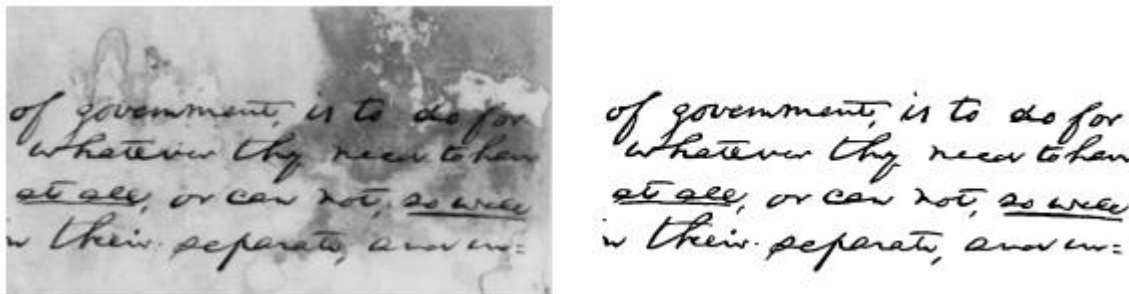
### 3.3.1 Binarização

Uma imagem binária é uma imagem que possui apenas pixels que representam as cores preto e branco onde só é necessário 1 bit para representar o grau de amplitude do pixel, 0 cor preta e 1 cor branca (IKEDA, 2011).

A binarização ou também chamada de limiarização é um método de transformar uma imagem em escala de cinza em uma imagem binária. Existem diversos métodos de transformar o valor desses pixels de escala de cinza para binários 0 e 1. Muitos métodos utilizam de um valor de intensidade chamado de limiar, o limiar é o valor que delimita que os valores de cinza acima dele serão definidos como 1 e os abaixo dele definidos como 0, aplicando essa regra em uma matriz de uma imagem em escala de cinza, transforma a mesma em uma matriz com células binárias 0 e 1 (IKEDA, 2011).

A Figura 13 demonstra esse processo de binarização, onde a Figura 13-a está em escala de cinza e a Figura 13-b está binarizada.

Figura 13 - Binarização de uma imagem.



Fonte: (IKEDA, 2011)

(a)

(b)

O processo de binarização pode ser implementado por diversos métodos, um dos mais comumente utilizadas é o método Otsu.

Esse método supõe que uma imagem pode ser dividida em duas classes pelas intensidades dos pixels. Para isso o método calcula um limiar  $k$  por meio da frequência que cada intensidade aparece em um pixel da imagem e por meio desse limiar os pixels são separados em antes de  $k$  e depois de  $k$  (OTSU, 1979).

Para calcular o limiar o método Otsu trabalha com o teste de todos os possíveis valores para  $k$  calculando uma nota para o determinado limiar por meio das variâncias (OTSU, 1979).

A fórmula para esse cálculo é demonstrada na Equação 1, onde  $\sigma_G^2$  representa a variância global e  $\sigma_B^2(k)$  a variância entre as classes que estão separadas pelo determinado limiar  $k$ . Após aplicar todos os testes o valor de  $k$  que apresentou o maior resultado é escolhido como o melhor limiar (OTSU, 1979).

$$n(k) = \frac{\sigma_B^2(k)}{\sigma_G^2} \quad (1)$$

A Equação 2 demonstra a fórmula para o cálculo da variância onde  $\mu$  é a média da intensidade da imagem e  $M$  e  $N$  representam a altura e largura da imagem.

$$\sigma^2 = \frac{1}{MN} \sum_{x=0}^N \sum_{y=0}^M (f(x, y) - \mu)^2 \quad (2)$$

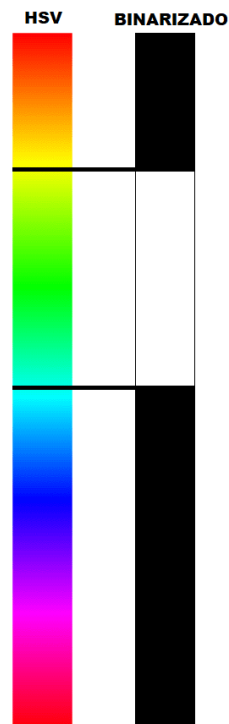
O método utilizado para a binarização nesse trabalho diferente do método de Otsu utiliza um intervalo de intensidade determinado e fixo, que não precisa ser calculado para aplicação em cada imagem, pois a cor verde sempre vai estar delimitada da mesma maneira dentro da medida de intensidade de cor.

Trabalhando no canal Hue do modelo de cor HSV delimitar o que é verde é uma tarefa muito simples, é necessário somente determinar um limiar inicial e um limiar final para o intervalo que representa essa cor.

Posteriormente para aplicar a binarização por esse método, é necessário separar o que foi selecionado como verde do restante da imagem deixando somente as informações que apresentam essa cor.

A Figura 14 demonstra graficamente como é esse intervalo escolhido para binarizar a imagem.

**Figura 14 - Binarização da cor verde pelo canal HSV**



Fonte: autoria própria.

### 3.3.2 Componentes Conexos

A rotulação de componentes conexos de uma imagem é o processo de atribuir um valor único para os pixels que pertencem a uma mesma região conexa, esse processo só pode ser aplicado após ser feito a binarização da mesma imagem, para que assim todos os pixels vizinhos entre si pertençam ao mesmo objeto (DO MONTE LIMA, 2008).

Segundo Do Monte Lima (2008) o processo do algoritmo de rotulação trabalha escaneando a imagem binária a partir do pixel do topo superior esquerdo, analisando os vizinhos de cada um dos pixels e categorizando por meio das seguintes instruções:

- Se nenhum vizinho pertence a um rótulo de objeto, é criado um novo rótulo e é atribuído ao pixel atual.
- Se existem vizinhos que pertencem a apenas um rótulo de objeto, é atribuído o valor do rótulo dos vizinhos ao pixel atual.
- Se existem vizinhos que pertencem a mais que um rótulo de objeto, é atribuído o valor de um dos rótulos dos vizinhos ao pixel atual e é criada uma equivalência entre os diferentes rótulos.

Posteriormente ao escaneamento os rótulos equivalentes são agrupados em classes de equivalência e é feito um segundo escaneamento para resolver essas equivalências e definir os componentes conexos finais (DO MONTE LIMA, 2008).

A Figura 15 demonstra a imagem binarizada Figura 15-a e o resultado da rotulação Figura 15-b, onde cada cor representa um rótulo.

**Figura 15 - Rotulação de componentes conexos**



Fonte: (DO MONTE LIMA, 2008)

(a)

(b)

Sendo essa uma das técnicas aplicadas na primeira metodologia desenvolvida.

### 3.4 INTELIGÊNCIA ARTIFICIAL

O conceito básico de inteligência é a capacidade de um indivíduo compreender o mundo ao seu redor e aplicar um conhecimento prévio para resolver problemas que lhe forem apresentados. Assim inteligência consiste nas capacidades de perceber, entender e analisar informações para uma tomada de decisão, aprendendo com suas decisões passadas (SILVA et al., 2018).

Já inteligência artificial são aplicações computacionais que tem a capacidade de analisar e desempenhar tarefas que são consideradas inteligentes em casos de execução das mesmas por humanos (AKERKAR, 2014).

Segundo Ertel (2011) a abrangência do campo de aplicação da inteligência artificial varia dentre as mais diversas aplicações do cotidiano de um ser humano e algumas das características que compõe suas definições básicas são:

- Capacidade de raciocínio: aplicar regras lógicas a um conjunto de dados disponíveis para chegar a uma conclusão.
- Aprendizagem: aprender com os erros e acertos para maximizar os futuros resultados.
- Reconhecer padrões: detectar e classificar objetos, texturas e padrões de comportamento.
- Inferência: capacidade de aplicar raciocínio em situações cotidianas.

#### 3.4.1 Aprendizagem de Máquina

Aprendizagem de máquina é um subcampo da inteligência artificial que desenvolve programas que possibilitam o computador a resolver problemas sem a explicação lógica do problema feita pela programação (SILVA et al., 2018).

Existem diferentes técnicas para fazer o computador aprender a interagir com o ambiente em que ele atua e de acordo com como é feita essa interação para a descoberta da informação inteligente podemos classificar em três tipos de

aprendizagens: supervisionada, não supervisionada e por reforço (RUSSEL et al., 2013).

Aprendizado supervisionado é aquele que aprende por meio de um supervisor que informa ao programa o quão bem ele está atuando, essa técnica é aplicada quando existe alguma maneira de validar a saída final do programa (RUSSEL et al., 2013). Usualmente é apresentado na entrada um conjunto de treinamento e os resultados esperados, com base na diferença dos resultados obtidos e os esperados o programa verifica o que pode ser feito para corrigir o seu funcionamento (RUSSEL et al., 2013).

Aprendizado não supervisionado é aquele que não possui as informações sobre os resultados esperados, não existe uma maneira de validar a sua saída. Nesse caso o programa tenta aprender por meio de relações entre os dados, conjuntos similares de agrupamentos ou prevendo resultados futuros por meio de informações passadas (RUSSEL et al., 2013).

Aprendizado por reforço é aquele que interage com o ambiente de forma dinâmica e recebe uma pontuação de atuação baseado em uma função de avaliação, a cada interação o programa ajusta seus pesos e vai melhorando a sua pontuação. Esse tipo de aprendizagem pode atuar em ambientes desconhecidos e por meio de repetitivos processos de tentativa e erro é gerado um conhecimento para se adaptar para aquele ambiente (RUSSEL et al., 2013).

Nesse trabalho será aplicado um aprendizado supervisionado, por meio do treinamento de uma rede neural para detecção de plantas de milho em imagens digitais.

### 3.4.2 Redes Neurais Convolucionais

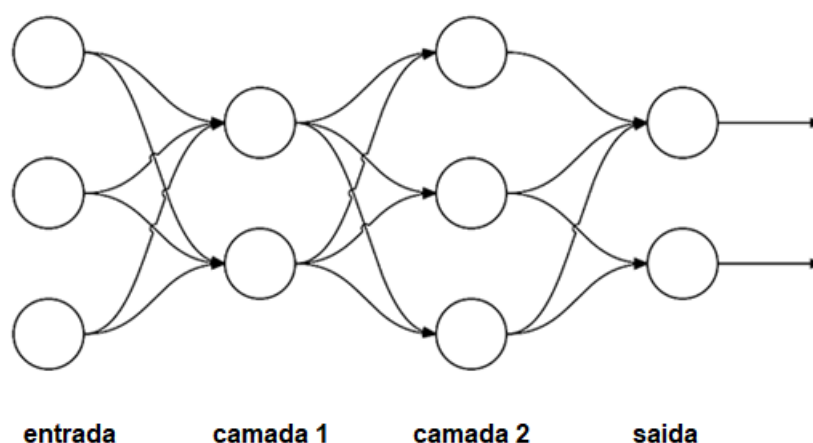
Redes Neurais Convolucionais ou em inglês *Convolutional Neural Networks*, são um modelo matemático de classificação criado com inspiração no sistema biológico do funcionamento de neurônios. Essas redes apresentam resultados promissores em aplicações nas áreas de classificações e detecção de objetos em imagens digitais (GU, 2018).

Segundo Lecuns et al. (2015) as CNN's são redes multicamadas projetadas para o reconhecimento de padrões por meio da análise dos pixels sem a necessidade

de um pré-processamento complexo. Normalmente para reconhecer esses padrões um algoritmo precisaria ser implementado com as informações sobre o que categoriza esses padrões de forma manual, mas em uma CNN os neurônios da rede se adaptam a cada interação do treinamento, aprendendo assim de forma automatizada as informações necessárias para encontrar esses padrões. O que torna essa tecnologia extremamente vantajosa na sua aplicação são a independência de um conhecimento específico e a independência do esforço humano no desenvolvimento de suas funcionalidades básicas.

Na Figura 16 pode-se observar a estrutura de camadas de uma rede, que trabalha em etapas até chegar em um resultado final.

**Figura 16 - Estrutura de uma Rede Neural com múltiplas camadas.**



Fonte: Adaptado de (ROSEBROCK, 2016)

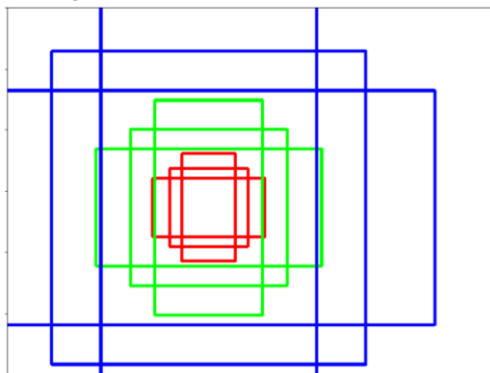
Redes de detecção CNN trabalham com âncoras que são um conjunto de caixas delimitadoras com altura e largura pré-definidas, essas são dispostas de diversas maneiras sobre a imagem da detecção e todas passam por um processo de classificação (LIN et al., 2017).

Dependendo do resultado que é obtido na classificação ela é categorizada ou não como um objeto da imagem, posteriormente são feitos ajustes nas dimensões e posicionamento das caixas que foram definidas como objetos, para verificar se a classificação melhora por meio desses ajustes, melhorando assim a dimensão e a posição final das detecções (LIN et al., 2017).

O uso dessas âncoras permite a rede fazer detecções rápidas e precisas para um grande número de regiões, a Figura 17 demonstra diversas âncoras de tamanhos

e dimensões diferentes sobrepostas em regiões do que seria uma possível imagem (MARTINS, 2019).

**Figura 17 - Exemplos de âncoras posicionadas em torno do centro da imagem.**



**Fonte: (MARTINS, 2019).**

A rede CNN escolhida para realização deste trabalho foi a rede RetinaNet. Ela é uma rede CNN de um estágio que trabalha com uma função de *Loss* chamada de perda focal, os detalhes dessa rede são apresentados na seção 3.4.5.

### 3.4.3 Redes de Dois e Um Estágio

Redes CNN de detecção podem ser classificadas em dois diferentes tipos as que utilizam dois estágios e as que utilizam um estágio para detecção (MARTINS, 2019).

Durante muito tempo as redes de detecção em dois estágios eram a tecnologia que apresentavam resultados com melhor assertividade de detecções disponíveis. Essas redes trabalham de forma onde no primeiro estágio é feito uma procura de possíveis regiões de interesse e no segundo cada uma dessas possíveis regiões é recategorizada por meio de uma rede de classificação mais precisa (MARTINS, 2019).

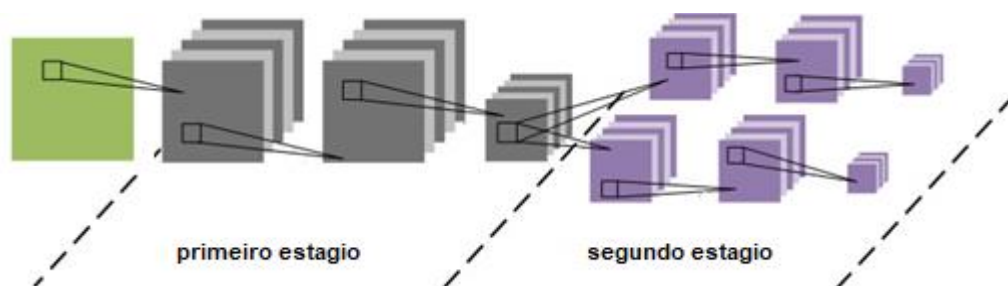
Redes de um estágio são redes que também apresentam resultados promissores, mas na maioria de suas aplicações elas possuem com complexidade computacional muito menor, mas apresentam mais erros em comparação às redes com dois estágios, essa diferença ocorre devido a essas redes não possuírem uma segunda classificação para validar a classe (SOVIANY; IONESCU, 2018). Devido a esses fatores de complexidade computacional redes de um estágio são usadas



principalmente em problemas em tempo real onde a velocidade de detecção é importantíssima (SOVIANY; IONESCU, 2018).

A Figura 18 demonstra a delimitação de cada um dos estágios de uma CNN de dois estágios.

**Figura 18 - Exemplo da estrutura de uma CNN de dois estágios.**



Fonte: Adaptado de (YU et al., 2014).

Analisando a Figura 18 é possível verificar como o segundo estágio aumenta a quantidade de processos necessários para chegar no resultado final.

#### 3.4.4 Loss

A função de *Loss* ou no português perda é uma medida de análise utilizada durante o treinamento de redes para analisar como as âncoras estão sendo classificadas, quanto menor o valor de perda melhor estão sendo a classificação das âncoras.

A função de perda pode ser implementada de diversas maneiras diferentes, como exemplo podemos citar as funções de entropia cruzada binária e entropia cruzada categórica (KETKAR; SANTANA, 2017).

A rede RetinaNet implementa a sua função para calcular a perda pela função de perda focal demonstrada nas Equações 3 e 4 da secção 3.4.5.

#### 3.4.5 RetinaNet

RetinaNet é uma rede de detecção de um estágio, mas diferente da grande maioria das redes de um estágio ela busca obter um grande desempenho na sua taxa

de assertividade de detecções semelhante às redes de dois estágios, mas tentando ainda manter uma complexidade computacional menor (MARTINS, 2019).

A grande diferença dessa rede é a utilização de uma função de perda focal que tem como objetivo diminuir o peso do desbalanceamento entre as âncoras dos objetos de interesse e as do fundo durante o treinamento da rede, essa função trabalha por meio de um fator exponencial que diminui a perda de âncoras que já foram bem classificadas (LIN et al., 2017).

Para uma âncora com classificação  $y$  e predição  $p$ , perda focal é definida Conforme as Equações 3 e 4:

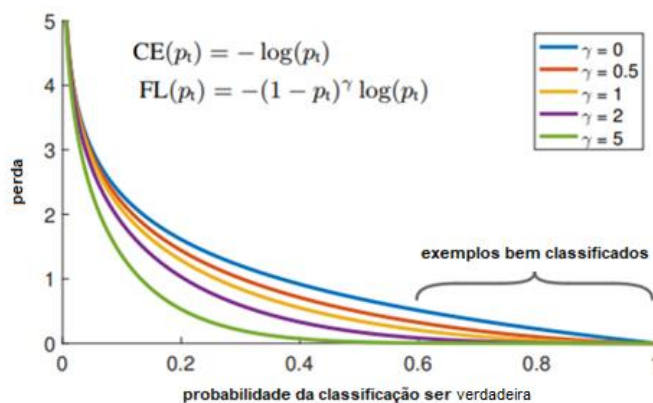
$$pt = \begin{cases} p, & \text{se } y = 1 \\ 1 - p, & \text{se } y = 0 \end{cases} \quad (3)$$

$$FL(pt) = -(1 - pt)^\gamma \log(pt) \quad (4)$$

Analisando a Equação 3 pode-se verificar que com um fator exponencial igual a zero a função de perda se equivale a função de perda normalmente utilizada chamada de perda entropia cruzada, mas aumentando o valor do exponencial essa perda vai diminuindo (LIN et al., 2017).

Os gráficos da Figura 19 demonstram como esse fator exponencial influencia na perda, onde o eixo X demonstra a probabilidade de uma classificação ser verdadeira e o eixo Y o valor de perda relativo.

**Figura 19 - Perda Focal para diferentes valores de  $\gamma$ .**



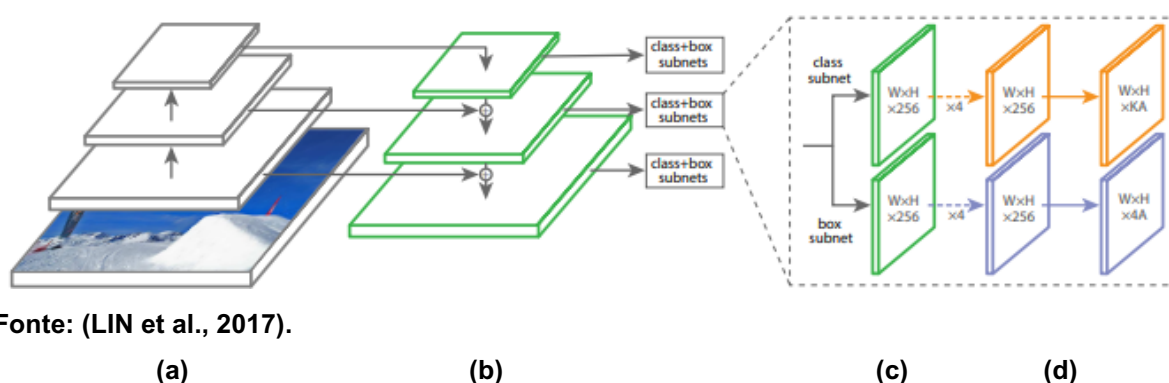
Fonte: (LIN et al., 2017).

Segundo Lin et al. (2017) a rede RetinaNet é composta por 4 componentes principais sendo eles:

- *Backbone*: É uma rede convolucional *feedforward*, na qual todos os neurônios de uma camada são conectados com todas as entradas da camada. Essa é responsável por processar a imagem original, aprendendo a retirar os descritores necessários para descrever a imagem.
- *Feature Pyramid Network (FPN)*: Componente responsável por aumentar a capacidade da rede de trabalhar com objetos em escalas diferentes.
- Sub-rede de classificação: Componente que utiliza as ativações de cada nível do FPN para classificar cada uma das âncoras correspondentes entre as possíveis classes. A mesma sub-rede de classificação é utilizada para os diferentes níveis da FPN, onde cada nível corresponde à um conjunto de âncoras com tamanho específico.
- Sub-rede de refinamento: realiza uma função similar à sub-rede de classificação, porém o resultado é a regressão de um vetor de 4 elementos. Cada uma das âncoras recebe 4 coordenadas de refinamento para realocar suas posições.

A Figura 20-a está representado o *Backbone*, a Figura 20-b está representado FPN, a Figura 20-c está representado a Sub-rede de classificação e a Figura 20-d está representado a Sub-rede de refinamento.

**Figura 20 - Componentes principais de uma rede RetinaNet**



### 3.5 MÉTRICAS DE AVALIAÇÃO

Após aplicação de uma detecção de objetos em uma imagem essas detecções podem ser analisadas para poder validar se sua aplicação está identificando corretamente os objetos desejados na detecção. Devido a isso foram criadas várias métricas de avaliação para analisar a qualidade das detecções encontradas em uma imagem.

#### 3.5.1 Mean Average Precision

Uma medida comumente utilizada para analisar desempenho do treinamento de uma CNN é a função de mAP do inglês *Mean Average Precision*, que indica a média da precisão das classificações entre todas as imagens da base de dados do treinamento (MANNING; SCHÜTZE; RAGHAVAN, 2008).

A fórmula para calcular essa média está demonstrada na Equação 5, onde  $Q$  é o tamanho total da nossa base de dados de avaliação,  $m_j$  é o número total de indivíduos verdadeiros em uma determinada imagem e  $P(R_{jk})$  é a precisão do indivíduo na imagem.

$$mAP(Q) = \frac{1}{Q} \sum_{j=1}^Q \frac{1}{m_j} \sum_{k=1}^{m_j} P(R_{jk}) \quad (5)$$

A precisão do indivíduo na imagem é uma medida de avaliação para a classe que lhe foi atribuída. Esta é calculada diretamente no resultado da CNN (SILVA et al., 2018).

#### 3.5.2 Contagem de Indivíduos

Métricas como mAP que trabalham diretamente com a precisão geral das detecções, nem sempre apresentam resultados que demonstram a real qualidade da rede para a aplicação que lhe foi proposta (SILVA et al., 2018). Devido a esse fator além da métrica de mAP foram utilizadas outras métricas que estão diretamente ligadas a contagem das detecções de plantas.

Comparando os resultados da rede diretamente com o real obtém-se uma informação de avaliação muito mais concisa e confiável para a aplicação real da rede (MATOS et al., 2009). Uma maneira de fazer essa comparação é analisando a distância entre o ponto central de cada caixa da detecção com cada um dos pontos reais demarcados manualmente na base de teste e caso essa distância entre dois desses pontos seja abaixo de um limiar considerado pequeno, os mesmos são selecionados e marcados como correspondentes.

Para chegar em uma medida de precisão algumas informações tem que ser preestabelecidas como por exemplo:

- Verdadeiro Positivo: Quando uma planta é detectada corretamente.
- Falso Negativo: Quando uma planta existe na base de teste, mas não é detectada pela rede.
- Falso Positivo: Quando uma planta é detectada pela rede, mas não existe na base de teste.

Tendo em mão todas essas variáveis é possível calcular as medidas de precisão e revocação da detecção, como proposto no trabalho de (MATOS et al., 2009). Onde  $VP$  que representa número verdadeiros positivos,  $FP$  representa o número de falsos positivos e  $FN$  representa o número de falsos negativos.

A fórmula da precisão adaptada de (MATOS et al., 2009) é demonstrada na Equação 6, essa fórmula é utilizada para analisar a quantidade de falsos positivos.

$$precisão = \frac{VP}{VP + FP} \quad (6)$$

A fórmula da revocação adaptada de (MATOS et al., 2009) é demonstrada na Equação 7, essa fórmula é utilizada para analisar a quantidade de falsos negativos.

$$revocação = \frac{VP}{VP + FN} \quad (7)$$

A seguir serão apresentadas as metodologias utilizadas no desenvolvimento deste trabalho.

## 4 METODOLOGIA

Este trabalho tem como objetivo criar um detector automático de plantas para conseguir quantificar as mesmas dentro de um cultivo. Foram estudadas duas metodologias diferentes para analisar as imagens. A metodologia inicial trabalha somente com técnicas de processamento de imagem e será apresentada na seção 4.1. A segunda metodologia trabalha com aprendizagem de máquina utilizando as Redes Neurais Convolucionais conforme seção 4.2.

A primeira etapa de ambas metodologias consiste na utilização de um VANT para obter as imagens do cultivo (seção 4.3), posteriormente a detecção das plantas na imagem são realizadas de formas distintas em cada uma das metodologias aplicadas. E ao final são verificadas as quantidades de plantas detectadas.

### 4.1 OBTENÇÃO DAS IMAGENS

As imagens que compõem a base de dados deste trabalho foram obtidas por autoria própria utilizando um VANT de fotografia profissional para mapeamento. Este método de mapeamento possibilita imagens de maior resolução que os métodos tradicionais de sensoriamento remoto, como imagens de satélite. O VANT utilizado foi o modelo PHANTOM 4 ADVANCED, o qual é mostrado na Figura 21.

**Figura 21 - VANT utilizado no mapeamento.**



Fonte: autoria própria.

As fotos foram tiradas com uma câmera embarcada de 20MP a uma altura de 55 metros em um período entre duas a três semanas após o início do plantio, a Figura 22 é um exemplo de uma das imagens obtidas.

**Figura 22 - Foto tirada da plantação com o VANT a 55 metros de altura**



**Fonte: autoria própria.**

A escolha da altura do VANT para obter as imagens foi determinada para viabilizar a aplicação dessa técnica em um meio comercial, tentando obter um meio termo entre qualidade e viabilidade. Quanto maior a altura em que o VANT obtém as imagens, maior é a área de abrangência e menor é a qualidade.

Já a escolha do período foi relacionada ao tamanho da planta, onde entre 2 a 3 semanas após o início do plantio, as plantas já estavam com tamanhos consideravelmente bons para serem identificadas e ainda não estavam grandes o suficiente para sobrepor umas as outras.

A partir da obtenção das imagens foram utilizadas as duas metodologias para realizar a análise das imagens e detecção das plantas de milho.

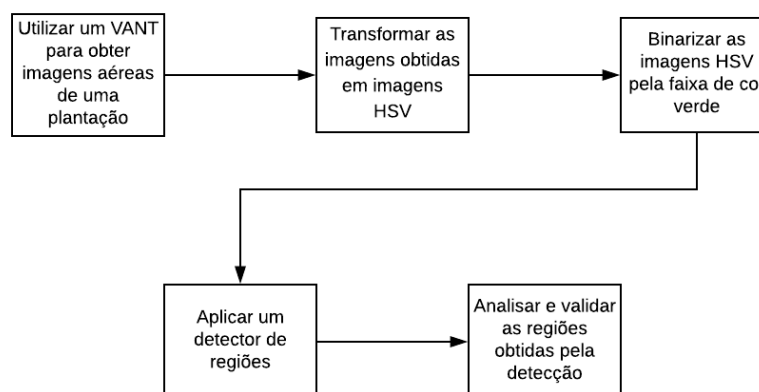
#### 4.2 ABORDAGEM UTILIZANDO TÉCNICAS DE PROCESSAMENTO DE IMAGEM

Para a execução do algoritmo de detecção por meio da análise de espectro de cores e binarização foi utilizado a linguagem de programação Python juntamente a biblioteca de visão computacional OpenCV. A biblioteca OpenCV possibilitou uma manipulação extremamente ampla das imagens devido à facilidade na utilização de

métodos já implementados pela biblioteca, além de possuir uma grande comunidade para auxiliar com suporte em fóruns de debate.

A Figura 23 demonstra o passo a passo dessa metodologia.

**Figura 23 - Etapas de metodologia para detecção com uso técnicas de processamento de imagem.**



Fonte: autoria própria.

#### 4.2.1 Transformação para HSV e Binarização

O canal HSV trabalha com uma componente de cor *Hue* que representa o espectrograma de cores por meio dos ângulos, assim separando facilmente uma cor da outra por intervalos dentro desses ângulos. Além da facilidade de separar cores o canal HSV trabalha com a componente de iluminação separada da cor na componente chamada de *Value*, possibilitando analisar a imagem para detectar cores específicas sem analisar o brilho da imagem, tentando minimizar o impacto que a iluminação do dia pode gerar na cor da imagem.

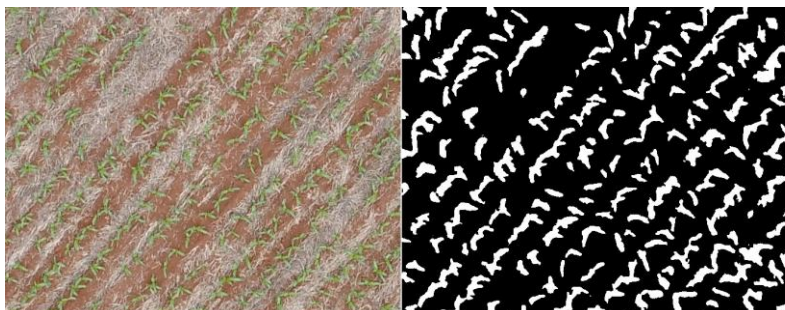
Na aplicação deste trabalho faz-se necessário encontrar na imagem os elementos que representam uma planta e separá-los do resto da imagem. Para encontrar o aspecto da cor verde foi analisado o intervalo da componente *Hue* na biblioteca OpenCV de 20 a 60, sem levar em consideração os elementos de *Saturation* e *Value*.

Uma análise completa de uma imagem em todos os pixels é realizada, e categoriza-se aqueles que estão no intervalo da cor verde como uma planta, foi gerada uma imagem binária como apresentado na Figura 24-b que é a binarização feita pela



delimitação da cor verde da Figura 24-a. Pode se analisar também que a cor branca representa as áreas onde se encontram plantas e o preto o fundo da imagem.

**Figura 24 - Binarização das plantas.**



Fonte: autoria própria.

(a)

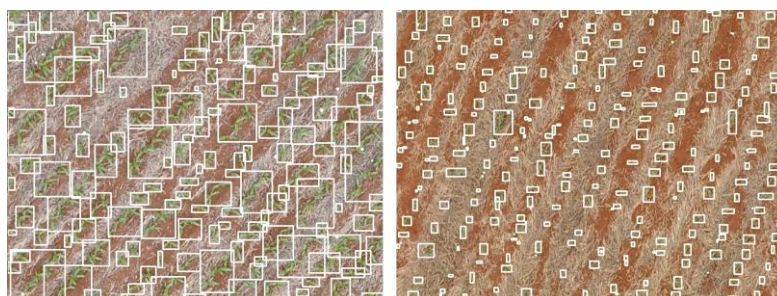
(b)

Como visto nas Figuras 24-a e 24-b o resultado apresentado separou bem os pixels das plantas do restante da imagem, mas também apresentou pequenos pontos falhos que interferem no resultado dependendo do tipo da imagem que for apresentada para a binarização.

#### 4.2.2 Detecção de Regiões

Após separar as áreas que representam plantas, a próxima etapa é delimitar e separar cada uma das plantas para poder fazer a contagem das mesmas. Aplicando um detector de regiões é possível delimitar as áreas de cada uma das representações da cor branca, separando pelas regiões que os pixels e seus vizinhos ocupam como é demonstrado nas Figuras 25-a e 25-b.

**Figura 25 - Regiões detectadas nas imagens binárias.**



Fonte: autoria própria.

(a)

(b)

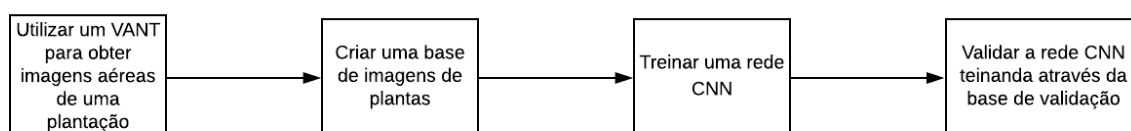
Os resultados obtidos nas Figuras 25-a e 25-b demonstram como o tamanho das plantas influencia diretamente nas detecções e que a aplicação dessa metodologia ainda depende de muitas variáveis que na grande maioria das vezes não se tem controle, como o próprio tamanho das plantas e o distanciamento entre elas, que acabam interferindo diretamente nas regiões que serão detectadas.

### 4.3 ABORDAGEM COM USO DE CNN

Devido aos resultados ruins obtidos na metodologia utilizando apenas técnicas de processamento de imagem (seção 4.1) foi proposta uma segunda metodologia que trabalha com aprendizagem de máquina treinando redes neurais convolucionais CNN. Para essa abordagem foi utilizado a linguagem de programação Python juntamente às bibliotecas Keras e TensorFlow e seu desenvolvimento foi feito por meio da ferramenta de processamento em nuvem do *Google Colab*, que disponibiliza máquinas equipadas com os requisitos necessários para o treinamento de redes CNN.

A figura 26 demonstra o passo a passo dessa metodologia.

**Figura 26 - Etapas de metodologia para detecção com uso de CNN.**



**Fonte: autoria própria.**

#### 4.3.1 Criação da Base de Imagens

Toda aplicação que trabalha com aprendizagem de máquina, utilizando redes neurais geralmente trabalha com três conjuntos de dados, um para treinamento, um para validação e um para teste. Na aplicação deste trabalho o conjunto de validação e treinamento estão sendo utilizados juntos. Devido a forma de como a aplicação trabalha com a entrada dos dados esses conjuntos não precisam ser separados,

gerando assim a necessidade de separar manualmente somente dois conjuntos um para o treinamento e outro para teste.

Para que um treinamento de uma rede tenha bons resultados é necessário amostras suficientes que representem uma boa parte dos casos presentes em uma análise. Devido a isso para a base de treinamento foram utilizadas imagens com variações diversas, como exemplos: variações nos tamanhos das plantas, iluminação, fundo, distanciamento entre as plantas e quantidades diferentes de plantas por imagem.

Para separar cada uma das plantas contidas nas imagens originais da plantação, foi utilizada a ferramenta de código livre LabelImg, disponível no repositório (GITHUB, 2018), que possibilitou rotular todas as plantas contidas nas imagens tiradas pelo VANT.

A interface desta ferramenta é demonstrada na Figura 27.

**Figura 27 - Ferramenta LabelImg.**



**Fonte: autoria própria.**

Para separação foram selecionadas cada uma das plantas delimitando sua área total e as categorizando dando a elas o rótulo da classe (Planta). Todas as informações dos rótulos de cada uma das imagens foram salvas no formato Pascal VOC, onde cada delimitação de uma planta é salva em uma lista de rótulos dentro de um arquivo no formato XML como demonstrado na Figura 28.

**Figura 28 - Formatação dos rótulos salvos no formato Pascal VOC.**

```

<annotation>
  <folder>TCC</folder>
  <filename>DJI_0011_0092.jpg</filename>
  <path>D:\labelImg-master\TCC\DJI_0011_0092.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>695</width>
    <height>541</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>Planta</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>288</xmin>
      <ymin>428</ymin>
      <xmax>317</xmax>
      <ymax>440</ymax>
    </bndbox>
  </object>
  <object>
    <name>Planta</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>227</xmin>
      <ymin>191</ymin>
      <xmax>251</xmax>
      <ymax>214</ymax>
    </bndbox>
  </object>
</annotation>

```

**Fonte: autoria própria.**

Para a base de treinamento e validação foram delimitados 3129 rótulos de plantas dentro de 7 imagens aproximadamente 80% do total de rótulos. Já para a base de teste foram delimitados 757 rótulos dentro de 4 imagens aproximadamente 20% do total de rótulos, que serão utilizados para fazer os testes finais dos resultados obtidos pela rede treinada.

#### 4.3.2 Seleção da CNN

A rede CNN escolhida para realização deste trabalho foi a rede RetinaNet, a escolha dessa rede deu-se com base em pesquisas sobre trabalhos, artigos e outras publicações que tratavam sobre detecção de objetos de dimensões pequenas em imagens obtidas de fontes aéreas. Como exemplos cita-se o estudo apresentado por (XIANG et al., 2019) que trabalhou com a rede RetinaNet utilizando VANTs para detectar pessoas andando nas ruas. O trabalho apresentado por (DOUILLARD, 2018) trabalha com rede RetinaNet utilizando imagens de visões aéreas de cidades para detectar carros circulando e estacionados na cidade. O trabalho de (CULMAN; DELALIEUX; VAN TRICHT, 2020) que utiliza rede RetinaNet para detectar palmeiras em imagens aéreas.

Todas as publicações citadas são pesquisas recentes, entre os anos de 2018 a 2020 e obtiveram bons resultados trabalhando com a rede RetinaNet em imagens de observação aérea.

#### 4.3.3 Treinamento da Rede

O treinamento e configurações da rede CNN foram feitos na plataforma do *Google Colab* utilizando como base do código para treinamento em Keras da rede RetinaNet implementada por Hans Gaiser disponível em repositório no *GitHub* (GAISER, 2017).

Para a realização dos experimentos as configurações da rede não foram alteradas, foram modificados somente os parâmetros referentes ao número de épocas de treinamento.

O treinamento de uma CNN necessita de hardware com alto desempenho devido à grande complexidade dos modelos utilizados. Atualmente o método mais eficiente para treinar redes neurais é utilizando placas de vídeo, que possuem centenas de unidades de processamento e aceleram o processo de aprendizado de redes com diversas camadas e coeficientes, além de suprirem também a necessidade de altas capacidades de memória para realizar os treinamentos e avaliações (SILVA et al., 2018).

Devido a necessidade de um hardware com um bom desempenho para realizar o treinamento de uma rede CNN, optou-se em realizar o treinamento das redes em máquinas remotas em nuvem utilizando a plataforma do *Google Colab* que disponibiliza máquinas específicas para esses treinamentos equipadas com placas de vídeo NVIDIA Tesla.

O sistema utilizado pelo *Google Colab* é um Linux e foram instaladas as seguintes ferramentas:

- Python 3.6.9
- TensorFlow 2.2.0
- Keras 2.3.1
- NVIDIA CUDA 10.1

#### 4.3.4 Configuração e Entradas

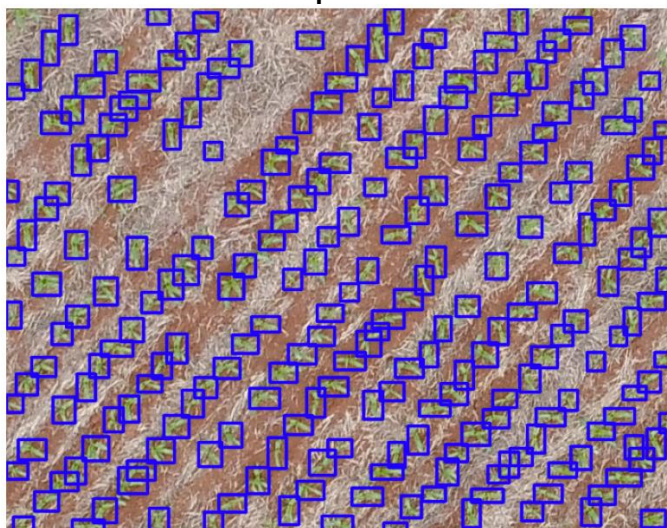
Foram realizados experimentos com configurações de rede diferentes e entrada de dados distintas. Tais experimentos são apresentados a seguir.

##### 4.3.4.1 Experimento 1

O primeiro treinamento realizado neste trabalho utilizou uma base de imagens inicial que continha somente 1613 rótulos e foi treinada durante 10 épocas.

A Figura 29 demonstra os resultados obtidos nesse experimento em uma das imagens da base de teste.

**Figura 29 - Aplicação da rede RetinaNet no experimento 1.**



**Fonte: autoria própria.**

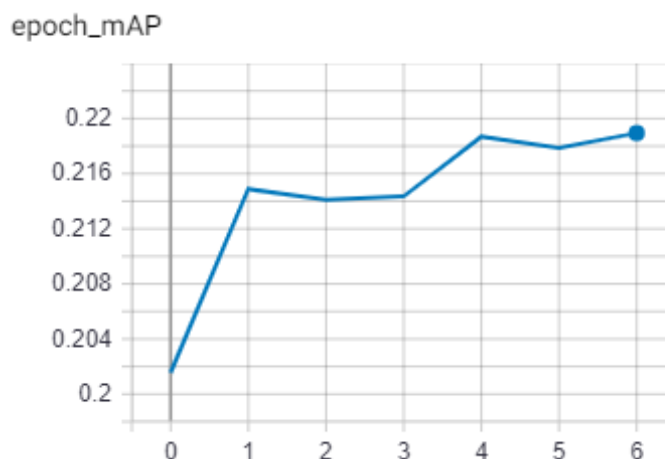
O treinamento desta rede já apresentou um bom resultado com taxa média de precisão de 97,0% e revocação de 97,2%. Também possibilitou demonstrar que era possível obter as informações necessárias para a contagem das plantas.

##### 4.3.4.2 Experimento 2

O segundo treinamento realizado neste trabalho utilizou a base de imagens já finalizada que continha todos os 3129 rótulos e foi aplicado o parâmetro de auto avaliação do mAP para determinar quantas épocas deveriam ser executadas.

A Figura 30 demonstra a curva da função mAP para este treinamento realizado com o parâmetro de auto avaliação, onde o eixo X demonstra o número de épocas e o eixo Y a média das precisões das classificações encontradas na detecção.

**Figura 30 - Gráfico da função de mAP do treinamento da rede RetinaNet no experimento 2.**



**Fonte: autoria própria.**

Os parâmetros utilizados para o cálculo do mAP no gráfico da Figura 30 foram diferentes dos aplicados na avaliação dos resultados dos experimentos, devido a essa configuração de parâmetros não estar presente no treinamento. Por isso foram usados os parâmetros padrões de avaliação sendo estes: Máximo de detecções em 100, limiar de pontuação em 5% e limiar de interseção em 50%. Estes parâmetros são detalhados na seção 4.3.

Por fim este treinamento rodou por 7 épocas com 500 passos e demorou 4 horas e 29 minutos para ser concluído, e o melhor valor de perda encontrado para esse treinamento foi de 0,7748 em sua época final. A taxa média obtida pela precisão foi de 95,8% e pela revocação de 97,8%.

#### 4.3.4.3 Experimento 3

O último experimento utilizou a base de imagens já finalizada que continha todos os 3129 rótulos igual ao segundo experimento. Não foi utilizado o parâmetro de auto avaliação e a rede foi treinada por 20 épocas com 500 passos. Esse treinamento demorou 14 horas e 56 minutos para ser concluído.

A Figura 31 demonstra a curva da função de perda para o treinamento realizado com 20 épocas, onde o eixo X demonstra o número de épocas e o eixo Y o valor encontrado pela função da perda focal.

**Figura 31 - Gráfico da função de perda do treinamento da rede no experimento 3.**



**Fonte: autoria própria.**

Analisando a Figura 31 percebe-se que o gráfico ficou muito próximo de uma curva exponencial, demonstrando que nas épocas finais o treinamento estava tendo melhorias pouco significativas. Dentro do gráfico menor valor de perda encontrado foi na sua última época, onde a rede encontrou um valor de 0,6169.

Essa representação do gráfico como exponencial demonstra que o resultado final foi bem sucedido, apresentando um valor melhor que o demonstrado no segundo experimento que obteve um valor de perda igual a 0,7748. Possibilitando poucas melhorias a partir desse ponto.

Para este experimento a taxa de precisão média foi de 99,0% e de revocação 97,6%.

### 4.3 ANÁLISE DOS RESULTADOS

As análises dos resultados foram divididas em quatro seções cada uma para analisar as aplicações de detecção apresentados neste trabalho. Sendo a detecção utilizando técnicas de processamento de imagem (seção 4.1), os treinamentos iniciais utilizando a rede neural RetinaNet com entradas e parâmetros diferentes (seção 4.2.4)



e o treinamento final que apresentou os melhores resultados para o treinamento da rede neural RetinaNet (seção 4.2.3).

As métricas utilizadas para avaliação das detecções neste trabalho foram o cálculo da média das precisões das classificações pelo mAP (seção 3.5.1) e a avaliação da contagem de indivíduos verdadeiros pelas medidas de precisão e revocação das detecções (seção 3.5.2), que estão diretamente ligadas aos falsos positivos e falsos negativos respectivamente.

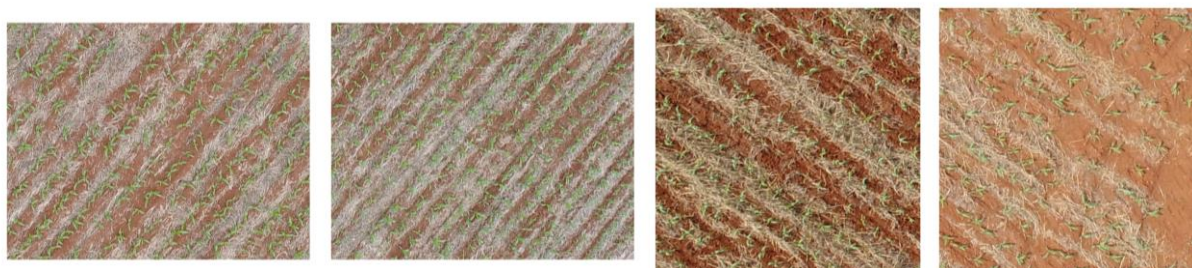
Alguns parâmetros da detecção precisaram ser configurados para serem feitas as avaliações pelas métricas, sendo estes:

- Máximo de detecções: Delimita um número máximo de detecções a serem encontradas na imagem.
- Limiar de pontuação: Delimita as classificações consideradas como corretas pela pontuação recebida pelas mesmas.
- Limiar de interseção: Delimita as classificações consideradas como corretas comparando a classificação com o resultado real apresentado, verificando o tamanho da interseção entre as duas caixas delimitadoras.

Para a avaliação do método de mAP foram utilizados os parâmetros de limiar de pontuação em 50%, limiar de interseção em 50% e máximo de detecções em 300. Já para as avaliações de precisão e revocação o único parâmetro a ser configurado foi o limiar de pontuação em 50%, mas além desse parâmetro outra diferença foi que para o cálculo da precisão e da revocação foi utilizada a base de imagens de testes citada na seção 4.2.1.

A Figura 32 demonstra as 4 imagens que foram utilizadas para gerar os 757 rótulos presentes na base de teste.

**Figura 32 – Imagens utilizadas na base de teste.**



Fonte: autoria própria.

(a)

(b)

(c)

(d)

#### 4.4 RESULTADO TÉCNICAS DE PROCESSAMENTO DE IMAGEM

Como já discutido anteriormente na seção 4.1 a metodologia proposta utilizando técnicas de processamento de imagem onde as plantas foram separadas pelo espectro da cor verde, não apresentou bons resultados variando muito os valores dos resultados de imagem para imagem.

Para analisar os resultados obtidos as únicas métricas que podem ser aplicadas para essa metodologia são as medidas de precisão e revocação das detecções, devido ao fato dessa detecção não utilizar uma medida para avaliar a qualidade das suas classes, impossibilitando o cálculo de mAP.

A Tabela 1 demonstra os valores obtidos por esse resultado onde cada entrada representa uma das imagens da base de teste demonstradas na Figura 32.

**Tabela 1 - Avaliação do método utilizando técnicas de processamento de imagem.**

	Real Detecção		Verdadeiro Positivo	Falso Negativo	Falso Positivo	precisão	revocação
<b>Figura32-a</b>	235	178	149	86	29	83,7%	63,4%
<b>Figura32-b</b>	296	397	249	47	148	62,7%	84,1%
<b>Figura32-c</b>	121	488	120	1	368	24,5%	99,1%
<b>Figura32-d</b>	105	162	100	5	62	61,7%	95,2%
<b>TOTAL</b>	757	1225	618	139	607	50,4%	81,6%

Fonte: autoria própria.

O resultado apresentou muitos erros relacionados ao grande número de falsos positivos encontrados na detecção. Devido a isso a porcentagem da precisão é ruim atingindo média de precisão igual a 50,4% um valor muito abaixo do esperado para uma boa detecção.

Além dos erros relacionados aos falsos positivos o número de falsos negativos também foi consideravelmente alto, apresentando uma média de revocação de 81,6% que não foi considerado satisfatório.

#### 4.5 RESULTADO RETINANET – EXPERIMENTO 1

Dentro das avaliações das CNN's a fórmula do mAP é uma medida que pode ser calculada facilmente para apresentar uma estimativa da sua avaliação. Aplicando o cálculo da média das precisões das classificações para este treinamento encontrou se um valor de 0,5910 para o mAP.

Igualmente a seção 4.4 foram aplicadas as fórmulas de precisão e revocação utilizando a base de imagens de teste e os seus resultados estão dispostos da mesma maneira na Tabela 2.

**Tabela 2 - Avaliação da rede RetinaNet treinada por 10 épocas com a base de 1613 rótulos.**

	Real Detecção		Verdadeiro Positivo	Falso Negativo	Falso Positivo	precisão	revocação
<b>Figura32-a</b>	235	235	227	8	8	96,5%	96,5%
<b>Figura32-b</b>	296	290	287	9	3	98,9%	96,9%
<b>Figura32-c</b>	121	122	117	4	5	95,9%	96,6%
<b>Figura32-d</b>	105	111	105	0	6	94,5%	100%
<b>TOTAL</b>	757	758	736	21	22	97,0%	97,2%

Fonte: autoria própria.

Comparando os resultados obtidos nessa seção com os da seção 4.4 pode se verificar que a utilização de uma CNN melhorou muito o desempenho da aplicação, diminuindo drasticamente o número de falsos positivos e negativos e assim aumentado a precisão média para 97,0% e a revocação média para 97,2%.

#### 4.6 RESULTADO RETINANET – EXPERIMENTO 2

Aplicando o cálculo da média das precisões das classificações para este treinamento encontrou se um valor de 0,6029 para o mAP.

Calculando a precisão e revocação das detecções utilizando a base de imagens de teste foram encontrados os resultados dispostos na Tabela 3.

**Tabela 3 - Avaliação da rede RetinaNet treinada com auto a avaliação por mAP.**

	Real Detecção		Verdadeiro	Falso	Falso	precisão revocação	
			Positivo	Negativo	Positivo		
<b>Figura32-a</b>	235	240	232	3	8	96,6%	98,7%
<b>Figura32-b</b>	296	290	284	12	6	97,9%	95,9%
<b>Figura32-c</b>	121	133	120	1	13	90,2%	99,1%
<b>Figura32-d</b>	105	110	105	0	5	95,4%	100%
<b>TOTAL</b>	757	773	739	16	32	95,8%	97,8%

Fonte: autoria própria.

Após a análise dos resultados foi possível verificar que a auto avaliação do mAP apresentou bons resultados para a revocação média chegando em 97,8%, melhorando seu desempenho se comparado a análise da seção 4.5, mas por outro lado a precisão média diminuiu chegando em 95,8%, mesmo trabalhando com uma base de imagens maior e mais variada que a utilizada no resultado da rede da Seção 4.5.

#### 4.7 RESULTADO RETINANET – EXPERIMENTO 3

Para o treinamento final os resultados encontrados foram os melhores para a aplicação de contagem das plantas.

Aplicando o cálculo da média das precisões das classificações foi encontrado um valor de 0,6055 para o mAP, sendo o segundo maior dentre os treinamentos.

Calculando a precisão e revocação das detecções utilizando a base de imagens de teste foram encontrados os resultados dispostos na Tabela 4.

**Tabela 4 - Avaliação do método utilizando a rede RetinaNet com o melhor treinamento.**

	Real Detecção		Verdadeiro	Falso	Falso	precisão revocação	
			Positivo	Negativo	Positivo		
<b>Figura32-a</b>	235	233	231	4	2	99,1%	98,2%
<b>Figura32-b</b>	296	286	286	10	0	100%	96,6%
<b>Figura32-c</b>	121	120	117	4	3	97,5%	96,6%
<b>Figura32-d</b>	105	107	105	0	2	98,1%	100%
<b>TOTAL</b>	757	746	739	18	7	99,0%	97,6%

Fonte: autoria própria.

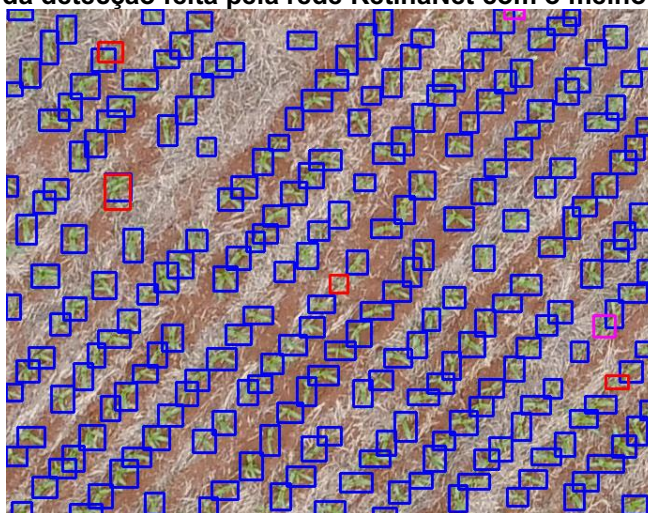
Conforme observado na Tabela 4, o resultado encontrado pela média da aplicação nas quatro entradas obteve uma precisão de 99,0% e uma revocação de 97,6%.

Essa aplicação foi a que apresentou a melhor precisão e o melhor valor de mAP dentre todas as avaliadas, tendo 0,026 a mais de mAP que o experimento 2 que ficou em segundo lugar na métrica de mAP e 2,0% a mais de precisão que o experimento 2 que também ficou em segundo lugar na métrica de precisão. Além de apresentar o segundo melhor valor na métrica de revocação tendo apenas 0,2% a menos que o experimento 2 que foi o melhor resultado.

Analisando este resultado verifica-se que a rede treinada pelo auto avaliação do mAP apresentou resultados melhores em duas das três métricas. Mas considerando que a diferença na precisão foi de 2% a mais e na revocação foi de apenas 0,2% a menos, a quantidade de erros obtidos por esse treinamento foi a menor dentre todos, sendo esse o resultado que apresentou o valor mais próximo da contagem real.

A Figura 33 demonstra a aplicação dessa rede em uma das imagens da base de testes. As caixas delimitadoras em azul representam seus verdadeiros positivos, em roxo seus falsos positivos e em vermelho seus falsos negativos.

**Figura 33 - Avaliação da detecção feita pela rede RetinaNet com o melhor treinamento.**



Fonte: autoria própria.

A Tabela 5 demonstra a comparação dos resultados encontrados pelos experimentos neste trabalho com os resultados do trabalho de (XIANG et al., 2019)

que também utilizam da rede RetinaNet para detecções em imagens aéreas. A métrica utilizada como avaliação na Tabela 5 foi o cálculo do mAP.

**Tabela 5 – Comparação de treinamentos.**

<b>Trabalhos</b>	<b>mAP em %</b>
<b>Experimento 1</b>	59,1%
<b>Experimento 2</b>	60,2%
<b>Experimento 3</b>	60,5%
<b>XIANG RetinaNet</b>	51,2%
<b>XIANG Modelo Modificado</b>	84,6%

**Fonte: autoria própria.**

Analisando a Tabela 5 a taxa mAP encontrada no Experimento 3 pode ser comparada ao resultado encontrado em (XIANG et al., 2019). Ambos os experimentos trabalharam com a rede RetinaNet, sem aplicar modificações nas suas estruturas internas e utilizam a rede para detecção de pequenos objetos em imagens aéreas. É importante observar que a comparação direta das taxas de detecção (precisão e revocação) não podem ser consideradas equivalentes, pois as bases de dados e os objetos de detectados são distintos entre esses experimentos.

## 5 CONCLUSÕES

Com os resultados obtidos, foi possível validar a utilização de tecnologias de processamento de imagens e aprendizagem de máquina para a aplicação de contagem e detecção de plantas de milho. Os modelos treinados na CNN ofereceram boas taxas de acerto e as taxas de erros são próximas às apresentadas pela literatura. O resultado médio obtido pela aplicação com o melhor desempenho obteve 99% de precisão e 97,6% de revocação, quando utilizadas as configurações do experimento 3.

A utilização de duas metodologias possibilitou demonstrar como o uso de CNN's apresentam, na sua maioria, resultados superiores que os métodos de detecção por técnicas de processamento de imagem utilizando cores. Os resultados obtidos pelas CNN's neste trabalho melhoraram a qualidade das detecções.

A aplicação da RetinaNet para imagens aéreas foi uma boa escolha, pois obteve bons resultados utilizando as imagens obtidas pelo VANT conforme o demonstrado nos resultados dos experimentos apresentados nas seções 4.5, 4.6 e 4.7.

No que diz respeito a aquisição das imagens, foi verificado que a melhor fase a se obtê-las foi entre 2 a 3 semanas após o plantio, onde as plantas de milho já estão grandes o suficiente para serem detectadas, mas ainda estão bem separadas umas das outras. As variações de iluminação, cor de fundo e espaçamento entre as plantas influenciaram pouco as imagens obtidas nessa fase.

Para futuros trabalhos uma possível melhoria para implementar na aplicação seria a modificação da rede RetinaNet, reformulando alguns de seus componentes internos para focar o seu escopo diretamente para aplicação de detecção de milho em imagens aéreas. Semelhante ao que foi feito no trabalho de (XIANG et al., 2019) onde ele modificou os componentes internos da rede para focar sua aplicação de detecções de pedestres em imagens aéreas, assim Xiang conseguiu melhorar o seu mAP de 51,2% utilizando a rede RetinaNet normal para 84,0% com a sua rede modificada.

Por fim, este trabalho demonstrou a viabilidade na aplicação de detecção e contagem de plantas sendo o primeiro na literatura a fazer essa contagem diretamente para plantas de milho. Essa informação auxilia no desenvolvimento da produtividade agrícola aumentando assim a precisão da informação anteriormente prevista por métodos estatísticos.

## REFERÊNCIAS

- Agricultura de precisão**, Brasília, Ministério da Agricultura, Pecuária e Abastecimento, 2011, 2ª edição revisada e atualizada.
- CAPELLI, N. L. **Agricultura de precisão-Novas tecnologias para o processo produtivo**. LIE: DMAQAG: FEAGRI: UNICAMP, 1999.
- COELHO, Antônio Marcos. **Agricultura de precisão: manejo da variabilidade espacial e temporal dos solos e culturas**. Embrapa Milho e Sorgo-Documents (INFOTECA-E), 2005.
- IKEDA, Denis T. **Seleção Não-Supervisionada de Métodos de Binarização para Documentos Históricos**. Departamento de Ciência da Computação, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2011.
- JORGE, Lúcio A. de C.; INAMASU, Ricardo Y. **Uso de veículos aéreos não tripulados (VANT) em Agricultura de Precisão**. São Carlos, SP: Embrapa, 2014.
- MARQUES FILHO, Ogê; NETO, Hugo Vieira. **Processamento digital de imagens**. Brasport, 1999.
- OLIVEIRA, Henrique Cândido de et al. **A NEW METHODOLOGY FOR CROP FAILURE DETECTION IN UAV IMAGERY BY USING IMAGE PROCESSING BASED ON MATHEMATICAL MORPHOLOGY**. In: IEEE Geoscience and Remote Sensing Letters, Volume: 15, 2018, tópico temático, IEEE, 2018.
- EMBRAPA. **Perdas invisíveis na cultura do milho**. Jornal Eletrônico da Embrapa Milho e Sorgo, Sete Lagoas, set. 2014. Disponível em: <<http://gao.cnpms.embrapa.br/index.php?ed=Nzc=>>>. Acesso em: 02 mai. 2019.
- ROZA, D. Novidade no campo: Geotecnologias renovam a agricultura. **Revista InfoGEO**, n, 2000.
- SANCHES, Ionildo José. **Sobreposição de imagens de termografia e ressonância magnética: uma nova modalidade de imagem médica tridimensional**. 2009.



SHIRATSUCHI, L. S. O avanço dos drones. **Embrapa Agrossilvipastoril-Artigo de divulgação na mídia (INFOTECA-E)**, 2014.

SOARES, Guilherme Afonso et al. **Identificação de linhas de plantio por meio da transformada de Hough**. 2017.

BALLARD, D. H.; BROWN, C. M. **Computer vision**, 1982. Prentice-Hall, Englewood Cliffs, NJ, 1982.

SILVA, Giovanni Cimolin da et al. **Detecção e contagem de plantas utilizando técnicas de inteligência artificial e machine learning**. Repositório institucional UFSC. 2018.

FEITOSA, Rafael Divino Ferreira et al. **Modelos matemáticos para redução do espectro provável e detecção de tons de pele humana em imagens coloridas representadas nos espaços de cores RGB e HSV**. 2015.

KAKUMANU, Praveen; MAKROGIANNIS, Sokratis; BOURBAKIS, Nikolaos. **A survey of skin-color modeling and detection methods**. Pattern recognition, v. 40, n. 3, p. 1106-1122, 2007.

DO MONTE LIMA, João Paulo Silva et al. **Reconhecimento de Padrões em Tempo Real Utilizando a Biblioteca OpenCV**. CEP, v. 50670, p. 901, 2008.

AKERKAR, Rajendra. **Introduction to artificial intelligence**. PHI Learning Pvt. Ltd., 2014.

ERTEL, Wolfgang. **Introduction to artificial intelligence**. [S.l.]: Springer Science & Business Media, 2011.

RUSSEL, Stuart et al. **Artificial intelligence: a modern approach**. Pearson Education Limited, 2013.

GU, Jiuxiang et al. **Recent advances in convolutional neural networks**. Pattern Recognition, v. 77, p. 354-377, 2018.

LECUN, Yann et al. **Lenet-5, convolutional neural networks**. 2015. Disponível em: <<http://yann.lecun.com/exdb/lenet>>. Acesso em: 06 out. 2020.

ROSEBROCK Adrian. **A simple neural network with Python and Keras**. 2016. Disponível em: <<https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras/>>. Acesso em: 06 out. 2020.

MARTINS, Roberto Augusto Philippi et al. **Aplicação de Redes Convolucionais Profundas para Detecção de Massas em Mamografias**. 2019.

SOVIANY, Petru; IONESCU, Radu Tudor. **Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction**. In: 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). IEEE, 2018. p. 209-214.

LIN, Tsung-Yi et al. **Focal loss for dense object detection**. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2980-2988.

EMMI, Luis et al. **Fleets of robots for precision agriculture: a simulation environment**. *Industrial Robot: An International Journal*, 2013.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento de imagens digitais**. Editora Blucher, 2000.

YU, Dingjun et al. **Mixed pooling for convolutional neural networks**. In: International conference on rough sets and knowledge technology. Springer, Cham, 2014. p. 364-375.

GITHUB. **labellmg-master**. 2018. Disponível em: <<https://github.com/tzutalin/labellmg>>. Acesso em: 07 nov. 2019.

GAISER, Hans. **keras-retinanet**. 2017. Disponível em: <<https://github.com/fizyr/keras-retinanet>>. Acesso em: 02 nov. 2019.

DOUILLARD, Arthur. **Object Detection with Deep Learning on Aerial Imagery**. 2018. Disponível em: <<https://medium.com/data-from-the-trenches/object-detection-with-deep-learning-on-aerial-imagery-2465078db8a9>>. Acesso em: 02 abr. 2020.

XIANG, Chao et al. **Pedestrian Detection Under Unmanned Aerial Vehicle an Improved Single-Stage Detector Based on RetinaNet**. In: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2019. p. 1-6.

CULMAN, María; DELALIEUX, Stephanie; VAN TRICHT, Kristof. **Palm Tree Inventory From Aerial Images Using Retinanet**. In: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). IEEE, 2020. p. 314-317.

KETKAR, Nikhil; SANTANA, Eder. **Deep Learning with Python**. Berkeley, CA: Apress, 2017.

MANNING, Christopher D.; SCHÜTZE, Hinrich; RAGHAVAN, Prabhakar. **Introduction to information retrieval**. Cambridge university press, 2008.

OTSU, Nobuyuki. **A threshold selection method from gray-level histograms**. IEEE transactions on systems, man, and cybernetics, v. 9, n. 1, p. 62-66, 1979.

MATOS, Pablo Freire et al. **Relatório técnico “métricas de avaliação”**. Universidade Federal de São Carlos, 2009.