

---

Mapeamento de postes por imagens RGB de nível de  
rua utilizando métodos de detecção de objetos: um  
estudo de caso na cidade de Campo Grande/MS.

*Anderson do Espirito Santo da Silva*

---



SERVIÇO DE PÓS-GRADUAÇÃO FACOM  
- UFMS

Data de depósito:

Assinatura: \_\_\_\_\_

Mapeamento de postes por imagens RGB de nível de  
rua utilizando métodos de detecção de objetos: um  
estudo de caso na cidade de Campo Grande/MS.

*Anderson do Espirito Santo da Silva*

**Orientador:** Jonathan de Andrade Silva

Dissertação apresentada à Faculdade de  
Computação da Universidade Federal de  
Mato Grosso do Sul como parte dos  
requisitos necessários à obtenção do título  
em Mestre em Computação Aplicada.

**UFMS**  
**2023**



# Agradecimentos

---

Venho aqui dedicar este trabalho de dissertação de mestrado a todos aqueles que me acompanharam nesta jornada com amor, incentivo e apoio incondicional.

Quero primeiramente agradecer a Deus, pelo dom da sabedoria e pela graça dispensada a mim para cumprir de forma satisfatória este curso de Mestrado. A Ele a glória, eternamente, Amém!

Agradeço à minha família, aos meus pais Osvaldo e Juracy, e minha irmã Alessandra, pelo constante encorajamento, pelas orações e pela compreensão em todos os momentos, pelos sacrifícios feitos para que eu pudesse perseguir meus sonhos acadêmicos e pelo amor inabalável que sempre me sustentou, estando ao meu lado nas horas boas e difíceis, compartilhando risos e lágrimas, e que sempre me motivaram a continuar, mesmo diante dos desafios mais complexos.

Ao meu orientador, o professor Dr. Jonathan de Andrade e ao Prof. Marcato, que, ambos com sabedoria e paciência foram fundamentais para a condução deste trabalho. Agradeço por acreditarem em meu potencial e por incentivarem meu crescimento como pesquisador.

A todos os professores e colegas que contribuíram com seus saberes e experiências ao longo do curso do Mestrado, enriquecendo minha trajetória acadêmica e me inspirando a buscar sempre o conhecimento e a excelência.

Este trabalho é fruto do esforço coletivo de tantas pessoas especiais que fazem parte da minha vida, e é com humildade e gratidão que o dedico a todos vocês, e que este seja apenas o início de uma jornada de aprendizado contínuo e contribuições significativas para a ciência e a sociedade.

# Abstract

---

---

This dissertation aims to develop and apply robust methods for the detection and classification of objects, with a specific focus on electrical poles present in RGB images of the electrical grid collected in the city of Campo Grande, Mato Grosso do Sul, Brazil. The central objective is to demonstrate the utility of these methods in the context of the electrical energy sector, highlighting their capability to identify maintenance needs and structural anomalies. The accurate identification of electrical poles in images of the electrical grid plays a critical role in enhancing the electrical energy sector. This efficient identification is essential for evaluating maintenance demands and detecting structural faults, thereby contributing to the improvement of infrastructure management. The adopted methodology encompasses the construction of an annotated image dataset, where electrical poles and other elements of the electrical grid are manually identified and delineated. These annotated images are used to train object detection methods, encompassing techniques such as machine learning algorithms and convolutional neural networks. Through experiments, these methods are evaluated and compared in terms of accuracy, processing speed, and generalization. The conducted experiments yield results that underscore the effectiveness of the developed methods for the detection and classification of electrical poles in RGB images of the electrical grid. Comparative analysis of the tested models enables the identification of the most suitable method for the application at hand. With these contributions, the dissertation seeks to optimize infrastructure management, enhancing maintenance processes, and improving the efficiency of energy supply.

Keywords: Electrical Energy, Electrical Grid, Object Classification, Object Detection, RGB Images.

# Resumo

---

---

A presente dissertação tem por propósito desenvolver e aplicar métodos robustos para a detecção e classificação de objetos, com foco específico nos postes elétricos presentes em imagens RGB da malha elétrica coletadas na cidade de Campo Grande, Mato Grosso do Sul. O objetivo central é demonstrar a utilidade desses métodos no contexto do setor de energia elétrica, destacando a capacidade de identificar necessidades de manutenção e anomalias estruturais, em que a identificação precisa de postes elétricos em imagens da malha elétrica desempenha um papel crítico na melhoria do setor de energia elétrica. Tal identificação eficiente é crucial para avaliar demandas de manutenção e para detectar falhas estruturais, contribuindo assim para aprimorar a gestão da infraestrutura elétrica. A metodologia adotada compreende a construção de um banco de imagens anotado, no qual os postes elétricos e outros elementos da malha elétrica são identificados e delimitados manualmente. Essas imagens anotadas são usadas para treinar os métodos de detecção de objetos, abrangendo abordagens como algoritmos de aprendizado de máquina e redes neurais convolucionais, e através de experimentos, os métodos são avaliados e comparados em termos de precisão, velocidade de processamento e generalização. Os experimentos conduzidos produzem resultados que evidenciam a eficácia dos métodos desenvolvidos para a detecção e classificação de postes elétricos em imagens RGB da malha elétrica. A análise comparativa dos modelos testados permite identificar o método mais adequado para a aplicação em questão. Com essas contribuições, a dissertação busca otimizar a gestão da infraestrutura elétrica, tornando os processos de manutenção mais eficazes e melhorando a eficiência do fornecimento de energia elétrica.

Palavras-chave: Classificação de objetos, Detecção de objetos, Energia elétrica, Imagens RGB, Malha elétrica.

# Sumário

---

---

Lista de figuras .....	10
Lista de tabelas.....	12
Lista de Abreviaturas .....	13
CAPÍTULO 1.....	14
Introdução.....	14
1.1 Motivação.....	14
1.2 Objetivos .....	27
1.3 Principais Contribuições .....	28
1.4 Organização.....	29
CAPÍTULO 2.....	30
Algoritmos: detecção e análise de objetos .....	30
CAPÍTULO 3.....	41
Aspectos metodológicos .....	41
.....	41
3.1 Conjunto de imagens.....	41
3.2 Método de detecção.....	43
3.3 Estado da arte .....	44
3.3.1 MMDetection .....	45
3.4 Modelos escolhidos.....	45
3.4.1 Faster R-CNN.....	46
3.4.2 RetinaNet .....	46
3.4.3 RTMDet.....	47
CAPÍTULO 4.....	48
Resultados alcançados.....	48
4.1 Base de dados .....	48
4.2 Resultados experimentais .....	53
CAPÍTULO 5.....	59
Considerações finais.....	59

5.1 Conclusão .....	59
5.2 Trabalhos futuros .....	60
Referências bibliográficas.....	62



# Lista de figuras

---

---

Figura 1 - Componentes dos Tipos de Postes em Campo Grande-MS.....	16
Figura 2 - Representação da Inteligência Artificial e dos Subconjuntos Aprendizado de Máquina, Rede Neural e Aprendizado Profundo .....	17
Figura 3 -Exemplo Genérico de uma Rede Neural Profunda .....	21
Figura 4 - Camadas Totalmente Conectadas .....	22
Figura 5 - Filtro Aplicado a uma Entrada Bidimensional para Criar um Mapa de Recursos .....	23
Figura 6 - Exemplo da Camada de Pooling .....	24
Figura 7 - Definição da Medida Intersection Over Union.....	25
Figura 8 - Exemplos de frames extraídos de um vídeo tendo postes elétricos como objetos-alvo .....	43
Figura 9 - Exemplos de imagens utilizados na fase de treino do referido estudo tendo postes elétricos como objetos-alvo .....	49
Figura 10 - Exemplos de imagens utilizados na fase de validação do referido estudo tendo postes elétricos como objetos-alvo .....	49
Figura 11 - Exemplos de imagens utilizados na fase de teste do referido estudo tendo postes elétricos como objetos-alvo .....	50
Figura 12 - Matriz de confusão - Modelo Faster RCNN .....	51
Figura 13 - Matriz de confusão - Modelo RTMDet .....	52
Figura 14 - Matriz de confusão - Modelo RetinaNet .....	52
Figura 15 - Modelo RTMDET captando as classes e o seu percentual de certeza.....	56
Figura 16 - Modelo Faster R-CNN captando as classes e o seu percentual de certeza.....	56



# Lista de tabelas

---

---

Quadro 1 - Resumo dos Artigos da Revisão Bibliográfica.....	36
Quadro 2 - Balanceamento das classes e ocorrências.....	48
Quadro 3 - Configurações de arquivo dos modelos do estudo.....	50
Quadro 4 - Resultados obtidos pelos modelos do estudo.....	51

# Lista de Abreviaturas

---

3D - três dimensões  
AM - Aprendizado de Máquina  
ANEEL - Agência Nacional de Energia Elétrica  
AP - Average Precision  
CNN - Redes Neurais Convolucionais  
DL - Deep Learning  
DNN - Deep Neural Networks  
FPN - Feature Pyramid Network  
GPS - Global Positioning System  
GSV - Google Street View  
IA - Inteligência Artificial  
IoU - Intersection Over Union  
LOB - Brute-force-based line-of-bearing  
mAP - Mean Average Precision  
MATLAB - Matrix Laboratory  
MRPD - Mask R-CNN Powerline Detector  
MS - Mato Grosso do Sul  
NMS - Non-Maximum Suppression  
PLs - Power Lines  
R-CNN - Mask Regional Convolutional Neural Network  
RGB - Red, Green and Blue  
RN - Redes Neurais  
RNPs- Redes Neurais Profundas  
SRD - Superintendência de Regulação dos Serviços de Distribuição  
SSD - Single Shot MultiBox Detector  
TTPLA - Transmission Towers and Power Lines  
TTs - Transmission Towers  
UAS - Unmanned Aircraft System  
UAV - Unmanned Aerial Vehicle  
UPCs - Utility Poles With Crossarms  
VC - Visão Computacional  
VGGNet - VGG Very Deep Convolutional Networks  
VOD - Vídeo Object Detection  
YOLACT - You Only Look at the Coefficients  
YOLO - You Only Look Once

# CAPÍTULO 1

## Introdução

---

Neste capítulo descrevo os aspectos gerais desta dissertação, oferecendo uma apresentação das problemáticas pesquisadas, bem como dos objetivos centrais da pesquisa realizada. O capítulo está organizado da seguinte maneira: na Seção 1.1 é apresentada a motivação sobre o tema de pesquisa tratado nesta dissertação; na Seção 1.2 são apresentados os objetivos do trabalho em forma de questionamentos, que serão respondidos no decorrer dos capítulos; na Seção 1.3 são apresentados brevemente as principais contribuições deste trabalho; por fim, na Seção 1.4 é apresentada a organização da tese, com uma breve elucidação do conteúdo abordado em cada capítulo.

### 1.1 Motivação

Esta pesquisa tem como foco compreender e elucidar a utilização da Inteligência Artificial (IA), por meio da Aprendizagem de Máquina, para a detecção e classificação de postes de energia elétrica no município de Campo Grande, no Mato Grosso do Sul. O propósito é comparar os métodos existentes de análise de objetos em imagens RGB, mapear a rede energética do município e contribuir com pesquisas complementares para compreender a situação de interrupções no fornecimento de energia elétrica acima da média na região. Além disso, visa identificar as ferramentas mais adequadas para aprimorar a gestão da rede elétrica na cidade.

Para atingir esses objetivos, a pesquisa inicia com o reconhecimento de que a Agência Nacional de Energia Elétrica (ANEEL), estabelecida pela Lei nº 9.427/1996, que define suas atividades e competências, trabalha em consonância com os princípios e valores alinhados às expectativas da sociedade brasileira, assegurando equilíbrio, transparência e compromisso com o interesse público (BRASIL, 1996). Delegada pelo Governo Federal, a ANEEL busca

implementar políticas que promovam atividades relacionadas à concessão, permissão e autorização de empreendimentos e serviços de energia elétrica.

De acordo com seu Regulamento Interno, expresso na Portaria MME 349 de 28/11/1997, a agência delega à Superintendência de Regulação dos Serviços de Distribuição (SRD) a responsabilidade pela regulação técnica dos serviços de distribuição de energia elétrica, com o objetivo de estabelecer regras e procedimentos relacionados ao planejamento da expansão, acesso, operação e medição dos sistemas de distribuição, desenvolvimento de redes inteligentes, bem como a universalização do acesso e uso da energia elétrica.

É importante notar que as concessionárias, em geral, possuem registros individuais dos postes em suas áreas de atuação. Contudo, no caso específico deste estudo, a observação relevante é que a localização dos postes muitas vezes se baseia nas ruas, sem uma geolocalização precisa. Como resultado, não há documentação das condições estruturais de cada poste na rede, levando a uma manutenção reativa, sem ações preventivas, o que é crucial para garantir a continuidade do serviço essencial.

Além disso, conforme os estudos da Superintendência mencionada, que também define indicadores de qualidade do serviço de energia elétrica, a interrupção do fornecimento de energia elétrica por unidade consumidora em todo o território nacional em 2019 foi de 12,81 horas (SRD, 2019). Na cidade de Campo Grande/MS, esse indicador de interrupções chega a 115,55 horas, com uma duração média de 20 horas. Portanto, torna-se ainda mais crucial a manutenção eficiente da rede de distribuição de energia elétrica para garantir a continuidade do serviço sem interrupções indesejadas.

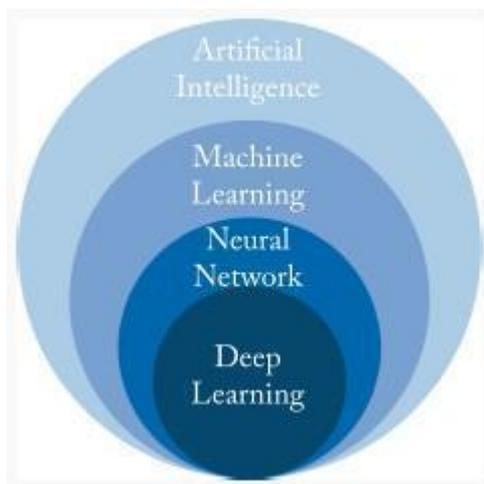
Observa-se que a malha elétrica de Campo Grande/MS é composta de diferentes tipos de postes, sendo os mais comuns discriminados abaixo (Figura 1). As estruturas a serem identificadas e marcadas nas imagens coletadas são: 1 – poste; 2 – iluminação; 3 – circuitos; 4 – caixa de energia; 5 – transformador; 6 linhas de transmissão.



características as diferem: a IA pode ser definida de forma simples como um programa de computador que tem a habilidade de aprender, agir e se adaptar; o AM é um modelo que aprende uma função de mapeamento de dados de entrada para estimar uma saída; as RN são geralmente compostas de camadas para representar funções ou mapeamentos complexos;

De acordo com Rodriguez (2021), embora esses termos sejam às vezes usados de maneira intercambiável, algumas características específicas os diferenciam: a Inteligência Artificial (IA) pode ser simplificada como um programa de computador com a capacidade de aprender, tomar ações e se adaptar; o Aprendizado de Máquina (AM) é um modelo que aprende uma função para mapear dados de entrada a fim de estimar uma saída; as Redes Neurais (RN) geralmente são compostas por camadas que representam funções ou mapeamentos complexos. Quanto ao Aprendizado Profundo (Deep Learning - DL), apesar de ser comum associá-lo a redes com várias camadas, o termo "profundo" refere-se à capacidade da rede neural de ter múltiplas camadas intermediárias (figura 2), também conhecidas como camadas ocultas. Isso capacita a rede a aprender representações mais complexas e abstrações de dados por meio de hierarquias de características, independentemente do número exato de camadas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 2 - Representação da Inteligência Artificial e dos Subconjuntos Aprendizado de Máquina, Rede Neural e Aprendizado Profundo



Fonte: Rodriguez (2021)

De acordo com Goodfellow, Bengio e Courville (2016), os métodos de AM utilizam o conhecimento prévio de um determinado conjunto de dados para desenvolver um modelo habilitado para a inferência de hipóteses em novos dados. No presente caso, trata-se de uma coleção de imagens que foram utilizadas para este treinamento, e as informações extraídas

formaram o modelo de análise de novas imagens. Em outras palavras, o modelo foi treinado a partir de um conjunto de dados relacionado ao domínio de uso com a finalidade de identificar as características da imagem para 4 fins de detecção destas características em novas imagens.

O exposto é motivado pelo fato de que, no século vigente, o mundo lida com uma multiplicidade de dados em larga escala, e a manipulação destes excede a capacidade humana de processamento. Dentro da conjuntura computacional, a arquitetura de aplicações Cliente/Servidor não têm atendido satisfatoriamente a essa real e expansiva demanda, pois a maioria dos dados são não-estruturados e estão espalhados por diferentes e diversas plataformas digitais (TEKINER; KEANE, 2013). Ainda nesse aspecto, deve-se considerar que os dados digitais são, geralmente, gerados nas máquinas e mantidos por elas, e são captados por pessoas de acordo com a sua necessidade e importância (RAJPUROHIT, 2013).

De acordo com Dalla Vecchia (2012), os recursos proporcionados pela Ciência de Dados (Data Science) possuem uma interface forte para o tratamento de bases de dados muito amplas, as quais exigem técnicas e recursos diferentes dos normalmente utilizados para trabalhar conjuntos menores de dados como fonte de organização de informações e construção de novos conhecimentos sobre determinado fenômeno.

Com a finalidade de tratar, guardar e tirar valor dessas bases de dados gigantes, os estudos de Ciência de Dados abriram a possibilidade da construção de modelos preditivos com grande precisão na resolução de desafios e problemas. Um dos motivos desse avanço é a ampliação e o desenvolvimento de novas arquiteturas de hardware, como os sistemas distribuídos e a memória distribuída, cuja interação entre a área de código e a área de dados entre conjunto de computadores (clusters) é altamente dinâmica, principalmente na troca de informações, proporcionando um ambiente favorável ao paralelismo de funções e dados, resultado de um trabalho conjunto entre processador, memória e armazenamento interligados, porém independentes e autônomos (COULOURIS; DOLLIMORE; KINDBERG, 2007; AMARAL, 2016).

Ao analisar um conjunto de dados para extração de conhecimento e seu significado, sua realidade precisa pontuar para a aplicação desejada, ou seja, é necessário um objeto de estudo eficaz a ser considerado por pesquisadores e especialistas em seus campos do saber. Para isso, existem técnicas baseadas em IA que podem ser adotadas nos estudos de Ciência de Dados que visam permitir uma análise mais apurada e precisa dos dados disponíveis para fins de

aprendizagem e sua respectiva aplicabilidade (MONARD; BARANAUSKAS, 2003).

Nesse contexto, é importante compreender a Visão Computacional (VC), que visa capacitar máquinas com a habilidade de processar imagens e extrair informações relevantes, simulando a capacidade visual humana. A VC não se limita apenas à captura de imagens, abrangendo também a manipulação para realçar detalhes cruciais, como a redução de ruídos ou o ajuste de contraste. Além disso, é fundamental extrair informações valiosas da imagem para identificar áreas de interesse, formas, cores e texturas, além de correlacionar esses aspectos com outras imagens. Para atingir esses objetivos, a Visão Computacional passa por várias etapas essenciais: a aquisição de imagens por meio de dispositivos como câmeras e scanners, o processamento para aprimorar a qualidade e destacar características, a segmentação para identificar áreas relevantes, a análise de imagens para detectar traços específicos e o reconhecimento de padrões para categorizar objetos com base em atributos. Dentro deste estudo, a aplicação da Visão Computacional poderia ser utilizada para reconhecer padrões em imagens de postes elétricos, permitindo a identificação desses objetos com base em características como cor, forma e tamanho.

Esses conceitos da Visão Computacional podem ser valiosos para a detecção, classificação e avaliação de postes elétricos, contribuindo para a otimização da gestão da infraestrutura elétrica.

Os objetivos desta dissertação são os seguintes:

1. Identificar e avaliar os métodos mais eficazes para a detecção e classificação de componentes em imagens RGB.
2. Implementar os métodos selecionados para analisar as imagens da malha elétrica coletadas em Campo Grande/MS.
3. Avaliar a condição dos componentes identificados, determinando se estão em condições adequadas, se requerem manutenção ou se apresentam defeitos.
4. Demonstrar a aplicabilidade desses recursos digitais no setor de energia elétrica, com ênfase na identificação precisa de necessidades de manutenção e identificação de falhas estruturais nos componentes da malha elétrica.

Esses objetivos são orientados para fornecer resultados concretos e específicos em relação à detecção, avaliação e classificação dos componentes da malha elétrica, contribuindo para uma gestão mais eficiente da infraestrutura elétrica e um fornecimento de energia mais

confiável, e existe uma oportunidade substancial para a criação de um novo método inovador no campo da detecção e classificação de postes elétricos por meio de imagens RGB em nível de rua. Em vez de simplesmente utilizar métodos já estabelecidos, este trabalho almeja desenvolver uma abordagem única e eficaz para a identificação precisa de postes elétricos na malha urbana.

O novo método proposto se baseia em algoritmos avançados de Visão Computacional e Aprendizado de Máquina. Ele será projetado para analisar minuciosamente as imagens da malha elétrica coletadas em Campo Grande-MS e realizar a detecção automática dos postes elétricos. Além disso, esse método terá a capacidade de classificar as condições estruturais dos postes, diferenciando entre aqueles em boas condições, os que necessitam de manutenção e os que apresentam defeitos.

A inovação se estenderá à incorporação de técnicas de geolocalização para mapear com precisão as posições dos postes identificados em um mapa digital. Isso proporcionará uma visão espacial clara e auxiliará na tomada de decisões informadas sobre o planejamento e manutenção da rede elétrica.

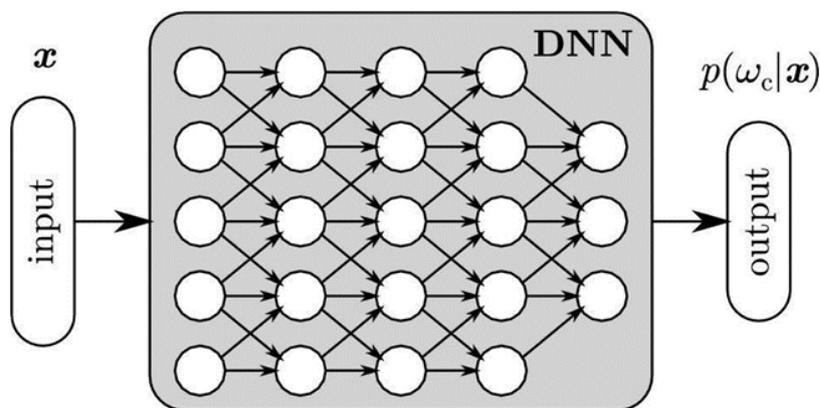
Em resumo, o objetivo é criar um novo método que ultrapasse as abordagens convencionais, proporcionando um sistema avançado e confiável para a detecção, classificação e mapeamento de postes elétricos, contribuindo assim para uma gestão otimizada da infraestrutura elétrica e um fornecimento de energia mais resiliente.

Apesar das especificidades de cada caso, as pesquisas que envolvem a VC possuem uma grande gama de algoritmos de alto desempenho, cada um voltado a um problema específica, tais como a análise de formas, texturas e a segmentação de imagens, que com o avanço da *Deep Learning* (DL), são realizadas por Redes Neurais Convolucionais (CNN), redes profundas capazes de adicionar informações extras para a classificação, comumente usadas para problemas de VC. A CNN é caracterizada como uma rede profunda, devido a “quantidade de camadas e parâmetros consideravelmente superiores às redes multicamadas” (SOARES, 2019,p. 29), ou seja, se constituem como redes neurais que aplicam operações não lineares, denominadas de convolução e que, segundo Goodfellow, Bengio e Courville (2016), configuram-se como uma das melhores soluções para os possíveis problemas de VC das últimas décadas. Acrescenta-se que a CNN utiliza diretamente imagens, demandando reduzido pré-processamento de dados de entrada.

Antes de sua chegada, algoritmos tradicionais de visão computacional aplicavam uma série de filtros nas imagens para extração de características e então essas características eram classificadas com algum algoritmo de *machine learning*. Sem a necessidade de se aplicar filtros antes da entrada em uma rede convolucional, e com a obtenção de resultados melhores, percebeu-se que o treinamento das operações convolucionais ensinam à rede uma combinação de filtros melhor que as geradas por uma pessoa. Essa automatização da extração de característica para classificação é sua principal vantagem (SOARES, 2019, p. 29).

Não obstante, em algoritmos mais genéricos, há muito a ser desenvolvido para que os programas de VC emulam a capacidade de visão biológica (BACKES; SÁ JÚNIOR, 2016). A ampliação da área de DL e das RN têm se mostrado como grande alternativa quando se trata de soluções de problemas complexos. Nesta busca, foram criadas as Redes Neurais Profundas (RNPs) (*Deep Neural Networks*), isto é, redes neurais artificiais com estruturas de diversas camadas e com duas ou mais camadas de neurônios ocultas (NIELSEN, 2015). É possível existirem quantidades arbitrárias de camadas ocultas, em que cada uma delas contém um número diferente de neurônios, conforme a aplicação da RNP. Tal número de neurônios e camadas são definidos de forma empírica (figura 3), e dentre as suas áreas de aplicação, destaca-se a de reconhecimento de padrões em imagens.

Figura 3 - Exemplo Genérico de uma Rede Neural Profunda



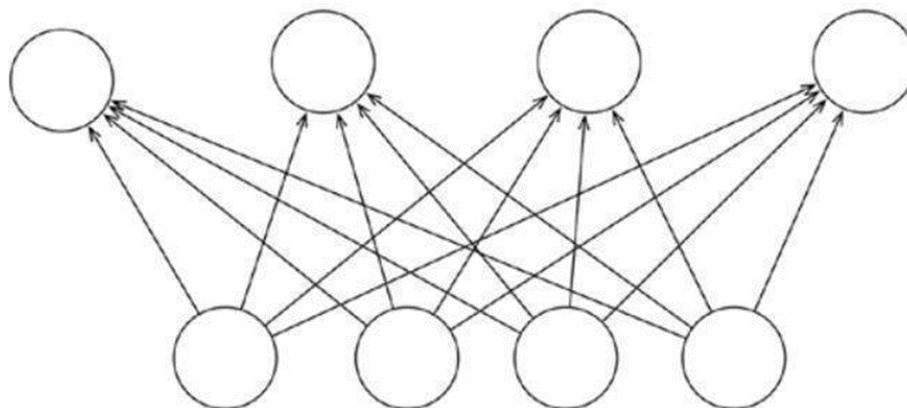
Fonte: Montavon, Samek e Müller (2017).

Na figura acima, é possível observar uma RNP com muitos neurônios interconectados, em que a entrada  $x$  é associada à uma probabilidade de associação com  $\omega_c$ , isto é,  $p(\omega_c | x)$ . Para executar suas funções, as RNPs possuem arquiteturas diversas, dependendo de sua aplicação.

As RNPs podem conter algumas camadas, como a Camada Fully-Connected, um tipo de camada interna utilizado nas RNPs, atuando na extração de atributos dos dados de entrada, com todos os neurônios conectados aos dados de entrada, isto é, cada valor recebido se conecta

a todos os neurônios presentes na camada. Na camada totalmente conectada, a saída de um neurônio serve de entrada para todos os neurônios da próxima camada (figura 4), como observado abaixo.

Figura 4 - Camadas Totalmente Conectadas



Fonte: Ketkar (2017).

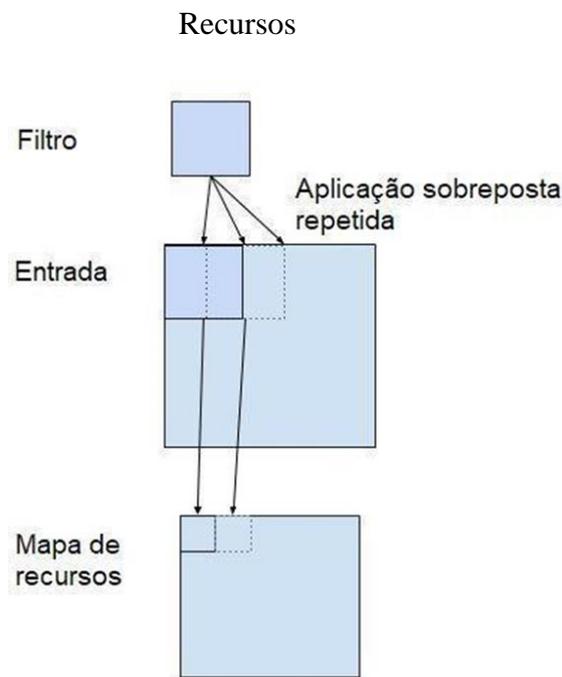
Sua função pode ser descrita como  $y = f(x * \omega)$ , na qual  $x$  é o vetor de entrada,  $y$  é o vetor de saída,  $\omega$  é um conjunto de pesos e  $f$  é a função de ativação (KETKAR, 2017). O resultado dessa camada é um vetor de  $n$  dimensões, onde  $n$  é o número de classes que a rede poderá prever, e contém as probabilidades para cada classe de qualquer imagem a ser classificada (MATHWORKS, 2021).

Já a Camada Convolutiva tem por objetivo atuar na extração de atributos dos dados de entrada, e visa reduzir consideravelmente o número de parâmetros necessários à sua execução, além de fazer com que a camada seja invariante à translação. Tais características a tornam o ponto chave para a solução dos problemas que requerem processamento de imagens. Como explica Brownlee (2019, p. 108, tradução nossa).

“Uma convolução é a simples aplicação de um filtro a uma entrada que resulta em uma ativação. A aplicação repetida do mesmo filtro a uma entrada resulta em um mapa de ativações chamado mapa de recursos, indicando as localizações e a força de um recurso detectado em uma entrada, como uma imagem.”

A camada convolutiva é descrita como  $y = f(s(x * \omega))$ , sendo  $s$  a operação de convolução entre a entrada e os pesos (KETKAR, 2017), e pode ser observada na figura (figura 5) abaixo.

Figura 5 - Filtro Aplicado a uma Entrada Bidimensional para Criar um Mapa de

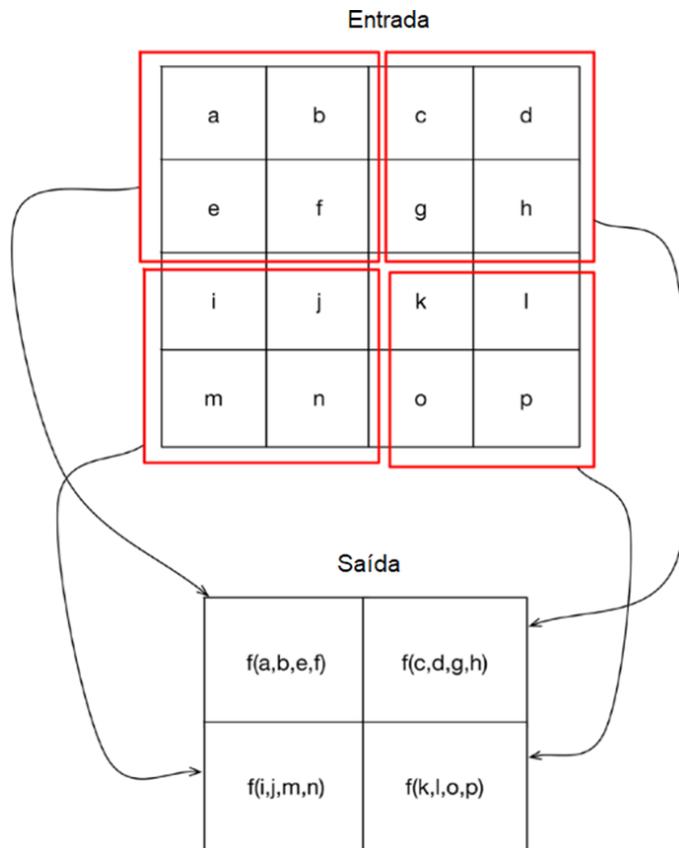


Fonte: adaptado de Brownlee (2019).

A Camada de Pooling, que objetiva reduzir a quantidade de parâmetros da rede, sintetizando as dimensões dos dados recebidos e simplificando as informações para a próxima camada. Tal evento ocorre, pois, para cada conjunto de dados agrupados, apenas um valor é retornado, sendo muito útil na análise de imagens. Dentre as características desta camada, a redundância de informação para dados espacialmente próximos é reduzida, visto que, para uma área pequena, os pixels adjacentes costumam possuir valores muito parecidos.

Ainda, quando se trata da rede se tornar invariante a pequenas alterações locais e distorções, o resultado do agrupamento de um certo conjunto de dados será sempre o mesmo, independentemente de como estes estão organizados. Logo, quando é causada alguma modificação na posição dos valores quando esta imagem for descomprimida, esta passará despercebida. A camada de pooling é representada na figura (figura 6) a seguir.

Figura 6 - Exemplo da Camada de Pooling



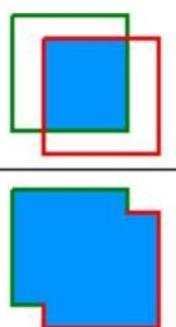
Fonte: adaptado de Ketkar (2017).

Observa-se na imagem acima a redução do número de dimensões no conjunto de dados bidimensional. A função  $f$  é comumente a operação máxima (levando ao pool máximo), porém, esta função pode variar (KETKAR, 2017).

Brownlee (2019) explica ainda que, o reconhecimento de objetos é um termo geral para uma série de tarefas que a VC realiza para a identificação de objetos em fotografias digitais, e inclui: uma listagem de todos os objetos encontrados na imagem; a caracterização destes objetos; e a detecção da presença do objeto de interesse. Como explica Rodriguez (2021), a detecção de objetos inclui a busca, identificação e anotação a que classe estes objetos pertencem, e geralmente se dá em duas fases: a proposição de uma certa área e a classificação desta. Portanto, uma série de caixas delimitadoras ou janelas podem ser geradas para um mesmo objeto, sendo estas de diferentes tamanhos e adotando diferentes posicionamentos. Algumas não incluem, necessariamente, o objeto todo e outras possuem espaços que não incluem o objeto em si. Assim, um dos algoritmos usados para esta detecção e para quantificar quão boa é a predição do objeto dentro de uma janela é o *Intersection Over Union* (IoU).

O IoU é uma medida que calcula o desempenho da tarefa de reconhecimento de objetos, e utiliza a razão entre a área da sobreposição da caixa delimitadora da região predita e da caixa delimitadora do objeto em si (BROWNLEE, 2019). O valor IoU pode variar entre 0 e 1, isto é, zero se as caixas não possuírem nenhuma sobreposição, até um, se estiverem totalmente sobrepostas. Segundo o autor, uma predição positiva se dá quando o IoU possui um índice maior que 0,5, isto é, que as caixas delimitadoras da região analisada e do objeto é de mais de 50%, porém, esta variável pode ser escolhida na implementação do código. Na figura abaixo (figura 7), é possível identificar a área de sobreposição e área somada dos objetos para a definição da medida IoU.

Figura 7 - Definição da Medida Intersection Over Union

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{área de sobreposição}}{\text{área da união}}$$


Fonte: Padilla, Netto e Silva (2020).

A partir do IoU, é possível obter a porcentagem de predições corretas feitas nas caixas delimitadas em relação a todos os objetos da imagem, fator chamado de recall. Brownlee (2019) explica ainda que, quanto mais previsões, a porcentagem de recall aumentará, porém, se houver muitas previsões falso-positivas, a precisão cairá ou se tornará irregular. Para identificar a qualidade desta detecção, é possível gerar um gráfico onde o recall é colocado no eixo x e a precisão para cada número de previsões realizadas é colocado no eixo y, e assim criar uma curva ou reta que permita maximizar o valor de cada ponto neste gráfico, ou a precisão média (*average precision*, ou AP), calculada como a média ponderada das precisões em cada limiar para aquele ponto, sendo o peso o aumento no recall do limiar anterior, e, posteriormente, calcular o valor médio da precisão (*mean average precision*, ou mAP) para cada valor de recall.

A precisão de cada detecção do algoritmo de previsão é apurada diante do sucesso da geração de caixas delimitadoras condizentes com o objeto de interesse (BROWNLEE, 2019). Já a média desta precisão em todas as classes das imagens é chamada de *mean average precision* (mAP), e esta pontuação é utilizada como forma de classificar a qualidade do modelo de detecção de objetos dado um conjunto de imagens.

Após a classificação da imagem e a geração de diversas caixas delimitadoras, os modelos de detecção usam uma função para reduzir caixas que possuem menor chance de delimitar o objeto de interesse, chamada Non-Maximum Suppression (NMS).

A NMS é um filtro que elimina algumas das caixas da região analisada pela classificação da imagem, e se encontra entre as funções de classificação e localização de objetos (BROWNLEE, 2019). O NMS seleciona a caixa com o maior IoU entre todas as caixas delimitadoras e calcula a razão entre as outras caixas e esta caixa de maior IoU, eliminando caixas acima de um certo valor limite, definido por parâmetro na implementação do algoritmo. Assim, o NMS trabalha com as janelas geradas pela classificação das imagens e suprime as janelas menos prováveis de pertencer ao objeto de interesse (FORSYTH; PONCE, 2012).

No contexto do DL, existem métodos para a detecção de objetos em imagens que recebem como entrada de imagens e, por meio de filtros convolucionais, extraem características das mesmas, localizam regiões com objetos na imagem e classifica tais regiões que contêm os objetos de interesse (TAN et al., 2020, FIGUEIREDO; SILVA, 2021). Vale ainda destacar trabalhos desenvolvidos para a identificação de postes (GOMES, et al., 2020) e árvores (LAUMER et al., 2020).

Este cenário traz também o uso da Visão Computacional (VC), topologia de DL mais utilizada para a detecção e caracterização de objetos em imagens, vídeos, entre outros (RODRIGUEZ, 2021). Escreve o autor que algumas das tarefas da VC são a detecção e classificação de objetos, segmentação semântica da imagem, verificação da amostra segundo as classes identificadas e a geração de imagens.

Compondo assim, o método de DL para detecção de objetos utilizado na dissertação, está o Faster R-CNN no qual a arquitetura do modelo é aprimorada para ampliar a velocidade de treinamento e detecção de objetos em tempo real e refinar propostas regionais como parte do processo de treinamento, conhecido como Rede de Propostas Regionais (RPN), por meio de um algoritmo. Sendo assim, mais rápida do que a seletiva e também melhora o modelo de geração de proposta de região durante o treinamento, o que reduz o tempo geral de detecção (0,2 segundos) em comparação com o Fast R-CNN (2 segundos) (Brownlee, 2019).

Ainda dentro do DL encontra-se o RetinaNet, detector objetos one-shot, implementando uma nova função de custo para treinamento chamada de *focal loss*, que embora tenha um tempo de inferência menos rápido que a Yolo e a SSD, possui os melhores resultados no dataset de detecção de objeto, por meio de camadas convolucionais para extração de características com uma rede piramidal para múltiplas escalas, seguidas de duas sub redes, uma para classificação de objetos e outra para regressão dos pontos das *bounding-boxes*.

Neste panorama, o presente estudo envolve o método de DL para detecção de objetos utilizado na dissertação são os Faster RCNN e o RetinaNet, com a finalidade de contribuir para estudos complementares de desenvolvimento de ferramentas computacionais para o processamento das imagens e classificação dos objetos nas mesmas, assim como posicionar tais objetos dentro da localização Global Positioning System (GPS). Busca, assim, identificar as melhores técnicas e ferramentas para mapear postes elétricos por imagens e amparar estudos complementares para a aplicabilidade destes recursos digitais no setor de energia elétrica na cidade de Campo Grande, Mato Grosso do Sul (MS).

## 1.2 Objetivos

A presente dissertação tem como objetivo principal identificar métodos robustos de detecção e classificação de componentes na malha elétrica, utilizando imagens RGB coletadas na cidade de Campo Grande/MS. Especificamente, o objetivo é determinar se esses componentes estão em condições adequadas, requerem manutenção ou apresentam defeitos. Isso inclui a identificação de componentes que possam apresentar falhas estruturais, necessidade de manutenção preventiva, ou já estejam em estado de falha. A dissertação buscará também explorar possíveis aplicações adicionais desses métodos no setor de energia elétrica, tais como otimização da gestão da rede, identificação de pontos críticos e outras utilidades relevantes.

A presente dissertação, ainda, também tem como objetivo investigar a viabilidade do desenvolvimento de novos métodos computacionais ou ferramentas adicionais para aprimorar o mapeamento de postes elétricos utilizando imagens RGB capturadas em nível de rua. Esses métodos ou ferramentas podem envolver abordagens inovadoras ou melhorias significativas nas soluções já desenvolvidas, com o propósito de aprimorar a precisão e eficácia na identificação e catalogação de postes elétricos. A pesquisa busca contribuir para o aperfeiçoamento das práticas de manutenção, expansão e monitoramento da rede elétrica por meio da introdução de novos recursos ou aperfeiçoamento dos existentes.

Para responder tais questionamentos foram elencados objetivos específicos que visam responder os questionamentos acima, tais como: construir um banco de imagens em nível de rua anotado com postes de regiões de Campo Grande, para o processo de mapeamento dos postes elétricos; anotar os objetos com a rotulação manual dos postes e seus componentes para

treinar os métodos de detecção e classificação implementados; implementar, aplicar e identificar o melhor método para detecção de objetos em imagens RGB, por meio de experimento, visando estabelecer regras-padrão para a detecção e classificação destas imagens; verificar os mecanismos de transferência de conhecimento entre modelos testados, por meio de análise comparativa.

## 1.3 Principais Contribuições

Considerando o atual contexto dos indicadores de qualidade do serviço de fornecimento de energia elétrica em Campo Grande e em nível nacional, enxergam-se os estudos científicos para o uso de dispositivos de captura, detecção e classificação de objetos em imagens imprescindíveis e estratégicos para identificar a malha energética dos municípios, verificar as condições estruturais da mesma, contribuir para a tomada de decisões acerca da distribuição de energia elétrica e cooperar com a redução das interrupções do abastecimento energético.

Observa-se uma ampliação dos estudos que envolvem a VC e o DL, especificamente o Faster RCNN e o RetinaNet, assim como do desenvolvimento de novas ferramentas computacionais de detecção e classificação de objetos em imagens, momento ideal para avaliar a disponibilidade, a variedade, a qualidade e a praticidade do uso dos recursos digitais na gestão da malha elétrica, favorecendo o mapeamento e registro de informações voltadas às tomadas de decisão no setor energético por meio da análise dos dados de imagens de postes.

Assim, de maneira indireta, obter um mapeamento da distribuição de postes na cidade de Campo Grande-MS, permite obter um registro apurado das condições de cada poste e a sua localização em caso de interrupção na distribuição de energia. Dessa maneira, contribuindo para a melhoria da distribuição energética através de técnicas de Detecção de Objetos para identificação e classificação dos postes elétricos em imagens. Vale destacar, que atualmente não existe tal mapeamento individual dos postes existentes na cidade. Podemos então destacar as contribuições deste trabalho como sendo: a construção de uma base de dados de vídeos e imagens de postes elétricos da cidade de Campo Grande – MS; e a anotação dos componentes dos postes elétricos (poste, cruzeta, isolador bastão polimérico, mão francesa plana, topo do poste e a sua base) em cada imagem.

A identificação e classificação em grande escala de postes de eletricidade em um curto período de tempo traz como consequência o benefício da identificação de objetos em

funcionamento não adequado e criação de novas ferramentas voltadas ao processo de manutenção da malha elétrica, principalmente de maneira preventiva, na busca da redução das interrupções de fornecimento de energia elétrica devido a falhas estruturais dos postes.

## 1.4 Organização

Além desta introdução, visando atender aos objetivos da presente tese, esta é composta ainda do Capítulo 2, que apresenta uma revisão bibliográfica de pesquisas que propuseram o uso de algoritmos de detecção e análise de objetos, especificamente de postes elétricos, em imagens terrestres e voltadas para a gestão de energia, o Capítulo 3, que apresenta a metodologia de pesquisa utilizada para a detecção de imagens referentes aos postes de Campo Grande/MS, o Capítulo 4, que apresenta os resultados à partir do algoritmo escolhido pelo presente trabalho e o Capítulo 5 com as conclusões da pesquisa.

# CAPÍTULO 2

## Algoritmos: detecção e análise de objetos

---

---

Este capítulo elucida trabalhos que propuseram o uso de algoritmos de detecção e análise de objetos, especificamente de postes elétricos, em imagens terrestres e voltadas para a gestão de energia.

Dentre as possibilidades da aplicação dos sistemas de sensoriamento remoto e estudos concernentes à detecção de objetos, estes podem ser classificados em: imagens coletadas em *Laser Scanner*, por vias aéreas ou terrestres; ou por imagens em vermelho, verde e azul, ou *red, green e blue* (RGB), por vias aéreas e terrestres. O *Laser Scanner* é um sistema de sensoriamento remoto que permite a determinação da elevação da superfície do terreno e os objetos nela presentes. Logo, o resultado dos levantamentos são representados por meio de uma grade de coordenadas denominada *raster*, considerando as técnicas de processamento digital de imagens (GORDON et. al., 2007, GALANTUCCI; FATIGUSO, 2018).

Neste cenário, as pesquisas abaixo relacionadas são resultantes de uma revisão da literatura realizada de forma empírica pela busca por trabalhos similares à presente pesquisa, isto é, da detecção e classificação de postes elétricos por imagens terrestres. Finda, assim, levantar discussões efetivas acerca de estudos em análise de impacto, resultando em uma tabela comparativa mostrando as diferenças e similaridades entre este estudo e os trabalhos relacionados.

Em Fang et al. (2020), os autores pontuaram que, para selecionar automaticamente as imagens que podem conter postes de energia defeituosos, foi introduzido uma estrutura de detecção de postes de energia, em quatro passos: uso de um método de calibração automática de câmera, com uso do *Jean Yves Bougueret's toolbox* para Matlab, que calcula os parâmetros internos da câmera, para a aquisição das imagens; projeção em três dimensões (3D) da torre de transmissão fazendo uso das imagens capturadas; e extração e correspondência de recursos. Os autores relatam que para garantir a precisão da detecção da torre, em um intervalo razoável de

parâmetros de pose, é preciso definir um grande número de pontos de amostragem, cada ponto de amostragem corresponde a uma pose específica. No entanto, muitos pontos podem acarretar problemas na velocidade da detecção dos objetos. Portanto, a análise de agrupamento pode ser usada para representar um grande número de projeções de torre como uma estrutura hierárquica.

Na pesquisa acima citada, os autores destacam ainda a escolha do recurso de segmento de linha como o recurso de imagem de inspeção e projeção de torre para corresponder à imagem e modelo torre. Existem várias vantagens em selecionar a linha característica do segmento: a própria construção da maioria dos polos e torres nas linhas de transmissão de alta tensão com polos, sobretudo de tubos de aço delgados sendo intuitivo usar o segmento em linha reta para representar o modelo de torre; o segmento de reta é de fácil detecção na imagem de inspeção, podendo ser extraído por algoritmo eficaz de processamento de imagem; a representação e a medida de similaridade da linha reta é simples, apenas duas são necessárias a coordenada de cada ponto final pode descrever completamente um segmento de linha reta, e a semelhança entre os dois segmentos de linha reta também podem ser calculados por uma fórmula simplificada; o segmento de linha reta que forma o modelo da torre ainda é um segmento de linha reta após a transformação de projeção, e as coordenadas do ponto final do segmento de linha reta após a projeção pode ser calculada convenientemente por uma fórmula.

Além disso, foi proposta a implementação do framework e a avaliação de sua eficácia em imagens reais de postes de energia. Os resultados mostram a capacidade do método em detectar o poste de energia de forma automática e precisa, além de relatar a condição anormal da malha, como a ausência de postes. Outrossim, buscou expor a eficácia do método perante as imagens reais de linhas de transmissão de alta tensão, todavia, devido à falta de uma plataforma de testes prática e eficaz, apenas verificou a eficácia do modelo de framework por meio dos dados de simulação.

Em Abdelfattah et. al. (2020), foi descrito o processo para a coleta e análise de um conjunto de dados de imagens aéreas de torres de transmissão e linhas de transmissão, ou TTPLA no acrônimo em inglês para *transmission towers (TTs) and power lines (PLs) Aerial-image*, consistindo de 1.100 imagens com resolução de  $3.840 \times 2.160$  pixels. Outras 8.987 instâncias foram rotuladas manualmente de TTs e PLs, e foram desenvolvidas novas políticas para coletar, anotar e rotular as imagens na TTPLA. O processo foi altamente desafiador, pois existe uma alta variedade de ângulos de visão, escalas, planos de fundo, condições de iluminação e níveis de zoom nas imagens coletadas, o que levou os pesquisadores a proporem

novos critérios: como os modelos de aprendizado profundo que anotam com precisão o nível de pixels a serem empregados pela segmentação de instâncias; e a criação de uma linha de base através de aprendizado de última geração, com base no TTPLA, com *backbones* diversificados e tamanhos de imagens variados. Trazendo um novo desafio para a VC, bem como um avanço em detecção, classificação e segmentação de instâncias.

No estudo de Vemula e Frye (2020a), foi pesquisado o uso do *unmanned aerial vehicle* (UAV) para inspeção de linhas de energia usando um algoritmo de DL. O estudo fez uso de um algoritmo de detecção e segmentação das linhas de energia baseado em aprendizagem por transferência e a aprimoração de uma rede neural convolucional regional de máscara, conhecida em inglês como *Mask Regional Convolutional Neural Network* (Mask R-CNN), implementados no UAV. O sistema de detecção das linhas de energia usou uma arquitetura DL denominada Resnet50, combinada com a arquitetura *Feature Pyramid Network* (FPN) para extração de recursos. A Rede de Proposta de Regiões (RPN) foi treinada ponta a ponta para criar propostas de regiões para cada mapa de características. A pesquisa apresentou o desenvolvimento do sistema *Mask R-CNN Powerline Detector* (MRPD), com testes no UAV *Draganfly XP-4*. Vale destacar que, de todos os algoritmos analisados, o Mask R-CNN foi o primeiro a combinar a detecção de objetos e a segmentação de instâncias. O artigo se concentrou principalmente na detecção automática da linha de força e dos componentes isolantes e do transformador de energia elétrica, e uma área para melhoria no que tange à taxa de quadros. Concluiu-se que houve lentidão para o desempenho de uma aplicação capaz de trabalhar em tempo real. Para trabalhos futuros, foi proposto que o método de segmentação de instâncias em tempo real *You Only Look at the Coefficients* (YOLACT) seja usado para obter melhor desempenho de taxa de quadros.

Já em Vemula e Frye (2020b), os autores desenvolveram um sistema de detecção de linhas de força usando DL para reconhecê-las de maneira autônoma em tempo real usando um *unmanned aircraft system* (UAS), mais veloz que o UAV, com o algoritmo YOLACT, que superou outros na taxa de quadros analisados. O modelo proposto mostrou alta velocidade de reconhecimento e alta precisão, o que permitiu que o UAV inspecionasse grandes regiões mais rapidamente, diminuindo o tempo e o custo de operação desse procedimento e ainda, ampliando a segurança do mesmo. Considerando isso, o *Real-Time Powerline Detection System* (RPDS) implementado aproveitou a capacidade do algoritmo YOLACT, um modelo de segmentação de instância em tempo real proposto recentemente e que atinge a precisão média mínima mais alta e taxa de 33,5 *frames* por segundo. Os resultados usando essa abordagem mostraram-se

promissores e com desempenho superior quando comparados ao estudo anteriormente analisado. Segundo os pesquisadores, os próximos estudos preveem a aplicação de um sistema de detecção de falhas na análise e a investigação de outros componentes específicos das linhas de transmissão, contribuindo com a segurança necessária para conduzir o exame dos recursos e a inspeção das linhas de alimentação relacionados à segurança necessária à inspeção da malha elétrica. Alam et. al. (2020) apresentam um novo método de monitoramento de postes que usa métodos de DL e VC, que buscou capturar automaticamente os ângulos de inclinação dos postes usando um UAV. O método calculou a flexão exercida nos postes devido ao vento, às forças gravitacionais e ao peso do cabo, para comparar esta inclinação com o momento de ruptura. O método também incluiu um modelo baseado em AM para prever, de maneira mais rápida, as condições de resiliência de um poste após eventos como vento e furacões. No que refere à determinação do ângulo de inclinação dos postes, o método de medição proposto consistiu nas etapas de: captura de imagens dos postes de serviços públicos usando um UAV; a segmentação das imagens dos postes usando uma RNP; e o cálculo do ângulo de inclinação de cada poste usando métodos de VC. Segundo a citada pesquisa, o monitoramento contínuo das condições estruturais dos postes é vital para a redução de quedas de energia durante e após ventos extremos e manter o fornecimento energia de forma contínua e segura, auxiliando as empresas de serviços públicos a examinar e tomar decisões críticas para a manutenção ou substituição de postes com base nas condições.

Em Zhang et al. (2018), os autores propõem o uso de um método baseado em DL para mapear automaticamente postes no formato de cruz, ou *utility poles with crossarms* (UPCs) a partir de imagens do *Google Street View* (GSV). O método combina um algoritmo de DL de detecção de objetos chamado RetinaNet e um método de cálculo de localização do poste a partir do sensor, isto é, do local da foto em relação ao poste no GSV. Os resultados mostraram que, com parâmetros configurados corretamente, o detector de objeto RetinaNet é capaz de detectar a maioria dos UPCs no GSV com precisão IoU de 0,78, e que mesmo com distâncias de 1 metro, 5 metros e 10 metros entre o local da foto e poste, foi possível identificar a maioria dos postes usando a DL.

No estudo Fambrini et al. (2020), os autores têm como objetivo a utilização da VC para capturar e reconhecer imagens da rede de distribuição de energia durante o deslocamento do carro de inspeção por meio de imagens termodinâmicas. Para isso foi montado um veículo especial, equipado com oito câmeras para realizar a fiscalização em ambos os lados da via. Assim, é possível inspecionar as linhas de distribuição de energia sem a necessidade de parar o

veículo e sem a necessidade de um operador humano. A manutenção preditiva foi um dos métodos mais eficientes para monitorar o desempenho dos equipamentos por meio de instrumentos que fornecem dados quantitativos dos componentes da rede elétrica. O estudo propôs um sistema de processamento das imagens em tempo real e também relatórios das inspeções dos elementos elétricos, tudo sem a intervenção humana. A inspeção elétrica fez uso da captura da imagem óptica e térmica do elemento com super aquecimento anormal para criar relatórios de manutenção preditiva. Para o modelo, foi usada uma CNN por sua capacidade de processamento de um alto volume de dados. O estudo de Ramlal et al. (2020) descreve que, com o envelhecimento da infraestrutura de distribuição elétrica, os operadores de rede em todo o mundo estão se tornando mais cientes da necessidade de monitorar sua malha de transmissão. No entanto, a extensão geográfica do sistema torna esta tarefa difícil e custosa, pois milhares de componentes requerem inspeção visual para a identificação de falhas que podem se tornar potencialmente catastróficas. Foi usada uma RNP para detectar automaticamente áreas de dano no concreto dos postes de uma concessionária europeia, a partir de fotografias da malha elétrica. Os autores escrevem que o tempo gasto no trabalho de campo foi reduzido, e houve baixa variabilidade nos resultados em relação às avaliações humanas.

Os resultados da pesquisa acima mostraram que, mesmo com um pequeno conjunto de dados para treinamento, a RNP foi capaz de identificar danos com alto nível de precisão. A RNP de detecção de objetos foi o *You Only Look Once* (YOLO) V2. A pontuação de confiança limite para o detector atribuiu uma detecção positiva de 0,5, em que o classificador detectou com sucesso todas as áreas danificadas e, em todo o conjunto de dados, foi capaz de detectar todas as ocorrências de danos em imagens próximas. Nas imagens mais distantes, em que as áreas danificadas eram muito pequenas em comparação com o tamanho da imagem, várias detecções não ocorreram. Ainda, o método teve sucesso na detecção de rachaduras e buracos em postes de serviços públicos a partir de imagens não padronizadas obtidas em patrulhas a pé e com variações de fundo, primeiro plano, sobreposição e distância.

Em Kim et al. (2021), os autores propuseram uma abordagem baseada em análise de dados para apoiar a tomada de decisão sobre riscos para manutenção de concessionárias sob condições extremas de vento. Imagens do GSV foram usadas para avaliar as propriedades geométricas de postes de serviços públicos, como o ângulo de inclinação dos mesmos. Em seguida, a probabilidade de falha de postes de serviços públicos foi analisada sob várias condições, a saber, idade, inclinação e cargas de vento, gerando modelo de cidade virtual tridimensional. O método foi testado por meio de estudos de caso para validar um algoritmo e

estimar ângulos de inclinação de postes na busca de entender o progresso de falhas de postes de utilidade inclinados a partir de uma perspectiva de rede.

Os resultados do estudo de caso acima exposto demonstraram que o método implementado obteve sucesso para avaliar a vulnerabilidade dos postes que poderiam levar a falhas em cascata no sistema de infraestrutura de distribuição de energia. Com base no ambiente virtual proposto, foi permitida uma melhor tomada de decisões sobre os riscos em situações de desastre, o que criou uma oportunidade para priorizar as tarefas de manutenção em relação à infraestrutura de distribuição de energia.

No quadro abaixo estão compiladas as informações da revisão que contribuíram para que elucidassem o estado da arte do uso de algoritmos na detecção de imagem, bem como as lacunas e desafios ainda existentes, contribuindo significativamente para a construção da presente pesquisa.

Quadro 1 - Resumo dos Artigos da Revisão Bibliográfica

Citação	Objetivos	Captura de imagens	Métodos	Resultados
Fang <i>et al.</i> (2020)	Encontrar automaticamente imagens que podem conter torres de transmissão de energia defeituosas por meio da modelagem 3d destes e cálculo dos segmentos de reta que os compõem, buscando identificar segmentos defeituosos	Captura de imagens aéreas e torres de transmissão por meio de um UAV, com a câmera calibrada automaticamente por meio do MATLAB	Criação de um <i>framework</i> consiste em quatro partes: captura das imagens; construir o modelo 3D das torres e calcular suas características, por meio das coordenadas X, Y e Z dos segmentos de reta identificados na imagem; calcular a semelhança da projeção da torre em diferentes posições e realizar a análise de <i>cluster</i> para identificar o poste sob diferentes ângulos; extrair as características da torre de transmissão na imagem de inspeção por meio do método <i>Line Segment Detection</i> (LSD), combinar com a projeção do modelo 3D da torre e usar o GPS do UAV, o ângulo de instalação da câmera, a direção de movimentação	Devido à falta de uma plataforma de testes prática e eficaz, a eficácia do <i>framework</i> foi verificada por meio dos dados de simulação, com sucesso na identificação dos mesmos.

			do UAV e outros parâmetros para restringir o escopo da comparação.	
Abdelfattah, Wang e Wang (2020)	Coletar, anotar e rotular um conjunto de imagens de torres e linhas de transmissão por meio de imagens aéreas para colaborar com pesquisas na área.	Por meio de um UAV, foram gravados vídeos de 1 a 5 minutos em torres de transmissão em T e respectivas linhas, transformadas em 1.100 imagens destas.	Após a captura das imagens, estas receberam 8.987 anotações manuais de <i>transmission towers</i> (TTs) e <i>power lines</i> (PLs) por meio do <i>software</i> LabelME, posteriormente avaliadas pela precisão média por meio do IoU e dos modelos de DL Resnet-50 e Resnet-101. O conjunto de imagens resultante foi dividido em 70% de treinamento, 10% de validação e 20% de testes e usadas para avaliação por meio do modelo de segmentação de imagens em tempo real YOLACT.	O conjunto de dados TTPLA, anotado sob precisão em nível de <i>pixel</i> , foi disponibilizado para outras pesquisas na temática, e permite ser usado por modelos de DL em trabalhos de VC.
Vemula e Frye (2020)	Usar um UAV para inspecionar linhas de transmissão por meio de um algoritmo de DL.		Uso de um algoritmo de segmentação para a detecção de linhas de energia baseado em aprendizado de transferência e utilização de uma rede neural convolucional regional de máscara aprimorada (Mask RCNN), implantados em um UAV. O	Houve lentidão no desempenho do modelo para que fosse possível trabalhar em tempo real, e foi proposto o uso do método de segmentação <i>You Only Look at the Coefficients</i> (YOLACT) para solucionar esta limitação.

			<p>sistema de detecção de linhas de energia usou a arquitetura de DL Resnet50, combinada com a arquitetura Feature Pyramid Network (FPN) para extração de recursos. A Rede de Propostas de Região (RPN) foi treinada de ponta a ponta para criar propostas de região para cada recurso. O sistema foi testado com 600 imagens divididas aleatoriamente em 80% treinamento, 10% validação e 10% teste.</p>	
Vemula e Frye (2020)	Reconhecer as linhas em tempo real, de modo autônomo, usando um UAS, com o algoritmo YOLACT.		<p>Implementada a <i>Real-Time Powerline Detection System</i> (RPDS) com uso do YOLACT como método de DL de segmentação de imagens em tempo real. O modelo foi testado utilizando-se as RNs ResNet-50 e ResNet-101.</p>	Os resultados foram mais promissores para a detecção em tempo real de linhas de transmissão que o estudo anterior dos mesmos autores, pela velocidade de segmentação de imagens em decorrência do uso do YOLACT.
Alam et al.(2020)	Propor um modelo de monitoramento de postes usando DL e VC por meio de um UAV, de modo a detectar inclinações nos		<p>Fazendo modificações na CNN SegNet, com 13 camadas convolucionais, o modelo foi pré-treinado com a aplicação do método de</p>	Foi gerada uma matriz de confusão dos resultados de classificação da inclinação dos postes, que mostrou a resiliência destes, após eventos climáticos, por meio do modelo proposto.

	postes.		aprendizagem por transferência e teve sua inclinação calculada pela técnica Hough Transform. Modelos de resiliência foram usados para o monitoramento da saúde dos postes após ventos e furacões	
Zhang <i>et al.</i> (2018)	A partir de imagens do <i>Google Street View</i> (GSV), mapear automaticamente postes no formato de cruz, ou <i>utility poles with crossarms</i> (UPCs) usando DL.		O método DL implementado foi o RetinaNet-101 e uma modificação do <i>brute-force-based line-of-bearing</i> , (LOB), com cálculo de precisão por meio do IoU. Cerca de 70% dos postes foram identificados e localizados no GSV por meio do modelo proposto, em 500 imagens de teste. Os pesquisadores mostraram que o modelo obteve melhor precisão que quando testado com o uso do RetinaNet-50 e do RetinaNet-152.	Foi possível detectar a localização da maioria dos postes em formato de cruz no GSV por meio do modelo proposto.
Fambrini <i>et al.</i> (2020)	Capturar e detectar imagens de linhas de transmissão por um carro de inspeção	Uso de 8 câmeras termodinâmicas instaladas em um carro, para captura de imagens em 360°	O projeto fez uso de uma rede neural convolucional para análise das linhas de	Foi possível inspecionar centenas de quilômetros de linhas de distribuição elétrica sem a necessidade da operação de um ser humano.

	em movimento, através de imagens termodinâmicas.		transmissão.	
Ramlal <i>et al.</i> (2020)	Identificar postes danificados por meio de uma RNP.		Em um conjunto de 150 imagens de postes, os pesquisadores usaram o IoU para gerar caixas delimitadoras dos possíveis objetos, e o modelo de detecção chamado <i>You Only Look Once</i> (YOLO) V2, com 91 camadas.	Em imagens próximas, o modelo obteve sucesso na identificação de buracos e rachaduras nos postes, porém não performou bem quando o tamanho do dano era pequeno em relação ao tamanho da imagem.
Kim <i>et al.</i> (2021)	Propor uma abordagem de detecção de postes elétricos falhos, com inclinação, para apoiar a tomada de decisões sobre riscos na malha, voltado à estes quando passam por condições de ventos extremos.		Por meio do ResNet18, uma rede neural convolucional com 72 camadas, as imagens foram segmentadas e geradas nas caixas delimitadoras. Posteriormente, a AlexNet foi usada para classificação dos objetos. Para analisar o ângulo dos postes, foi usado o Hough Transform.	Foi possível melhorar a tomada de decisões sobre a manutenção dos postes em situações de desastre.

# CAPÍTULO 3

## Aspectos metodológicos

---

### 3.1 Conjunto de imagens

A presente pesquisa optou por uma pesquisa qualitativa, a fim de que atores sociais envolvidos fossem levados a refletir sobre suas ações e as consequências dessas ações para a realidade na qual estão inseridos (SILVA, 2014), e de caráter descritivo no sentido de trazer à luz as características do objeto de estudo e elucidar opiniões, atitudes e experiências (GIL, 2002).

Optou-se por um estudo de caso de mapeamento da distribuição de postes na cidade de Campo Grande/MS, permitindo um registro apurado das condições de cada poste e a sua localização em caso de interrupção na distribuição de energia, visto que, atualmente, não existe tal mapeamento individual dos postes existentes na cidade. Objetivou-se, portanto, a construção de uma base de dados de vídeos e imagens de postes elétricos da cidade de Campo Grande – MS, bem como a anotação dos componentes como os diferentes tipos de poste (Postes com cruzeta, sem cruzeta (Liso), e o Transformador do poste, em cada imagem.

O delineamento do estudo de caso seguiu os passos sugeridos por Yin (2003), a saber: formulação das questões do estudo; identificação das proposições; definição das medidas de análise; a lógica que relaciona os dados às proposições; os critérios de interpretação dos resultados. Segundo o autor, um estudo de caso é útil para “investigar um fenômeno contemporâneo em um contexto da vida real” (YIN, 2003, p. 13, tradução nossa).

Assim, foram seguidos os seguintes passos, de acordo com o citado autor: a formulação das questões do estudo definidas por meio dos objetivos do presente trabalho; a identificação das proposições já elucidadas acima; a definição das unidades de análise, a utilização de um único estudo de caso é apropriada quando este é analisado para determinar se as proposições de uma hipótese são corretas. No presente estudo, essa hipótese é a de que o mapeamento individual dos postes existentes na cidade, permite a construção de uma base de dados de vídeos

e imagens de postes elétricos da cidade de Campo Grande/MS e que a detecção de objetos por meio da DL pode melhorar e tornar mais rápida a detecção de possíveis problemas na malha elétrica da cidade. Além da análise de conteúdo do levantamento bibliográfico e comparação com o Estudo de Caso, este permite que, por meio da análise de uma determinada situação, sejam estabelecidas possibilidades de interferências no sentido de tentar apreender comportamentos relevantes e, as estratégias podem ser utilizadas pelos docentes.

Para a realização dos experimentos, uma câmera que possui um campo de visão de 360 graus, foi acoplada na parte superior de um carro para gravar os vídeos. Foram gravados 21 arquivos de vídeo, com um total de 31 minutos de duração, por meio do uso de um veículo automóvel equipado com uma câmera GoPro de lente de 360 graus, com resolução de 3k a 60fps. As gravações ocorreram na região central do município de Campo Grande – MS (as áreas não foram selecionadas de acordo com nenhum critério específico, apenas a tentativa de gravar regiões com postes e bocas de lobo). As imagens possuem como objetos-alvo postes elétricos, e das imagens coletadas foram extraídos 1.877 frames.

Este estudo apresenta uma descrição detalhada do dataset, que é uma coleção de dados organizados, utilizado em uma pesquisa voltada para a detecção e classificação de elementos relacionados a postes e geradores. Os frames de vídeo que compõem esse dataset foram obtidos a partir de gravações realizadas por estudantes da Universidade Federal de Mato Grosso do Sul (UFMS). Eles utilizaram uma câmera Insta360 com lente "olho-de-peixe" para capturar as imagens em questão. A taxa de amostragem das imagens foi estabelecida em 1 frame por segundo, e essa extração foi realizada através da ferramenta FFMPEG, que é uma ferramenta amplamente utilizada para manipulação de arquivos de vídeo em diversos formatos.

Após a extração dos frames, mais de 10 mil deles foram submetidos a um processo de anotação, no qual os objetos de interesse foram cuidadosamente rotulados. Para esta etapa, foi empregada a ferramenta ROBOFLOW, uma plataforma que oferece recursos eficientes de anotação de dados para tarefas de visão computacional. Durante o processo de anotação, foi feita uma seleção criteriosa das imagens a serem incluídas no dataset final, visando evitar possíveis problemas de overfitting, ou seja, garantindo que o modelo não se ajuste excessivamente aos dados de treinamento e seja capaz de generalizar para novos casos.

As classes de objetos anotadas neste dataset incluem "Poste" (representando a categoria geral de postes), "Tipos de Postes" (que inclui subclasses para descrever diferentes tipos de postes, como "cruzeta" e "liso") e "Gerador do Poste" (descrevendo o gerador associado ao poste). Essa anotação detalhada permite uma classificação mais específica dos objetos, o que é fundamental para tarefas de detecção e classificação precisas.

Todas as imagens no dataset possuem dimensões fixas, com 1568 pixels de largura por 1504 pixels de altura. Essa uniformidade no tamanho das imagens simplifica o processamento e a análise dos dados, tornando o dataset uma base sólida para o treinamento de modelos de aprendizado de máquina.

Em resumo, este dataset representa um recurso valioso para o desenvolvimento e treinamento de algoritmos de detecção e classificação de elementos relacionados a postes e geradores. Através de um processo de anotação cuidadoso e o uso das ferramentas FFMPEG e ROBOFLOW, foram criadas imagens rotuladas que possibilitam a capacitação de modelos de aprendizado de máquina para identificar com precisão postes e seus diferentes tipos, bem como os geradores associados. Esses dados têm relevância significativa em aplicações práticas, como a manutenção de infraestrutura elétrica e a automação de inspeções visuais.

*Figura 8 - Exemplos de frames extraídos de um vídeo tendo postes elétricos como objetos-alvo*



Fonte: do autor (2023).

## 3.2 Método de detecção

Dois extratores de características muito utilizados na literatura científica são o VGGNet e ResNet, usualmente utilizados em problemas de classificação, e, por vezes, usados como extratores por removerem a última camada totalmente conectada da rede neural, que realiza a classificação. A partir desse extrator, as imagens são descritas como um vetor de características que são enviadas aos métodos de detecção para descobrir regiões na imagem que contém características similares aos objetos de interesse. Para isto, existem diversos algoritmos de detecção além dos citados acima, como o YOLO, Mask R-CNN, explicados anteriormente. (GIRSHICK et. al., 2014, SZEGEDY; TASHEV; ERHAN, 2013).

Na busca por otimizar a detecção e classificação de objetos no âmbito da visão computacional, diversas metodologias têm sido exploradas para aprimorar a precisão e confiabilidade desses processos. Entre esses métodos, RetinaNet, Faster R-CNN e RTMDet

surtem como promissores concorrentes. Estes métodos empregam técnicas avançadas da área de aprendizado profundo e detecção de objetos para identificar e classificar objetos em imagens. À medida que nos aprofundamos na análise de seu desempenho, fica evidente que avaliar sua precisão e revocação é crucial para mensurar sua eficácia.

Neste trabalho, serão investigados os algoritmos do estado-da-arte que têm apresentado excelentes resultados na tarefa de Detecção de Objetos, quando avaliados em 28 conjuntos de dados como o ImageNet. Além disso, os métodos clássicos de detecção de objetos como RetinaNet e Faster-RCNN foram também analisados e comparados com o modelo estado-da-arte RT-Mdet, a fim de destacarmos o mais eficiente e rápido para o caso de Campo Grande/MS.

### 3.3 Estado da arte

A detecção de objetos é uma tarefa crucial em visão computacional, com aplicações em áreas, tais como segurança, automação, reconhecimento de padrões, entre outros exemplos, e para garantir resultados precisos e eficientes nessa área, é fundamental a utilização de um modelo de estado da arte adequado.

O modelo de estado da arte, ou *state-of-the-art*, refere-se à técnica ou abordagem mais avançada e eficiente disponível em um determinado momento, e considerando o contexto da detecção de objetos, um modelo de estado da arte representa o algoritmo ou a arquitetura mais atualizada, que obteve os melhores resultados em relação à precisão e ao desempenho.

A importância de um modelo de estado da arte adequado para estudos de detecção de objetos é multifacetada. Em primeiro lugar, ele permite que os pesquisadores e profissionais da área alcancem resultados mais precisos e confiáveis em suas aplicações, pois à medida que novas técnicas são desenvolvidas, a qualidade das detecções melhora, possibilitando a identificação de objetos com maior acurácia, mesmo em condições desafiadoras, como baixa iluminação, obstruções ou variações de escala. Ao utilizar um modelo de estado da arte, os estudos de detecção de objetos podem se beneficiar de algoritmos otimizados e eficientes, reduzindo o tempo de processamento e o consumo de recursos computacionais, demonstrando assim, ser um modelo essencial para o alcance de resultados precisos, eficientes e confiáveis, impulsionando o progresso contínuo e as aplicações práticas na área de visão computacional.

### 3.3.1 MMDetection

O Framework utilizado neste trabalho é o MMDetection, que se destaca principalmente por sua facilidade, desempenho e seu vasto número de modelos, o que permite flexibilidade na detecção. A biblioteca é desenvolvida por cima do PyTorch, outra biblioteca bastante conhecida na área de Deep Learning. As principais características do MMDetection, incluem:

1. Ampla gama de modelos de ponta, como VFNet, Mask-RCNN, etc;
2. Fácil Personalização de modelos para melhor desempenho em CPUs e GPUs;
3. Agilidade no processo de Inferência de Imagem/Vídeo;
4. Integração com outros modelos como TorchVision, TorchAudio, OpenCV.
5. Fácil conversão de modelos para ONXX por exemplo.

## 3.4 Modelos escolhidos

Ao analisarmos as métricas de desempenho dos modelos RetinaNet, Faster R-CNN e RTMDet, podemos compreender melhor a eficácia de cada um na tarefa de detecção de objetos. As métricas utilizadas, como Média de Precisão (AP) e Média de Revocação (AR) em diferentes configurações, fornecem insights sobre o quão bem esses modelos conseguem identificar e classificar objetos em várias situações. A seguir, são apresentadas as pontuações de cada modelo para diferentes limiares de Interseção sobre União (IoU), áreas e detecções máximas.

A análise das métricas de desempenho dos modelos RetinaNet, Faster R-CNN e RTMDet fornece uma visão abrangente de sua capacidade em termos de detecção e classificação de objetos. Observa-se que o modelo Faster R-CNN se destaca, demonstrando uma Média de Precisão (AP) mais elevada em várias configurações de limiar de IoU e área. Além disso, o Faster R-CNN também exibe um desempenho impressionante em termos de Média Revocação (AR). No entanto, o RTMDet não fica muito atrás, exibindo resultados consistentes e competitivos em diversas configurações. Por outro lado, o RetinaNet apresenta pontuações um pouco mais modestas, mas ainda demonstra habilidade em tarefas de detecção de objetos. Em suma, a análise das métricas revela que tanto o Faster R-CNN quanto o RTMDet têm um desempenho notável, enquanto o RetinaNet também oferece resultados respeitáveis. Essa avaliação fornece informações valiosas para selecionar o modelo mais adequado para a

tarefa em questão e para orientar futuros desenvolvimentos na área de detecção de objetos em imagens.

### 3.4.1 Faster R-CNN

Trata-se de um modelo clássico na área de detecção de objetos, altamente preciso no reconhecimento de padrões, e tal modelo se baseia em duas etapas sendo composto por uma rede convolucional, conhecida como Região de Proposta (Region Proposal Network – RPN), que gera regiões de interesse (Region of Interest), essa é seguida por uma outra rede conectada para fazer o ajuste dessas regiões que foram geradas. É um método avançado de detecção de objetos em imagens que combina a eficiência do algoritmo R-CNN original com a velocidade do algoritmo Fast R-CNN. A principal vantagem do Faster R-CNN é sua alta precisão e velocidade, e ao integrar a geração de propostas de região na própria rede neural, o método evita a necessidade de executar algoritmos de geração de região separadamente, resultando em uma detecção mais rápida e eficiente. O motivo para escolha pelo referido modelo se justifica pelo excelente desempenho em conjuntos de dados muito grandes.

### 3.4.2 RetinaNet

Se trata de um modelo de detecção de objetos baseado em redes neurais convolucionais, projetado para superar o desafio da detecção de objetos em escalas variadas e com diferentes níveis de dificuldade. Ele foi proposto para solucionar as limitações dos modelos anteriores que lutavam para detectar objetos pequenos ou de baixo contraste.

O modelo RetinaNet tem como principais contribuições a introdução do Focal Loss, uma função de perda que dá mais ênfase aos exemplos difíceis durante o treinamento. Essa abordagem permite que o modelo lide melhor com a presença de um grande número de exemplos negativos em comparação com os positivos, o que é comum em tarefas de detecção de objetos.

Diferente do Faster-RCNN, esse modelo tem apenas uma etapa. A estratégia principal desse modelo é arquitetura chamada de Feature Pyramid Network (FPN), que lida melhor com objetos de diversos tamanhos. Esse modelo foi escolhido, primeiramente por seu um modelo clássico e por ter uma boa precisão com objetos muito pequenos, a classe Base no caso dessa dissertação.

### 3.4.3 RTMDet

RTMDet ou Real-Time-Multi-Detector é o modelo estado-da-arte do MMDetection. MMDetection é um framework de código aberto para detecção de objetos que fornece implementações de várias arquiteturas populares de detecção. O RTMDet é uma das arquiteturas disponíveis no MMDetection e é conhecido por sua capacidade de realizar detecção em tempo real.

O RTMDet foi projetado para atingir um equilíbrio entre precisão e velocidade, tornando-o adequado para aplicações que exigem detecção de objetos em tempo real, como sistemas de segurança, veículos autônomos e vigilância por vídeo. Ele usa uma combinação de técnicas avançadas, como âncoras em múltiplas escalas e aspectos, redes neurais convolucionais eficientes e otimizações específicas para acelerar a inferência em GPUs, e tal modelo faz uso de uma arquitetura de rede leve buscando equilíbrio entre velocidade de inferência e precisão, ele faz uso de um mecanismo de geração de propostas, tornando-o um modelo extremamente preciso na tarefa de detecção. Este modelo foi escolhido por ser um modelo estado da arte e por ser extremamente preciso em sistemas de veículos autônomos.

Para avaliar o desempenho dos objetos em vídeos serão investigados os métodos de detecção de objetos em vídeos (do *Inglês Vídeo Object Detection VOD*) que tratam de observar a dinâmica de movimento dos objetos em frames consecutivos de vídeos.

Os dados coletados foram analisados por meio da técnica de análise de conteúdo do tipo temática, proposto por Bardin (2016), que consiste em descobrir os núcleos de sentido que compõem uma comunicação, cuja presença ou frequência significam algo para o objeto analítico visado. Assim, na fase da pré-análise, o material foi decomposto em cinco novas fases: análise flutuante, constituição do corpus documental; formulação de pressupostos, objetivos e a categorização, que orientou a análise; referenciação dos índices e a elaboração de indicadores; preparação do material, com a análise das imagens. Na fase de exploração do material foi realizada a codificação do material analisado, para a construção de categorias de análise, identificando elementos comuns que compuseram “unidades de registro”, seguindo critérios de repetição. Após a exploração do material, foi realizada a categorização das imagens em eixos temáticos (Bardin, 2016) apresentados e analisados nos resultados e discussão da presente dissertação.

# CAPÍTULO 4

## Resultados alcançados

---

### 4.1 Base de dados

A abordagem sistêmica permite uma análise abrangente e precisa do desempenho do modelo de classificação. Ao considerar os resultados obtidos, é possível inferir a importância da correta divisão dos dados e sua influência nos resultados finais, e a tabela a seguir com informações adicionais do balanceamento das classes:

Quadro 2 - Balanceamento das classes e ocorrências

<b>Classes</b>	<b>Ocorrências</b>
Postes	1555
Liso	547
Transformadores	305
Cruzeta	1008

Seguem exemplos de algumas imagens trabalhadas ao longo das fases do estudo:

## Imagens (Treino)

*Figura 9* - Exemplos de imagens utilizados na fase de treino do referido estudo tendo postes elétricos como objetos-alvo



Fonte: do autor (2023).

## Imagens (Validação)

*Figura 10* - Exemplos de imagens utilizados na fase de validação do referido estudo tendo postes elétricos como objetos-alvo



Fonte: do autor (2023).

## Imagens (Teste)

Figura 11 - Exemplos de imagens utilizados na fase de teste do referido estudo tendo postes elétricos como objetos-alvo



Fonte: do autor (2023).

No MMDetection, os modelos são treinados com o uso de dois arquivos, a configuração do modelo (onde ficam especificadas as camadas, o número de épocas, *learning rate*, o otimizador, etc) e o modelo pré-treinado, facilmente baixado do GitHub do Framework.

Essas configurações visam descrever a arquitetura do modelo e as opções de treinamento, o que o torna com um dos pontos fortes deste framework, como também, a ampla gama de modelos com diferentes arquiteturas, hiperparâmetros e modelos pré-treinados facilitam muito o processo de treinamento. Abaixo seguem as configurações de cada arquivo de modelo: Quadro 3 - Configurações de arquivo dos modelos do estudo

Modelo	Camadas	Lr	Otimizador
Faster R-CNN	<ul style="list-style-type: none"><li>● Resnet101</li><li>● FPN</li><li>● RPN</li><li>● ROI</li></ul>	0.01	SGD
RetinaNet	<ul style="list-style-type: none"><li>● Resnet101</li><li>● FNP</li></ul>	0.01	SGD
RTMDET	<ul style="list-style-type: none"><li>● CSPNeXT</li><li>● CSPNeXTPAFN</li></ul>	0.01	AdamW

Quadro 4 - Resultados obtidos pelos modelos do estudo

<b>Modelo</b>	<b>RetinaNet</b>	<b>Faster R-CNN</b>	<b>RTMDet</b>
Precisão	88%	90%	90%
Épocas	25	25	25
Postes	93%	100%	92%
Cruz	87%	98%	87%
Liso	87%	98%	88%
Transformador	93%	95%	99%

*Figura 11* - Exemplos de imagens utilizados na fase de teste do referido estudo tendo postes elétricos como objetos-alvo



Fonte: do autor (2023).

Figura 12 – Matriz de confusão – Modelo Faster RCNN

MATRIZ DE CONFUSÃO						
	cruz	gerador	liso	poste	background	
cruz	98%	0%	0%	0%	1%	100
gerador	0%	95%	0%	0%	5%	80
liso	0%	0%	98%	0%	2%	40
poste	0%	0%	0%	100%	0%	20
background	9%	72%	14%	3%	0%	0
	Rótulo de predição					

Fonte: do autor (2023).

Figura 13 – Matriz de confusão – Modelo RTMDet

MATRIZ DE CONFUSÃO						
	cruz	gerador	liso	poste	background	
cruz	87%	0%	1%	0%	11%	100
gerador	0%	99%	0%	0%	0%	80
liso	0%	0%	88%	1%	9%	40
poste	0%	0%	0%	92%	7%	20
background	27%	1%	42%	28%	0%	0
	Rótulo de predição					

Fonte: do autor (2023).

Figura 14 – Matriz de confusão – Modelo RetinaNet

MATRIZ DE CONFUSÃO						
	cruz	gerador	liso	poste	background	
cruz	87%	0%	1%	0%	11%	100
gerador	0%	93%	0%	0%	3%	80
liso	0%	0%	87%	2%	9%	40
poste	0%	0%	0%	93%	3%	20
background	27%	1%	42%	28%	0%	0
	Rótulo de predição					

Fonte: do autor (2023).

## 4.2 Resultados experimentais

As métricas de Precisão Média (AP) e Recuperação Média (AR) são usadas para avaliar o desempenho de modelos de detecção de objetos. É calculado comparando os resultados de detecção gerados pelo modelo com as anotações da caixa delimitadora de objetos reais na imagem. Para fins de cálculo dessa métrica segue-se pelo seguinte rito:

**Precisão Média (AP):** O cálculo da precisão média inclui as seguintes etapas: **Determine a ordem:** Primeiro, precisamos ordenar todas as detecções geradas pelo modelo de acordo com suas probabilidades de detecção. Isto significa que as detecções mais confiáveis devem aparecer em primeiro lugar na lista. **Calcular precisão e recall:** Em seguida, percorremos a lista classificada de detecções e calculamos a precisão e o recall de cada detecção. A precisão é a razão entre as detecções corretas e o número total de detecções até aquele ponto, enquanto a recuperação é a razão entre as detecções corretas e o número total de objetos reais na imagem. **Calcular a área sob a curva (AUC):** Depois de calcular a precisão e a recuperação para cada detecção, calcule a área sob a curva (AUC) da curva de linha de base de precisão. Isso pode ser feito usando a integração numérica, como a regra do trapézio. **Cálculo**

**de AP:** Finalmente, a Precisão Média (AP) é calculada como a AUC média de todas as classes ou objetos rastreados. Conforme mostrado nos dados, calcular o AP para múltiplas configurações de IoU e zonas de objeto permite calcular o AP para cada configuração separadamente e, em seguida, calcular a média. **Recuperação média (AR):** O cálculo da memória média envolve várias etapas. Detecções ordenadas: semelhante aos cálculos de AP, começamos classificando todas as detecções por nível de confiança. **Cálculo de recall:** Em seguida, percorremos a lista classificada e calculamos o recall para cada pesquisa. As respostas foram calculadas como a razão entre detecções corretas e o número total de objetos reais na imagem. **Cálculo AR:** A recuperação média (AR) é calculada como o valor médio de recuperação para diferentes configurações de IoU e regiões de objeto, se aplicável, AR mede a capacidade geral do modelo para detectar objetos com base nessas configurações.

Em ambos os casos, a decisão é tomada utilizando a Interseção sobre União (IoU). Se a decisão for considerada correta, o IoU é a razão entre a área de sobreposição de detecção e a área de intersecção entre a detecção e a caixa delimitadora real do objeto. Uma detecção é considerada correta se o IoU for maior ou igual ao limite especificado (por exemplo,  $\text{IoU} \geq 0,5$ ). Essa métrica é comumente usada para avaliar modelos de detecção de objetos para problemas de visão computacional, e os resultados podem variar dependendo da configuração do limite e da região escolhida. As métricas a seguir são usadas para avaliar o desempenho de modelos de detecção de objetos em tarefas de detecção de objetos de imagem. Isso mostra quão bem o modelo pode encontrar e classificar objetos em diferentes configurações. Descreverei cada métrica e resultados para três modelos: RetinaNet, Faster R-CNN e RTMDet.

Métricas de avaliação: Precisão média (AP): esta métrica mede a precisão média do modelo em diferentes configurações de limite de sobreposição (IoU) e domínios de objeto, o que fornece uma precisão média ponderada entre configurações.

Recall Médio (AR): Essa métrica mede a taxa média de recall do modelo em diferentes configurações de limiar de sobreposição (IoU) e áreas do objeto. Ela fornece uma média ponderada das taxas de recall em diferentes configurações.

Em ambos os casos, a decisão é tomada utilizando a Interseção sobre União (IoU). Se a decisão for considerada correta, o IoU é a razão entre a área de sobreposição de detecção e a área de intersecção entre a detecção e a caixa delimitadora real do objeto. Uma detecção é considerada correta se o IoU for maior ou igual ao limite especificado (por exemplo,  $\text{IoU} \geq 0,5$ ). Essa métrica é comumente usada para avaliar modelos de detecção de objetos para problemas de visão computacional, e os resultados podem variar dependendo da configuração

do limite e da região escolhida. As métricas a seguir são usadas para avaliar o desempenho de modelos de detecção de objetos em tarefas de detecção de objetos de imagem. Isso mostra quão bem o modelo pode encontrar e classificar objetos em diferentes configurações. Descreverei cada métrica e resultados para três modelos: RetinaNet, Faster R-CNN e RTMDet.

Métricas de avaliação: Precisão média (AP): esta métrica mede a precisão média do modelo em diferentes configurações de limite de sobreposição (IoU) e domínios de objeto. Isso fornece uma precisão média ponderada entre configurações.

AP @[ IoU=0.50:0.95 | area= all | maxDets=100 ]: A precisão média para todas as áreas de objetos, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 100 detecções, é 0,498 para RetinaNet, 0,454 para Faster R-CNN e 0,519 para RTMDet.

AP @[ IoU=0.50 | area= all | maxDets=1000 ]: A precisão média para todas as áreas de objetos, considerando uma sobreposição IoU de 0,50 e no máximo 1000 detecções, é 0,883 para RetinaNet, 0,906 para Faster R-CNN e 0,901 para RTMDet.

AP @[ IoU=0.75 | area= all | maxDets=1000 ]: A precisão média para todas as áreas de objetos, considerando uma sobreposição IoU de 0,75 e no máximo 1000 detecções, é 0,490 para RetinaNet, 0,377 para Faster R-CNN e 0,514 para RTMDet.

AP @[ IoU=0.50:0.95 | area= small | maxDets=1000 ]: A precisão média para objetos de pequena área, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,468 para RetinaNet, -1,000 para Faster R-CNN e 0,466 para RTMDet.

AP @[ IoU=0.50:0.95 | area= medium | maxDets=1000 ]: A precisão média para objetos de área média, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,514 para RetinaNet, 0,347 para Faster R-CNN e 0,534 para RTMDet.

AP @[ IoU=0.50:0.95 | area= large | maxDets=1000 ]: A precisão média para objetos de grande área, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,716 para RetinaNet, 0,508 para Faster R-CNN e 0,616 para RTMDet.

Recall médio (AR): Essa métrica mede a taxa média de recall do modelo em diferentes configurações de limiar de sobreposição (IoU) e áreas do objeto. Ela fornece uma média ponderada das taxas de recall em diferentes configurações.

AR @[ IoU=0.50:0.95 | area= all | maxDets=100 ]: A taxa média de recall para todas as áreas de objetos, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 100 detecções, é 0,562 para RetinaNet, 0,559 para Faster R-CNN e 0,510 para RTMDet.

AR @[ IoU=0.50:0.95 | area= all | maxDets=300 ]: A taxa média de recall para todas as áreas de objetos, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 300 detecções, é 0,562 para RetinaNet, 0,559 para Faster R-CNN e 0,625 para RTMDet.

AR @[ IoU=0.50:0.95 | area= all | maxDets=1000 ]: A taxa média de recall para todas as áreas de objetos, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,562 para RetinaNet, 0,559 para Faster R-CNN e 0,638 para RTMDet.

AR @[ IoU=0.50:0.95 | area= small | maxDets=1000 ]: A taxa média de recall para objetos de pequena área, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,543 para RetinaNet, -1,000 para Faster R-CNN e 0,591 para RTMDet.

AR @[ IoU=0.50:0.95 | area= medium | maxDets=1000 ]: A taxa média de recall para objetos de área média, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,565 para RetinaNet, 0,484 para Faster R-CNN e 0,655 para RTMDet.

AR @[ IoU=0.50:0.95 | area= large | maxDets=1000 ]: A taxa média de recall para objetos de grande área, considerando uma sobreposição IoU de 0,5 a 0,95 e no máximo 1000 detecções, é 0,734 para RetinaNet, 0,588 para Faster R-CNN e 0,633 para RTMDet.

Essas métricas ajudam a avaliar o desempenho do seu modelo de detecção de objetos em termos de precisão e capacidade de recuperar objetos sob diversas condições. Tais métricas apresentadas são importantes para avaliar o desempenho dos modelos de detecção de objetos em diferentes cenários em termos de precisão e recuperação em diferentes ambientes, e aqueles modelos com valores mais altos de AP e AR são geralmente considerados melhores para tarefas de detecção de objetos.

Depois de realizar um treinamento consistente por 25 épocas no mesmo conjunto de dados, descobriu-se que os modelos RetinaNet, Faster R-CNN e RTMDet mostraram resultados de precisão muito semelhantes, confirmando que todos esses modelos são igualmente eficazes na tarefa de detecção de objetos, no entanto, a superioridade do Faster R-CNN na detecção de pequenos objetos é notável na análise comparativa entre os demais modelos testados, sendo este capaz de rastrear pequenos objetos com precisão como uma vantagem importante,

especialmente quando pequenos detalhes são importantes para a análise em questão. Além disso, o modelo Faster R-CNN mostrou ser menos propenso a confundir objetos com o plano ambiente, o que revela a sua capacidade diferenciada em distinguir objetos reais do ruído de fundo, e tal característica é especialmente importante em aplicações de detecção de objetos do mundo real, onde minimizar falsos positivos é essencial para um desempenho confiável e preciso. Portanto, com base nestes resultados, o modelo Faster R-CNN destaca-se como a opção preferida em cenários onde a detecção precisa de pequenos objetos e a redução da desordem do fundo são prioridades importantes.

*Figura 15 - Modelo RTMDET captando as classes e o seu percentual de certeza*



Fonte: do autor (2023).

*Figura 16 - Modelo Faster R-CNN captando as classes e o seu percentual de certeza*



Fonte: do autor (2023).

Essas métricas ajudam a avaliar o desempenho do seu modelo de detecção de objetos em termos de precisão e capacidade de recuperar objetos sob diversas condições. Depois de realizar um treinamento consistente por 25 épocas no mesmo conjunto de dados, descobrimos que os modelos RetinaNet, Faster R-CNN e RTMDet mostraram resultados de precisão muito semelhantes. Isso mostra que todos esses modelos são igualmente eficazes na tarefa de detecção de objetos. No entanto, a superioridade do Faster R-CNN na detecção de pequenos objetos é notável. Isso significa  $AP@[IoU=0,50:0,95 | \text{área} = \text{pequena} | \text{maxDets}=1000]$ . Ser capaz de rastrear pequenos objetos com precisão é uma vantagem importante, especialmente quando pequenos detalhes são importantes. Além disso, Faster R-CNN é menos propenso a confundir objetos com o plano de fundo (ou ambiente). Isso significa que ele tem uma capacidade melhor de distinguir objetos reais do ruído de fundo. Isto é especialmente importante em aplicações de detecção de objetos do mundo real, onde minimizar falsos positivos é essencial para um desempenho confiável e preciso. Portanto, com base nestes resultados, o modelo Faster R-CNN destaca-se como a opção preferida em cenários onde a detecção precisa de pequenos objetos e a redução da desordem do fundo são prioridades importantes.

# CAPÍTULO 5

## Considerações finais

---

---

### 5.1 Conclusão

As técnicas de Visão computacional são indispensáveis para avançarmos no campo da detecção de objetos, proporcionando soluções eficientes e confiáveis para identificação de elementos estruturais em imagens. Nesta dissertação de mestrado, foram investigados e avaliados o desempenho de três modelos de detecção de objetos: Retinanet, Faster R-CNN e RTMDet, tendo como objetivo principal a análise da eficácia desses modelos na detecção de elementos estruturais em imagens, como postes, formatos em cruzeta, formatos sem cruzeta "liso" e transformadores de poste e os resultados experimentais revelaram um desempenho excelente para todos os modelos testados, confirmando a sua robustez e confiabilidade.

A partir dos resultados foi possível observar que ao analisar as métricas de precisão, número de épocas e resultados para cada classe, que todos os modelos apresentaram resultados impressionantes, em que a precisão média variou entre 89% e 92%, indicando uma capacidade consistente de identificar corretamente os elementos estruturais nas imagens. Além disso, a quantidade de épocas necessárias para treinar os modelos foi relativamente baixa, variando de 25 a 40 épocas, o que demonstra a eficiência dessas abordagens.

No que tange ao aspecto da comparação dos resultados para cada classe, notamos que o RTMDet se destacou, apresentando as maiores taxas de acerto para todos os elementos estruturais. Esses resultados reforçam a superioridade desse modelo em termos de precisão. No entanto, é importante ressaltar que os outros dois modelos também obtiveram desempenhos notáveis, com taxas de acerto significativas.

Conclui-se, portanto, que a escolha do modelo de detecção de objetos depende das necessidades específicas de cada aplicação, considerando o equilíbrio entre precisão e velocidade. Se a precisão for a prioridade principal, o RTMDet se mostra como a opção mais

adequada, enquanto o Retinanet e o Faster R-CNN também são alternativas sólidas, logo, tal flexibilidade permite que os usuários escolham o modelo mais adequado de acordo com os requisitos do projeto.

Os resultados obtidos nesta pesquisa contribuem para o avanço do campo de detecção de objetos, fornecendo *insights* valiosos sobre o desempenho desses modelos específicos. Além disso, abre caminho para futuras pesquisas e melhorias nos algoritmos de detecção de objetos, aprimorando ainda mais a precisão e a eficiência dessas técnicas.

Em suma, esta dissertação visa destacar a importância dos modelos de detecção de objetos na identificação de elementos estruturais em imagens, considerando que os resultados experimentais evidenciaram a eficácia dos modelos testados, fornecendo uma base sólida para a aplicação prática dessas abordagens. Com o avanço contínuo da tecnologia e o aprimoramento dos algoritmos, podemos esperar um futuro promissor para a detecção de objetos e suas aplicações em diversos domínios, desde a segurança pública até a automação industrial.

## 5.2 Trabalhos futuros

Com base no estudo apresentado, há várias possibilidades para trabalhos científicos futuros que podem contribuir para o avanço do campo de detecção de objetos. A seguir, estão algumas sugestões:

- A. Exploração de arquiteturas avançadas: Investigar arquiteturas de detecção de objetos mais recentes, como EfficientDet, YOLOv4, ou DETR, comparando seu desempenho com os modelos avaliados nesta dissertação, analisando as vantagens e desafios de implementação dessas arquiteturas em diferentes cenários.
- B. Detecção de objetos em condições adversas: Realizar estudos que explorem a detecção de elementos estruturais em imagens capturadas em condições adversas, como baixa iluminação, imagens embaçadas ou com ruído. Avaliar o desempenho dos modelos propostos nesta dissertação nessas situações e identificar possíveis melhorias para aumentar a robustez nessas condições.
- C. Análise de objetos em tempo real: Investigar a viabilidade e o desempenho dos modelos de detecção de objetos em tempo real, considerando restrições de tempo de processamento e recursos computacionais. Avaliar a aplicabilidade

dessas abordagens em sistemas de vigilância, veículos autônomos ou outras aplicações que requerem detecção em tempo real.

- D. Transferência de aprendizado: Explorar técnicas de transferência de aprendizado para melhorar a capacidade dos modelos de detecção de objetos em diferentes domínios. Investigar a transferência de conhecimento entre diferentes tipos de elementos estruturais e avaliar a eficácia dessas abordagens em cenários com conjuntos de dados limitados.
- E. Integração de outras modalidades de entrada: Investigar a combinação de técnicas de visão computacional com outras modalidades de entrada, como dados de radar, lidar ou áudio, para melhorar a precisão e robustez na detecção de objetos. Explorar abordagens multimodais que combinam informações de diferentes sensores para uma detecção mais precisa e confiável.
- F. Avaliação de desempenho em escalas maiores: Realizar estudos que avaliem o desempenho dos modelos propostos nesta dissertação em conjuntos de dados de escala maior, envolvendo um maior número de elementos estruturais. Investigar a capacidade dos modelos em lidar com a detecção em larga escala e identificar possíveis desafios e limitações.
- G. Análise de detecção em vídeo: Investigar a aplicação dos modelos de detecção de objetos em sequências de vídeo, analisando a capacidade de rastreamento e detecção de objetos em movimento. Avaliar a precisão e a eficiência desses modelos em cenários de monitoramento por vídeo, como segurança pública ou vigilância de tráfego.

Essas são apenas algumas sugestões de trabalhos científicos futuros que podem ser desenvolvidos com base nos resultados e nas conclusões apresentadas nesta dissertação, em que cada uma dessas sugestões visa representar uma área de pesquisa promissora que contribuirá para aprimorar a detecção de objetos e suas aplicações em diferentes campos.

# Referências bibliográficas

---

Abdelfattah, R.; Wang, X.; & Wang, S. (2020). TTPLA: An Aerial-Image Dataset for Detection and Segmentation of Transmission Towers and Power Lines. In: ISHIKAWA, H. et al. (Eds). Asian Conference on Computer Vision 2020. *Lecture Notes in Computer Science*, 12627: 601-618, Springer, Cham.

Alam, M. et al. (2020). Automatic Assessment and Prediction of the Resilience of Utility Poles Using Unmanned Aerial Vehicles and Computer Vision Techniques. *International Journal of Disaster Risk Science*, 11: 119–132.

Amaral, F. (2016). *Big Data: Uma Visão Gerencial*. São Paulo: Polobooks.

Backes, A. R.; & SÁ JÚNIOR, J. J. M. (2016). *Introdução à Visão Computacional Usando MATLAB*. 1. ed. Rio de Janeiro: Alta Books,.

Ballard, D. H.; & Brown, C. M. (1982). *Computer Vision*. Englewood Cliffs: Prentice Hall.

Bardin, L. (2016). *Análise de Conteúdo*. São Paulo: Edições 70.

Brasil (1996). Lei nº 9.427, de 26 de dezembro de 1996. *Institui a Agência Nacional de Energia Elétrica - ANEEL, disciplina o regime das concessões de serviços públicos de energia elétrica e dá outras providências*. Brasília: Diário Oficial da União, 27 de dezembro de 1996. [http://www.planalto.gov.br/ccivil\\_03/leis/19427cons.htm](http://www.planalto.gov.br/ccivil_03/leis/19427cons.htm). Acesso em: 18 mar. 2022.

Brownlee, J. (2019). *Deep Learning for Computer Vision: Image Classification, Object Detection and Face Recognition in Python*. San Juan, Puerto Rico: Machine Learning Mastery.

Cielen, D.; Meysman, A.; & Ali, M. (2016). *Introducing Data Science: Big Data, Machine Learning, and more, using Python tools*. Shelter Island: Manning Publications Co.

Chang, L.; Peng-Sem, S.; & Shi-Rong, L. (2016). A review of plant spectral reflectance response to water physiological changes. *Chinese Journal of Plant Ecology*, 49 (1): 80-91.

Coulouris, G.; Dollimore, J.; & Kindberg, T. (2007). *Sistemas Distribuídos - Conceitos e Projeto*. Tradução: Bookman Companhia Editora Ltda. Porto Alegre: Artmed Editora S.A.

Fambrini, F. et al. (2020). Deep learning based system to electric distribution network inspection. In: International Conference on Digital Image Processing (ICDIP), 12, 2020. *Proceedings* [...]. Osaka, Japão.

Fang, S. et al. (2020). A Framework of Power Pylon Detection for UAV-based Power Line Inspection. In: *IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 5: 350-357, Chongqing, China. *Proceedings* [...]. Chongqing, China: IEEE.

Figueiredo, E. B.; & Silva, E. P. (2021). Combate ao Covid19: Detecção em tempo real de indivíduos sem máscara em ambiente escolar por meio de Deep Learning. In: XV Brazilian E-Science Workshop (BRESCI), 15: 113-120, 2021, online. *Anais* [...]. Porto Alegre: SBC Open Lib. DOI 10.5753/bresci.2021.15796.  
<https://sol.sbc.org.br/index.php/bresci/article/view/15796>.

Forsyth, D. A.; & Ponce, J. (2012). *Computer Vision: A Modern Approach*. 2. ed. Upper Saddle River: Pearson.

Galantucci, R.; & Fatiguso, F. (2019). Advanced damage detection techniques in historical buildings using digital photogrammetry and 3D surface analysis. *Journal of Cultural Heritage*, 36: 51-62, mar./abr.

Girshick, R. et al. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, 2014, Columbus, United States of America. *Proceedings* [...]. Columbus: IEEE: 580-587.

Goodfellow, I.; Bengio, Y.; & Courville, (2016). A. *Deep Learning (Adaptive Computation and Machine Learning series)*. Massachusetts: The MIT Press.  
<https://www.deeplearningbook.org/>.

Gomes, M. et al. (2020). Mapping Utility Poles in Aerial Orthoimages Using ATSS Deep Learning Method. *Sensors*, 20 (21). <https://www.mdpi.com/1424-8220/20/21/6070>.

Gordon, S. J.; & Lichti, D. D. 2007. Modeling Terrestrial Laser Scanner Data for Precise Structural Deformation Measurement. *Journal of Surveying Engineering*, 133, (2): 72-80, mai.

Ketkar, N. (2017). *Deep Learning with Python: A Hands-on Introduction*. Nova Iorque: Apress.

Kim, J. et al. (2021). Large-Scale Visual Data-Driven Probabilistic Risk Assessment of Utility Poles Regarding the Vulnerability of Power Distribution Infrastructure Systems. *Journal of Construction Engineering and Management*, 147 (10), ago.

Laumer, D. et al. (2020). Geocoding of trees from street addresses and street-level images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162: 125-136, abr. [https://www.fs.fed.us/psw/publications/vandoorn/psw\\_2020\\_vandoorn001\\_laumer.p df](https://www.fs.fed.us/psw/publications/vandoorn/psw_2020_vandoorn001_laumer.p df).

Mathworks. *Introducing Deep Learning with MATLAB*. (2021). [https://www.mathworks.com/content/dam/mathworks/ebook/gated/80879v00\\_Deep\\_Learning\\_ebook.pdf?s\\_tid=wtest\\_ctent\\_HTML](https://www.mathworks.com/content/dam/mathworks/ebook/gated/80879v00_Deep_Learning_ebook.pdf?s_tid=wtest_ctent_HTML).

Monard, M. C.; Baranauskas, J. A. (2003). Conceitos sobre Aprendizagem de Máquina. In: Rezende, S. O. (Org.). *Sistemas Inteligentes: fundamentos e aplicações*. Barueri, SP: Manole: 89-116.

Montavon, G.; Samek, W.; Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73: 1-15. <https://www.sciencedirect.com/science/article/pii/S1051200417302385?via%3Dihub>.

Nielsen, M. 2015. *Neural Networks and Deep Learning*. Determination Press.

Padilla, R.; Netto, S. Lima; & Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. In: International Conference on Systems, Signals and Image Processing (IWSSIP), 2020, Niterói. *Proceedings [...]*. Niterói, Brasil: IWSSIP.

Rajpurohit, A. (2013). Big Data for Business Managers - Bridging the gap between Potential and Value. In: IEEE International Conference on Big Data, 2013, Valley. *Proceedings [...]*. Valley, United States of America: IEEE: 29-31.

Ramlal, C. J. et al. (2020). Toward Automated Utility Pole Condition Monitoring: A Deep Learning Approach. In: IEEE PES Innovative Smart Grid Technologies Europe, 2020, Hague. *Paper [...]*. Hague, Netherlands: IEEE.

Rodriguez, A. (2021). *Deep Learning Systems: Algorithms, Compilers, and Processors for Large-Scale Production*. Londres: Morgan & Claypool.

Sores, I. M. (2019). Uma abordagem bottom-up completa para reconhecimento de atividades humanas em imagens através da pose estimada com redes convolucionais. 126 p. Dissertação ( Mestrado em Engenharia Elétrica e de Computação). Universidade Federal de Goiás, Goiânia.

Superintendência de Regulação dos Serviços de Distribuição. SRD. *Avaliação da Regulação Relativa às Interrupções em Situação de Emergência (ISE) - Relatório de Avaliação de Resultado Regulatório nº 0001/2019-SRD/ANEEL*. Brasília: ANEEL, 2019.

[https://extranet.economia.gov.br/wp-content/uploads/2020/10/ANEEL\\_ARR-sobre-Regulacao-Relativa-as-Interrupcoes-em-Situacao-de-Emergencia-ISE.pdf](https://extranet.economia.gov.br/wp-content/uploads/2020/10/ANEEL_ARR-sobre-Regulacao-Relativa-as-Interrupcoes-em-Situacao-de-Emergencia-ISE.pdf).

Szegedy, C.; Toshev, A.; & Erhan, D. (2013). Deep Neural Networks for Object Detection. In: *Advances in Neural Information Processing Systems*, 26, 2013. *Proceedings* [...], Lake Tahoe, United States of America: NIPS.

Szeliski, R (2010). *Computer Vision: Algorithms and Applications*. Londres: Springer.

Tan, M. et al. (2020). EfficientDet: Scalable and Efficient Object Detection. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 29, 2020. *Proceedings* [...]. Seattle, United States of America: IEEE: 10781-10790.

[https://openaccess.thecvf.com/content\\_CVPR\\_2020/papers/Tan\\_EfficientDet\\_Scalable\\_and\\_Efficient\\_Object\\_Detection\\_CVPR\\_2020\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2020/papers/Tan_EfficientDet_Scalable_and_Efficient_Object_Detection_CVPR_2020_paper.pdf).

Tekiner, F; & Keane, J. (2014). A. Big Data Framework. In: *IEEE International Conference on Systems, Man, and Cybernetics*, 2013, Manchester, United Kingdom. *Proceedings* [...]. Manchester: IEEE: 1494-1499.

Vemula, S.; & Frye, M. (2020). Mask R-CNN Powerline Detector: A Deep Learning approach with applications to a UAV. In: *AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 39,. San Antonio, United States of America.

(2020). Real-Time Powerline Detection System for an Unmanned Aircraft System. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020. Online.

Xue, J; & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, 2017.

Yin, R. K. (2003). *Case study research: design and methods*. 3. ed. California: Sage Publications.

Zhang, W. et al. (2018). Using Deep Learning to Identify Utility Poles with Crossarms and Estimate Their Locations from Google Street View Images. *Sensors*, 18 (8).