UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

ANATELLI ANNE FAGUNDES HERINGER

DETECÇÃO DE COMPONENTES DE REDE DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA POR VISÃO COMPUTACIONAL

VITÓRIA
2022

ANATELLI ANNE FAGUNDES HERINGER

DETECÇÃO DE COMPONENTES DE REDE DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA POR VISÃO COMPUTACIONAL

Parte manuscrita do Projeto de Graduação da aluna **Anatelli Anne Fagundes Heringer**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Orientadora: Profa. Dra. Raquel Frizera Vassallo
Coorientador: Dr. Clebeson Canuto dos Santos

VITÓRIA
2022

ANATELLI ANNE FAGUNDES HERINGER

DETECÇÃO DE COMPONENTES DE REDE DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA POR VISÃO COMPUTACIONAL

Parte manuscrita do Projeto de Graduação do aluno **Anatelli Anne Fagundes Heringer**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Aprovada em 16 de agosto de 2022.

COMISSÃO EXAMINADORA:

Raquel Frizera Vassallo

Profa. Dra. Raquel Frizera Vassallo
Universidade Federal do Espírito Santo
Orientadora

Clebeson Canuto dos Santos

Dr. Clebeson Canuto dos Santos
Coorientador

Fabio Ricardo O Bento

Msc. Fabio Ricardo Oliveira Bento
Instituto Federal do Espírito Santo
Examinador

Eng. Msc. Filiphe Oliveira Louback
Empresa EDP
Examinador

A Ele acima de tudo. A elas (saudade minha),
mamãe e vovó, partes de mim.

AGRADECIMENTOS

A cada dia e a cada feito meu sou grata àquele que me escolheu e me amou primeiro. É para Ele as primícias do meu agradecimento por todos os dias do meu viver.

Àquelas (em memória) que me amaram da forma mais pura e sincera ao longo de nossa jornada juntas. Foram elas que sempre acreditaram em mim quando nem eu o fazia. Minha mãe Ana e minha avó Neuza, donas dos maiores corações, dos mais lindos sorrisos e dos mais doces olhares, donas da minha saudade diária. Obrigada pelo cuidado, pela doação e pelo amor incondicional. Vocês estão eternizadas em meu coração para sempre.

Àqueles que ainda em vida posso agradecer por todo o amor e a dedicação, agradeço ao meu pai e avô, os "Josés" da minha vida, pela motivação, carinho, cuidado e pelas abdicações. Espero que sejam felizes por todos os dias que ainda estão por vir, pois vocês merecem muitos dias de muitos sorrisos (espero poder contribuir um pouco para isso).

À minha irmã querida Netalianne, minha inspiração. Obrigada pelo apoio, pelo incentivo, por buscar me compreender, mesmo nos momentos mais difíceis. A ela também dedico minhas conquistas, pois as dificuldades do caminho são menos duras de serem vividas com ela ao meu lado. A ela eu desejo a leveza e as coisas mais lindas da vida, afinal, por muito já passamos até hoje.

Aos amigos que a Ufes me presenteou, obrigada pelos momentos de ajuda e descontração, pelas conversas, pelos conhecimentos compartilhados. Levo comigo histórias que sempre serão contadas por onde eu for. Aos professores e educadores que amam o que fazem, e que em primeiro lugar são seres humanos especiais: obrigada pelas lições para toda a vida.

Aos meus colegas e amigos da EDP que tanto me ensinaram de tantas formas possíveis. Obrigada por cada momento de aprendizado e partilha, inclusive para desenvolvimento do presente trabalho. Espero que nossos caminhos se cruzem novamente no futuro.

Aos meus amigos de vida e a cada pessoa especial, aqueles mais próximos hoje, que vivem minhas alegrias e também as dificuldades, sem vocês eu não conseguiria. E aos que ao longo

do caminho me despedi de alguma forma: eu os levarei em minha memória por todos os dias, vocês também contribuíram em meu caminho até aqui.

Sou grata àqueles que mesmo por um dia sequer de muitas formas me ensinaram, me acompanharam, me acolheram e me amaram. Espero poder continuar contribuindo por onde eu estiver com o que aprendi com todos aqueles que foram mencionados e lembrados ao escrever esse texto. A todos eles e a cada um deles: obrigada mais uma vez.

RESUMO

A quarta revolução industrial, também conhecida como Indústria 4.0, vem ganhando aceleração desde as últimas décadas, motivada pela crescente necessidade de incorporação de máquinas mais inteligentes em processos industriais, como forma de garantir a demanda de novos produtos e serviços. A realidade aumentada e visão computacional em conjunto com métodos de inteligência artificial tem apresentado grande investimento em diversas aplicações nos mais variados setores empresariais, como o automobilístico, o de construção e o de energia, dentre outros. Em se tratando especificamente do setor de energia, uma de suas premissas básicas consiste no fornecimento desse bem (energia) aos usuários finais com qualidade, segurança e integridade. Nesse contexto, a busca por inovação e tecnologia em seus processos é cada dia mais necessária e presente. Algumas das aplicações de inteligência artificial estão relacionadas à sua utilização como ferramenta de auxílio em manutenções de campo e acompanhamento operacional. Sendo assim, o presente trabalho propõe o desenvolvimento de um sistema de detecção de componentes de redes de distribuição da empresa Energias de Portugal (EDP) no Espírito Santo, o qual foi baseado em visão computacional e técnicas de aprendizado profundo utilizando o *framework Darknet* e método Yolov4. Além disso, utilizou-se da técnica de *data augmentation* a fim de elevar a capacidade de generalização da rede. Utilizando as premissas anteriores, foi possível alcançar um mAP de 91,39% (liminar de 0,5) e um IOU médio de 67,26%. Portanto, para os conjuntos de dados disponibilizados, o sistema alcançou sucesso no objetivo proposto, indicando as práticas adotadas durante as etapas de desenvolvimento do trabalho. Para os próximos trabalhos, aponta-se incremento nas classes, incorporando a avaliação de falhas de equipamentos e modelos, por exemplo, além da integração da solução proposta a tecnologias de realidade aumentada para auxílio de técnicos de campo durante a realização de atividades de manutenção.

Palavras-chave: Indústria 4.0. Inteligência artificial. Detecção de objetos. Visão computacional. Aprendizado profundo.

ABSTRACT

The fourth industrial revolution, also known as Industry 4.0, has been accelerating in recent decades, motivated by the growing need to incorporate smarter machines into industrial processes, as a way of guaranteeing the demand for new products and services. Augmented reality and computer vision together with artificial intelligence methods have shown great investment in several applications in the most varied business sectors, such as automotive, construction and energy, among others. When dealing specifically with the energy sector, one of its basic premises is the supply of this good (energy) to end users with quality, safety and integrity. In this context, the search for innovation and technology in its processes is increasingly necessary and present. Some of the artificial intelligence applications are related to its use as an aid tool in field maintenance and operational monitoring. Therefore, the present work proposes the development of a component detection system for the distribution networks of Energias de Portugal (EDP) in Espírito Santo, based on computer vision and deep learning techniques using the Darknet framework and method Yolov4. In addition, the data augmentation technique was used in order to increase the generalization capacity of the network. Using the previous assumptions, it was possible to reach a mAP of 91.39% (threshold of 0.5) and an average IOU of 67.26%. Therefore, for the available data sets, the system was successful in the proposed objective, indicating the practices adopted during the stages of development of the work. For the next works, it is pointed out an increase in the classes, incorporating the evaluation of equipment failures and models, for example, in addition to the integration of the proposed solution with augmented reality technologies to help field technicians during maintenance activities.

Keywords: Industry 4.0. Artificial intelligence. Object detection. Computer vision. Deep learning.

LISTA DE FIGURAS

Figura 1 – Linha do tempo das revoluções industriais 15
Figura 2 – Reconhecimento de objetos embarcado em *hardware* de AR 16
Figura 3 – Fatores que impulsionam o campo da visão computacional na atualidade 20
Figura 4 – Redes neurais artificiais baseadas em aprendizado de máquina e aprendizado profundo 21
Figura 5 – Representação de um modelo de redes neurais, obtendo a saída esperada 22
Figura 6 – Diferenças entre classificação, localização, detecção e segmentação 23
Figura 7 - Evolução do número de pesquisas referentes ao tema detecção de objetos. 24
Figura 8 - Representação do sistema de detecção do YOLO. 26
Figura 9 - Representação da supressão não-máxima. 27
Figura 10 - Exemplo de saída da detecção realizada pelo YOLOv4. 28
Figura 11 - Fluxograma de etapas de elaboração do projeto. 31
Figura 12 - Representação de amostra da base de dados, apresentando grande diversificação de imagens. 33
Figura 13 – Imagem de fundo apresentando troncos de árvores, semelhantes a postes 34
Figura 14 – Exemplo de imagens que compõe base final estratificada. 35
Figura 15 - Exemplo de imagem rotulada à esquerda e seu respectivo arquivo gerado à direita. 36
Figura 16 - Frame retirado de filmagem feita por *drone* para exemplificar postes e cruzetas à longa distância. 37
Figura 17 - Rótulos de classes conforme considerações adotadas. 37
Figura 18 - Exemplo de rotação, alteração de brilho e saturação, respectivamente. 39
Figura 19 - Representação da definição matemática do conceito de IoU. 41
Figura 20 – Representação da curva suavizada (C) e curva real (B). 42
Figura 21 - Gráfico ilustrativo do comportamento do mAP e *Loss*, representando o desempenho de um modelo, frente ao número de iterações realizadas para treinamento da rede. 43
Figura 22 - Gráficos do comportamento do mAP e *Loss*, representando o desempenho do modelo proposto, frente ao número de iterações realizadas durante Treinamento 1 e 2, respectivamente, da rede. 48
Figura 23 – Desempenho da rede em imagens com visadas inferiores dos equipamentos 49
Figura 24 – Desempenho da rede em imagens com oclusão. 50

Figura 25 – Desempenho da rede para imagens com baixo brilho e resolução. 51
Figura 26 – Desempenho da rede para caso extremo de abalroamento. 51
Figura 27 - Exemplos de imagens de frames de *drones*. 52

LISTA DE TABELAS

Tabela 1 - Resumo de tipos de imagens selecionadas. 44
Tabela 2 - Divisão de imagens por etapa. 45
Tabela 3 - Representação do balanceamento de classes 45
Tabela 4 - Configuração para os diferentes treinamentos. 46
Tabela 5 - Métrica AP para diferentes classes e treinamentos. 48
Tabela 6 - Métrica de desempenho mAP e IOU médio para os treinamentos com e sem *data augmentation.* 49

LISTA DE QUADROS

Quadro 1 – Comparativo dos métodos Faster R-CNN, YOLO versão 2 e SSD 25
Quadro 2 - Referências de classes adotadas durante a rotulação dos objetos. 36
Quadro 3 - Parâmetros padrões da rede e utilizados para o modelo proposto. 46

LISTA DE ABREVIATURAS E SIGLAS

ABDi	Agência Brasileira de Desenvolvimento da Indústria
AI	Artificial Intelligence
AP	Average Precision
AR	Augmented Reality
CSDs	Centros de Serviço da Distribuição
DO	Detecção de Objetos
EDP	Energias de Portugal
FP	False Positive
FN	False Negative
IoU	Intersection Over Union
mAP	Mean Average Precision
PAnet	Path Aggregation Network
R-CNN	Region-based Convolutional Neural Network
SAS	Statistical Analysis System
SSD	Single Shot Detector
TP	True Positive
UFES	Universidade Federal do Espírito Santo
VC	Visão Computacional
YOLO	You Only Look Once

SUMÁRIO

1 INTRODUÇÃO 14

2 OBJETIVO 18

2.1 Objetivo Geral 18

2.2 Objetivos Específicos 18

3 REFERENCIAL TEÓRICO 19

3.1 Visão Computacional 19

3.2 Aprendizado Profundo 20

3.3 Reconhecimento de Objetos 22

3.3.1 Técnicas de Detecção de Objetos 23

3.4 YOLO 25

3.5 Trabalhos correlatos 28

4 METODOLOGIA 30

4.1 Metodologia 30

4.2 Etapas de Desenvolvimento 30

4.3 Obtenção e estratificação de dados 32

4.4 Rotulação 35

4.5 Data Augmentation 38

4.6 Implementação do treinamento 39

4.6.1 Métricas de Desempenho 40

5 RESULTADOS E DISCUSSÕES 44

5.1 Treinamento 44

5.2 Desempenho da rede 46

6 CONCLUSÕES E TRABALHOS FUTUROS 54

REFERÊNCIAS BIBLIOGRÁFICAS 56

1 INTRODUÇÃO

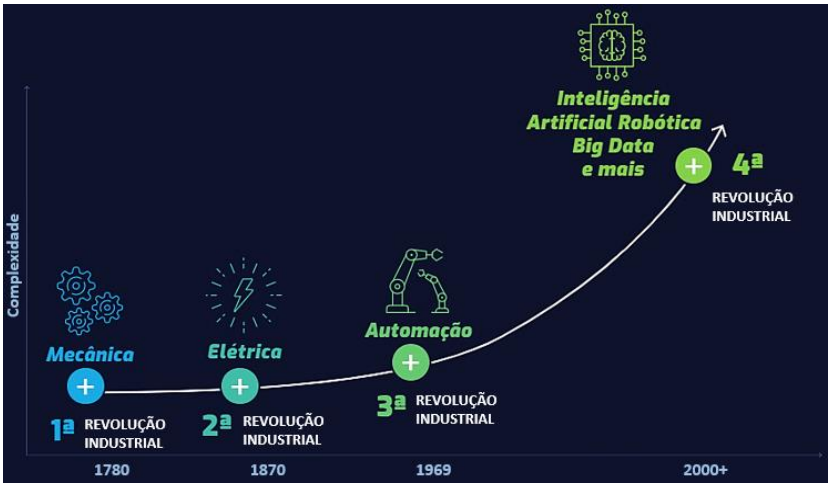
A história é marcada por diversas revoluções que ocasionaram significativos impactos no desenvolvimento da sociedade contemporânea. A primeira revolução industrial, que ocorreu no século XVIII, com início na Inglaterra, proporcionou a substituição de métodos artesanais por máquinas e ferramentas e trouxe o carvão como energia alternativa à madeira, além de outros biocombustíveis (COELHO, 2016). No final do século XIX, a eletricidade, as máquinas a vapor e o uso da força hidráulica revolucionaram novamente a indústria, surgindo assim a segunda revolução industrial (BORLIDO, 2017). O período entre 1960 e 1990 foi marcado por aquela que é conhecida como a terceira revolução industrial, caracterizada pela automação e robotização em linhas de produção, assim como pelo advento do computador, da telefonia e da Internet (COELHO, 2016).

Na última década, no ano de 2011, surge o conceito da Indústria 4.0 na Alemanha, expressão a qual daria nome à quarta revolução industrial da história, que é marcada fortemente por seus pilares, também conhecidos pelas atuais tecnologias disruptivas. Tais tecnologias apresentam como objetivo facilitar a visão e a tomada de decisão, tornando os processos mais robustos, seguros e inteligentes. A Agência Brasileira de Desenvolvimento Industrial (ABDi) apresenta a seguinte afirmação quanto ao impacto gerado por cada uma das revoluções mencionadas:

> As 3 primeiras revoluções industriais trouxeram as linhas de montagem, a eletricidade e a tecnologia da informação, [...] fazendo da competição tecnológica o cerne do desenvolvimento econômico. A quarta revolução industrial, que terá um impacto mais profundo e exponencial, se caracteriza por um conjunto de tecnologias que permitem a fusão do mundo físico, digital e biológico (ABDi, 2019).

Por meio dessa afirmação, percebe-se a crescente necessidade de pesquisas de inovação envolvendo tecnologias, que estejam relacionadas à Indústria 4.0, e que as empresas as introduzam em seus processos, de forma a acompanhar a evolução apresentada. A Figura 1 ilustra uma linha do tempo, exibindo as contribuições de cada um dos marcos em relação a sua complexidade.

Figura 1 – Linha do tempo das revoluções industriais



Fonte: Agência Brasileira de Desenvolvimento Industrial (2019).
Nota: Traduzido pela autora.

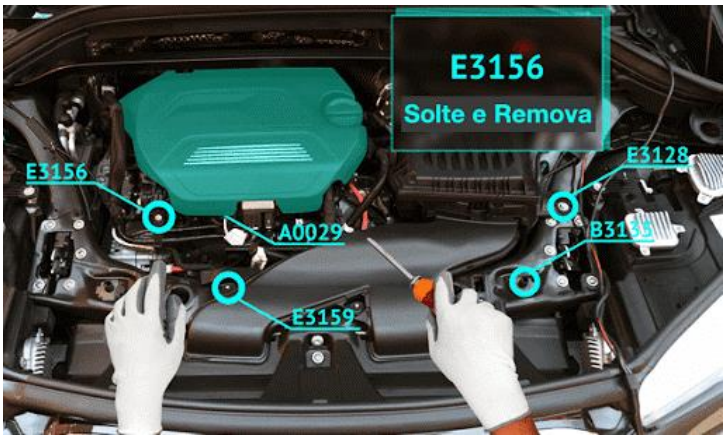
Segundo a ABDi a realidade aumentada (AR, do inglês *augmented reality*) e a inteligência artificial (AI, do inglês *artificial intelligence*) são duas tecnologias que impulsionam a Indústria 4.0. O termo AR é utilizado desde a década de 1960, quando Ivan Sutherland escreveu um artigo vislumbrando a evolução da realidade virtual e os seus reflexos no mundo real (SUTHERLAND, 1965). No que se refere ao potencial de aplicação de tais tecnologias, desde a década de 1990, já eram apontadas pelo menos três classes de aplicações potenciais da realidade aumentada quando associadas a recursos de AI, como: manutenção e reparação; planejamento de trajetória de robôs; e segmentação de instâncias (AZUMA, 1997).

No contexto industrial, a inteligência artificial e a realidade aumentada têm sofrido significativos avanços, objetivando a qualidade de vida do trabalhador, assim como a qualidade dos produtos, sistemas, treinamentos, manutenção e educação (GRAVE et al., 2001). Sua aplicação se tornou fundamental para a Indústria 4.0, uma vez que integram funcionalidade tecnológica à experiência humana para o aumento da produtividade e competitividade de mercado, colaborando com a segurança e a eficiência de serviço, por meio da redução de tempo e consequentemente dos custos de interrupção das atividades em campo.

No que tange à segurança, a inteligência artificial, quando também embarcada a *hardwares* de AR, permite que responsáveis diretos pelo operador verifiquem, por exemplo, em quais condições ele está realizando a operação. Da mesma forma, profissionais em campo são orientados a seguir com exatidão uma sequência de passos, por meio do reconhecimento de

objetos específicos mostrados em uma tela (de óculos inteligentes ou celulares por exemplo, como ilustra a Figura 2). Isto garante uma maior certificação quanto à prevenção de possíveis falhas ou esquecimentos humanos, que podem trazer riscos à integridade do trabalhador.

Figura 2 – Reconhecimento de objetos embarcado em *hardware* de AR



Fonte: Vrmat (2017).

O uso dessas tecnologias tem agregado valor às empresas que optam por integrar inovação e otimização aos seus processos (IND4.0, 2020). Tal iniciativa representa vantagem competitiva em um cenário no qual a diferenciação de mercado se dá fortemente pela inovação tecnológica. Alguns setores já apresentam crescentes investimentos nessa área de tecnologia, como por exemplo, o setor automobilístico (FORD, 2019), o de construção (AECOM, 2020), o de embalagens (TETRAPARK, 2018) e empresas do setor de energia elétrica, como a EDP (2019), a qual este trabalho utilizará como estudo de caso.

No cenário atual, a energia elétrica é considerada um bem essencial, uma vez que sua importância e as possibilidades geradas por ela vem crescendo de maneira acelerada nas últimas décadas. Isto se dá pelo seu papel nos processos de produção industrial, comercial e de prestação de serviços, como de saúde por exemplo, além da qualidade de vida que entrega aos domicílios da população. Sendo assim, é de suma importância estabelecer parâmetros mínimos de qualidade e continuidade de seu fornecimento, garantindo o direito e bem-estar do consumidor. E para que seja possível manter tais indicadores, se faz essencial, dentre outras práticas, a realização de manutenções de equipamentos da rede de distribuição.

No estado do Espírito Santo, a EDP detém da concessão referente a 70 municípios dos 78 que compreendem o estado, numa área equivalente a 41,2 mil km² e apresenta mais de 1,7 milhões

de clientes faturados (EDP, 2021). As manutenções corretivas de equipamentos e componentes da rede de distribuição, sob responsabilidade da distribuidora, são planejadas por meio de análises de engenheiros. Estes se baseiam em inspeções executadas por eletricistas de campo, que capturam um grande conjunto de fotos e utilizam *drones* para gravação de vídeos de componentes danificados ao longo de toda a extensão da rede, além de realizarem a classificação dos equipamentos e falhas encontradas no local.

Além disso, em situações que apresentem a necessidade de rápida intervenção nos equipamentos em subestações, por exemplo, e que coloquem, de alguma forma, a confiabilidade do sistema em prova, faz-se necessária a orientação à distância por meio de operadores. Isto se deve ao fato de os eletricistas ou técnicos plantonistas mais próximos ao local da ocorrência poderem não apresentar conhecimento específicos sobre aquela gama de equipamentos. Em tais condições, atualmente os especialistas orientam os técnicos por telefone, seguindo as instruções de trabalho [1], sua experiência e seus conhecimentos, e tirando dúvidas por meio de uma descrição via áudio.

Em tal contexto, o problema de pesquisa a ser abordado no presente Projeto de Graduação é: como otimizar e automatizar o processo de planejamento de manutenção e orientação aos técnicos de campo da EDP durante a realização de atividades corretivas, adotando como base os pilares da Indústria 4.0?

A justificativa para este trabalho é baseada em toda a conjuntura apresentada. O reconhecimento de ativos, em meio ao contexto de planejamento e realização da manutenção de falhas das redes de distribuição da EDP Espírito Santo, possibilitaria uma melhor eficiência nos processos de análise e de inspeções de equipamentos. Assim como afirmam Correa e Junior (2007), há relação direta entre a ocorrência de falhas humanas e o nível de conhecimento ou recursos de informação disponíveis para uso. Sendo assim, o maior conhecimento, por meio das informações oferecidas com praticidade, traz consigo a elevação da eficiência na execução do serviço, além de ser uma forma de auxílio à garantia da segurança e vida do trabalhador, de maneira a preservar sua integridade física

[1] Instruções de Trabalho consistem em documentos elaborados por colaboradores da EDP, os quais dominam aquela atividade específica, de forma a buscar uma maior disseminação de informações.

2 OBJETIVO

2.1 Objetivo Geral

O objetivo geral deste trabalho consiste no desenvolvimento de um sistema baseado em visão computacional para o reconhecimento de equipamentos em imagens de rede de distribuição de energia elétrica.

2.2 Objetivos Específicos

Os objetivos específicos traçados neste trabalho, os quais serviram como guia para alcançar o objetivo geral ao final de sua conclusão, estão expostos a seguir:

Investigar métodos de visão computacional para a escolha daquele mais adequado ao reconhecimento do componente de redes de distribuição de energia elétrica escolhido;

Rotular um conjunto de dados para o treinamento do modelo de visão computacional por meio das imagens fornecidas pela EDP;

Implementar um modelo de reconhecimento e avaliar o seu desempenho, validando-o por meio da realização de testes.

3 REFERENCIAL TEÓRICO

3.1 Visão Computacional

A visão computacional (VC) é uma área da inteligência artificial em que se utilizam elementos de processamento digital de imagens e de reconhecimento de padrões para extrair e interpretar as informações (BALLARD; BROWN, 1982, p. 13). Segundo Wiley e Lucas (2018), há uma sobreposição da VC com o processamento digital de imagens nas técnicas básicas, e alguns autores utilizam de ambos os termos indistintamente (BABATUNDE et al., 2015). Porém, o objetivo principal das técnicas de visão computacional é criar modelos de extração de dados e informações de imagens, enquanto o processamento de imagens busca, dentre outras coisas, tratar da implementação de transformações computacionais, como nitidez, contraste, entre outros.

A visão computacional pode ser considerada uma área de pesquisa extremamente complexa e diversa, sendo vastas as aplicações de extração dessas informações, a destacar: detecção de objeto (DO); classificação; rastreamento; segmentação de instâncias e análise de movimento (MONTANARI, 2015). O reconhecimento de padrões e o aprendizado profundo são comumente utilizados juntamente com a VC em diversas aplicações, como por exemplo, robôs que realizam ações a partir da captura de imagens (FREIRE, 2009). Outras aplicações, como aquelas que necessitem do fornecimento de informações pertinentes para o ser humano, buscam reproduzir aquilo que lhe seria capaz de realizar. Do ponto de vista da engenharia, a visão computacional visa construir sistemas autônomos que possam realizar algumas das tarefas que o sistema visual biológico é capaz de executar.

Freire (2009) mostra que, desde o século passado, várias extensas pesquisas, focadas no processo de estimulação visual, foram realizadas acerca dos olhos, neurônios e cérebro. Ele também apresenta as evoluções e impactos no campo de estudo da VC. Dessa forma, percebe-se que a visão computacional não é exatamente uma área de pesquisa nova. Em meados da década de 1970, pesquisadores já teriam desenvolvido métodos para o reconhecimento de padrões em imagens. Por outro lado, foi após o amadurecimento da *Internet*, nos anos 1990, quando grandes volumes de imagens foram disponibilizados *online* para análises, que o desenvolvimento de programas de reconhecimento de pessoas e objetos apresentou maior

crescimento (IND4.0, 2020). Esses grandes conjuntos de dados ajudaram a possibilitar que as máquinas fossem capazes de identificar pessoas e objetos específicos em fotos e vídeos.

Atualmente, inúmeros fatores convergem para impulsionar as pesquisas em visão computacional, os quais podem ser visualizados na Figura 3, como o desenvolvimento de novos hardwares e o surgimento de novos algoritmos de redes neurais. Segundo a *Statistical Analysis System* (2020):

> Os efeitos desses avanços no campo da visão computacional têm sido surpreendentes. As taxas de precisão de identificação e classificação de objetos foram de 50 para 90% em menos de uma década — e os sistemas de hoje são ainda mais precisos que seres humanos nas rápidas detecção e reação a estímulos visuais (Statistical Analysis System, 2020).

Figura 3 – Fatores que impulsionam o campo da visão computacional na atualidade

Fonte: Statistical Analysis System (2020).

Além disso, com o avanço do conceito da Indústria 4.0, a visão computacional, auxiliada pelas técnicas de aprendizado profundo, tem mostrado uma crescente utilização nesse contexto. Isto tem se percebido por meio de inovações tecnológicas, pesquisas e aplicações, buscando a detecção de equipamentos, pessoas, anomalias e falhas (IND4.0, 2020).

3.2 Aprendizado Profundo

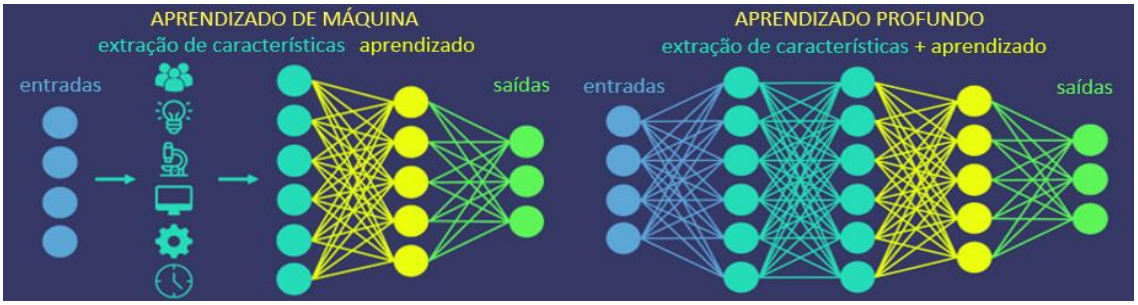
Nos últimos anos, os métodos de aprendizado profundo (do inglês, *deep learning*) têm demonstrado desempenho superior entre as técnicas de aprendizado de máquina (do inglês *machine learning*) em várias áreas, sendo a visão computacional uma das mais proeminentes (VOULODIMOS et al., 2018). Diversas tarefas como a detecção de objetos, o reconhecimento

de face e de ação são exemplos daquelas que utilizam métodos e conceitos do aprendizado profundo.

Segundo Haykin (1999, p. 2), suas técnicas permitem que modelos computacionais de múltiplas camadas aprendam e representem dados com vários níveis de abstração, imitando como o cérebro humano percebe e compreende informações multimodais, capturando implicitamente estruturas intrincadas de dados em grande escala. Segundo Fischler e Firschein (1987 apud HAYKIN, 1999, p. 25), as técnicas de aprendizado profundo são uma rica família de métodos, abrangendo redes neurais, modelos probabilísticos hierárquicos e uma variedade de algoritmos de aprendizado de características não supervisionadas e supervisionadas.[2]

O recente surto de interesse em métodos de aprendizagem profunda se deve ao fato de que eles têm demonstrado desempenho superior às técnicas clássicas em várias tarefas, bem como à abundância de grandes conjuntos de dados contendo diferentes tipos de diferentes fontes. A ambição de criar um sistema que simule o cérebro humano alimentou o desenvolvimento inicial das redes neurais artificiais (RNA) (LECUN; BENGIO; HINTON, 2016). A ideia em um sistema baseado em redes neurais artificiais é processar sinais enviando-os por uma rede de nós simulados, inspirados em neurônios no cérebro humano. Os sinais passam de nó em nó ao longo de conexões, ou links (semelhantes às junções sinápticas entre os neurônios). Na Figura 4, é ilustrada a arquitetura de uma rede neural simples e uma profunda, a qual apresenta mais de uma camada oculta, como é possível observar na imagem.

Figura 4 – Redes neurais artificiais baseadas em aprendizado de máquina e aprendizado profundo

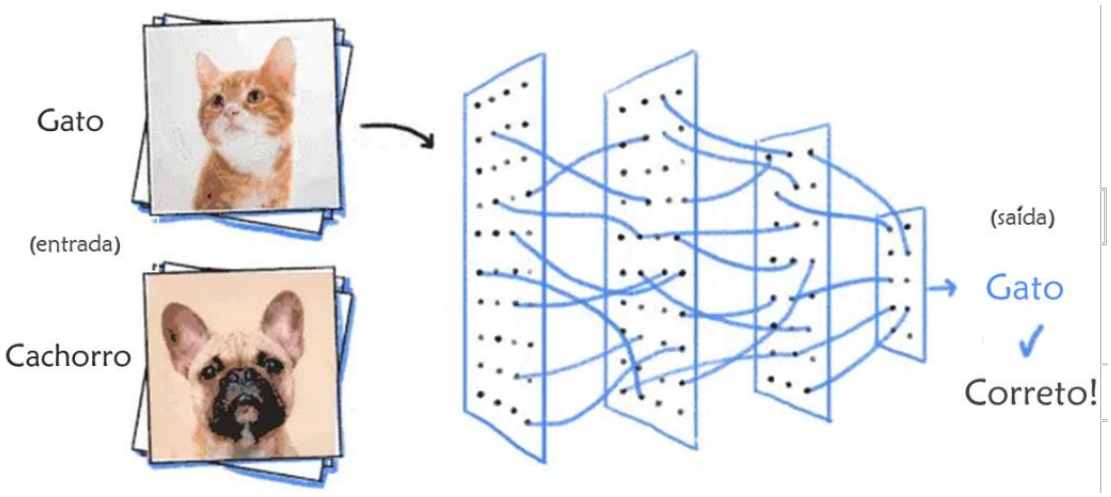


Fonte: *Quantdare* (2019).
Nota: Traduzido pela autora.

[2] Aprendizados supervisionados e não supervisionados são formas de treinamento. O primeiro deles, compreende o aprendizado a partir de dados pré-definidos, apresentando uma referência daquilo que está certo ou errado. O segundo, também conhecido como *clustering*, não apresenta um dado de referência.

A maioria das RNA's organiza os nós como uma série de camadas que são aproximadamente análogas a diferentes centros de processamento do córtex cerebral. Portanto, uma rede convolucional especializada em imagens teria uma camada de entrada que responde a *pixels* individuais, da mesma maneira que os bastonetes e cones respondem à luz que atinge a retina na visão humana. Uma vez ativados, os nós propagam seus níveis de ativação através das conexões ponderadas para a próxima camada. Isso continua até que alcancem a saída, onde é fornecido uma resposta, estimando se a entrada é referente a uma determinada classe, que, no exemplo da Figura 5, é um gato. Mas, caso a resposta esteja errada, considerando, por exemplo, que a imagem de entrada seja um cachorro, um algoritmo ajusta os parâmetros da rede para que a mesma alcance um melhor resultado na próxima iteração.

Figura 5 – Representação de um modelo de redes neurais, obtendo a saída esperada

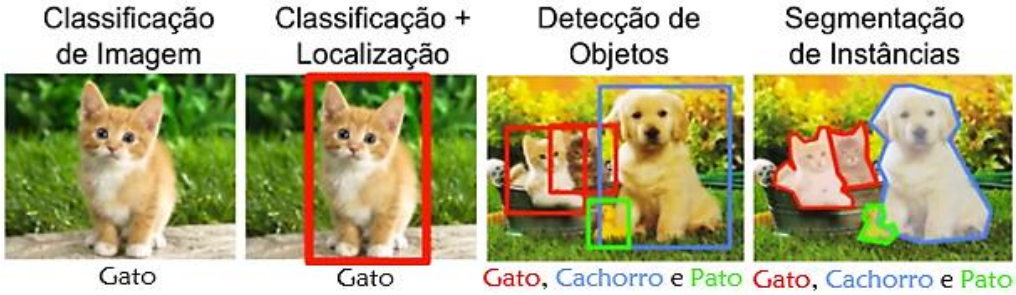


Fonte: *Medium* (2018).
Nota: Traduzido pela autora.

3.3 Reconhecimento de Objetos

O reconhecimento de objetos é um termo geral para descrever uma coleção de métodos e tarefas, relacionadas à visão computacional que envolvem a identificação de objetos em imagens digitais (RUSSAKOVSKY et al., 2015).Dentre estes, destacam-se a detecção de objetos e a segmentação. Na Figura 6, é apresentada uma ilustração da utilização das técnicas do reconhecimento de padrões de imagens, sendo possível visualizar suas diferenças.

Figura 6 – Diferenças entre classificação, localização, detecção e segmentação



Fonte: Ouaknine (2019).
Nota: Traduzido pela autora.

A seguir são apresentadas breves definições a respeito de cada uma das técnicas:

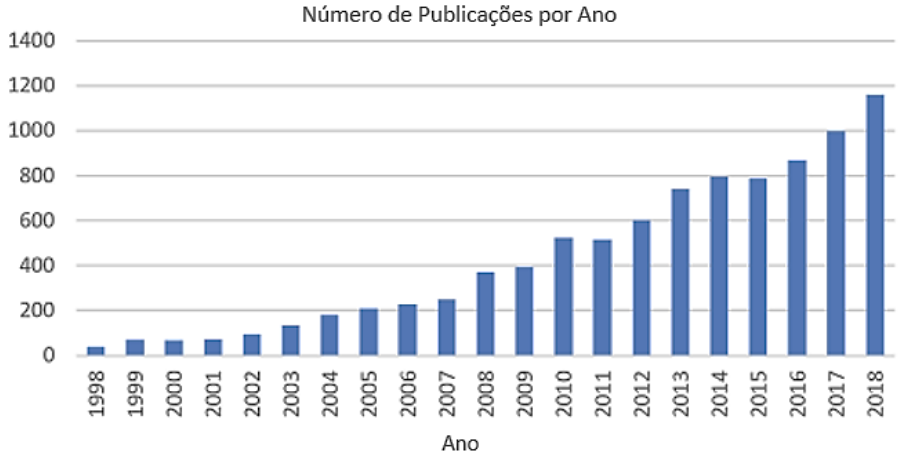
a) **Classificação:** envolve a atribuição de um rótulo de classe a uma determinada imagem. É normalmente utilizada quando se tem um único objeto e deseja-se classificá-lo;
b) **Localização**: envolve a verificação de onde um objeto se encontra naquela imagem;
c) **Detecção**: envolve a localização e classificação de vários objetos em uma imagem;
d) **Segmentação:** envolve a delimitação das regiões de ocorrências de objetos, criando uma máscara de pixels.

Sendo o objetivo do presente trabalho o reconhecimento de componentes situados em redes de distribuição de energia elétrica, o método utilizado é a detecção de objetos. Tal escolha se deve ao fato de as imagens digitais capturadas apresentarem mais de um equipamento em sua grande maioria, não havendo a necessidade de contorno de seus limites.

3.3.1 Técnicas de Detecção de Objetos

Nos últimos anos, o rápido desenvolvimento de técnicas de aprendizagem profunda atraiu pesquisadores de todo o mundo, trazendo notáveis descobertas referente à detecção de objetos e levando-a para um patamar de atenção sem precedentes (LECUN; BENGIO; HINTON, 2016). Ela tem sido amplamente utilizada em muitas aplicações do mundo real. A Figura 7 mostra o número crescente de publicações que estão associadas à detecção de objetos nas últimas duas décadas.

Figura 7 - Evolução do número de pesquisas referentes ao tema detecção de objetos.



Fonte: Zou e outros (2018).
Nota: Traduzido pela autora.

Atualmente, detectores de domínio pré-existentes são divididos em duas categorias: de múltiplos estágios, como a *Faster Region-Convolutional Neural Network* (*Faster* R-CNN); e de único estágio, como o *You Only Look Once* (YOLO) e o *Single Shot Detector* (SSD).

A rede *Faster* R-CNN foi publicada pela primeira vez em 2015. Na família R-CNN, a evolução entre as versões era geralmente em termos de eficiência computacional, redução no tempo de teste e melhoria no desempenho (KANG et al., 2016). Por outro lado, Kang e outros (2016) afirmam que ela continua sendo mais lenta quando comparada às de único estágio, apesar de obter uma precisão satisfatória ao serem analisadas por meio de algumas bases de dados em relação a outras redes.

A YOLO é uma técnica de detecção de objetos que tem ganhado muita atenção nos últimos anos. Após o seu lançamento em 2015, ela foi logo reconhecida como uma técnica inovadora. Por meio da utilização de uma nova abordagem, foi capaz de obter uma precisão igual ou superior aos outros métodos de detecção de objetos da época, porém com uma velocidade maior. Tal modelo realiza a regressão dos pontos correspondentes às caixas delimitadoras de cada objeto na imagem (localização) e realiza a predição a qual classe o objeto detectado pertence (classificação) com uma única rede em um estágio, permitindo previsões em tempo real (REDMON et al., 2016).

Semelhante ao modelo YOLO, Liu e outros (2016) desenvolveram o SSD para prever em apenas um estágio as caixas delimitadoras e as probabilidades de classe com uma arquitetura

de rede convolucional ponta a ponta. Ele apresenta alta precisão quando testado em alguns *datasets*, e mais vantagens em imagens de resoluções inferiores. No Quadro 1 é apresentado um comparativo entre os três métodos, a saber, o Faster R-CNN, a segunda versão do YOLO (YOLOv2) e o SSD, utilizando a base de treinamento *Pascal Voc* 2007[3].

Quadro 1 – Comparativo dos métodos Faster R-CNN, YOLO versão 2 e SSD

Método	Treinamento	Estágio	Precisão (%)	Tempo Real	Frames por segundo
Faster R-CNN	Voc2007	Múltiplos	73,2	Não	5
YOLO 2	Voc2007	1	76,8	Sim	67
SSD	Voc2007	1	79,2	Não	25,2

Fonte: Jeong, Park e Kwak (2017).
Nota: Traduzido pela autora.

Logo, por meio das informações apresentadas, ao considerar as limitações de processamento de um dispositivo de *hardware*, dentre as redes abordadas as de único estágio (YOLO e SSD) se mostram mais viáveis do que a Faster R-CNN. Dentre elas, a rede YOLO também apresenta vantagem quanto à utilização em aplicações de tempo real. Além disso, sua quarta versão lançada em 2020, ainda se mostra como o estado da arte sobre o conjunto de dados *Microsoft Coco*. Portanto, concluiu-se que, por tais motivos, a técnica YOLO seria a mais adequada para aplicação neste projeto.

3.4 YOLO

Assim como o próprio nome diz o Você Só Olha Uma Vez (do inglês, *You Only Look Once*), apresenta um método de detecção de objeto de passada única (do inglês, *single pass*), ou seja, se faz necessária olhar para uma mesma imagem uma única vez, diferentemente de outros métodos como a família R-CNN. Antes do surgimento do YOLO, sistemas de detecção utilizavam classificadores ou localizadores para encontrar as coordenadas do objeto na imagem, aplicando diferentes modelos e podendo realizar a divisão da imagem em várias partes, executando um classificador para cada uma por vez.

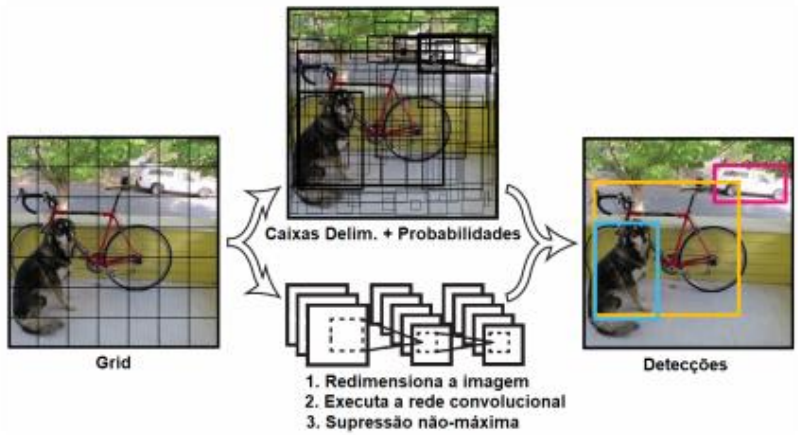
[3] O *Pascal Voc* 2007 assim como o *Microsoft Coco* são bases de dados que contém um conjunto de imagens utilizadas comumente para treinamento e avaliação das técnicas de detecção de objetos.

Na YOLO, inicialmente o algoritmo divide a imagem em um grid de S S células, sendo cada uma dessas células responsáveis pela predição de uma quantidade B de caixas delimitadoras (do inglês, *bounding box*), e possuem uma nota de confiança associada a cada uma das C classes que o modelo é capaz de detectar, nota esta obtida por uma composição da probabilidade com o IoU. Um dos principais conceitos da YOLO é construir uma rede para reduzir o volume de entrada de acordo com a Equação (1) (REDMON et al., 2016).

$$S \times S \times (B * 5 + C) \tag{1}$$

Além disso, as imagens de entrada são redimensionadas e uma única rede convolucional passa a ser executada, conforme Figura 8. Em relação aos parâmetros da caixa delimitadora, a mesma pode ser descrita a partir de 4 dimensões: centro da imagem (bx, by), largura (bw), altura (bh) e confiança citada anteriormente (REDMON et al., 2016). As coordenadas (bx,by) representam o centro da caixa relativa à delimitação da célula. A largura e altura da imagem são expressas pelo bw, bh (w e h, do inglês, *width* e *height*).

Figura 8 - Representação do sistema de detecção do YOLO.

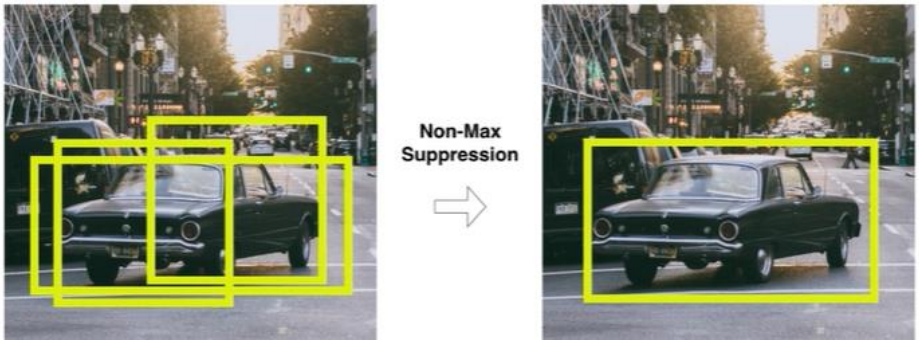


Fonte: Redmon et. al (2016).
Nota: Traduzido pela autora.

Partindo-se desse fato e levando em conta a geração de várias *bounding boxes* por célula, há a possibilidade de um mesmo objeto ser detectado mais de uma vez. Para contornar tal situação, o modelo utiliza a supressão de não-máximos (do inglês, *Non-Maximum Suppression*), conforme ilustrado nas Figura 9, a qual consiste na ordenação das predições com base em suas notas de confiança, realizando inicialmente as detecções de maior confiança ignorando as

subsequentes com sobreposição maior que 50% na mesma classe e que possuírem confiança menor.

Figura 9 - Representação da supressão não-máxima.



Fonte: Redmon et. al (2016).

Para esse trabalho foi escolhida a versão 4 do YOLO (YOLOv4), o qual tem se mostrado ser um detector promissor por equilibrar precisão com velocidade e por apresentar uma melhora de AP e FPS em 10% e 12%, respectivamente, em relação à sua versão anterior (BOCHKOVSKIY; WANG; LIAO, 2020). Há uma atualização ainda não oficial, YOLOv5, contudo, resultados preliminares indicam que a precisão média da versão YOLOv4 é ainda superior (NELSON; SOLAWETZ, 2020).

Segundo Bochkovskiy, Wang e Liao (2020), sua arquitetura consiste em:

***Backbone*:** *CSPDarknet53* (ou apenas *Darknet*);

***Neck*:** SPP e *Path Aggregation Network* (PANet);

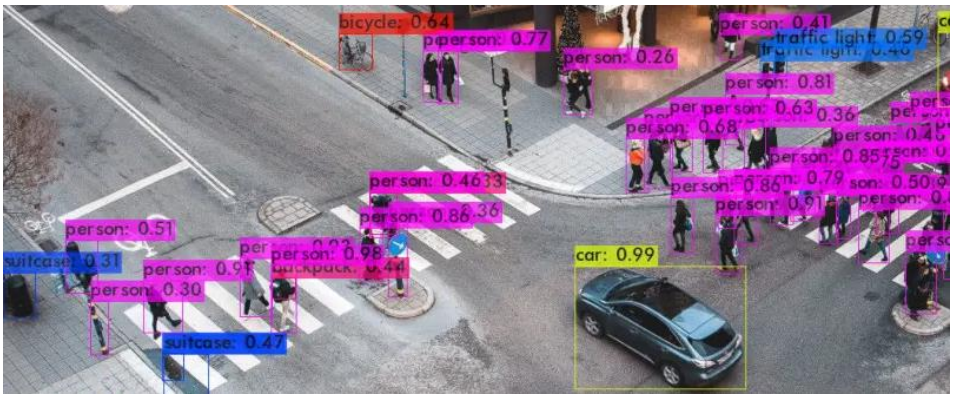
***Head*:** YOLOv3.

Sendo assim, a YOLOv4 utiliza o *backbone Darknet* para extrair o mapa de características das imagens de entrada. Com o intuito de diminuir o número de parâmetros, o *Darknet* realiza a separação do mapa de características em duas partes para que apenas uma delas passe por uma camada de convolução densa (CHEN; ZHONG; ZHANG, 2021). Com o objetivo de melhorar a detecção de objetos em diferentes escalas, há o uso dos blocos SSP e PANet, sendo o primeiro responsável pela extração de características multi-níveis e o segundo pela concatenação dos mapas de características vinda do fluxo do *backbone* (CHEN; ZHONG; ZHANG, 2021).

Para que se obtenha as coordenadas resultantes das *bounding boxes*, assim como a pontuação de confiança por classe correspondente, a YOLOv4 incorpora o modelo YOLOv3 como bloco

de predição (BOCHKOVSKIY; WANG; LIAO, 2020). A Figura 10 apresenta um exemplo de saída de detecção utilizando o modelo YOLOv4.

Figura 10 - Exemplo de saída da detecção realizada pelo YOLOv4.



Fonte: IAAcademy (2020).
Nota: Os objetos detectados na figura são referentes a carros (do inglês *cars*), pessoas (do inglês *persons*), bicicletas (do inglês *bicycles*), semáforos (do inglês *traffic lights*), malas de viagem (do inglês, *suitcases*), e uma mochila (do inglês, *backpack*) conforme mostrado.

3.5 Trabalhos correlatos

Diversos estudos que utilizam técnicas de visão computacional e aprendizado profundo vêm sendo desenvolvidos (ARAI; KAPOOR, 2019). Especificamente no âmbito do setor de energia a aplicação de detectores de objetos pode ser encontrado cada vez mais frequente na última década. Em Fu et al. (2018), por exemplo, os autores aplicaram o método para detecção e reconhecimento inteligente de quadros de distribuição de alta tensão. O sistema proposto tem como objetivo identificar o posicionamento de uma chave, assim como reconhecer e determinar seu estado.

Na pesquisa desenvolvida por Lin et al. (2016) foi discutido que o aumento substancial de imagens de infravermelho no sistema elétrico de subestações apresenta um novo desafio, exemplificando a forma como a avaliação do status dos dispositivos é feita tradicionalmente. Para superar esse problema, o trabalho de Lin et al. (2016) propõe um método de análise automática de imagens de infravermelho. O algoritmo implementado primeiro segmenta imagens em *superpixels* e, em seguida, adota redes neurais convolucionais para classificação dos objetos. Foram alcançados bons resultados em comparação com outros métodos de treinamento não supervisionado testados.

Oliveira et al. (2020) aplicaram o YOLO no contexto de monitoramento para gestão de obras em alimentadores de subestação, realizando a identificação de quatro componentes, sendo eles: suporte de concreto para painel, painel, grade de proteção e transformador de corrente. Os resultados obtidos em torno de 98% para o valor de mAP em dados de validação demonstraram a eficácia do detector Yolo para o estudo e aplicação em questão.

Olivo, Junior e Filho (2020) abordaram a importância das manutenções em equipamentos de redes de transmissão de energia e utilizaram o YOLOv3 para detecção desses equipamentos, apresentando resultados com mAP referente a 84,16% mesmo usando uma base de 1268 imagens, capturadas por *drones*, e objetos com diferentes modelos, cores e tamanhos. Em trabalhos futuros concluíram que a composição de uma base de imagens maior e a utilização de técnicas de pré-processamento, poderiam aprimorar os resultados.

Nota-se, a partir dos exemplos apresentados, que é demonstrado grande avanço do emprego da inteligência artificial nesse contexto, haja vista a importância que tecnologias como o reconhecimento de objetos vem ganhando na melhoria na qualidade nas redes elétricas. Sendo assim, a elaboração do presente trabalho é embasada no intuito de contribuir com tal evolução.

4 METODOLOGIA

4.1 Metodologia

No que tange à natureza da presente pesquisa, a mesma pode ser classificada como aplicada, uma vez que “objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos” (PRODANOV; FREITAS, 2013, p. 51). Do ponto de vista de seus objetivos, é classificada como uma pesquisa exploratória, visto que serão gerados conhecimentos sobre o assunto com o intuito de avançar no desenvolvimento de um trabalho de maior abrangência. Quanto à forma de abordagem do problema, é caracterizada por uma pesquisa quantitativa, pois serão utilizados modelos e análises matemáticas para verificar a sua assertividade. Igualmente, do ponto de vista dos procedimentos técnicos, observa-se que esta é uma pesquisa experimental, por estudar profundamente acerca da problemática abordada de forma a contribuir em sua resolução.

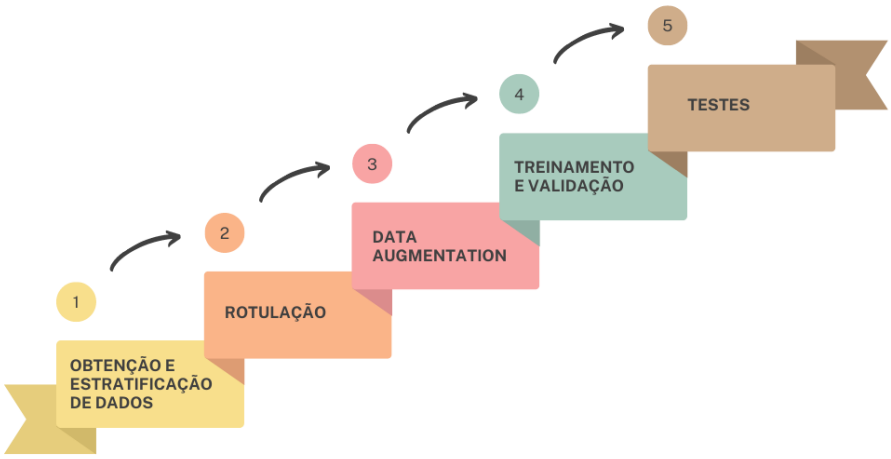
A metodologia proposta neste trabalho, como forma da implementação da detecção de componentes de rede de energia elétrica, partirá do aprofundamento dos campos da ciência, métodos e técnicas já abordados, em busca da obtenção de um embasamento consistente para atingir o objetivo da pesquisa e de suas etapas de desenvolvimentos, as quais também serão abordadas nessa mesma seção. Serão utilizadas opções metodológicas referentes à abordagem hipotética-dedutiva, uma vez que “(...) inicia-se com um problema ou uma lacuna no conhecimento científico, passando (...) por um processo de inferência dedutiva, o qual testa a predição da ocorrência de fenômenos abrangidos pela referida hipótese. ” (PRODANOV; FREITAS, 2013, p. 32). Quanto aos procedimentos, serão utilizados métodos experimentais, pois serão levantados dados para análise dos experimentos, assim como para validação dos resultados obtidos.

4.2 Etapas de Desenvolvimento

Em primeira instância, foi realizada a análise bibliográfica em busca da fundamentação técnica, para aprofundamento com relação ao estado da arte, assim como capacitação para utilização das ferramentas necessárias para implementação de todo o trabalho. Para tanto, foi reunido um grande conjunto de material de apoio, outrossim foi utilizado o conhecimento adquirido por

meio da realização da disciplina de Visão Computacional, ministrada pela orientadora deste trabalho, assim como de conhecimentos difundidos por meio de seminários realizados *online* semanalmente pelos participantes do laboratório LabVisio, os quais abordaram diversos temas e trabalhos pertinentes à referente pesquisa. Os conhecimentos adquiridos foram essenciais durante a consolidação do conjunto de dados, na escolha da melhor técnica de rotulação e sua execução, assim como na utilização do *Darknet* como o *framework.* Por fim, foi possível avaliar a qualidade e a assertividade da metodologia adotada, por meio da utilização de métricas de validação e teste do modelo. A Figura 11 mostra um resumo das etapas para implementação do modelo proposto.

Figura 11 - Fluxograma de etapas de elaboração do projeto.



Fonte: Produzido pela própria autora.

É válido salientar que as etapas de trabalho apresentadas a seguir foram precisamente documentadas e registradas, no intuito de relatar de forma fidedigna todo o processo de desenvolvimento e implementação do projeto, bem como na obtenção dos resultados os quais estão apresentados neste documento.

4.3 Obtenção e estratificação de dados

Para que seja possível o treinamento, teste e validação da RNA, e ter como resultado a detecção de objetos, faz-se necessária a utilização de um conjunto de dados. Tais dados são compostos por imagens as quais passam informações pertinentes ao aprendizado da rede. As fotos utilizadas para esses estudos podem ser reunidas por meio de três tipos de processos, os quais são listados abaixo:

Criação a partir de outros conjuntos de dados disponíveis em grande escala, como o ImageNet (RUSSAKOVSKY et al., 2015), Common Objects in Context (COCO) (LIN et al., 2014) e o Open Image Dataset (KUZNETSOVA et al., 2018);

Coleta dos dados a partir de técnicas de rastreamentos na Web (OLSTON, 2010);

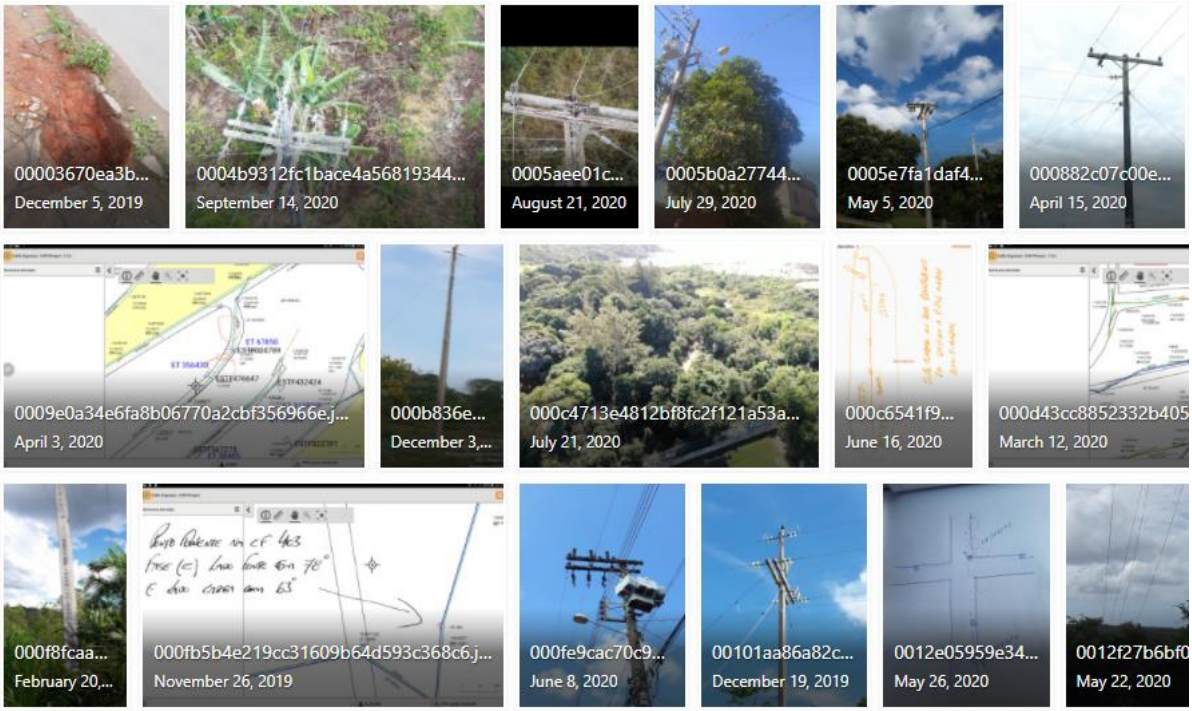
Realização da aquisição dos dados em campo, capturados no local da aplicação pela autora, pelos engenheiros ou pelos técnicos.

Tendo em vista inicialmente que os objetos a serem detectados nas imagens tratam-se de equipamentos de rede, concluiu-se que a utilização de *datasets* genéricos disponíveis não seria viável. Sendo assim, foram considerados os outros dois tipos de processo de obtenção desse conjunto. A partir da premissa de que a utilização de imagens próprias da rede de distribuição de energia elétrica, a qual é estudo de caso do presente trabalho, podem contribuir para a elevação da qualidade dos resultados, foi realizado contato com a área de Planejamento da Manutenção da EDP Espírito Santo. A área é responsável pela realização das inspeções dos equipamentos ao longo das linhas e pelo planejamento e execução das correções das falhas encontradas. Um conjunto de dados com um total de cerca de 60.000 imagens e 11 vídeos gravados por *drones*, com média de cerca de 4 minutos de duração cada, foram disponibilizados pela EDP para implementação do método proposto.

Iniciou-se a análise para determinação das classes, sendo escolhidos os objetos cruzeta, poste e transformador como elementos a serem detectados nas imagens ao longo do desenvolvimento da pesquisa. Tal escolha partiu da análise do conjunto e verificação de ocorrência dos objetos, contando também com a contribuição do time EDP.

Após a análise dos dados, foi verificado que apresentavam muitas repetições, fotos sem grande definição e muitas imagens de diagramas da rede, como é possível verificar na Figura 12.

Figura 12 - Representação de amostra da base de dados, apresentando grande diversificação de imagens.



Fonte: Produzido pela própria autora.

De acordo com Yao et al. (2019) e Russakovsky et al. (2015), o processo de estratificação das imagens para o banco de dados deve levar em consideração três fatores: escalabilidade, diversidade e acurácia.

A escalabilidade diz respeito a uma distribuição homogênea do número de imagens de acordo com as diferentes classes, sendo esse critério adotado na seleção das imagens. Já o conceito de diversidade está intimamente ligado à variabilidade de posicionamento, aos diferentes fundos e pontos de vistas dos objetos. Sendo assim, na filtragem realizada manualmente do conjunto foram considerados os seguintes critérios qualitativos de diversificação na seleção das fotos: nível de iluminação; qualidade da imagem; distância dos objetos; sobreposições de mais de um componente na mesma imagem; posição e angulação; corte de objetos e diferentes modelos e estado (com avarias ou não) do equipamento.

Durante o processo de análise verificou-se que não haviam muitas imagens à noite. Dessa forma, foram então coletadas em campo fotos no período noturno para compor a base. Além disso, também foram selecionadas e coletadas imagens de diferentes fundos, as quais não foram

rotuladas na próxima etapa, de forma a fornecer à rede informações de outros objetos que não devem ser identificados, buscando a redução da incidência de falsos positivos nos resultados. Como exemplo, a Figura 13 foi escolhida por conter imagens de árvores e vegetações, que comumente são encontradas próximas a postes de rede distribuição, podendo gerar confusão à rede.

Figura 13 – Imagem de fundo apresentando troncos de árvores, semelhantes a postes

Fonte: Produzido pela própria autora.

Por sua vez, a acurácia das fotos que compõem o banco pode ser entendida como um erro de indexação nos sistemas de busca de imagens (Yao et al., 2019). Todavia, como as técnicas utilizadas para levantamento das mesmas não envolveram buscas em sites de pesquisa, no método escolhido a acurácia consiste em selecionar imagens que contenham os objetos a serem detectados entre as disponíveis previamente. Ao final dessa etapa, foram reunidas 2000 fotos, que compõe a base estratificada. Na Figura 14, é possível visualizar algumas imagens exemplo da base estratificada.

Figura 14 – Exemplo de imagens que compõe base final estratificada.

Fonte: Produção da própria autora.

4.4 Rotulação

No que tange à área do aprendizado de máquina, a rotulação consiste no processo de identificar dados brutos (imagens, áudios, vídeos, etc.) e adicionar um ou mais rótulos significativos e informativos para fornecer contexto de forma que um modelo possa aprender com eles. Por exemplo, os rótulos podem indicar se uma foto contém um determinado objeto, ou quais palavras foram pronunciadas em uma gravação de áudio.

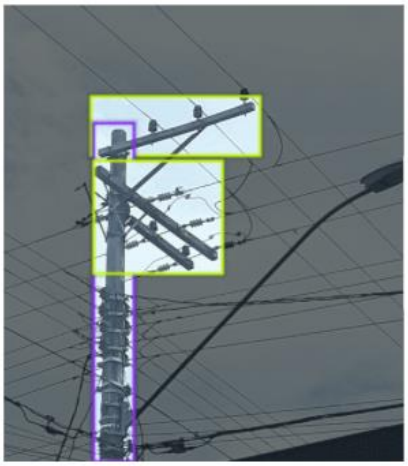
Sendo assim, após reunir o conjunto de imagens, foi dado início ao processo de rotulação dos dados. A ferramenta gráfica de rotulagem escolhida para a implementação dessa etapa foi a *LabelImg*, desenvolvida em Python, a qual permite a criação das caixas delimitadoras ao redor das ocorrências dos objetos. Ela gera um arquivo contendo a indicação da classe associada e as respectivas coordenadas do rótulo criado (TZUTA, 2018). Originalmente, o *LabelImg* gera um arquivo de rótulo para *Imagenet*, do tipo XML, mas é possível escolher o tipo TXT compatível com o YOLO *Darknet*. O arquivo de saída possui o mesmo nome atribuído à imagem passada e apresenta informação das coordenadas dos vértices dos *bounding boxes* criados e atribuídos a cada um dos objetos presentes na imagem, sendo adotada uma referência de nomenclatura para distinção das classes, conforme o Quadro 2. O formato do arquivo TXT gerado é possível visualizar na Figura 15, assim como a respectiva imagem rotulada.

Quadro 2 - Referências de classes adotadas durante a rotulação dos objetos.

Classe	Referência para o modelo
Cruzeta	0
Transformador	1
Poste	2

Fonte: Produção da própria autora.

Figura 15 - Exemplo de imagem rotulada à esquerda e seu respectivo arquivo gerado à direita.



```
0 0.431577 0.374694 0.428513 0.119485
0 0.385825 0.541054 0.331291 0.215686
2 0.275123 0.680147 0.109886 0.639706
```

Fonte: Produzida pela própria autora.

A etapa de rotulação vale-se de um processo manual o qual demanda uma interferência humana e os resultados podem variar a cada nova rotulação, sendo assim buscou-se manter o padrão no momento de criação das caixas delimitadoras. Diante disto, a fim de diminuir as divergências, algumas medidas foram adotadas como forma de padronização.

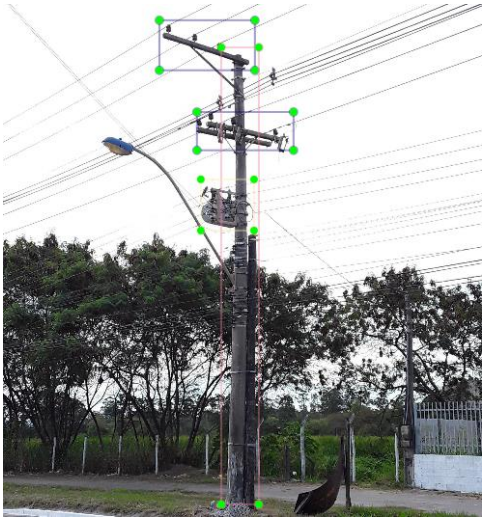
Em imagens que apresentavam a ocorrência de equipamentos que estivessem extremamente distantes da câmera, ou que apresentassem forma não definida pela baixa resolução, e que com isso se distanciavam muito do aspecto do restante do conjunto, foram desconsiderados no processo de rotulação. Isto se deve ao fato de no contexto de inspeção dos equipamentos, para que seja feita análise de uma falha é necessária uma mínima aproximação. O uso de imagens muito distantes pode elevar a possibilidade de identificação errônea de objetos. Sendo assim, nessas situações foi realizada uma análise qualitativa, com relação à qualidade da informação que o equipamento prestava. Na Figura 16, é possível visualizar exemplos desses casos.

Figura 16 - Frame retirado de filmagem feita por *drone* para exemplificar postes e cruzetas à longa distância.

Fonte: Produzido pela própria autora.

Além disso, vale ressaltar que para os casos de cruzetas, foi considerado como partes rotuláveis a própria cruzeta e os isoladores nela presentes uma vez que, com a variação dos ângulos das imagens, em certos casos eram impossíveis a rotulação sem enquadrar em alguns momentos alguns isoladores. Assim como no caso das cruzetas, para o transformador foi considerado como parte rotulável o próprio transformador e as buchas de ligações neles presentes. Não obstante, para o poste, toda a sua extensão foi considerada, mesmo que pudesse se sobrepor aos *bounding boxes* criados para as outras classes. A Figura 17 ilustra exemplos das rotulações, conforme padronizações adotadas.

Figura 17 - Rótulos de classes conforme considerações adotadas.



Fonte: Produzido pela própria autora.

4.5 Data Augmentation

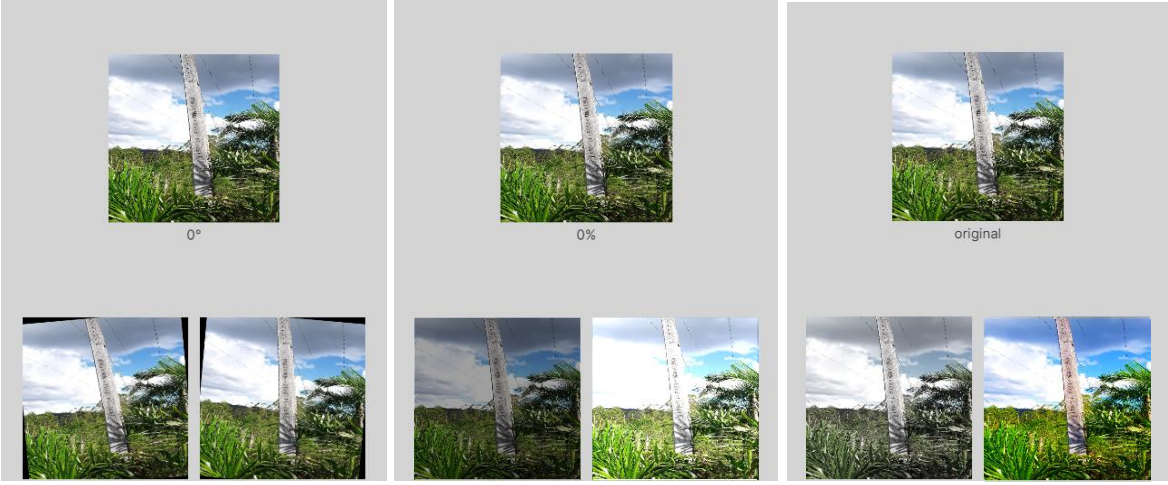
Para treinamentos de RNA são amplamente utilizadas técnicas de pré-processamento. Uma dessas técnicas, quase universalmente utilizada em modelos de *deep learning* aplicado a imagens, denominada *data augmentation*, apresenta como finalidade garantir maior diversificação ao conjunto de dados e, segundo Chollet (2018), mostra-se como uma forma poderosa para minimizar a possibilidade de ocorrência de sobreajuste (do inglês, *overfitting*). Os casos em que *data augmentation* normalmente é aplicado são aqueles em que o modelo apresenta dificuldades para generalizar a aprendizagem para novos dados, apresentando desempenho inferior ao medido durante o treinamento.

O emprego dessa abordagem consiste na expansão do conjunto de treinamentos, gerando novos dados a partir dos originais, por meio de transformações aleatórias, quanto ao tipo e número, durante o treinamento, resultando em imagens ainda não conhecidas pelo modelo. O emprego da técnica de *data augmentation online* foi escolhido de forma a garantir maior aleatoriedade durante o fornecimento de dados na aprendizagem do modelo. Ao avaliar as técnicas elencadas por Shorten e Khoshgoftaar (2019), considerando a base de dados existente e suas peculiaridades, além do contexto de inspeções, foram selecionados métodos de rotação, brilho e saturação.

O método de rotação se resume em girar a imagem num ângulo pré-definido a fim de tornar o modelo menos sensível à orientação dos objetos na imagem de acordo com a posição em que a lente da câmera se encontrou posicionada, por exemplo. No caso de postes, a rotação também se mostra interessante pelo fato de comumente postes apresentarem abalroamento. Para o caso do brilho, ao adicioná-lo às imagens, o modelo tende a ser mais eficiente para detectar objetos com luminosidades variadas no momento da captura.

Por último, o ajuste da saturação de uma imagem pode ajudar o modelo a ter um melhor desempenho quando as cores são diferentes (por exemplo, se um balanço de branco diferente for definido a partir da variação entre diferentes câmeras, ou se a iluminação ambiente estiver diferente, como seria o caso em fotos contra o sol, por exemplo, ou com alto nível de reflexo de luz. Na Figura 18, é possível visualizar exemplos das transformações a serem aplicadas aos dados de treinamento.

Figura 18 - Exemplo de rotação, alteração de brilho e saturação, respectivamente.



Fonte: Produção da própria autora.

Para realizar a implementação do *data augmentation* descrito, utilizam-se os parâmetros *angle* (referente à rotação), *exposure* (referente ao brilho) e *saturation* (referente à saturação), que geram imagens aplicando transformações aleatórias, dentro dos limites definidos, as quais são aplicadas durante o treinamento, a cada lote de dados retirado do banco original, antes de ser inserido na rede. Assim, a cada época de treinamento, a rede tem acesso a um conjunto de imagens modificado, favorecendo a qualidade de generalização e limitando as possibilidades de sobreajuste do modelo.

4.6 Implementação do treinamento

Após a composição dos bancos de imagens rotuladas, inicia-se a etapa de implementação do treinamento da RNA. Para tanto, foi utilizada a YOLOv4, apresentando o *backbone* CSN*Darknet*, de Bochkovskiy, Wang e Liao (2020b), a qual consiste em uma versão aprimorada da *Darknet* original desenvolvida pelos criadores da YOLOv3 (VAHTRA; ANBARJAFARI, 2019). O algoritmo YOLOv4 pode ser implementado de diferentes formas, sendo a *Darknet* uma das estruturas de rede neural de código aberto mais utilizada como *framework* dessa aplicação. Ao todo, foram utilizadas 53 camadas de rede neural *Darknet*, além de 53 camadas adicionais para detecção (REDMON; FARHADI, 2018).

Iniciando o processo de implementação do treinamento, em um primeiro momento é necessário fornecer à rede as informações das classes de objetos, o número total de classes, além dos

diretórios necessários a serem utilizados. O próximo passo é referente à configuração e aos ajustes da rede *Darknet*, sendo a seguir apresentado alguns dos parâmetros ajustados:

Largura e Altura (do inglês, *Width* e *height*)*:* Referente ao tamanho das imagens do conjunto de dados, em *pixels*;

Subdivisões (do inglês, *Subdivision*): Divide os dados de treinamento em pequenos lotes para manejo da memória computacional;

Tamanho do lote (do inglês, *Batch Size*): Usado para referir-se ao número de exemplos de treinamento utilizados em uma iteração do algoritmo;

Número de classes (N): Referente ao número total de classes utilizadas no modelo;

Número de filtros: Obtido a partir da equação

$$Filtros = (N + 5) \times 3$$

Lote Máximo (do inglês, *max_batches*): Referente ao número máximo de iterações de treinamento da rede, podendo o limite mínimo obtido a partir da seguinte equação:

$$max_batches = 2000 \times N$$

Passos (do inglês, *Steps*): Definido pela expressão

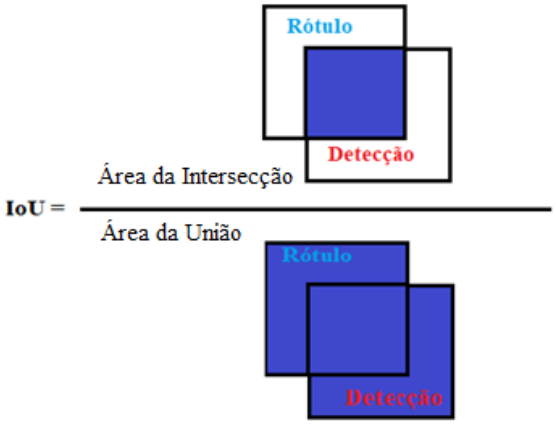
$$Steps = (0{,}8 \times max_batches), (0{,}9 \times max_batches)$$

4.6.1 Métricas de Desempenho

Para implementar um modelo de aprendizado de máquina é necessário realizar a validação enquanto ainda é treinado de forma a acompanhar sua estabilidade e aprendizado, buscando evitar subajustes ou sobreajustes à base de treino. Além disso, após o fim da aprendizagem do modelo, é indispensável realizar sua avaliação, a partir da implementação de um teste em um novo conjunto de dados totalmente desconhecidos, com o objetivo de verificar o desempenho e eficiência do modelo.

A Interseção sobre União (IoU, do inglês, *Intersection over Union*) está diretamente relacionada à capacidade de o detector localizar bem os objetos (PINTO,2019). A Figura 19 ilustra a determinação do IoU por meio de uma representação das *bounding boxes* bem como da sua definição matemática (WU; SAHOO; HOI, 2020).

Figura 19 - Representação da definição matemática do conceito de IoU.



Fonte: Olivatto (2021).

As outras métricas surgem a partir das informações produzidas pela IoU. A partir de um determinado valor de *threshold*, comumente utilizados os valores 0,50 e 0,75 (PADILLA; NETTO; DA SILVA, 2020), é possível determinar se uma detecção é considerada correta (IoU ≥ *threshold*) ou incorreta (IoU < *threshold*). A partir dessa informação é possível obter as seguintes divisões:

Verdadeiro Positivo (TP, do inglês, *True Positive*): uma detecção correta de uma *bounding box* verdadeira;

Falso Positivo (FP, do inglês, *False Positive*): uma detecção incorreta de um objeto inexistente ou a detecção errada de um objeto existente;

Falso Negativo (FN, do inglês, *False Negative*): uma não detecção de uma *bounding box* verdadeira. [4]

A partir disso, é possível compreender matematicamente, os conceitos de *precision* e *recall*, necessárias para o entendimento e cálculo da métrica de Precisão Média (AP, do inglês, *Average Precision*). A *Precision* está diretamente relacionada à capacidade de o modelo de treinamento identificar apenas objetos relevantes. Qualitativamente, pode ser compreendido como o percentual das detecções que são corretas (PADILLA; NETTO; DA SILVA, 2020). Matematicamente, é dado por:

[4] Nota: Para as métricas não são considerados os Verdadeiros Negativos (TN, do inglês, *True Negative*), uma vez que serão obtidos infinitos resultados (PADILLA; NETTO; DA SILVA, 2020).

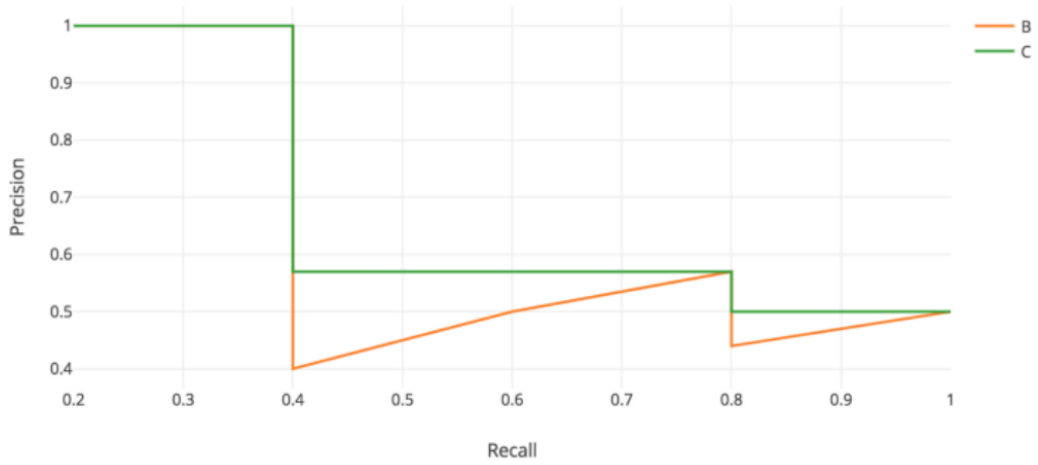
$$P = \frac{TP}{TP+FP} = \frac{TP}{Todas\ detecções\ realizadas} \tag{1}$$

Em contrapartida, o conceito de *Recall* está relacionado com a capacidade do modelo identificar todos os objetos corretamente. Qualitativamente, pode ser compreendido como o percentual dos objetos que foram realmente detectados (PADILLA; NETTO; DA SILVA, 2020). Matematicamente, é dado por:

$$R = \frac{TP}{TP+FN} = \frac{TP}{Todos\ bounding\ boxes\ presentes} \tag{2}$$

Mediante as equações supracitadas, obtém-se uma curva de relação entre o *precision* e o *recall*, a qual, segundo Saito e Rehmsmeier (2015), permite uma avaliação precisa e interpretação intuitiva do desempenho do classificador. O cálculo de precisão média por classe pode ser visto como a área de um gráfico suavizado de *precision* por *recall*, um exemplo deste gráfico mostrado na Figura 20.

Figura 20 – Representação da curva suavizada (C) e curva real (B).



Fonte: Nogueira (2018).

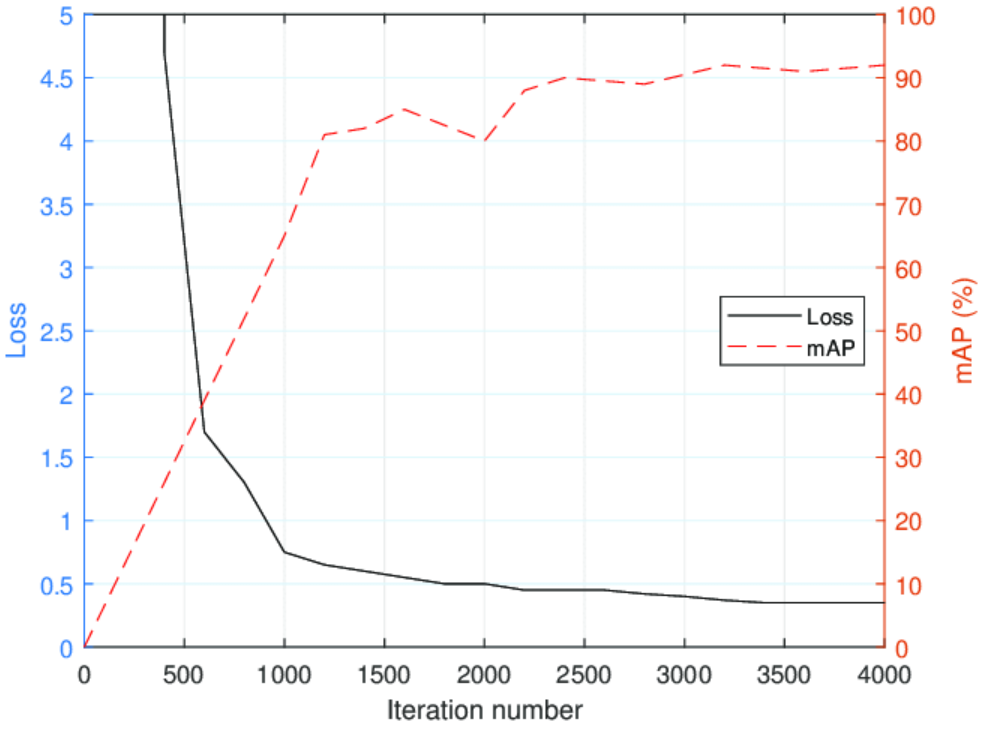
A partir da curva de precisão suavizada calcula-se a precisão média (AP) por classe com base na Equação 4, sendo N o número de predições e P o valor de *precision* para cada r (*recall*).

$$AP = \frac{1}{N} \sum_{r \in (0,\ldots,1)} P(r) \tag{4}$$

A AP apresenta uma extensão, denominada Média das Precisões Médias (mAP, do inglês, *Mean Average Precision*), a qual é utilizada para se obter a acurácia do detector de objeto levando em consideração todas as classes treinadas, e é dada pela média simples das AP's de cada classe.

Durante o processo de treinamento e validação do modelo, utilizou-se a mAP e a *Loss* (perda) para acompanhamento do modelo, como exemplificado no gráfico na Figura 21. Tal gráfico representa a evolução do mAP, à medida que o modelo evolui no número de iterações do treinamento, apresentando comportamento inverso à perda medida. A análise deste gráfico permite encontrar o melhor momento em número de iterações, para se finalizar o processo de aprendizado, minimizando a possibilidade de ocorrência do fenômeno de *overfitting*.

Figura 21 - Gráfico ilustrativo do comportamento do mAP e *Loss*, representando o desempenho de um modelo, frente ao número de iterações realizadas para treinamento da rede.



Fonte: Kim et al. (2020).

5 RESULTADOS E DISCUSSÕES

5.1 Treinamento

Para o desenvolvimento destas etapas foi utilizado um computador com placa gráfica (GPU, do inglês *Graphics Processing Unit*) NVIDIA TITAN V, disponibilizado pelo LabVISIO, por meio de acesso remoto via SSH (do inglês, *Secure Shell*). Para a rotulação dos dados o *software LabelImg* foi escolhido. Os códigos foram desenvolvidos em Python 3.6.9, utilizando o *framework Darknet* e o método de detecção YOLOv4.

Após a estratificação de imagens, o conjunto de dados final ficou composto por 2000 imagens com alto nível de diversificação de características. Foram levantadas ao todo 1900 que incluem imagens de *frames* selecionados de vídeos de *drones*, imagens coletadas pela própria autora no período noturno e as demais selecionadas que compunham o banco. As 100 imagens restantes apresentam diferentes fundos não rotulados para compor o conjunto de treinamento, buscando a redução da possibilidade de falsos positivos em outros objetos presentes das imagens, uma vez que os fundos apresentam grande variabilidade no contexto de aplicação do trabalho. Na Tabela 1, é apresentado um resumo do que foi descrito.

Tabela 1 - Resumo de tipos de imagens selecionadas.

Imagens	Número de imagens	Representação (%) do total
Período noturno	100	5%
Imagens de fundo	100	5%
Frames de vídeos de *drone*	400	20%
Demais imagens diversas	1400	70%
Total	2000	100%

Fonte: Produzida pela própria autora.

A base de dados foi então dividida em dados de treinamento, validação e teste, buscando manter a variabilidade dos dados, na proporção de 80% para treinamento, 10% para validação e 10% para teste (Tabela 2). Ao todo foram escolhidas 1600 imagens para treinamento, 200 para validação e 200 para teste, totalizando as 2000 imagens.

Tabela 2 - Divisão de imagens por etapa.

Etapa	Número de imagens	Representação (%) do total
Treinamento	1600	80%
Validação	200	10%
Teste	200	10%

Fonte: Produzida pela própria autora.

Após a conclusão da rotulação das imagens, conforme mostrado na Tabela 3, foi verificado um desbalanceamento entre a quantidade de rótulos utilizados na base de treinamento. A partir disso, resultados inferiores poderiam ser gerados para a classe de transformador, possibilitando a presença de maiores ocorrências de falsos positivos, ou falsos negativos para essa classe. Isso deve ao fato do total de imagens adquiridas apresentar menor incidência de transformadores, uma vez que existem uma quantidade muito maior de cruzetas e postes ao longo da rede de distribuição de energia.

Tabela 3 - Representação do balanceamento de classes

Classe	Número de rótulos	Pesos de balanceamento
Cruzeta	1674	1,07
Transformador	558	3,20
Poste	1785	1,00

Fonte: Produzido pela própria autora.

Sendo assim, foi realizada a inserção de parâmetros responsáveis pela minimização dos efeitos gerados pelo desbalanceamento de classes. Tal parâmetro denominado *focal_loss* é responsável pela modificação da função custo, de forma que a classe com maior peso receba maior ajuste de nível de importância, por meio da inserção dos pesos presentes na Tabela 3. Os valores são obtidos partindo de uma normalização considerando como referência a classe dominante no conjunto.

Com base nas informações apresentadas no item 4.6 deste trabalho, foram adequados os parâmetros da rede para atender as condições para o modelo proposto, conforme indicado no Quadro 3. Além da configuração de tais parâmetros, foi também utilizada a incorporação de pesos convolucionais pré-treinados, uma vez que podem otimizar o tempo de treinamento e elevar a acurácia final, assim como afirmaram He, Girshick e Dollar (2019).

Os valores teóricos e ajustados referentes a cada um dos parâmetros adequados ao modelo proposto, bem como os valores padrões para o arquivo disponível por AlexeyAB (BOCHKOVSKIY; WANG; LIAO 2020b) são exibidos no Quadro 3. O valor de *max_batches* foi ajustado para 9000, haja vista que o valor mínimo não pode ser inferior a 6000, valor este teoricamente adotado para 3 classes. Consequentemente, também houve alteração no parâmetro *steps* devido à relação com o *max_batches*.

Quadro 3 - Parâmetros padrões da rede e utilizados para o modelo proposto.

Parâmetro	Valor Padrão (AlexeyAB)	Valor teórico para modelo
Width e Height	416	608
Subdivisions	1	64
Batch Size	1	64
Número de Classes	80	3
Números de Filtros	255	24
Max_Batches	500500	9000
Steps	400000, 450000	7200, 8100
Focal_loss	-	[1.07 3.2 1]

Fonte: Produção da própria autora.

Considerando as informações supracitadas em conjunto com as descritas na Seção 4 deste trabalho, foram realizados 2 treinamentos distintos considerando as diferentes técnicas empregadas, os quais são descritos na Tabela 4. Para o treinamento aplicando-se a técnica de *data augmentation online*, foi utilizado um valor de 5º para ângulo de rotação e, com base nos experimentos conduzidos por Zaworski (2018), adotou-se variação de 20% no brilho e saturação.

Tabela 4 - Configuração para os diferentes treinamentos.

Nº Treinamento	*Data Augmentation*	Balanceamento de Classes
1	Não	Sim
2	Sim	Sim

Fonte: Produzida pela própria autora.

5.2 Desempenho da rede

Como citado na subseção anterior, foram realizados dois diferentes treinamentos variando-se parâmetros e configurações da rede. Durante a implementação do treinamento do modelo, a

cada 100 iterações, eram realizados testes de validação intermediários que forneciam as métricas de validação que apresentam informações referentes à eficiência do treinamento, sendo possível assim determinar em qual momento do aprendizado o modelo estava obtendo melhor desempenho. Além disso, ao final do processo, a rede fornecia um arquivo de pesos com os melhores resultados, além do modelo obtido a cada 1000 iterações.

Para etapa de testes e verificação do desempenho da rede, foram utilizadas 200 imagens ainda não vistas pela mesma. As métricas utilizadas para validação do modelo proposto neste trabalho foram AP, mAP e IoU médio. Todavia, a métrica AP representa pouco o desempenho da rede como um todo, sendo mais interessante para avaliação das classes individualmente. Além disso, as métricas *Precision* e *Recall*, citadas também no Capítulo 4 para entendimento do AP, também não demonstram grandes análises distintas, uma vez que, a área suavizada sob o gráfico que relaciona as duas métricas fornece o valor do mAP, gerando uma análise redundante.

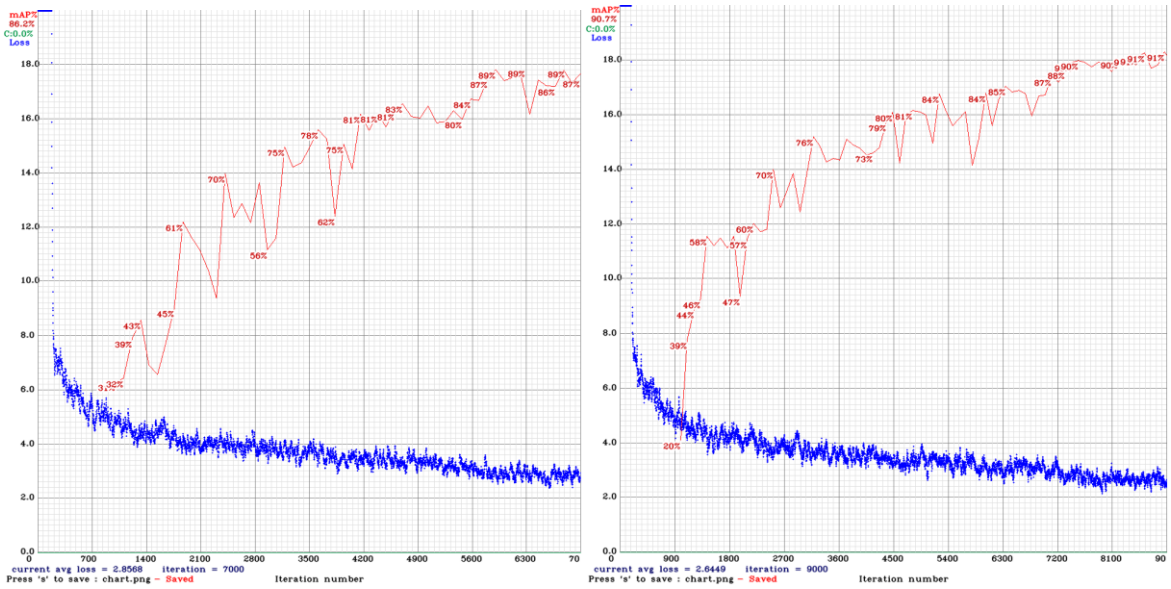
No modelo do *framework Darknet* o cálculo para os valores de mAP são realizados a cada 4 épocas. Considerando que o banco de dados para treinamento continha 1600 imagens e que o valor do *Batch size* era igual a 64, tem-se que uma época é igual a 25 iterações, justificando assim o valor de 100 iterações citado no começo dessa subseção.

A partir dos gráficos de curvas de mAP e *Loss* gerados durante o treinamento (vide Figura 22), foi realizado o acompanhamento do desempenho do modelo. Ao acompanhar o comportamento de evolução do mAP sobre o conjunto de validação e de *Loss* do conjunto de treinamento, é possível verificar o momento no qual a rede pode passar a apresentar indícios de *overfitting*. No caso do Treinamento 1, tal momento é referente a região a partir das 5800 iterações, uma vez que valores de mAP começam a apresentar estagnação, enquanto a curva de *Loss* apresenta valores inferiores a 3.

Não obstante, para o segundo treinamento, foi realizado processo análogo. Nota-se que através do emprego de técnicas de *data augmentation* a rede apresentou o mesmo comportamento citado anteriormente a partir de valores de mAP em 7500 iterações, retardando a ocorrência de sobreajustes e elevando o desempenho do modelo. Dessa forma, é possível afirmar que a possibilidade de *overfitting* é reduzida. Além disso, após a finalização dos treinamentos, por meio dos resultados obtidos a partir do conjunto de dados de testes, verifica-se que o modelo

proposto foi capaz de generalizar a detecção dos objetos, ao apresentar o desempenho, conforme exibido na Tabela 5 e 6 em dados nunca vistos anteriormente.

Figura 22 - Gráficos do comportamento do mAP e *Loss*, representando o desempenho do modelo proposto, frente ao número de iterações realizadas durante Treinamento 1 e 2, respectivamente, da rede.



Fonte: Produzido pela própria autora.

A Tabela 5 - Métrica AP para diferentes classes e treinamentos.mostra os resultados de AP obtidos para os dois treinamentos realizados. Como citado anteriormente, a métrica AP é mais interessante para avaliação das classes de forma individual, sendo assim, é possível afirmar que o modelo proposto apresenta uma dificuldade um pouco maior na detecção de poste, seguido de cruzeta para o primeiro treinamento. Já utilizando o *data augmentation* o modelo apresentou maior dificuldade na detecção de cruzetas se comparada às demais classes. O fato de não ter apresentado grande melhora no AP da cruzeta pode estar relacionado a maior variabilidade de imagens quanto ao ângulo de visada, por exemplo, no caso dessa classe.

Tabela 5 - Métrica AP para diferentes classes e treinamentos.

Treinamento	AP Cruzeta	AP Transformador	AP Poste
1	87,93%	88,56%	86,21%
2	88,38%	93,64%	92,17%

Fonte: Produzido pela própria autora.

A Tabela 6 apresenta o desempenho da rede por meio da utilização do mAP e IoU médio para os treinamentos realizados. Para o primeiro (sem *data augmentation* e com o balanceamento

das classes) em específico, foi alcançado um IoU médio de 63,53% e um mAP de 87,56%. Já para o segundo caso, onde foi empregada a técnica de *data augmentation*, houve uma evolução nas métricas de avaliação, obtendo-se um valor para o IoU médio de 67,26% e um mAP de 91,39%. Além disso, é possível notar pelos valores mostrados pelas Tabelas 5 e 6 (referentes à base de teste) que o modelo proposto foi capaz de responder bem a uma base nunca vista anteriormente, mostrando robustez quanto à generalização para detecção dos objetos e sendo mais um indício de não ocorrência de sobreajustes.

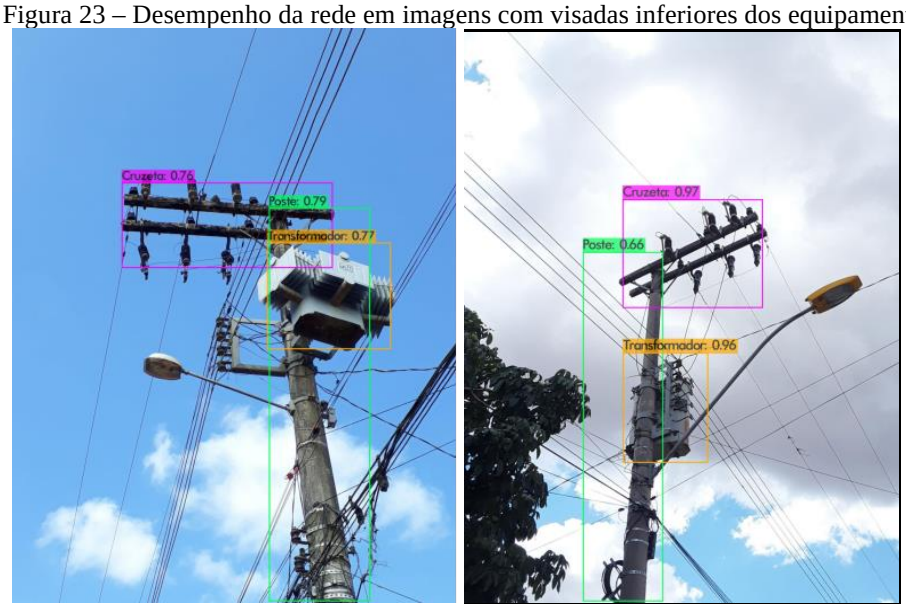
Tabela 6 - Métrica de desempenho mAP e IOU médio para os treinamentos com e sem *data augmentation*.

Treinamento	mAP	IOU médio
1	87,56%	63,53%
2	91,39%	67,26%

Fonte: Produzido pela própria autora.

Ao final, foram selecionadas algumas imagens da base de teste que apresentavam exemplos das diferentes considerações adotadas ao longo do desenvolvimento do trabalho, de forma a exemplificar e visualizar o desempenho do modelo proposto em situações diversas. Na Figura 23, verifica-se uma situação de detecção de duas imagens com as três classes presentes, apresentando visada inferior e leve angulação e posicionamento dos equipamentos, diferentes condições de iluminação e saturação, e resolução da imagem favoráveis. Nota-se que detecções que se aproximam de 80% para todas as classes na imagem esquerda, e o poste com um percentual de 66% se diferenciando dos outros dois equipamentos.

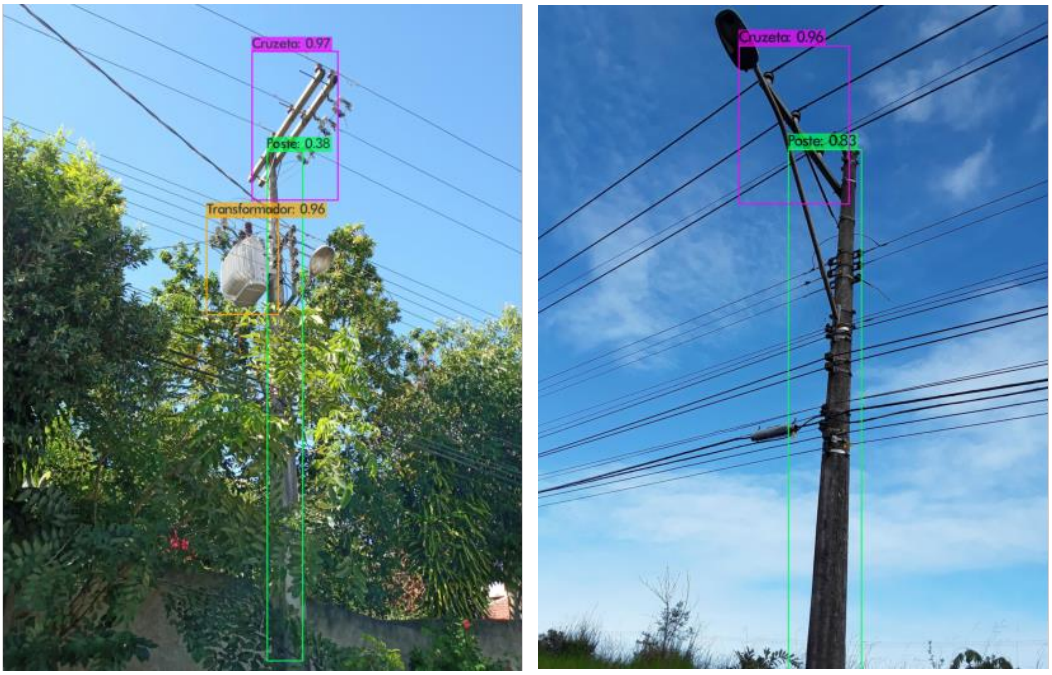
Figura 23 – Desempenho da rede em imagens com visadas inferiores dos equipamentos



Fonte: Produzido pela própria autora.

Na Figura 24, outro exemplo é dado, sendo possível verificar o desempenho da rede em situações de oclusão de parte dos objetos. Na imagem da esquerda é mostrado um exemplo em que o transformador, apesar do fundo de vegetações, apresenta alta taxa de detecção. Por outro lado, é interessante notar que o poste apresenta alto nível de oclusão pela vegetação, sendo difícil de visualizar a olho humano, mas mesmo assim apresentou uma *score* de 38%. Já na imagem da direita, tem-se que apesar de o suporte da luminária se encontrar próximo à cruzeta e poder gerar confusão até mesmo para o ser humano, a rede se mostra eficaz, não apresentando confusão entre ela e a classe cruzeta, da mesma forma como garante uma detecção de 96% para a mesma classe nesse caso.

Figura 24 – Desempenho da rede em imagens com oclusão.



Fonte: Produzido pela própria autora.

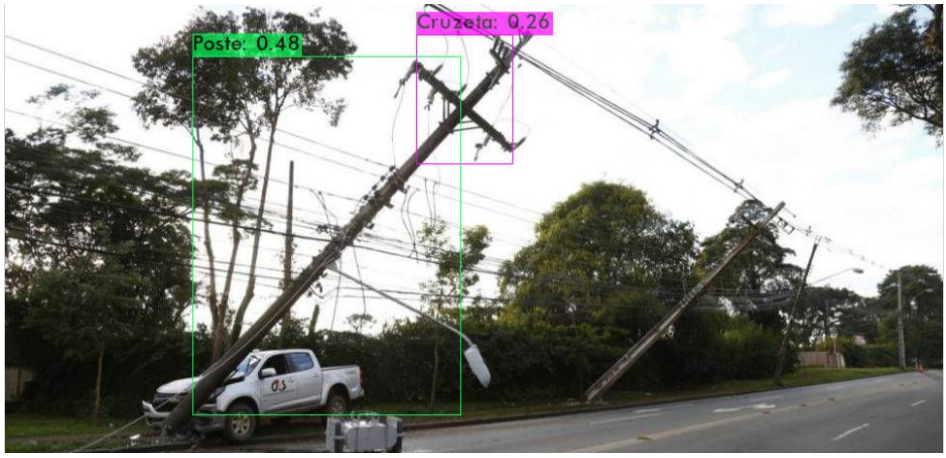
Outra análise realizada, foi referente a situações desfavoráveis de iluminação (baixo brilho), resoluções inferiores, reflexos de luz (alteração na saturação), além de apresentarem um ângulo de rotação. Na Figura 25, verifica-se uma exemplificação das condições anteriores descritas, apresentando valores acima de 74% para os objetos de primeiro plano e ainda a detecção de objetos em segundo plano, como foi o caso da cruzeta e poste na imagem da esquerda com *score* de 52% e 25%, respectivamente.

Figura 25 – Desempenho da rede para imagens com baixo brilho e resolução.

Fonte: Produzido pela própria autora.

Um caso extremo selecionado, o qual representa um poste com alto nível de abalroamento, e, portanto, com objetos em uma grande alteração na angulação é apresentado na Figura 26. Nota-se a detecção das classes, com *score* entre 25% e 50%, presentes para os equipamentos mais nítidos.

Figura 26 – Desempenho da rede para caso extremo de abalroamento.

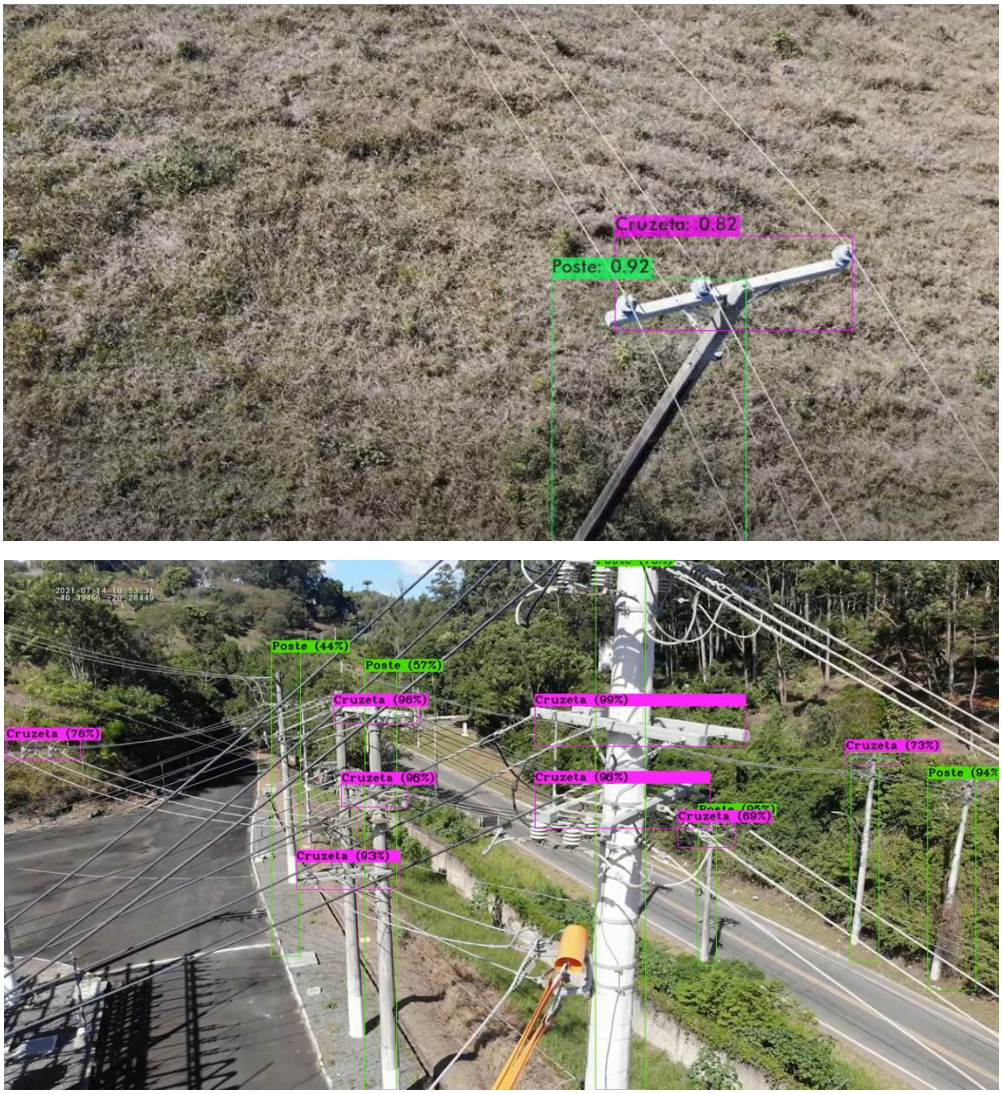


Fonte: Produzido pela própria autora.

A Figura 27 apresenta imagens obtidas por meio de frames de vídeos capturados por *drones*, equipamento cada vez mais difundida nas atividades de inspeções de redes aéreas, e por isso,

também foram consideradas nas análises. É possível observar *score* de 82% e de 92% para as classes cruzeta e poste, respectivamente, na imagem superior. A imagem inferior apresenta um cenário com vários objetos de duas classes em distância, posicionamentos e angulações diferentes. Além disso, o fundo da imagem apresentada é variado e as linhas de fases da rede se mostram como oclusão de alguns dos objetos, sendo mais um ponto que pode gerar confusão à rede. Todavia, apesar de todos os pontos desfavoráveis ao reconhecimento, grande parte dos equipamentos foram reconhecidos com alta pontuação, chegando a valores acima de 90% para vários objetos, mostrando mais uma vez o bom desempenho da rede.

Figura 27 - Exemplos de imagens de frames de *drones*.



Fonte: Produzido pela própria autora.

Sendo assim, diante do exposto, é possível afirmar que ambos os treinamentos realizados apresentam eficácia com relação ao objetivo do trabalho, ou seja, são capazes de detectar equipamentos de rede no contexto de distribuição de energia. Por outro lado, o treinamento empregando o *data augmentation* apresentou melhor desempenho. E ainda, a partir das análises de imagens selecionadas, pode-se verificar de forma geral que a rede se comporta bem às condições adversas consideradas ao longo do desenvolvimento do trabalho. Por fim, ao considerar o mAP, que representa mais adequadamente a qualidade da inferência como um todo, obtiveram-se medidas de desempenho relevantes que validam o modelo proposto.

6 CONCLUSÕES E TRABALHOS FUTUROS

Na atualidade é inevitável pensar em novas soluções aplicadas ao contexto industrial com o emprego de novas tecnologias e do conceito da Indústria 4.0, o que é valido também para o setor de energia. O contexto atual das manutenções e inspeções de equipamentos de campo em distribuidoras do setor ainda apresentam grandes limitações devido à intensa realização de análises e trabalhos manuais. Dessa forma, tem-se a elevação de HH (homem hora) no desempenho de atividades repetitivas, que acometem diferentes níveis de processo em tais organizações, as quais poderiam ser desempenhadas por tecnologias. Frente a tal cenário, o emprego de inteligência artificial, especificamente o aprendizado de máquina e a visão computacional, confirmaram-se como alternativa viável para subsidiar soluções de inspeção e manutenção.

Dada a complexidade da tarefa de detecção de equipamentos, houve a necessidade de se compreender os conceitos envolvidos neste processo, iniciando-se pela visão computacional, permeando pelo aprendizado profundo e redes neurais, até a adoção do YOLOv4 como modelo com características que atendessem à demanda inicial. O método proposto para a composição do banco de treinamento atendeu ao objetivo de inclusão de variados padrões de equipamentos existentes na rede de distribuidoras de energia, sem tornar inviável sua estruturação. A partir disso foi possível realizar a rotulação de dados do trabalho.

A rede neural foi avaliada utilizando algumas das principais métricas de detecção de objetos. O treinamento com a base de testes inicial sem a aplicação da técnica de *data augmentation* alcançou um IoU médio de 63,53% e um mAP de 87,56% nos dados de teste, resultados academicamente já satisfatórios e que validam o treinamento da rede. Enquanto o treinamento realizado empregando o pré-processamento de *data augmentation* apresentou valores de IoU 67,26% e um mAP de 91,39%, se mostrando como um aprimoramento dos resultados e desempenhando com eficiência e eficácia para o contexto proposto. Ao final, a abordagem desenvolvida se apresentou como uma solução satisfatória, mesmo considerando a alta variabilidade já citada nas características de iluminação, fundo, resolução de imagens, posicionamentos etc., para imagens aéreas por *drones*, tanto quanto para fotos capturadas por câmeras de celulares pelos técnicos de campo.

A partir dos resultados alcançados, pode-se afirmar que o processo de estratificação da base de treinamentos quanto à sua dimensão e qualidade apresentadas, o emprego de técnicas adequadas de rotulação, assim como a implementação e validação adequada da rede contribuíram nos resultados alcançados, se mostrando como práticas que podem ser indicadas durante o emprego do YOLOv4, para elevar a eficiência da rede. Além disso, o emprego da técnica de *data augmentation* foi capaz de aprimorar o desempenho, reduzindo a possibilidade de sobreajustes e melhorando o processo de aprendizado do modelo.

Consequentemente, considera-se que o principal objetivo proposto foi alcançado: o desenvolvimento de um detector capaz de identificar equipamentos de rede de distribuição de energia elétrica para agregação de valor ao processo de inspeções e manutenções no âmbito das distribuidoras de energia elétrica.

Para trabalhos futuros, sugere-se a ampliação do banco de imagens rotuladas, incluindo outras classes que comtemplem, por exemplo, o estado de conservação de equipamentos, casos de falhas existentes, ou ainda os modelos dos equipamentos. Com base nos resultados apresentados, pode-se também extrapolar a aplicação do método proposto para outros tipos de componentes de rede de distribuição como, por exemplo, isoladores, para-raios e religadores.

Outra linha possível seria integrar o sistema de detecção desenvolvido com hardwares de realidade aumentada de forma a aplicar essa tecnologia para orientação de eletricistas e técnicos em campo por especialistas. Desta forma, podem ser exploradas aplicabilidades, as quais vão desde as inspeções de equipamentos de distribuição em campo, até as realizações de manutenções e intervenção a ativos que compõem o sistema de energia, contribuindo no processo de orientação à distância, e reforçando os conceitos da Indústria 4.0 apresentados na introdução deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA BRASILEIRA DE DESENVOLVIMENTO INDUSTRIAL. **Agenda Brasileira para a Indústria 4.0**. 2019. Disponível em: http://www.industria40.gov.br. Acesso em: 24 out. 2020.

AECOM. **Virtual visions of the future**. 2020. Disponível em: https://aecom.com/without-limits/article/virtual-visions/. Acesso em: 29 out. 2020.

ARAI, K.; KAPOOR, S. **Advances in Computer Vision**: Proceedings of the 2019 Computer Vision Conference, v. 1. Springer. 2019.

AZUMA, R. T. A Survey of Augmented Reality. **Presence: Teleoperators and Virtual Environments,** v. 6, p. 355-385, 1997. Disponível em: https://www.mitpressjournals.org/doi/pdf/10.1162/pres.1997.6.4.355. Acesso em: 25 out. 2020.

BABATUNDE, O.; ARMSTRONG, L.; LENG, J.; DIEPEVEEN, D. A computer-based vision systems for automatic identification of plant species using kNN and genetic PCA. **Journal of Agricultural Informatics**, v. 6, n. 2, p. 32-44, 2015. Disponível em: http://journal.magisz.org/index.php/jai/article/view/164/164. Acesso em: 03 nov. 2020.

BALLARD, D. H.; BROWN, C. M. **Computer Vision**. 1. ed. New Jersey: Prentice Hall, 1982. Disponível em: http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/Ballard__D._and_Brown__C._M.__1982__Computer_Vision.pdf. Acesso em: 02 nov. 2020.

BOCHKOVSKIY, A.; WANG, C.; LIAO, H. **Yolov4: Optimal speed and accuracy of object detection**. 2020. Disponível em: https://arxiv.org/pdf/2004.10934.pdf. Acesso em: 15 jun. 2022.

BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. **Yolo v4, v3 and v2 for Windows and Linux**. 2020b. Disponível em: https://github.com/AlexeyAB/darknet. Acesso em: 15 mai. 2022.

BORLIDO, D. J. A. **Indústria 4.0** – Aplicação a Sistemas de Manutenção. 2017. Dissertação (Mestrado em Engenharia Mecânica) – Faculdade de Engenharia, Universidade do Porto, Porto, 2017. Disponível em: https://repositorio-aberto.up.pt/bitstream/10216/102740/2/181981.pdf. Acesso em: 18 out. 2020.

CHEN, W.; ZHONG, X.; ZHANG, J. **Optimization research and defect object detection of aeroengine blade boss based on YOLOv4**. Journal of Physics: Conference Series, IOP Publishing, v. 1746. 2021. Disponível em: https://www.researchgate.net/publication/348600290_Optimization_Research_and_Defect_Object_Detection_of_Aeroengine_Blade_Boss_Based_on_YOLOv4. Acesso em: 10 jul. 2022.

CHOLLET, F. **Deep Learning with Python**. 1. ed. Shelter Island: Manning, 2018.

COELHO P. M. N. **Rumo à Indústria 4.0**. 2016. Dissertação (Mestrado em Engenharia e Gestão Industrial) – Faculdade de Ciências e Tecnologia, Universidade de Coimbra, Coimbra, 2016. Disponível em: https://eg.uc.pt/bitstream/10316/36992/1/Tese%20Pedro%20Coelho%20Rumo%20%C3%A0%20Industria%204.0.pdf. Acesso em: 20 out. 2020.

CORREA R. P. C.; JUNIOR M. M. C. **Análise e classificação dos fatores humanos nos acidentes industriais**, v.17, p.186-198, 2007. Disponível em: https://www.scielo.br/pdf/prod/v17n1/12.pdf. Acesso em: 05 fev. 2021.

DAVIES, E. R. **Computer and Machine Vision:** Theory, Algorithms, Practicalities. 4. ed. Waltham: Elsevier, 2012. 871 p.

ENERGIAS DE PORTUGAL. **Área de Atuação**. 2021. Disponível em: https://ri.edp.com.br/pt-br/edp-energias-do-brasil/area-de-atuacao/. Acesso em: 06 jul. 2022.

ENERGIAS DE PORTUGAL. **Wearables:** Realidade Aumentada. 2018. Disponível em: https://www.edp.com/pt-pt/inovacao/wearables-realidade-aumentada. Acesso em: 01 nov. 2020.

FISCHLER, M. A; FIRSCHEIN, O. **Intelligence:** The Eye, The Brain and The Computer. Reading MA: Addison – Wesley, 1987 apud HAYKIN, S. **Neural Networks and Learning Machines**. 3 ed. Hamilton: Prentice Hall, 2009.

FORD. **Ford Será a Primeira na América Latina a Usar a Realidade Aumentada em Serviços Pós-Venda**. 2019. São Paulo. Ford. Disponível em: https://media.ford.com/content/fordmedia/fsa/br/pt/news/2019/02/12/ford-anuncia-projeto-piloto-inovador-com-uso-de-realidade-aument.html. Acesso em: 29 out. 2020.

FREIRE, J. F. A. C. F. **Aplicações de Visão em C++**. 2009. Dissertação (Mestrado Integrado em Engenharia Mecânica) – Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2009.

FU, C.-Z.; SI, W.-R.; HUANG, H.; CHEN, L.; GAO, Q.-J.; SHI, C.-B.; WANG, C**. Research on a Detection and Recognition Algorithm for High-Voltage Switch Cabinet Based on Deep Learning with an Improved YOLOv2 Network**. 2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA).

GRAVE, L.; ESCALEIRA, C.; SILVA, A. F.; MARCOS, A. A realidade virtual como ferramenta de treino para montagem de cablagens elétricas. *In*: 10º ENCONTRO PORTUGUÊS DE COMPUTAÇÃO GRÁFICA, 10., 2001, Lisboa. **Proceedings** [...]. Lisboa: Instituto de Engenharia de Sistemas e Computadores, 2001. p. 147-163. Disponível em: https://repositorioaberto.uab.pt/bitstream/10400.2/2296/1/ARVirtualComoFerramenta-etc.pdf. Acesso em: 28 out. 2020.

HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. Hamilton: Prentice Hall, 2009. 906 p.

HE, K.; GIRSHICK, R.; DOLLAR, P. **Rethinking imagenet pre-training**. *In*: Proceedings of the International Conference on Computer Vision. Seoul, Korea: IEEE, 2019. p. 4918–4927. Disponível em: https://ieeexplore.ieee.org/document/9010930. Acesso em: 10 jul. 2022.

IAACADEMY. **Detecção de Objetos com YOLO:** Uma abordagem moderna. 2020. Disponível: https://iaexpert.academy/2020/10/13/deteccao-de-objetos-com-yolo-uma-abordagem-moderna/. Acesso em: 25 out. 2020.

INDÚSTRIA4.0. **Como a visão computacional está revolucionando a gestão de qualidade na indústria**. 2019. Disponível em: https://www.industria40.ind.br/artigo/18753-como-a-visao-computacional-esta-revolucionando-a-gestao-de-qualidade-na-industria. Acesso em: 20 out. 2020.

JEONG, J.; PARK, H.; KWAK, N. Enhancement of SSD by concatenating feature maps for object detection. *In*: BRITISH MACHINE VISION CONFERENCE. 2017, London. **Proceedings** […] London: British Machine Vision Association, 2017. p. 76-88. Disponível em: https://arxiv.org/pdf/1705.09587.pdf. Acesso em: 07 nov. 2020.

KANG, K.; OUYANG, W.; LI, H.; WANG, X. Object Detection from Video Tubelets with Convolutional Neural Networks. *In*: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. 2016, Las Vegas. **Proceedings** [...]. Las Vegas: Institute of Electrical and Electronics Engineers, 2016. p. 817-825. Disponível em: https://openaccess.thecvf.com/content_cvpr_2016/papers/Kang_Object_Detection_From_CVPR_2016_paper.pdf. Acesso em: 07 nov. 2020.

KIM, W.; CHO, H.; KIM, J.; KIM, B.; LEE, S. **YOLO-Based Simultaneous Target Detection and Classification in Automotive FMCW Radar Systems**. Sensors, 20(10), 2897. 2020. Disponível em: https://www.researchgate.net/publication/341527656_YOLO-Based_Simultaneous_Target_Detection_and_Classification_in_Automotive_FMCW_Radar_Systems. Acesso em: 20 mai. 2022.

KIRNER, C.; SISCOUTTO, R. **Realidade Virtual e Aumentada**: Conceitos, Projetos e Aplicações. 9., ed. Porto Alegre: Sociedade Brasileira de Computação, 2007. Disponível em: http://de.ufpb.br/~labteve/publi/2007_svrps.pdf. Acesso em: 28 out. 2020.

KUZNETSOVA, A.; ROM, H.; ALLDRIN, N.; UIJLINGS, J.; KRASIN, I.; PONT-TUSET, J.; KAMALI, S.; POPOV, S.; MALLOCI, M.; KOLESNIKOV, A.; DUERIG, T.; FERRARI, V. The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale. **International Journal of Computer Vision**, Norwell, v. 128, p. 1956-1981, 2020.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, London, v. 521, p. 436-444, 2015. Disponível em: https://www.researchgate.net/publication/277411157_Deep_Learning. Acesso em: 09 nov. 2020.

LIN, T.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft COCO: Commom Object in Context. *In*: EUROPEAN CONFERENCE ON COMPUTER VISION. 2014. **Proceedings** [...]. Springer, Cham, 2014. Disponível em: https://arxiv.org/pdf/1405.0312.pdf. Acesso em: 18 jun. 2022.

LIN, Y.; SUN, J.; LI, C.; MA, Y.; GENG, Y.; CHEN, Y. Deep learning for intelligent substation device infrared fault image analysis. *In:* MATEC WEB OF CONFERENCES, 55. 2016, Les Ulis. **Proceedings** […]. Les Ulis: EDP Sciences, 2016. Disponível em: https://www.scinapse.io/papers/2343204296. Acesso em: 20 jun. 2022.

LIU, W.; ANGUELOV, D.; SZEGEDY, C.; REED, S.; FU C. Y.; BERG, A. C. SSD: Single Shot MultiBox Detector. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 14., 2016, Amsterdam. **Proceedings** […]. Amsterdam: Springer Science, 2016. p. 21-37. Disponível em: https://arxiv.org/pdf/1512.02325.pdf. Acesso em: 08 nov. 2020.

MEDIUM. **Introduction to Deep Learning**. 2018. Disponível: https://medium.com/@tm2761/what-is-deep-learning-15b487adaaa3. Acesso em: 25 out. 2020.

MONTANARI, R. **Detecção e Classificação de Objetos em Imagens para Rastreamentos de Veículos**. 2015. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo, São Paulo, 2015. Disponível em: https://teses.usp.br/teses/disponiveis/55/55134/tde-08012016-113715/publico/RaphaelMontanari_dissertacao_revisada.pdf. Acesso em: 02 nov. 2020.

NELSON, J.; SOLAWETZ, J. **Responding to the Controversy about YOLOv5**. 2020. Disponível em: https://blog.roboflow.com/yolov4-versus-yolov5/. Acesso em: 20 abr. 2022.

NOGUEIRA, A. B. **Análise de viabilidade no uso de Deep Learning para contagem de pessoas com câmeras de segurança**. Monografia (Graduação em Engenharia de Controle e Automação) – Universidade Federal de Santa Catarina. Santa Cataria, 2018. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/200004/PFC%20Angelo%20Baruffi%20Nogueira_2018-2.pdf?sequence=1&isAllowed=y. Acesso em: 15 jul. 2022.

OLIVATTO, T. R. **Identificação automática de rampas de acessibilidade apoiada por visão computacional a partir de imagens panorâmicas *street-level***. Dissertação (Mestrado em Engenharia Urbana) **–** Universidade Federal de São Carlos. São Carlos. 2021.

OLIVEIRA, B. A. S.; NETO, A. P. F.; FERNANDINO, R. M. A.; COSTA, D. P.; GUIMARÃES, F. G. Deep Learning para Detecção de Componentes em Alimentadores de Subestações. *In:* XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS. 2020, Florianópolis. **Anais** [...]. Santa Catarina: Sociedade Brasileira de Telecomunicações, 2020. Disponível em: https://biblioteca.sbrt.org.br/articlefile/2138.pdf. Acesso em: 10 jun. 2022.

OLIVO, R. Y.; JUNIOR, A. C.; FILHO, P. L. P. Uma abordagem neural na identificação e objetos em imagens para auxílio na manutenção de redes elétricas. *In*: WORKSHOP DE TRABALHOS DA GRADUAÇÃO - CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES (SIBGRAPI), 33. 2020, Evento Online. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 179-182. DOI: https://doi.org/10.5753/sibgrapi.est.2020.13006.

OLSTON, C.; NAJORK, M. Web Crawling. **Foundations and Trends in Information Retrieval**. v. 4, p. 175-246, 2010.

OUAKNINE, A. Review of Deep Learning Algorithms for Object Detection. **Medium**. San Francisco, fev. 2018. Disponível em: https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852. Acesso em: 09 fev. 2021.
PADILLA, R.; NETTO, S. L.; DA SILVA, E. A. B. **A Survey on Performance Metrics for Object-Detection Algorithms**. 2020. International Conference on Systems, Signals and Image Processing (IWSSIP).

PINTO, P. D. C. C. **Implementation of Faster R-CNN Applied to the Datasets COCO and PASCAL VOC**. 69 p. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2019.

PRODANOV, C. C.; FREITAS, C. E. **Metodologia Do Trabalho Científico:** Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

QUANTDARE. **What is the difference between Deep Learning and Machine Learning?**. 2019. Disponível em: https://quantdare.com/what-is-the-difference-between-deep-learning-and-machine-learning/. Acesso em: 20 out. 2020.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection. *In*: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. 2016, Las Vegas. **Proceedings** [...]. New Jersey: Institute of Electrical and Electronics Engineers, 2016. p. 779-788. Disponível em: Acesso em: Disponível em: https://arxiv.org/pdf/1506.02640.pdf. Acesso em: 05 nov. 2020.

REDMON, J.; FARHADI, A. **YOLOv3: An Incremental Improvement. Washington**, p. 6, 2018. Disponível em: https://arxiv.org/abs/1804.02767. Acesso em: 17 jun. 2022.

RUSSAKOVSKY, O. DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S; MA, S.; HUANG, Z.; KARPATHY A.; KHOSLA, A; BERNSTEIN, M; BERG A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision**, Norwell, v. 115, n.3, p. 211-252, 2015.

SAITO, T.; REHMSMEIER, M. **The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets**. Plos One, [s. l.], v. 10, n. 3, p. 1-21, 4 mar. 2015. Disponível em: http://dx.doi.org/10.1371/journal.pone.0118432. Acesso em: 18 maio 2022.

SHORTEN, C.; KHOSHGOFTAAR, T. M. **A survey on Image Data Augmentation for Deep Learning.** Journal of Big Data, v. 6, n. 1, p. 60, 2019. Disponível em: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0. Acesso em: 22 jun. 2022.

STATISTICAL ANALYSIS SYSTEM. **Visão Computacional:** O que é e qual a sua importância. 2019. Disponível em: https://www.sas.com/pt_br/insights/analytics/computer-vision.html#:~:text=A%20SciSports%20usa%20vis%C3%A3o%20computacional,esportivos%20em%20fluxo%20s%C3%A3o%20realizadas. Acesso em: 03 nov. 2020.

SUTHERLAND, I. E. The Ultimate Display. *In:* INTERNATIONAL FEDERATION OF INFORMATION PROCESSING CONGRESS, 2., 1965, New York. **Proceedings** [...]. New York: International Federation of Information Processing, 1965. p. 506-508. Disponível em: http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=52AD83A881D2FA7D4C846CE00F55B973?doi=10.1.1.136.3720&rep=rep1&type=pdf. Acesso em: 05 nov. 2020.

TETRA PAK. **Tetra Pak traz a realidade aumentada para suas embalagens**. 2018. Disponível em: https://www.tetrapak.com/pt-br/about-tetra-pak/news-and-events/newsarchive/tetra-pak-traz-a-realidade-aumentada-para-suas-embalagens. Acesso em: 30 out. 2020.

TZUTA, Lin, LabelImg. Git code (2015). **https://github.com/tzutalin/labelImg**. Acesso em: 16 mar. 2021.

VAHTRA, R.; ANBARJAFARI, G. **Parking Space Monitoring and ID Based Car Tracking**. 51 p. Dissertação (Mestrado em Ciência da Computação) — University of Tartu. Tastu, Estônia, 2019.

VOULODIMOS, A.; DOULAMIS, N.; DOULAMIS, A.; PROTOPAPADAKIS, E. Deep Learning for Computer Vision: A Brief Review. **Computational Intelligence and Neuroscience**, v. 2018 , p. 1-13, fev. 2018. Disponível em: https://www.hindawi.com/journals/cin/2018/7068349/. Acesso em: 06 fev. 2020.

VRMAT. **Realidade Virtual para Indústria 4.0:** Indústria 4.0 – Uma nova realidade que já chegou. 2017. Disponível em http://www.vrmat.com.br/pt/industria-4-0/. Acesso em: 25 out. 2020.

WU, X.; SAHOO, D.; HOI, S. C. H. **Recent advances in deep learning for object detection**. Neurocomputing, 2020. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0925231220301430. Acesso em: 20 jun. 2022.

YAO, Y.; ZHANG, J.; SHEN, F.; LIU, L.; ZHU, F.; ZHANG, D.; SHEN, H. T. (2019). Towards Automatic Construction of Diverse, High-quality Image Datasets. **IEEE Transactions on Knowledge and Data Engineering**, 1–1. doi:10.1109/tkde.2019.2903036. Acesso em: 10 jul. 2022

ZAWORSKI, R. **Data Augmentation Techniques and Pitfalls for Small Datasets**. 2018. Disponível em: https://snow.dog/blog/data-augmentation-for-small-datasets. Acesso em: 15 jul. 2022.

ZOU, Z.; SHI, Z.; GUO, Y.; YE, J. Object Detection in 20 Years: A Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Toronto, v. 41, n. 11, p. 2709-2748, nov. 2019. Disponível em: https://arxiv.org/pdf/1905.05055.pdf. Acesso em: 04 nov. 2020.