

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

FACULDADE DE ENGENHARIAS, ARQUITETURA E URBANISMO E GEOGRAFIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

IMPLEMENTAÇÃO E AVALIAÇÃO DE MÉTODOS DE APRENDIZAGEM PROFUNDA PARA MAPEAMENTO DE POSTES EM ORTOIMAGENS AÉREAS

Matheus Bueno Gomes

Campo Grande - MS, Brasil

Junho de 2022



IMPLEMENTAÇÃO E AVALIAÇÃO DE MÉTODOS DE APRENDIZAGEM PROFUNDA PARA DETECÇÃO DE POSTES EM ORTOIMAGENS AÉREAS

Matheus Bueno Gomes

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul na Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Prof.Dr. Wesley Nunes Gonçalves
Coorientador: Prof. Dr. Edson Batista Barbosa

Campo Grande - MS, Brasil

Junho de 2022

IMPLEMENTAÇÃO E AVALIAÇÃO DE MÉTODOS DE APRENDIZAGEM PROFUNDA PARA MAPEAMENTO DE POSTES EM ORTOIMAGENS AÉREAS

Matheus Bueno Gomes

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul na Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

Examinada por:

Prof. Dr. Wesley Nunes Gonçalves

Prof. Dr. Edson Batista Barbosa

Prof. Dr. Raymundo Cordero Garcia

Prof. Dr. José Marcato Junior

Campo Grande - MS, Brasil
Junho de 2022

Agradecimentos

Primeiramente agradeço a Deus por todas as bençãos concedidas durante essa caminhada, por todas portas abertas e por todas as portas fechadas. Se aqui estou hoje é graças a Ele. Agradeço a Jesus por todos os acertos, mas principalmente pelos erros e aprendizados.

Gostaria de agradecer a meus pais, Marcelo e Bernadeth, que sempre me incentivaram a buscar mais e mais para minha vida, tanto profissional, acadêmica ou pessoal. Vocês sempre foram meus maiores apoiadores. Somente nós sabemos os sacrifícios que vocês fizeram para que eu pudesse chegar até aqui.

Um agradecimento especial ao meu orientador, Profº Drº Wesley, que me aceitou como seu orientado com o mestrado já em andamento. Obrigado por toda a paciência, disponibilidade e profissionalismo demonstradas durante a execução deste projeto. Espero que nossa caminhada acadêmica possa continuar se estendendo para novos horizontes.

Agradeço ao Profº Drº Edson pelas oportunidades que me proporcionou no meio acadêmico. É por grande influência tua que me encontro nesta etapa acadêmica e sou extremamente grato por isso. Você é uma grande inspiração para mim e espero que possamos continuar trilhando juntos em novos projetos e desafios.

Agradeço a minha namorada, Juliana Sayuri, por todos os momentos de incentivo e apoio durante esta trajetória. Obrigado por todos os momentos em que precisavamos estudar e ligavamos a chamada de video para estudarmos juntos. Desde quando você entrou na minha trajetória, minha vida só tem caminhado pra cima.

Por fim, gostaria de agradecer meus grandes amigos, Alisson, Gabriel Silvestri, Christoffer, Hideki e Lucas Vegini. A Familia Rox Feeders e seus crimes nos mais diversos jogos online contribuem muito para minha trajetória. As risadas e os momentos juntos são alivios para os momentos estressantes e responsabilidades da vida.

"Mas o conhecimento envaidece, é o amor que constrói. Quando alguém julga ter alcançado o saber, é por que ainda não sabe onde está o verdadeiro conhecimento."

1 Corintíos 8:1-2

IMPLEMENTAÇÃO E AVALIAÇÃO DE MÉTODOS DE APRENDIZAGEM PROFUNDA PARA DETECÇÃO DE POSTES EM ORTOIMAGENS AÉREAS

Matheus Bueno Gomes

Atualmente a gestão de ativos é um processo de grande importância dentro de uma concessionária de energia, visto que é fundamental para as tomadas de decisões da corporação e está diretamente ligada ao equilíbrio econômico financeiro da empresa. Um grande aliado neste processo são coordenadas georreferenciadas de cada ativo, auxiliando a corporação no planejamento de novos investimentos além de fornecer um panorama completo da infraestrutura da companhia. Vislumbrando a possibilidade de um maior controle regulatório através de dados geoespaciais, a partir de 2009 a ANEEL (Agência Nacional De Energia Elétrica) instituiu o Banco de Dados Geográfica da Distribuidora (BDGD), que impôs as concessionárias de energia o fornecimento anual do cadastro georreferenciado de seus ativos. A ANEEL não determina uma metodologia de geocadastramento específica para as companhias, que ainda frequentemente dispensam o uso de recursos tecnológicos e optam por metodologias de trabalho custosas e arcaicas, mapeando os ativos por meio de profissionais que inspecionam a rede elétrica rua a rua. Esta dissertação visa analisar duas propostas de metodologias de aprendizado profundo que auxiliem na georreferenciação automática de ativos elétricos através de ortoimagens aéreas. A primeira proposta é a utilização da arquitetura *Adaptive Training Sample Selection* (ATSS) nesta tarefa, um método novo e pouco utilizado no ramo de sensoriamento remoto. O desempenho do método foi comparado com as arquiteturas Faster R-CNN e RetinaNet, já comumente utilizadas no ramo de sensoriamento remoto, na execução da mesma tarefa. A segunda proposta é a implementação e análise dos métodos ATSS, TOOD (*Task-aligned One-stage Object Detection*), Varifocal-Net e Deformable DETR (*Detection Transformer*) na tarefa de detecção e classificação de seis classes de postes presentes no dataset.

Palavras-chaves: gestão de ativos, aprendizado profundo, georreferenciamento.

IMPLEMENTATION AND EVALUATION OF DEEP LEARNING METHODS FOR POST DETECTION IN AERIAL ORTOIMAGES

Matheus Bueno Gomes

Currently, asset management is a process of great importance within an energy concessionaire, since it is fundamental to the decision-making of the corporation and is directly linked to the company's economic and financial balance. A great ally in this process are georeferenced coordinates for each asset, assisting the corporation in planning new investments and also providing a complete overview of the company's infrastructure. Glimpsing the possibility of greater regulatory control through geospatial data, as of 2009, ANEEL (National Electric Energy Agency) instituted the Distributor's Geographic Database (BDGD), which imposed energy concessionaires to annually provide the georeferenced register of your assets. ANEEL does not determine a specific geo-registration methodology for companies, which still often dispense the use of technological resources and opt for costly and archaic working methodologies, mapping assets through professionals who inspect the electricity grid street by street. This dissertation aims to analyze two proposals for deep learning methodologies that assist in the automatic georeferencing of electrical assets through aerial orthoimages. The first proposal is to use the Adaptive Training Sample Selection (ATSS) architecture in this task, a novel method and still not often used in the field of remote sensing. The performance of the method will be compared with the Faster R-CNN and RetinaNet architectures, already commonly used in the field of remote sensing, in the execution of the same task. The second proposal is the implementation and analysis of ATSS, TOOD (Task-aligned One-stage Object Detection), Varifocal-Net and Deformable DETR (Detection Transformer) methods in the detection and classification task of six classes of poles present in the dataset.

Keywords: asset management, deep learning, georeferencing.

Sumário

Lista de ilustrações I

Lista de tabelas III

1 Introdução 1
- 1.1 Motivação 1
- 1.2 Objetivos 4
- 1.3 Contribuições 4
  - 1.3.1 Banco de Imagens 4
  - 1.3.2 Ferramenta para mapeamento de postes elétricos. 5
  - 1.3.3 Experimentos com métodos do estado da arte 5
- 1.4 Organização 5

2 Trabalhos Correlatos 7
- 2.1 Lasers e LiDAR 7
- 2.2 Imagens Terrestres 8
- 2.3 Imagens Aéreas 11

3 Mapeamento Automático de Postes 13
- 3.1 Introdução 14
- 3.2 Materiais e Métodos 18
  - 3.2.1 Área de Estudo 18
  - 3.2.2 Métodos Utilizados 18
  - 3.2.3 Configuração do Experimento 22
  - 3.2.4 Avaliação dos Métodos 23
- 3.3 Resultados 24
  - 3.3.1 Análise Quantitativa 24
  - 3.3.2 Análise Qualitativa 25
- 3.4 Custos Computacionais 27
- 3.5 Discussão 30

3.6 Considerações Finais 32

4 Classificação Automática de Postes 34

4.1 Materiais e Métodos 34

4.1.1 Área de Estudo 34

4.1.2 Postes Classificados 34

4.1.2.1 Poste com Iluminação Dupla 35

4.1.2.2 Poste de Iluminação Circular 35

4.1.2.3 Poste de Iluminação de 3, 4 e 5 pontas. 36

4.1.2.4 Poste de Transmissão tipo T 37

4.1.3 Métodos Experimentados 38

4.1.4 Configuração do Experimento 41

4.1.5 Avaliação dos Métodos 43

4.2 Resultados 43

4.2.1 Análise Quantitativa 44

4.2.2 Custos Computacionais 44

4.3 Discussão 45

4.4 Considerações Finais 55

5 Conclusão 56

Referências 59

Lista de ilustrações

Figura 1 – Exemplos de postes presentes em nosso conjunto de dados. . . . 17
Figura 2 – Localização da área de estudo: (**a**) América do Sul e Brasil; (**b**) Mato Grosso do Sul; (**c**) Campo Grande; e (**d**) Área de estudo. 18
Figura 3 – Passo a passo para criação dos rótulos com caixas delimitadoras. 19
Figura 4 – Curva de erro durante o treinamento dos três métodos de detecção de objetos. 23
Figura 5 – Resultados obtidos pelos métodos quando localizando postes oclusos por topos de árvores: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN. 26
Figura 6 – Exemplos de imagens com postes sombreados: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN. 27
Figura 7 – Exemplos de imagens com diferentes modelos e formatos de postes: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN. 29
Figura 8 – Exemplos de imagens com postes instalados em diferentes tipos de terreno: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN. 30
Figura 9 – Exemplos de postes que não são totalmente cobertos por uma caixa delimitadora com 40×40 de tamanho: (**a**) Exemplo 1; (**b**) Exemplo 2; e (**c**) Exemplo 3. 31

Figura 10 – Exemplos da classe Iluminação Dupla. A vista aérea é demonstrada em (a) e a vista lateral em (b). 35
Figura 11 – Exemplos da classe Poste Circular. A vista aérea é demonstrada em (a) e a vista lateral em (b). 36
Figura 12 – Exemplos de poste das classes Postes 3, 4 e 5 Pontas. 37
Figura 13 – Exemplos da classe Poste T. A vista aérea é demonstrada em (a) e a vista lateral em (b). 38
Figura 14 – Erro durante o treinamento dos métodos. 42

Figura 15 – Exemplos de nosso dataset demonstrando os desafios da classificação de diferentes tipos de postes. Nota-se nos exemplos que as principais características estão parcialmente cobertas por árvores. 46
Figura 16 – Postes Circulares detectados pelo ATSS. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 47
Figura 17 – Detecções errôneas da classe Iluminação Dupla pelo método ATSS. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. . 48
Figura 18 – Postes Iluminação Dupla detectados pelo Varifocal-Net. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 49
Figura 19 – Detecções errôneas da classe tipo T pelo método VRF-Net. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 49
Figura 20 – Detecções errôneas de Postes 3 Pontas pelo método Deformable DETR. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 51
Figura 21 – Postes 4 e 5 Pontas detectados pelo método Deformable-DETR. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 51
Figura 22 – Postes 5 Pontas e Circular detectados pelo método TOOD. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. 53
Figura 23 – Detecções errôneas da classe Iluminação Dupla pelo método TOOD. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método. . 54

Lista de tabelas

Tabela 1 – Descrição dos conjuntos utilizados para treinamento, validação e testes. 22
Tabela 2 – Precisão média obtida para cada método experiementado. Para a criação dos *ground-truths*, o tamanho da caixa delimitadora foi fixado em 40×40. 24
Tabela 3 – Precisão média para diferentes formatos de caixa delimitadora para o método ATSS. 25
Tabela 4 – Tempo de processamento dos métodos. 28
Tabela 5 – Descrição dos conjuntos utilizados para treinamento, validação e testes. 41
Tabela 6 – Descrição das configurações dos métodos. 42
Tabela 7 – Precisão média por classe para os métodos. 44
Tabela 8 – Custo Computacional dos Métodos. 45

1 Introdução

1.1 Motivação

O processo de gestão de ativos pelas concessionárias de energia vem sofrendo grandes evoluções nos tempos recentes. Este movimento é impulsionado graças ao crescente aumento e aprimoramento dos controles regulatórios por parte da Agência Nacional de Energia Elétrica – ANEEL [1] e também pelo contínuo crescimento de infraestrutura elétrica no Brasil, aumentando a exigência de informações geográficas para suporte ao processo de tomada de decisão [2]. A Resolução Normativa ANEEL n. 480, de 3 de abril de 2012 (Diário Oficial, de 13 abr. 2012, seção 1, p. 48) define um Ativo Imobilizado em Serviço (AIS) como:

> "O conjunto de todos os bens, instalações e direitos que, direta ou indiretamente, concorram, exclusiva e permanentemente, para manutenção das atividades da concessionária de serviço público de energia elétrica."

A gestão de ativos é um processo fundamental em empresas de energia, pois está diretamente ligada a tomada de decisões da corporação e impactam diretamente na valorização da empresa. A gestão de ativos visa à administração dos riscos com maior eficiência, buscando a redução de perdas e gastos, a maximização dos lucros da empresa em longo prazo [3] e a garantia da qualidade na prestação do serviço dentro de um mercado extremamente exigente, como é o setor de distribuição de energia elétrica [4].

Dados geo-espaciais são de grande utilidade para o processo de gestão de ativos. Este tipo de informação é vital para o crescimento sustentável da companhia, municiando-a de informações para o planejamento de novos investimentos, viabilizando um crescimento da infraestrutura elétrica de forma ordenada e programada. As coordenadas geográficas dos ativos também permitem o estudo a fundo das linhas de transmissão, possibilitando a simulação de fenômenos na rede e seus impactos na qualidade do serviço [5].

Apesar do aumento da demanda de informações geográficas no processo

de gestão de ativos [6], até o ano de 2009 não havia uma legislação regulatória que padronizasse a produção de informações geoespaciais do setor elétrico [7]. Visando a instituição de um controle regulatório que padronizasse e unificasse as informações dos ativos de todas as distribuidoras do Brasil em um sistema único [2], a ANEEL instituiu em Junho de 2009 através da Resolução Normativa 367/09 [8] o Manual de Controle Patrimonial do Setor Elétrico (MCPSE) [9], que impôs uma nova sistemática de controle dos ativos de rede imobilizados em serviço. Pela Resolução Normativa 395/09, foi criado o Banco de Dados Geográfico da Distribuidora (BDGD), instituído na Revisão 1 dos Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (Prodist), que determina a obrigatoriedade das concessionárias de distribuição a fornecer anualmente o cadastro georreferenciado completo das linhas e redes de distribuição [7].

O BDGD da distribuidora compõe o Sistema de Informações Geográficas Regulatórias (SIGR), uma base de dados cartográfica digital de sistemas auxiliares que permite a análise e extração de informações dos principais ativos informados pela distribuidora [2]. O SIGR é hoje uma ferramenta fundamental, pois aumenta a confiabilidade das informações fornecidas pelas distribuidoras e também promove uma maior rastreabilidade das informações dos ativos. É de grande importância para a revisão tarifária periódica, um dos mecanismos que definem o valor de energia paga pelo consumidor, onde os custos operacionais e a remuneração dos investimentos por parte das distribuidoras são reavaliados. Além disso, é essencial também para o reajuste tarifário anual com a realização da correção monetária e compartilhamento de ganhos de produtividade [2].

O Prodist atualmente não especifica uma metodologia obrigatória que as distribuidoras de energia devam seguir no levantamento dos dados necessários para a composição do BDGD, padronizando apenas os documentos referentes à estruturação, formato e forma de envio da base à ANEEL. Desta forma, as distribuidoras são livres para buscar quaisquer tipos de recursos tecnológicos que os auxiliem no cadastramento dos ativos elétricos e seus atributos, desde que atendam às exigências regulatórias em relação à qualidade de dados e posicionamento [10]. Apesar da possibilidade de utilização de recursos tecnológicos para o processo de cadastramento, o mapeamento dos ativos da rede elétrica ainda é frequentemente realizado

através de processos trabalhosos, com grande parte das informações sendo levantadas de forma manual através de pessoas andando pelas ruas e mapeando as redes elétricas. Tal metodologia gera uma baixa confiabilidade em relação à qualidade das informações cadastradas e exigidas pelas normas da ANEEL [2]. Além disso, o processo manual é demorado e custoso.

Caso a distribuidora não alcance a meta de assertividade exigida pela ANEEL, esta sofre repercussões na Base de Remuneração Regulatória (BRR)[11], que é o montante de investimentos realizados pelas distribuidoras na prestação dos serviços que será coberto pelas tarifas cobradas aos consumidores[12]. Tal sanção impacta diretamente no equilíbrio econômico-financeiro da empresa, obtido através do reposicionamento periódico da tarifa de energia a ser paga pelos consumidores, que constitui como parte do retorno dos investimentos realizados pela concessionária [13].

Apesar do recente incentivo por ferramentas tecnológicas que auxiliem a gestão e mapeamento de ativos [10], a área é ainda pouco explorada por metodologias de visão computacional e inteligência artificial. Quando explorada, poucas metodologias se utilizam de imagens aéreas, que facilitam a detecção de ativos da rede em comparação com métodos que se utilizam de imagens terrestres. Além disso, os trabalhos encontrados na literatura utilizam-se de métodos rasos que atualmente são superados em eficiência (e.g[14, 15, 16, 17]) e acurácia por métodos de aprenziagem profunda do estado da arte e que ainda não foram explorados para a tarefa proposta por este trabalho.

Visando o cenário descrito e as lacunas encontradas atualmente para obtenção de informações geográficas de ativos elétricos, esta dissertação propõe o estudo e implementação de métodos de aprendizagem profunda do estado da arte para mapeamento automático de postes elétricos através de ortoimagens aéreas. Esta proposta pode ser considerada como o ponto de início para a elaboração futura de uma ferramenta que mapeie também outros tipos de ativos imobilizados em serviço, provendo às concessionárias de energia informações geoespaciais confiáveis que amparem no processo de gestão de seus ativos.

1.2 Objetivos

O objetivo deste trabalho é a implementação de métodos que realizam o mapeamento de postes (elétricos ou de iluminação) em ortoimagens aéreas usando aprendizagem profunda e visão computacional.

A primeira etapa deste trabalho foi a implementação e análise do desempenho do método ATSS (*Adaptive Training Sample Selection*) para detecção de postes elétricos em imagens aéreas. Para melhor comparação, foram também implementados para a realização desta tarefa os métodos RetinaNet e Faster R-CNN, de forma que a performance destes métodos sirvam de fator comparativo para o desempenho do método ATSS. As duas abordagens foram analisadas e comparadas entre si do ponto de vista qualitativo e quantitativo. Os custos computacionais das abordagens estudadas também foram investigados.

O segundo método proposto foi uma abordagem para detecção e classificação de seis classes de postes. Para isto foram implementados os métodos ATSS, TOOD, Varifocal-Net e Deformable DETR, métodos do estado da arte que foram ainda pouco experimentados em projetos de sensoriamento remoto. Os resultados dos métodos foram analisados do ponto de vista qualitativo, quantitativo e de custos computacionais. Por fim, foram realizadas comparações entre os resultados obtidos em costraste com os recursos propostos por cada um dos métodos, afim de discutir quais das inovações do estado da arte performam melhor no contexto da nossa aplicação.

1.3 Contribuições

1.3.1 Banco de Imagens

Grande parte dos métodos do estado da arte necessita de um banco de dados amplo em troca de uma boa precisão na localização do objeto de interesse. A elaboração de um banco de dados de 1057 ortoimagens aéreas com os postes elétricos devidamente anotados abre caminho para o desenvolvimento e avaliação de novos métodos que executam tarefas semelhantes à proposta desta dissertação.

1.3.2 Ferramenta para mapeamento de postes elétricos.

O desenvolvimento de uma ferramenta que realize o mapeamento de postes elétricos com eficiência e precisão pode ser considerado como o ponto de início para o mapeamento de outros componentes da rede elétrica, como transformadores, chaves, linhas de transmissão, entre outros. A elaboração de uma ferramenta que alcance este propósito pode simplificar e baratear o processo anual de cadastro georeferenciado dos ativos das distribuidoras de energia elétrica.

1.3.3 Experimentos com métodos do estado da arte

Os métodos usando aprendizagem profunda (e.g., [18][19][20][21][22][23]) são métodos recentes no estado da arte e ainda são poucos os projetos que utilizam estas metodologias. Os resultados e informações contidas nesta dissertação contribuem para a utilização destes métodos em tarefas e contextos ainda inexplorados.

1.4 Organização

Visando atingir os objetivos, este trabalho está dividido em cinco capítulos.

O presente capítulo expõe o contexto regulatório da ANEEL no processo de gestão de ativos bem como as dificuldades encontradas para a captação de informações geoespaciais dos ativos imobilizados em serviços. Este capítulo também apresenta os objetivos e contribuições desta dissertação.

No Capítulo 2 são apresentados os trabalhos correlatos que auxiliaram na formulação e desenvolvimento desta proposta. Estes estudos apresentam ferramentas de teor tecnológico similares ao desta dissertação, com métodos de aprendizagem profunda visando à detecção em imagens aéreas ou terrestres de objetos com formato semelhante a postes elétricos.

O Capítulo 3 descreve os experimentos realizados com os métodos ATSS, Faster R-CNN e RetinaNet. São descritos detalhes sobre a arquitetura de cada método, a implementação, área de estudo, configurações de treinamento, as métricas de avaliação utilizadas e os resultados obtidos. Também é retratada a análise de

custos computacionais entre os métodos. Além disso o Capítulo 3 também relata a análise realizada da influência das dimensões das caixas delimitadoras na acurácia do método ATSS.

O Capítulo 4 demonstra os experimentos realizados para detecção e classificação de seis classes de postes, realizados com os métodos ATSS, Deformable DETR, TOOD e Varifocal-Net. São demonstrados neste capítulo os recursos e inovações propostos por cada método, área de estudo, configuração dos métodos, métricas de avaliação e os resultados obtidos. Os resultados dos métodos são comparados entre si, afim de averiguar quais das inovações propostas pelos métodos experimentados performa melhor no contexto deste experimento.

Por fim, o Capítulo 5 descreve as conclusões desta dissertação e recomendações para futuros trabalhos.

2 Trabalhos Correlatos

Atualmente existem alguns trabalhos de teor tecnológico semelhante ao objetivado por esta dissertação e que auxiliaram no desenvolvimento e formulação desta proposta. Os relatos encontrados na literatura e descritos neste tópico foram organizados pelo tipo de tecnologia empregado e ponto de vista (aérea ou terrestre) utilizado para detecção do objeto.

2.1 Lasers e LiDAR

No contexto de localização de postes elétricos, foram encontrados vários trabalhos que se utilizam de tecnologias como lasers e LiDAR (*Light Detection And Ranging*) para esta tarefa. O foco das pesquisas encontradas em sua maioria objetiva o emprego destas tecnologias em carros autônomos. Apesar de efetiva nesta tarefa, estas ferramentas não são indicadas no mapeamento e inventário de objetos em larga escala. Como o escaneamento é feito do nível da rua, o mapeamento de grandes áreas se torna um processo lento e pouco efetivo.

Em [24] é relatada uma proposta de utilização de dados obtidos através de mapeamento a laser para detecção de objetos com formato similares a postes (cilíndricos), sejam estes postes de energia, prédios, seres humanos ou quaisquer outros objetos neste formato. O trabalho propõe uma adaptação do método padrão de detecção de círculo utilizando *Hough Tables* para atribuir probabilidade de um objeto ser ou não no formato do objeto de interesse.

O trabalho de Lehtomaki et al.[25] propõe um método automático baseado em *clusters* para localização de objetos análogos a postes. A detecção é realizada a partir de nuvem de pontos oriundos de um laser scanner acoplado a um veículo. Um *cluster* é definido como um poste quando vários critérios relacionados ao formato do objeto são atendidos. O método proposto em [25] obteve 77.7% de assertividade nos testes executados pelo autor.

Também se utilizando de nuvem de pontos, em [26] é demonstrado um

método de classificação de pontos para detecção e classificação de objetos em formato de poste. O método proposto por Huang et al. [26] tem como entrada uma nuvem de pontos em alta escala de uma área urbana e produz como saída os candidatos a objetos análogos a postes. O método então classifica estes candidatos em três categorias: postes de energia, postes de iluminação, e placas. A precisão alcançada em uma área de teste com 45 quadras de dimensões 100 x 100 m foi de 52%.

A proposta de Javanmardi et al. [27] relata a utilização da tecnologia LiDAR na detecção de postes de luz e placas de trânsito. O método se utiliza do algoritmo de reconstrução de superfície para localização dos vetores normais como um conjunto de características. Neste conjunto é aplicado k-médias para a segmentação dos vetores em pontos que pertençam à rua ou não. Em seguida, aplica-se o algoritmo de cuboides deslizantes para detecção de objetos muito elevados ou próximos às bordas e topo da estrada. Por fim, uma remoção de outliers é realizada para eliminação de pontos ruidosos. O autor descreve a detecção de 137 placas de trânsito (94,48% de assertividade) e 33 postes de luz (89,19% de assertividade) nos testes realizados.

2.2 Imagens Terrestres

Parte das publicações encontradas pauta-se na utilização de imagens terrestres para detecção dos postes, o que para os objetivos desta dissertação, torna o processo de identificação complicado, visto que o objeto precisa ser identificado a partir de múltiplos pontos de vista, além de tornar o processo trabalhoso, pois para a identificação de todos os objetos é preciso executar o mapeamento percorrendo rua a rua. Ainda sim, a literatura demonstra que a utilização de imagens do nível de rua é uma ferramenta efetiva na tarefa de inspeção de postes elétricos, desenvolvimento de robôs para manutenção automática e carros autônomos.

O trabalho de Gutierrez et al. [28] apresenta uma metodologia para inspeção de postes elétricos. O método descrito atua através de um reconhecimento baseado em três estágios: segmentação por cor, segmentação por formato e por fim um último estágio que calcula o ângulo e diâmetro do poste. De acordo com o descrito

pelos autores, o trabalho obteve 100% de eficácia nos testes executados.

Semelhante, o trabalho descrito em [29] também teve como objetivo o desenvolvimento de uma proposta visando à construção de um sistema robotizado de inspeção e manutenção de postes elétricos. A proposta de Liao et al. [29] baseia-se na utilização de uma câmera de profundidade e na adoção de um algoritmo que combina duas estratégias de localização do poste. A primeira estratégia busca por bordas em cada linha de pixels da imagem. Assim que uma borda é detectada, é adotada uma nova estratégia para procura do poste elétrico. A segunda estratégia busca o objeto na região entre as bordas que foram localizadas.

Em [30] é proposto um método visando um custo computacional não tão elevado e que obtivesse boa precisão mesmo em condições visuais desfavoráveis. A aplicação proposta por Watanabe [30] se assemelha com os requisitos necessários para ferramentas de aprendizagem profunda em carros autônomos. Utiliza-se de um *Single Shot MultiBox Detector* (SSD) para realização desta tarefa. Após um treinamento com 600 imagens e um teste com 100 imagens, obteve precisão média de 72,2%.

O trabalho exposto em [31] visa à detecção de postes elétricos em aplicações para carros autônomos através de uma câmera estereoscópica. O método proposto por Blaga et al. [31] analisa primeiramente a intensidade das imagens para identificar possíveis regiões em que o objeto de interesse está localizado. Depois, o método cria mapas de disparidade de tipo U e V para estimar a posição dos postes elétricos. Os objetos candidatos são clusterizados e por fim refinados para eliminação de *outliers*. O trabalho demonstra uma acurácia do método acima de 80% em todos os testes realizados e também um bom desempenho em imagens noturnas.

Destacam-se também alguns trabalhos que aplicam métodos de segmentação de objetos. Zhu et al. [32] relata a utilização da rede SegNet para detectar e segmentar postes elétricos a partir de imagens de nível terrestre oriundas de um VANT (Veículo Aéreo Não Tripulado). Após a segmentação, foram aplicadas técnicas para o cálculo do ângulo do poste em relação ao solo. O método proposto foi avaliado a partir de um dataset com 84 postes expostos em 64 imagens. A acurácia

alcançada na segmentação foi de 93,74% e o erro médio do ângulo de inclinação foi de 0,59 graus.

Em trabalhos aplicando segmentação, Cheng et al. [15] propõem um novo método para segmentação de postes elétricos em imagens terrestres utilizando a técnica de *graph cut*. O funcionamento do método baseia-se na integração de um conhecimento prévio do formato geral de postes de energia com o *framework* de *graph cut*, visando menores custos computacionais e melhores resultados na segmentação. O método proposto busca primeiramente, através de métodos de detecção de linhas, por áreas que contenham objetos com formato semelhante ao formato geral inserido como conhecimento prévio. Após isso um gráfico ponderado é gerado e então os critérios de *graph cut* encontram a melhor segmentação para o objeto.

O trabalho de Tang et al. [33] propõem um método semi-supervisionado para mapeamento de linhas de distribuição elétrica a partir de anotações de imagens terrestres. Este trabalho propõe a utilização de duas redes Inception-V3 que recebem como entrada as imagens obtidas em céu diurno de um carro com uma câmera inclinada para cima. O método proposto obteve precisão de 97% e recall de 90%.

Alguns trabalhos encontrados relatam o uso de imagens oriundas do Google Street View (GSV) [34, 35]. Cabello et al. [34] descreve a utilização do método Bag of Visual Words (BoVW) para detecção de postes elétricos em imagens do Google Street View na cidade de Campinas/SP.

Em [35] também é proposto a utilização de imagens do Google Street View como base de treinamento e validação de um método. O trabalho propõe o mapeamento de postes elétricos em beiras de estrada através de imagens do GSV. A proposta de Zhang et al. [35] é de certa forma o trabalho que se utiliza de imagens terrestres que o objetivo mais se assemelha com a proposta desta dissertação, se afastando do contexto de inspeção de postes elétricos que esses trabalhos com imagens terrestres geralmente buscam. A proposta combina o método RetinaNet com um método modificado de força bruta de linha de bearing para estimação das localizações dos postes localizados. O método proposto alcançou precisão média

de 78% utilizando limiar de intersecção sobre a união (IoU) maior que 0.3, em um total de 500 imagens com 937 objetos denotados.

2.3 Imagens Aéreas

Os trabalhos correlatos descritos abaixo destacam-se por utilizarem o ponto de vista aéreo na execução de objetivos semelhantes com o desta dissertação. Ainda sim, essas propostas se utilizam de metodologias que não são tidas como métodos do estado da arte.

Em [16] propõem-se um método para localização de postes elétricos através da detecção da sombra causada pelo poste. Utiliza-se de transformada de Hough, transformada de Radon e de filtros detectores de borda para execução desta tarefa. Porém, a detecção da sombra não é a melhor maneira para localização do poste elétrico. A ferramenta proposta por Cetin et al. [16] é eficaz somente em cenários em que não exista algum outro objeto maior que oclua o poste, impossibilitando-o de gerar sombra, como copas de árvores por exemplo. Além disso, o método descrito torna a detecção suscetível à hora do dia em que as imagens são geradas, já que a sombra deriva do ângulo e intensidade da incidência solar no objeto.

Já o relato de Alam et al. [36] expõem o desenvolvimento de um método para detectar e calcular o ângulo de inclinação de postes em relação ao solo, visando predizer a resiliência destes objetos. A proposta utiliza uma versão modificada da SegNet para detecção do objeto e o algoritmo SVM (*Support Vector Machine*) para predição da resiliência do poste. Apesar de utilizar métodos do estado da arte, a proposta exige que as imagens, ainda que aéreas, sejam feitas próximas aos postes, esbarrando na mesma problemática encontrada em relação à utilização de imagens terrestres.

Chen et al. [37] propõem um projeto de detecção e contagem de postes elétricos utilizando um VANT, visando à melhoria no gerenciamento pós-desastres naturais da rede de distribuição de energia. Este VANT é equipado com uma câmera, que realiza a filmagem de todo o percurso do veículo. Em tempo real, uma rede convolucional do tipo YOLO V3 analisa os frames e localiza os postes elétricos. Apesar dos bons resultados relatados, o método utilizado nessa proposta

é um método desatualizado, criado no ano de 2016 e que não faz parte do estado da arte. Foram realizadas atualizações no *backbone* e no algoritmo de treinamento da arquitetura, que hoje se encontra em sua quinta versão (V5). Além disso, apesar de possuir menores custos computacionais, a arquitetura YOLO perde em termos de precisão e acurácia quando comparada a outros métodos do estado da arte.

O trabalho de Liu et al. [38] propõe um método em dois estágios para inspeção de postes elétricos através de imagens aéreas, objetivando a facilitação do processo de inspeção de postes elétricos. Utiliza o método Faster R-CNN com VGG16. Após a detecção do objeto, a parte superior do retângulo envolvente é então reprocessada pelo método para que as condições de conversação dos componentes do poste sejam descobertas automaticamente. O trabalho relata uma precisão média de 79%. Apesar disto, o método proposto por Liu [38] peca em termos de custos computacionais. Ainda que nada relacionado ao tema seja relatado em seu trabalho, métodos de dois estágios são comumente atrelados a um alto custo computacional na execução de suas tarefas.

3 Mapeamento Automático de Postes

Mapear postes elétricos usando imagens de visão lateral adquiridas com câmeras montadas em carros é uma tarefa custosa, principalmente em grandes áreas devido à necessidade do mapeamento rua a rua. Para grandes áreas, imagens aéreas são alternativas viáveis já que possibilitam a cobertura de áreas maiores, embora a detecção e o mapeamento de postes em ambientes urbanos usando imagens de vista superior seja um desafio. Assim, propomos o uso da *Adaptive Training Sample Selection* (ATSS) para a detecção de postes elétricos em áreas urbanas, uma vez que é um método novo e ainda não investigado em aplicações de sensoriamento remoto. Nesta proposta, comparamos o ATSS com Faster *Region-based Convolutional Neural Networks* (Faster R-CNN) e *Focal Loss for Dense Object Detection* (RetinaNet), atualmente usados em aplicações de sensoriamento remoto, para avaliar o desempenho da metodologia proposta. Usamos 99.473 imagens de 256×256 pixels com distância da amostra do solo (GSD) de 10 cm. As imagens foram divididas em conjuntos de dados de treinamento, validação e teste em proporções aproximadas de 60%, 20% e 20%, respectivamente. Como os rótulos dos postes são originalmente as coordenadas de um único ponto e os métodos de detecção de objetos requerem uma caixa delimitadora (*bounding box*), avaliamos a influência do tamanho da caixa delimitadora no método ATSS variando as dimensões de 30×30 a 70×70 pixels. Para a tarefa proposta, nossos resultados mostram que ATSS é, em média, 5% mais preciso do que o Faster R-CNN e RetinaNet. Para um tamanho de caixa delimitadora de 40×40 e com interseção sobre união (IoU) de 50% (AP_{50}), alcançamos precisão média (mAP) de 0,913 para ATSS, 0,875 para Faster R-CNN e 0,874 para RetinaNet. Em relação à influência do tamanho da caixa delimitadora no método ATSS, nossos resultados indicam que em AP_{50}, o *bounding box* de 60×60 é cerca de 6,5% mais preciso em comparação com o formato 30×30. Em AP_{75}, essa margem atinge 23,1% em favor do tamanho da caixa delimitadora de 60×60. Em termos de custos computacionais, todos os métodos testados permaneceram no mesmo nível, com um tempo médio de processamento em torno de 0,048 s por imagem. Nossos resultados mostram que o ATSS supera

outras metodologias e é adequado para o desenvolvimento de ferramentas para detecção e mapeamento automático de postes elétricos.

3.1 Introdução

A rede de distribuição de energia é um dos ativos mais importantes em nosso dia a dia. No entanto, postes elétricos são vulneráveis a situações que levam à deterioração, como por exemplo, eventos climáticos extremos [36], corrosão [39] e acidentes de carro [40]. Postes frágeis podem levar à interrupção do serviço de transmissão de energia, danos a veículos e propriedades alheias, gerando mais despesas para as concessionárias de energia. Em resposta, às concessionárias realizam constantes inspeções humanas e manutenções in loco para evitar a interrupção do serviço. Para isso, essas empresas mantêm bancos de dados georreferenciados de todos os ativos de rede. Porém, com o aumento da rede, manter esses bancos de dados torna-se uma tarefa desafiadora e cara. Portanto, há a necessidade de métodos menos onerosos e que possam abranger áreas maiores para o mapeamento desses ativos, como os postes elétricos.

Pesquisas anteriores se utilizaram de imagens ou nuvens de pontos 3D capturadas por dispositivos montados em sistemas de mapeamento móvel terrestre para localizar automaticamente postes usando métodos de aprendizagem máquina. Lehtomäki et al. [25] propôs um método baseado em *clusters* para encontrar agrupamentos em nuvens de pontos obtidas a partir de um sensor de varredura a laser montado no topo de um veículo. Um *cluster* é classificado como poste quando vários critérios baseados na forma do objeto são atendidos. Um desafio para este método são os postes obliquos e outros objetos de formato cilindrico na região mapeada, semelhante a postes de eletricidade, que levam à detecção de falsos positivos. Sharma et al. [17] também propôs um método baseado em *clustering*, com o algoritmo *Mean Shift* para segmentar postes de energia através de câmeras e imagens de smartphones em áreas rurais, onde o acesso de veículos não é viável. A maioria dos métodos de aprendizado de máquina se utiliza do formato do objeto original como premissa para encontrar outros objetos que possuam formatos semelhantes [25, 17, 34]. Porém, à diversidade de modelos de postes existentes, di-

ferentes distâncias dos objetos e possíveis oclusões, tornam crítica a configuração dos parâmetros desses algoritmos. Para lidar com essas tarefas desafiadoras, métodos de aprendizado profundo têm sido propostos em aplicações de sensoriamento remoto.

Métodos de aprendizado profundo têm sido aplicados significativamente à análise de imagens de sensoriamento remoto, principalmente usando arquiteturas de redes neurais convolucionais (CNNs) [41, 42, 43, 44, 45, 46, 14, 47]. O uso de métodos de detecção de objetos está emergindo na comunidade de sensoriamento remoto [48]. Zhang et al. [35] aplicou um método de aprendizagem profunda de detecção de objetos , denominado RetinaNet, para detecção de postes em imagens do Google Street View. Os autores também usaram uma técnica de linha de bearing (LOB) baseada em força bruta modificada para estimar as coordenadas posicionais. Tang et al. [33] propuseram um método semissupervisionado que requer apenas rótulos em nível de imagem (contendo linha de energia, imagem positiva ou não, imagem negativa) para detectar e localizar linhas de distribuição e postes elétricos. Duas redes Inception-v3 foram aplicadas nas imagens obtidas de uma câmera no nível da rua montada em um carro apontada para cima no céu diurno. No entanto, nesses métodos de imagens de rua e visão lateral do objeto, as informações dos postes precisam ser estimadas considerando vários pontos de vista [35], tornando esse método uma atividade demorada principalmente em áreas maiores devido à necessidade de mapeamento rua a rua.

Sensores de detecção a bordo de plataformas aéreas e orbitais aparecem como uma alternativa que torna viável a cobertura de áreas maiores. As imagens orbitais têm uma distância máxima decimétrica de amostra do solo (GSD), dificultando a detecção de postes. Por outro lado, os sensores aéreos fornecem imagens de alta resolução espacial. De fato, a área coberta nas imagens aéreas é maior do que o nível da rua e também é possível observar a distância entre os postes e a interligação de suas linhas de distribuição de energia. Além disso, imagens aéreas facilitam a captura desses objetos em estradas com trafégo intenso, economizando tempo e esforço nesta tarefa [36]. No entanto, para esta aplicação, poucos trabalhos têm considerado imagens aéreas no mapeamento de postes, as quais são comumente utilizadas em diversas aplicações de sensoriamento remoto [49].

Alam et al. [36] propôs uma variante do SegNet inicializada com uma rede VGG16 pré-treinada para detectar postes em imagens oriundas de um VANT (veículos aéreos não tripulados). As imagens foram capturadas em cenários urbanos voando perto dos postes (quase semelhantes às imagens verticais capturadas de veículos). O objetivo é detectar o poste com SegNet, estimar o ângulo dos postes detectados com Transformada Hough e prever a resiliência do poste com o algoritmo SVM (*Support Vector Machine*). Liu et al. [38] propôs um método com duas Faster-RCNN também inicializadas com a rede VGG16, visando igualmente a detecção de postes em imagens originárias de VANT's. A primeira Faster R-CNN detecta os postes e a segunda focaliza na metade superior dos objetos detectados para detectar a falta da tampa do pólo do poste. Além disso, a RetinaNet superou outros métodos de aprendizagem profunda em várias aplicações de sensoriamento remoto [50, 51, 48]. Apesar desses esforços iniciais, há uma falta de pesquisas com foco no mapeamento automático de postes usando imagens aéreas, principalmente utilizando novos métodos de aprendizagem profunda, como *Adaptive Training Sample Selection* (ATSS). O ATSS [18] é o método pertencente ao estado da tarde e que ainda não foi investigado em aplicações de sensoriamento remoto.

O ATSS propõe uma nova estratégia de seleção de amostras negativas e positivas para o treinamento de métodos de detecção de objetos. Diferente dos métodos que usam o limiar de intersecção sob a união (IoU) fixo para selecionar amostras positivas e negativas, como Faster-RCNN e RetinaNet, o ATSS propõe um algoritmo para obter um limiar de IoU individual para cada caixa delimitadora rotulada por um especialista (*ground-truth*). Portanto, essa abordagem pode melhorar a detecção de pequenos objetos ou objetos que ocupam uma pequena área da caixa delimitadora, como os postes. Além disso, esta proposta beneficia a detecção de objetos da mesma classe em diferentes escalas e formas, diferentemente da maioria dos métodos atuais. De acordo com os resultados apresentados por Zhang et al. [18], o ATSS superou os métodos atuais do estado da arte no conjunto de dados COCO. Apesar desses resultados, o ATSS ainda não foi investigado em aplicações de sensoriamento remoto.

O desenvolvimento de métodos automáticos de detecção de postes é funda-

mental para manter a sustentabilidade da crescente rede de distribuição de energia, enfrentar o aumento de eventos climáticos extremos e manter a qualidade do serviço de transmissão de energia. O objetivo deste estudo foi avaliar o desempenho do ATSS para detectar e mapear postes em imagens aéreas. Comparamos os resultados obtidos com o Faster R-CNN [20] e o RetinaNet [19], que são métodos comumente aplicados na análise de imagens de sensoriamento remoto. O RetinaNet, por exemplo, superou outros métodos de aprendizado profundo em diversas aplicações de sensoriamento remoto [50, 51, 48].

A maioria das pesquisas com imagens aéreas é focada na inspeção de postes, que requer vôos baixos para monitorar seus elementos. Para fins de mapeamento são preferidos voos mais altos, dificultando a identificação dos postes nas imagens, como pode ser verificado na Figura 1. Isso leva a alguns desafios durante a detecção, como a presença de sombras, postes ocluídos por outros objetos, diferentes fundos nas imagens, diferentes tipos e tamanhos de postes e o fato de que os postes representam uma pequena porção da caixa delimitadora.

Figura 1 – Exemplos de postes presentes em nosso conjunto de dados.



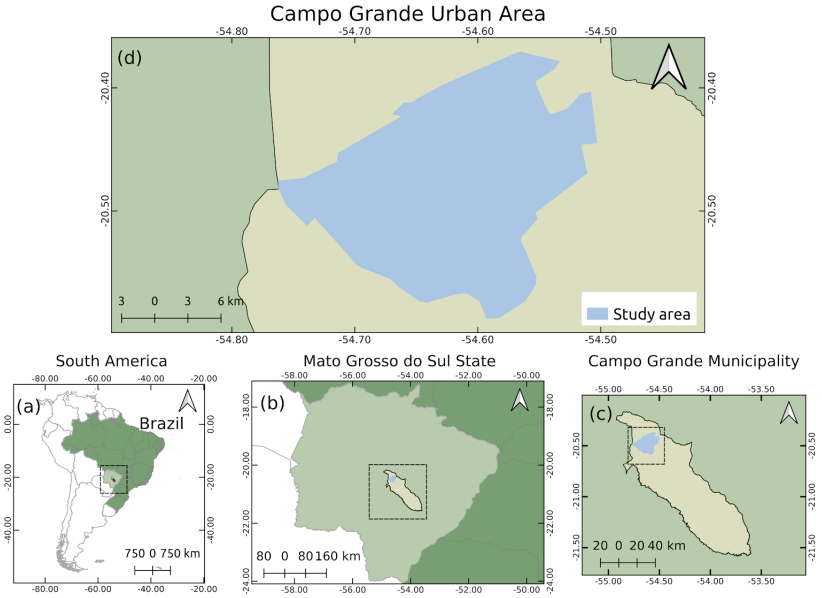
O restante deste trabalho está organizado da seguinte forma. A 3.2 demonstra o método desenvolvido por esta pesquisa. A seção 3.3 apresenta os resultados obtidos com nossa abordagem enquanto os compara com outros métodos. A seção 3.5 discute os resultados obtidos pela nossa abordagem. A seção 3.6 conclui nossa pesquisa.

3.2 Materiais e Métodos

3.2.1 Área de Estudo

Utilizamos como área de estudo a região urbana do município de Campo Grande, no estado de Mato Grosso do Sul, Brasil (Figura 2).

Figura 2 – Localização da área de estudo: (**a**) América do Sul e Brasil; (**b**) Mato Grosso do Sul; (**c**) Campo Grande; e (**d**) Área de estudo.



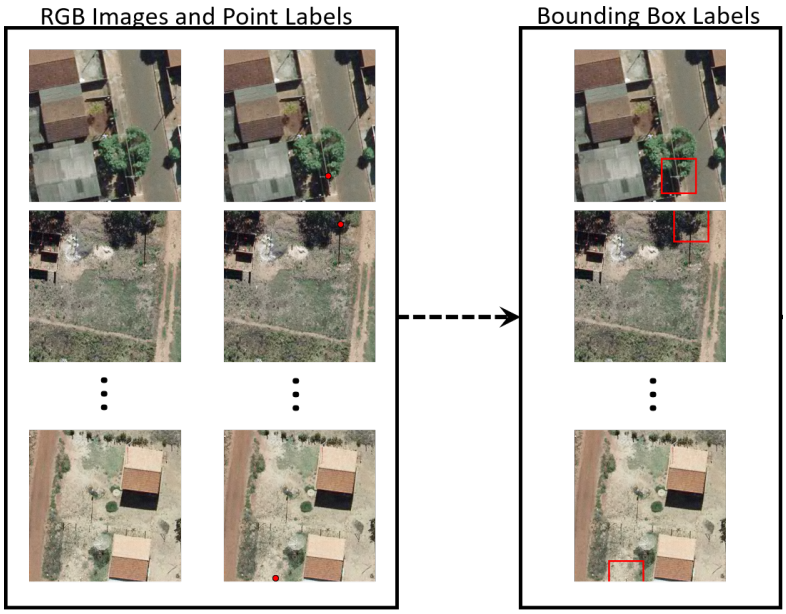
As ortoimagens aéreas RGB foram fornecidas pela Prefeitura de Campo Grande, Estado de Mato Grosso do Sul, Brasil. As ortoimagens têm uma distância de amostra do solo (GSD) igual a 10 cm. No total, 1057 ortoimagens com dimensões de 5619×5946 pixels foram utilizadas nos experimentos. Essas imagens foram divididas em 99.473 recortes, onde 111.583 postes foram rotulados. Detalhes a respeito da configuração experimental são apresentados na Seção 3.2.3.

3.2.2 Métodos Utilizados

Em geral, devido a bancos de dados georreferenciados, os rótulos fornecidos para os postes consistem apenas de suas coordenadas. O conjunto de dados foi

rotulado manualmente como um ponto (representado por uma coordenada x,y) na base de cada poste. Como os métodos de detecção de objetos precisam de uma caixa delimitadora em vez de um ponto, cada rótulo (ponto) do *ground-truth* foi convertido em um retângulo (representado por duas coordenadas). Com isso, uma caixa retangular (conhecida como caixa delimitadora) centralizada neste ponto foi construída definindo-se a largura w_b e altura h_b. Desse modo, o ponto médio de cada caixa delimitadora é representado pelo ponto (rótulo *ground-truth*) para determinar a localização da caixa delimitadora e os dois parâmetros w_b e h_b que definem seu tamanho. Nos experimentos, avaliamos a influência de diferentes tamanhos (valores de w_b e h_b) nos resultados. Na Figura 3 demonstra-se os rótulos representados por uma coordenada e as caixas delimitadoras correspondentes.

Figura 3 – Passo a passo para criação dos rótulos com caixas delimitadoras.



Dadas às caixas delimitadoras anotadas, os algoritmos de detecção de objetos podem ser treinados para aprender essas caixas delimitadoras possibilitando os de prever a localização dos objetos em uma imagem. Nesse sentido, os algoritmos pertencente ao estado da arte como RetinaNet [19] e Faster R-CNN [20] geralmente geram um grande número de caixas (chamadas de caixas de âncora), estimam quão boas são essas âncoras em relação às caixas delimitadoras e as dividem em exemplos

positivos e negativos. A qualidade das caixas de âncora é estimada considerando a razão da interseção da união entre a caixa de âncora e as caixas delimitadoras. Esta medida é conhecida como Intersecção sobre a União (IoU). Como o refinamento desse grande número de âncoras geradas, que possuem quatro parâmetros (coordenadas espaciais) cada, pode ser custoso computacionalmente, recentemente surgiram métodos que não se utilizam de caixas de âncora [52]. No entanto, Zhang et al. [18] concluiu que o principal impacto nos desempenhos dos métodos não é apenas considerar representações baseadas com a utilização ou não de âncoras, mas como esses métodos selecionam as amostras positivas e negativas. A ideia principal do algoritmo ATSS [18] é selecionar um pequeno conjunto (maiores K) de boas caixas de âncoras, baseado nas características estatísticas do objeto para as etapas de classificação e regressão.

Comparamos o método ATSS com os métodos RetinaNet e Faster R-CNN. Os métodos de detecção de objetos usados neste estudo são descritos resumidamente a seguir. Os métodos foram implementados com base nos códigos-fonte propostos pelo Multimedia Laboratory, por meio do projeto MMDetection. O projeto possui código aberto e está disponível em `https://github.com/open-mmlab/mmdetection`.

Faster R-CNN [20] é uma rede neural convolucional (CNN) usada para detectar objetos em dois estágios. Esta arquitetura pode ser descrita em três componentes: espinha dorsal (também chamado de *backbone*, destinado a extrair mapas de características das imagens), pescoço (tem como função refinar os mapas de características) e cabeça (executa tarefas de classificação e predição, por exemplo) [53]. Usando os mapas de características gerados pelo *backbone* (ResNet50 neste trabalho), o primeiro estágio, denominado *Region Proposal Network* (RPN) e presente no pescoço da arquitetura, recebe os mapas de características e gera como resultado caixas delimitadoras de objetos candidatos para identificação. O segundo estágio também recebe os mapas de características e visa por meio de uma camada de *Region of Interest Polling*, extrair características de cada uma das caixas delimitadoras candidatas propostas pelo primeiro estágio. Esta operação é baseada em *max polling*, visando obter um mapa de características de tamanho fixo que não dependa das dimensões das caixas delimitadoras propostas pelo RPN.

Por fim, uma camada *softmax* prediz a localização e classes das regiões propostas na arquitetura.

RetinaNet [19] é um detector de objetos de um estágio que resolve o desequilíbrio de classes através da redução da perda atribuída a imagens bem classificadas. O desequilíbrio de classe ocorre quando o número de exemplos de *backgrounds* (retângulos envolventes sem o objeto de interesse) é muito maior do que os exemplos do objeto de interesse. Por meio de um novo cálculo de erro, o treinamento é focado em exemplos concretos e evita que o grande número de exemplos de fundo de imagem dificulte o aprendizado do método pela rede. A arquitetura RetinaNet é baseada em um *backbone*, um pescoço e duas sub-redes específicas (componentes da cabeça da arquitetura). O *backbone* e o pescoço consistem no ResNet-50 (neste trabalho) e em uma *Feature Pyramid Network* de [54], respectivamente, responsáveis por obter um mapa de característica de toda a imagem recebida na entrada da arquitetura. A primeira sub-rede é uma pequena *Fully Convolution Network*, uma rede simples com cinco camadas convolucionais que são anexadas ao pescoço da arquitetura. Esta sub-rede é responsável por prever, em cada posição espacial por toda a imagem, uma probabilidade de presença do objeto. A segunda sub-rede tem uma estrutura idêntica à primeira, mas é responsável por calcular a regressão para estimar as coordenadas da caixa delimitadora. A primeira e a segunda sub-redes são paralelas.

ATSS (*Adaptive Training Sample Selection*) [18] é um método que introduz uma nova forma de definir amostras positivas e negativas durante a fase de treinamento. Ao contrário de outros métodos do estado da arte, que têm um grande número de hiperparâmetros, o ATSS tem apenas um hiperparâmetro K, correspondente ao número de âncoras mais próximas do centro do *ground-truth*, que são selecionados de cada camada da *Pyramid Feature Network*. Este hiperparâmetro é mantido fixo em 9. O ATSS também difere de outros métodos do estado da arte ao usar um limiar de IoU individual para cada *ground-truth* durante o treinamento, obtido pela adição da média e do desvio padrão dos IoU's das caixas de âncoras propostas em relação ao *ground-truth*. Caixas delimitadoras com IoU maior do que o limiar calculado para o respectivo *ground-truth* são consideradas amostras positivas. Em nossos experimentos, o ATSS foi configurado com ResNet-50 como

backbone e FPN como pescoço.

3.2.3 Configuração do Experimento

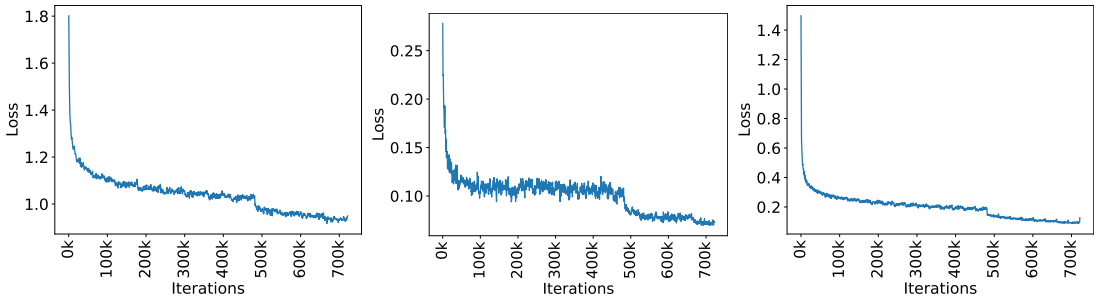
Para configuração do experimento, dividimos as ortoimagens RGB em conjuntos de treinamento, validação e teste. Usamos 634 (60%), 212 (20%) e 211 (20%) ortoimagens de diferentes áreas para treinamento, validação e teste, respectivamente. Cada ortoimagem tem 5619 × 5946 pixels (561,90 m × 594,60 m). Visto que os métodos de detecção de objetos recebem imagens menores como entrada, cada ortoimagem foi dividida em recortes não sobrepostos de 256 × 256 pixels (25,60 m × 25,60 m), cobrindo uma área de aproximadamente 655,36 m^2 por recorte. O número de ortoimagens e recortes são mostrados na Tabela 5.

Tabela 1 – Descrição dos conjuntos utilizados para treinamento, validação e testes.

Conjunto	**Número de imagens**	**Número de Ortoimagens**
Treinamento	60,124	634
Validação	19,869	212
Teste	19,480	211
Total	99,473	1057

Para o processo de treinamento, inicializamos o *backbone* de todos os métodos de detecção de objetos com pesos pré-treinados da ImageNet (`http://www.image-net.org/`). Aplicamos um otimizador Gradiente Descendente Estocástico com momento igual a 0.9. Para isso, utilizamos o conjunto de validação para ajustar a taxa de aprendizado e o número de épocas para reduzir o risco de overfitting. Avaliamos empiricamente as taxas de aprendizagem (0,0001, 0,001 e 0,01). Descobrimos que a função de erro converge melhor quando se utiliza 0,001 de taxa de aprendizagem e o erro tende a se estabilizar após 24 épocas. Na Figura 4 ilustra-se a curva de erro para todos os métodos. Como podemos ver, a função de perda diminui rapidamente nas primeiras iterações e se estabiliza no final, indicando que o número de épocas é suficiente e a taxa de aprendizado adequada. Depois de realizar os ajustes, definimos a taxa de aprendizado inicial para 0,001 e o número de épocas para 24 (∼700 k iterações).

Figura 4 – Curva de erro durante o treinamento dos três métodos de detecção de objetos.



(a) ATSS (b) Faster R-CNN (c) RetinaNet

A aplicação proposta foi desenvolvida utilizando o *framework* MMDetection [26] no sistema operacional Ubuntu 18.04. Os procedimentos de treinamento e teste foram realizados em um computador equipado com CPU Intel^{o} Xeon (E) E3-1270 @ 3,80 GHz, 64 GB de memória RAM em conjunto com placa de vídeo Titan V, GPU produzida pela NVIDIA contendo 5120 CUDA (*Compute United Device Architecture*) núcleos e 12 GB de memória gráfica.

3.2.4 Avaliação dos Métodos

O desempenho dos métodos é avaliado por meio da Precisão Média (AP). Para obter as métricas de precisão e *recall*, a Intersecção sobre a União (IoU) foi calculada como a área de sobreposição entre as caixas delimitadoras preditas e o *ground-truth* divididas pela área de união entre elas. Nos experimentos, usamos valores IoU comuns de 0,5 e 0,75. Se a previsão obtiver IoU maior do que o limiar, a previsão é considerada como verdadeiro positivo (TP); caso contrário, é um falso positivo (FP). Um falso negativo (FN) ocorre quando os *ground-truths* não são detectados por nenhuma previsão. Usando as métricas descritas acima, a precisão e *recall* são estimadas usando as Equações (4.1) e (4.2), respectivamente. A área sob a curva de precisão $\times$ *recall* representa a precisão média (AP).

$$P = \frac{TP}{TP + FP} \tag{3.1}$$

$$R = \frac{TP}{TP + FN} \tag{3.2}$$

3.3 Resultados

Esta seção descreve os resultados obtidos em nossos experimentos. A seção 3.3.1 apresenta uma análise quantitativa do resultado, enquanto a seção 3.3.1 discute as qualitativas. Por fim, a seção 3.4 relata os custos computacionais dos métodos avaliados.

3.3.1 Análise Quantitativa

Primeiro, avaliamos os métodos de detecção de objetos usando precisão média com dois limiares de IoU. No primeiro teste, a precisão média foi obtida usando um limiar IoU de 0,5 (AP_{50}). No segundo teste, aumentamos o limiar para 0,75 e a precisão média foi calculada (AP_{75}). Na Tabela 2 demonstra-se os resultados de cada teste usando uma caixa delimitadora de tamanho fixo de 40×40.

Tabela 2 – Precisão média obtida para cada método experiementado. Para a criação dos *ground-truths*, o tamanho da caixa delimitadora foi fixado em 40×40.

Método	$\boldsymbol{AP_{50}}$	$\boldsymbol{AP_{75}}$
Faster R-CNN	0.875	0.682
RetinaNet	0.874	0.673
ATSS	0.913	0.749

Observa-se que a arquitetura ATSS superou os outros dois métodos em cerca de 4% para AP_{50} e 7% para AP_{75}. Os resultados mostram a maior precisão do ATSS em comparação com outras arquiteturas do estado da arte. Também se observa que, para AP_{50}, Faster R-CNN e RetinaNet têm resultados semelhantes. Quando o limiar de IoU é aumentado para 0,75, o Faster R-CNN obteve um desempenho ligeiramente melhor em comparação com a RetinaNet. Considerando os

111.583 postes presentes em nosso conjunto de dados, a maior precisão em AP_{50} se traduz em cerca de 4464 postes identificados pela rede ATSS que os métodos Faster R-CNN e RetinaNet não foram capazes de detectar. No AP_{75}, esse número sobe para 7.811 postes localizados pelo ATSS que os dois outros métodos acabaram por não localizar.

Para avaliar a influência do tamanho da caixa delimitadora, a Tabela 3 apresenta os resultados usando cinco tamanhos na construção do *ground-truth*. Os resultados foram obtidos utilizando o ATSS como método detector de objetos, uma vez que este obteve os melhores resultados no experimento anterior. Os resultados indicam que aumentar o tamanho da caixa delimitadora de 30×30 para 60×60 melhoram a AP_{50} em 6,5% e a AP_{75} em 23,1%. Em termos de postes detectados, a arquitetura ATSS com caixas delimitadoras 60×60 em comparação com as caixas delimitadoras de 30×30 foi capaz de localizar 7523 postes a mais em AP_{50} e 25776 em AP_{75}.

Tabela 3 – Precisão média para diferentes formatos de caixa delimitadora para o método ATSS.

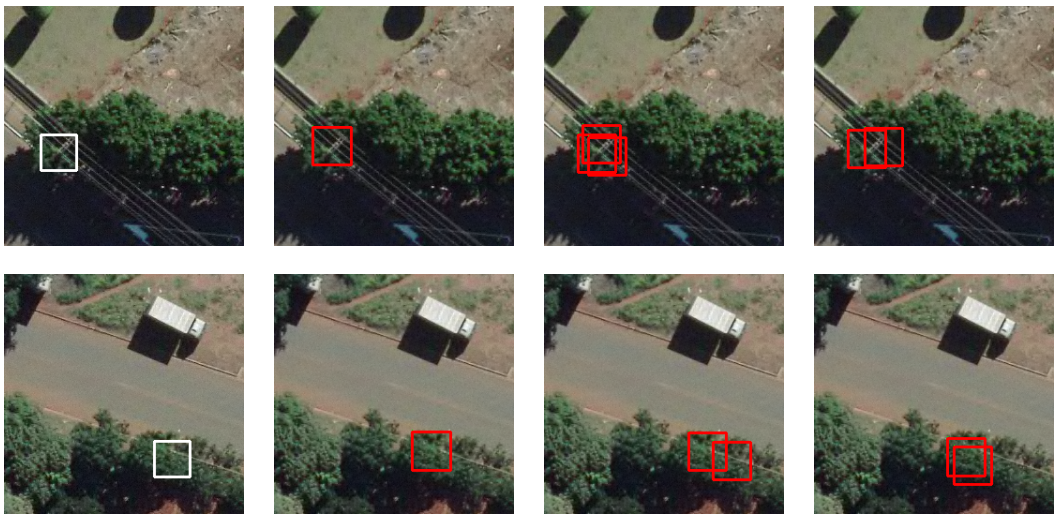
Dimensão	$\boldsymbol{AP_{50}}$	$\boldsymbol{AP_{75}}$
30×30	0.879	0.623
40×40	0.913	0.749
50×50	0.931	0.815
60×60	0.944	0.854
70×70	0.944	0.853

3.3.2 Análise Qualitativa

Os resultados também foram analisados do ponto de vista qualitativo. A prioridade foi analisar a assertividade dos métodos para postes em diferentes condições, como sombreado, coberto por árvores e com diferentes tipos de terreno ao fundo da imagem. Também avaliamos o desempenho dos métodos ao detectar diferentes tipos, tamanhos e formatos dos postes presentes em nosso conjunto de dados. Os resultados demonstrados nesta seção foram obtidos usando o tamanho da caixa delimitadora de 40×40 e AP_{50}.

A Figura 5 ilustra a detecção de postes ocluídos pela copa das árvores. Observa-se que o método mais assertivo para este cenário foi o ATSS. O Faster R-CNN e o RetinaNet até foram capazes de detectar a região da imagem na qual o objeto estava presente, mas ambos os métodos geraram várias caixas delimitadoras para o mesmo objeto quando este se encontra ocluso por árvores.

Figura 5 – Resultados obtidos pelos métodos quando localizando postes oclusos por topos de árvores: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN.



(a) ground-truth (b) ATSS (c) Retina Net (d) Faster R-CNN

A Figura 6 demonstra os postes que são visíveis, mas se encontram em uma região sombreada. Observa-se que a característica das previsões é semelhante ao cenário anterior. Quando os postes estão sombreados, os métodos Faster R-CNN e RetinaNet não tiveram uma boa assertividade. Vários exemplos nestas situações de iluminação foram analisados, e ambos os métodos acabam gerando várias caixas delimitadoras para o mesmo objeto, enquanto o ATSS novamente não teve problemas para localizar os postes.

Em relação aos diferentes tamanhos e modelos de postes, a Figura 7 demonstra as detecções e a sua imagem terrestre correspondente obtida pelo Google Street View. O ATSS obteve um bom desempenho, detectando postes dos mais diversos tipos e tamanhos. Os métodos Faster R-CNN e RetinaNet obtiveram boa

assertividade ao processar imagens de postes de menor escala, como postes de iluminação ou postes elétricos de baixa tensão, mas não foram tão eficazes em localizar postes maiores, como os modelos utilizados para transmissão de energia elétrica de alta tensão, conforme demonstrados na última linha da Figura 7.

No conjunto de dados, há exemplos de postes instalados em diferentes tipos de terreno, como grama, rua, areia, etc. A Figura 8 ilustra o comportamento dos métodos sob esta variação de fundo. Observou-se que a mudança do tipo de terreno não é um fator que interfere na assertividade dos métodos. Todos os métodos obtiveram bom desempenho no processamento de imagens de postes instalados em diferentes tipos de terreno.

Figura 6 – Exemplos de imagens com postes sombreados: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN.



(a) ground-truth (b) ATSS (c) Retina Net (d) Faster R-CNN

3.4 Custos Computacionais

Como o treinamento e a validação foram realizados no computador descrito na Seção 3.2.3, também utilizamos a mesma estação de trabalho para avaliar os custos computacionais dos métodos. Para isso, foi obtido o tempo de processamento de todas as 19.480 imagens do conjunto de teste para cada método. Considerando

que cada imagem cobre uma área de 655 m^2, também estabelecemos a velocidade de varredura como métrica, que corresponde à quantidade de área por segundo que cada método é capaz de processar. Com o tempo total de processamento, foi estimado o tempo médio de processamento para cada imagem e a velocidade de varredura da área, conforme detalhado na Tabela 4.

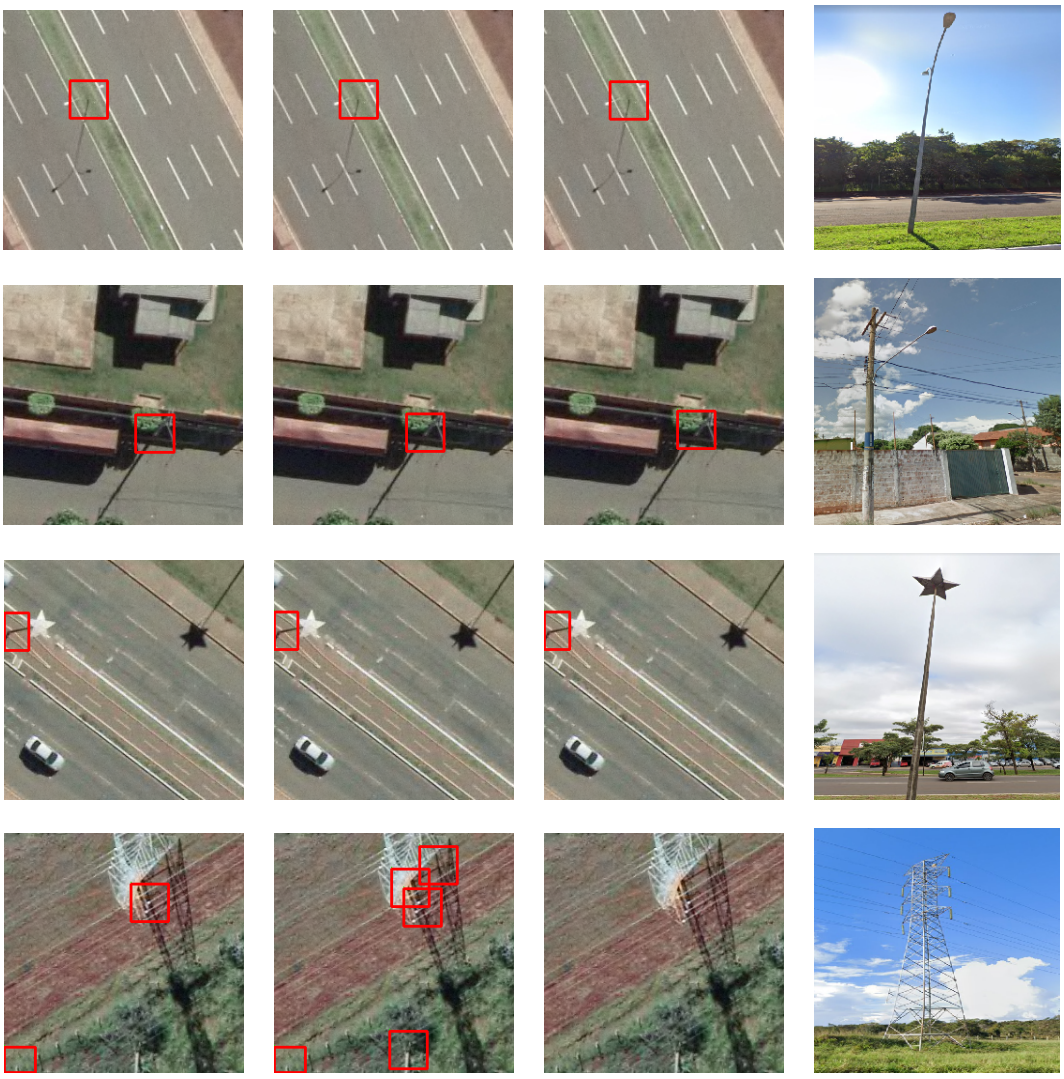
Tabela 4 – Tempo de processamento dos métodos.

Método	**Tempo Médio (s)**	**Velocidade de Varredura (m^2/s)**
Faster R-CNN	0.048	13,651.14
Retina Net	0.047	13,943.82
ATSS	0.048	13,651.14

Como esperado, os três métodos alcançaram um tempo médio de processamento quase igual, com o método RetinaNet sendo apenas 1 ms mais rápido em comparação com os outros métodos, uma diferença mínima que pode ser considerada insignificante. Essa igualdade em termos de custos computacionais é justificada pelo fato de que todos os três métodos testados têm a arquitetura ResNet50 como *backbone*.

Os custos computacionais obtidos demonstram que a estação de trabalho manipulou com eficiência o elevado número de imagens do conjunto de dados. O processamento das 19.480 imagens de teste levou cerca de 15 minutos para cada método. As etapas de treinamento e validação de cada método levaram em torno de 20 h.

Figura 7 – Exemplos de imagens com diferentes modelos e formatos de postes: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN.

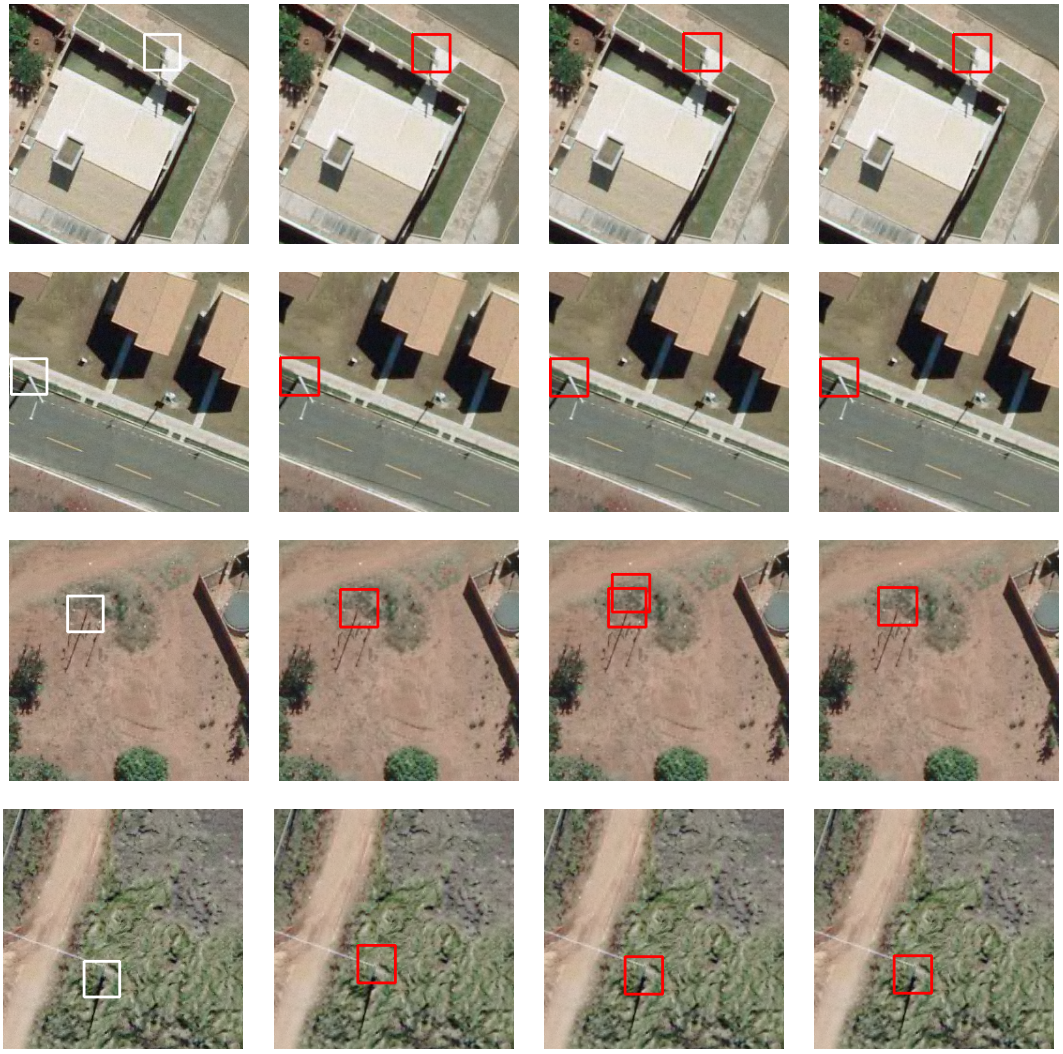


(a) ATSS (b) Retina Net (c) Faster R-CNN (d) Google Street View

A velocidade de varredura obtida demonstra a grande vantagem do uso de imagens aéreas em comparação com o mapeamento rua a rua, conforme descrito por Zhang [35], com nosso método proposto varrendo em apenas alguns segundos os postes de áreas do tamanho de um bairro. O tempo médio demonstra que os métodos têm potencial para serem utilizados em aplicações em tempo real e também podem no futuro serem embarcados em processadores com baixa capacidade

computacional.

Figura 8 – Exemplos de imagens com postes instalados em diferentes tipos de terreno: (**a**) *ground-truth*; (**b**) ATSS; (**c**) RetinaNet; e (**d**) Faster R-CNN.



(a) ground-truth (b) ATSS (c) Retina Net (d) Faster R-CNN

3.5 Discussão

Conforme descrito na seção 3.3, observou-se que o ATSS foi, em média, 5% mais preciso que o Faster R-CNN e RetinaNet usando caixa delimitadora com

dimensões de 40×40 pixels. Em relação aos custos computacionais, os três métodos permaneceram no mesmo nível.

O tamanho da caixa delimitadora se mostrou um parâmetro influente na precisão média obtida pelo ATSS. Ao aumentar o tamanho da caixa delimitadora de 30 × 30 para 60 × 60 identificamos que em AP_{50} a precisão foi elevada de 87,9% para 94,4% e de 62,3% para 85,3% em AP_{75}. Como no conjunto de dados os postes são vistos de diferentes ângulos, caixas delimitadoras menores podem não cobrir completamente o objeto, conforme exemplificado na Figura 9 com caixas delimitadoras de 40 × 40 pixels. O uso de caixas delimitadoras maiores não apenas elimina o risco de não cobrir totalmente o objeto de interesse, mas também permite que o método aprenda elementos inerentes ao objeto que podem auxiliar na sua detecção, como a sombra do poste no solo. No entanto, caixas delimitadoras muito grandes podem incluir informações irrelevantes para a detecção de objetos, conforme observado para caixas delimitadoras de tamanho 70 × 70.

Ao observar o conjunto de dados de imagens, nota-se que a detecção de postes em imagens aéreas é uma tarefa desafiadora, uma vez que existem diferentes terrenos, tamanhos e tipos de postes, postes cobertos por outros objetos como árvores, entre outras situações. Outro problema está relacionado ao tamanho de cada objeto no *ground-truth*. O objeto em si corresponde a uma pequena parte da caixa delimitadora, onde todo o resto corresponde ao fundo da imagem.

Figura 9 – Exemplos de postes que não são totalmente cobertos por uma caixa delimitadora com 40×40 de tamanho: (**a**) Exemplo 1; (**b**) Exemplo 2; e (**c**) Exemplo 3.



(a) (b) (c)

Considerando essas dificuldades, a relativa superioridade do ATSS em relação aos outros dois métodos pode ser creditada à proposta de cálculo de um limiar de IoU individual para cada *ground-truth* durante o treinamento. Como o poste ocupa uma pequena área da caixa delimitadora, ao usar um valor fixo de limiar de IoU para todos os *ground-truth*, como $> 0{,}5$ ou $> 0{,}75$, uma amostra pode ser considerada positiva, mesmo que a caixa delimitadora interseccione apenas o fundo apresentado no *ground-truth*, exclusivamente pelo fato de que a previsão tem uma IoU maior do que um valor predeterminado. O cálculo do limiar de IoU proposto pelo método ATSS evita a ocorrência deste cenário, prevenindo que o método aprenda recursos que não correspondem ao objeto de interesse.

O cálculo do limiar de IoU para cada *ground-truth* mostrou ser uma ferramenta poderosa também para conjuntos de dados em que os objetos têm tamanhos e formas diferentes, cenário este apresentado neste trabalho. A obtenção de um limiar de IoU para cada *ground-truth* possibilitou ao método aprender as características de diferentes modelos de postes. Os resultados obtidos demonstram que o cálculo de um limiar de IoU para cada objeto durante o treinamento contribuiu para que o método ATSS aprendesse características mais gerais dos postes presentes no conjunto de dados.

3.6 Considerações Finais

Neste trabalho propomos e avaliamos o desempenho do método ATSS para a detecção e localização de postes em ortoimagens aéreas. Dois métodos do estado da arte amplamente usados em aplicações de sensoriamento remoto, Faster R-CNN e RetinaNet, foram usados para fins de comparação.

A tarefa foi realizada de forma satisfatória por todos os métodos; no entanto, o ATSS obteve o melhor desempenho entre os três experimentados, com 91,3% de precisão média usando o limiar IoU de 0,5, sendo em média 4% mais preciso do que os métodos Faster R-CNN e RetinaNet, conforme demonstrado na Tabela 2. Esse resultado já era esperado devido às inovações propostas pelo método ATSS em relação a outros métodos do estado da arte. Os custos computacionais dos três métodos foram semelhantes, com processamento médio de 20,83 imagens

por segundo. A melhor configuração de tamanho de caixa delimitadora encontrada para o método ATSS foi com dimensões de 60×60 pixels. Esta configuração alcançou uma precisão média de 94,4% para AP_{50} e 85,4% para AP_{75}, um aumento de 6,5% para AP_{50} e 23,1% para AP_{75} quando comparado com caixas delimitadoras de dimensões 30×30 pixels.

Os resultados indicam que o método ATSS é um forte candidato para o desenvolvimento de uma ferramenta operacional para detecção e localização de postes. Em nossa aplicação obteve melhor desempenho em comparação a outras metodologias propostas e publicadas em outros artigos, como a localização de postes por imagens de nível de rua [35].

4 Classificação Automática de Postes

Neste capítulo serão descritos os experimentos realizados com os métodos ATSS, Deformable DETR, Varifocal-Net e TOOD na tarefa de classificação de seis classes de postes em ortoimagens aéreas. O dataset utilizado neste experimento é o mesmo apresentado no capítulo 3.

4.1 Materiais e Métodos

4.1.1 Área de Estudo

A área de estudo deste experimento é a região urbana de Campo Grande, estado de Mato Grosso do Sul, Brasil. A cidade de Campo Grande foi considerada ideal para os nossos experimentos, dada a gama de análises qualitativas que as imagens aéreas proporcionaram para este trabalho. Além de uma alta variedade de tipos de terreno, a área de estudo é considerada uma das cidades mais arborizadas do mundo [55], proporcionando análise da eficiência dos métodos na classificação dos postes quando rodeados ou cobertos por árvores.

4.1.2 Postes Classificados

As imagens utilizadas neste experimento são as mesmas pertencentes ao dataset utilizado no capítulo 4 e disponibilizadas pela Prefeitura Municipal de Campo Grande. No total foram disponibilizadas 1057 ortoimagens com dimensões de 5619 x 5946 pixels e *ground sample distance* (GSD) de 10 centimetros.

A partir destas ortoimagens, foram realizados 949 recortes de 256 x 256 pixels, de forma que somente os recortes que possuíssem os objetos de interesse para classificação foram utilizados para o treinamento dos métodos. Os rótulos foram gerados utilizando o software LabelMe (disponível em `https://github.com/wkentaro/labelme`). Os postes foram rotulados em seis classes, que serão brevemente descritas abaixo.

4.1.2.1 Poste com Iluminação Dupla

O poste com iluminação dupla é um poste de iluminação em formato de "V", possuindo um bulbo de luz em cada extremidade. Ele é geralmente posicionado nos canteiros centrais das avenidas da área de estudo deste experimento. Na Figura 10 é demonstrado um exemplo deste poste em conjunto com sua vista lateral através do Google Street View.

Figura 10 – Exemplos da classe Iluminação Dupla. A vista aérea é demonstrada em (a) e a vista lateral em (b).



(a) Vista Aérea (b) Vista Lateral

4.1.2.2 Poste de Iluminação Circular

Esta classe (Figura 11) corresponde a um poste de iluminação utilizado comumente para iluminação de ciclovias e pequenos bosques. Em nosso dataset, esta classe é corriqueiramente encontrada ao redor de árvores e outros objetos cilíndricos que dificultam sua detecção.

Figura 11 – Exemplos da classe Poste Circular. A vista aérea é demonstrada em (a) e a vista lateral em (b).

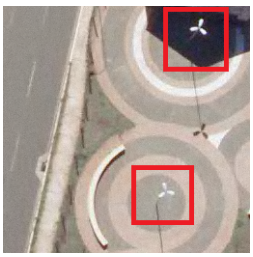


(a) Vista Aérea (b) Vista Lateral

4.1.2.3 Poste de Iluminação de 3, 4 e 5 pontas.

Estas três classes, representadas na 12 abaixo, de postes de iluminação possuem características físicas semelhantes, porém estão costumeiramente instalados em regiões diferentes, proporcionando diferentes *backgrounds*. Além disso, consideramos interessante para o experimento a avaliação do quão minucioso os métodos são na classificação de postes com características tão semelhantes entre si.

Figura 12 – Exemplos de poste das classes Postes 3, 4 e 5 Pontas.



(a) 3 Pontas Vista Aérea



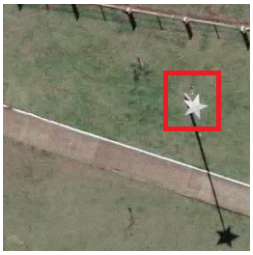
(b) 3 Pontas Vista Lateral



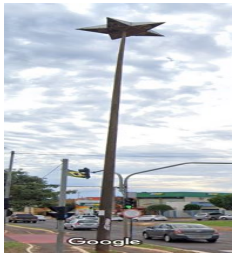
(c) 4 Pontas Vista Aérea



(d) 4 Pontas Vista Lateral



(e) 5 Pontas Vista Aérea



(f) 5 Pontas Vista Lateral

4.1.2.4 Poste de Transmissão tipo T

Esta classe de poste é o único utilizado para transmissão e distribuição de energia elétrica, podendo ou não possuir um braço de iluminação semelhante ao poste da classe iluminação dupla. Visualmente, esta classe destaca-se por possuir uma haste horizontal em seu topo, responsável pela fixação da fiação de energia elétrica no poste, que caracteriza um formato similar a letra “T” (Figura 13).

Figura 13 – Exemplos da classe Poste T. A vista aérea é demonstrada em (a) e a vista lateral em (b).



(a) Vista Aérea



(b) Vista Lateral

4.1.3 Métodos Experimentados

Pelas imagens acima, nota-se que o objeto de interesse de detecção ocupa uma porção pequena da imagem quando comparado ao *background*, estando suscetível a uma variedade de tipos de fundo como arenoso, cimentado, gramado, e etc, de acordo com o terreno em que o poste se encontra instalado.

Além disso, nota-se a complexidade na classificação destes objetos, visto que as dissemelhanças entre algumas classes são sutis e se dão em uma região limitada, geralmente no topo do poste. As diferenças entre as bases cilíndricas de cada tipo de poste não são claras, em grande parte devido a resolução das imagens aéreas.

Neste experimento comparamos os métodos do estado da arte ATSS, Deformable DETR, Varifocal-Net e TOOD (*Task-aligned One-stage Object Detection*) na tarefa de localizar e classificar postes em ortoimagens aéreas. Os métodos utilizados estão brevemente descritos abaixo. As arquiteturas foram implementadas baseadas no código proposto pela Multimedia Laboratory, através do projeto MMDetection. O projeto se encontra disponível em `https://github.com/open-mmlab/mmdetection`.

O método DETR (*Detection Transformer*) [56] é o primeiro detector de ob-

jetos de ponta a ponta, abrindo mão do uso de caixas de âncora, *non-maximum supression* (NMS), pós processamento, e etc, e se utilizando de uma arquitetura simples, combinando redes neurais convolucionais e *Transformers encoder-decoders*. Apesar disto, o método DETR possui alguns contratempos, como dificuldade na detecção de objetos pequenos e um elevado número de épocas para convergir. Observando estes problemas, Zhu et al. [57] criaram o método *Deformable DETR*, apresentando o conceito de *Deformable Attention Module*. Inspirado pela *deformable convolution*, o módulo *deformable attention* atende apenas a um pequeno conjunto de pontos-chave de amostragem em torno de uma referência, independente das dimensões dos mapas de recursos. Assim como grande parte dos métodos mais modernos de detecção de objetos, o Deformable DETR propõe o uso de mapas de recursos multiescalas, de forma que os módulos de atenção do método DETR original foram substituídos por encoder-decoders multiescala.

Outras melhorias também foram acrescentadas por Zhu. Visando aprimorar a performance e inspirado pelo refinamento iterativo desenvolvido em estimativa de fluxo óptico (Teed Deng, 2020) [58], acrescentou-se ao método um mecanismo de refinamento iterativo para cada *bounding box*. Além disso, inspirado pelos detectores de objeto de dois estágios, Zhu adicionou uma variante do Deformable DETR como primeiro estágio da arquitetura, responsável por gerar propostas de regiões passíveis de detecção. Estas regiões alimentam um decoder como consultas de objeto para refinamento adicional, formando um Deformable DETR de dois estágios. Para os experimentos realizados com o Deformable DETR, configuramos o método com as adições de refinamento e duplo estágio citadas acima.

A Varifocal-NET [23] é um método do estado da arte que propõe a melhoria na forma em que a qualidade das detecções são classificadas. Em grande parte dos métodos do estado da arte, o ranqueamento de detecções é feito através de parâmetros como pontuação de classificação e acurácia do retângulo envolvente. A problemática desta abordagem é que a pontuação de classificação não é um bom parâmetro para estimar a qualidade da detecção, implicando na remoção errônea de boas detecções durante o NMS. Para isto, Zhang et. al. apresenta a criação de um parâmetro chamado *IoU-Aware Classification Score* (IACS), como uma representação conjunta dos parâmetros de probabilidade de presença de objeto e

acurácia de localização da caixa envolvente. Propõem a criação de uma nova função de erro chamada *Varifocal Loss*, que supervisiona o método na regressão de IACs e adota um método assimétrico de definição dos pesos.

O método de Zhang também apresenta o conceito de representação das caixas de âncora no formato de *Star-Shaped Bounding Box*, propondo a utilização de nove pontos fixos de amostragem em um retângulo envolvente. Estas informações são então processadas por uma convolução deformável, através de uma camada denominada *Star-Shaped Convolution*. Esta representação permite ao método capturar a geometria do retângulo envolvente e o contexto em que este se encontra na imagem. Na cabeça do método, duas ramificações de *Star-Shaped Convolution* são responsáveis respectivamente pela obtenção do IACS e pela regressão de compensações na localização do retângulo envolvente, proporcionando um refinamento da detecção gerada. O método é construído baseado na arquitetura da rede ATSS. Em nossos experimentos utilizamos o método com o *backbone* ResNext-101 e *Feature Pyramid Network*.

A arquitetura TOOD (*Task-aligned One-stage Object Detection*), é um método de um estágio desenvolvido por Feng et. al. [22]. Propõe a melhoria de convergência entre localização e classificação de objetos, alinhando estas tarefas em uma abordagem baseada em aprendizado. Devido à divergência dos mecanismos de aprendizado para classificação e localização, os recursos aprendidos pelas duas tarefas podem ser diferentes, causando certo desalinhamento quando as regressões são realizadas por duas ramificações separadas na cabeça dos métodos. O método proposto por Feng propõe o conceito de Task Alignment Learning (TAL) para aproximação de âncoras ideais para as tarefas de localização e classificação. Para isto, apresenta um modelo de cabeça denominado *Task-Aligned Head*, em que tanto classificação e localização do objeto são obtidos pela mesma ramificação na cabeça do método, computando recursos interativos das tarefas e realizando predições através do *Task-Aligned Predictor* (TAP), uma camada de atenção para o processamento em conjunto de recursos de classificação e localização dos objetos. Em nossos experimentos utilizamos ResNet50 como *backbone* e com o método configurado para geração de caixas de âncoras durante o processo de detecção.

O ATSS (*Adaptive Training Sample Selection*) [59] é um método do estado

da arte que concluiu que um passo de grande importância para o treinamento é a seleção de amostras positivas e negativas, introduzindo um novo algoritmo para este processo. Para cada *ground-truth* g, k caixas de âncoras cujo o centro esteja mais perto do centro de g são selecionadas como amostras positivas. O IoU entre as amostras positivas e os *ground-truths* são calculados, obtendo então a média M_g e o desvio padrão σ_g das detecções. O método também propõe que para predição final dos candidatos o limiar de IoU para considerar um *bounding box* positivo seja individual para cada *ground-truth*. Este limiar é obtido como $T_g = M_g + \sigma_g$. Em nossos experimentos, utilizou-se ResNet50 e *Feature Pyramid Network* (FPN) como *backbone* e valor de k = 9.

4.1.4 Configuração do Experimento

As imagens foram divididas em conjuntos de treino, validação e teste. Dos 949 recortes, 565 (60%) imagens foram utilizadas para treinamento, 99 (10%) imagens para validação e 285 (30%) para teste. Através do GSD das imagens, é possível determinar a área coberta por cada imagem do nosso dataset. Cada recorte possui dimensões de 256×256 pixels ($25,6 \times 25,6$ m), cobrindo uma área de aproximadamente 655,36 m² por recorte. O número de objetos por classe em cada conjunto é demonstrado na Tabela 5 abaixo.

Tabela 5 – Descrição dos conjuntos utilizados para treinamento, validação e testes.

Conjunto	**T**	**Circular**	**3 Pontas**	**4 Pontas**	**5 Pontas**	**Ilu. Dupla**
Treino	139	57	138	113	102	47
Validação	24	9	23	21	18	9
Teste	78	25	71	57	55	25
Total	241	91	232	191	175	81

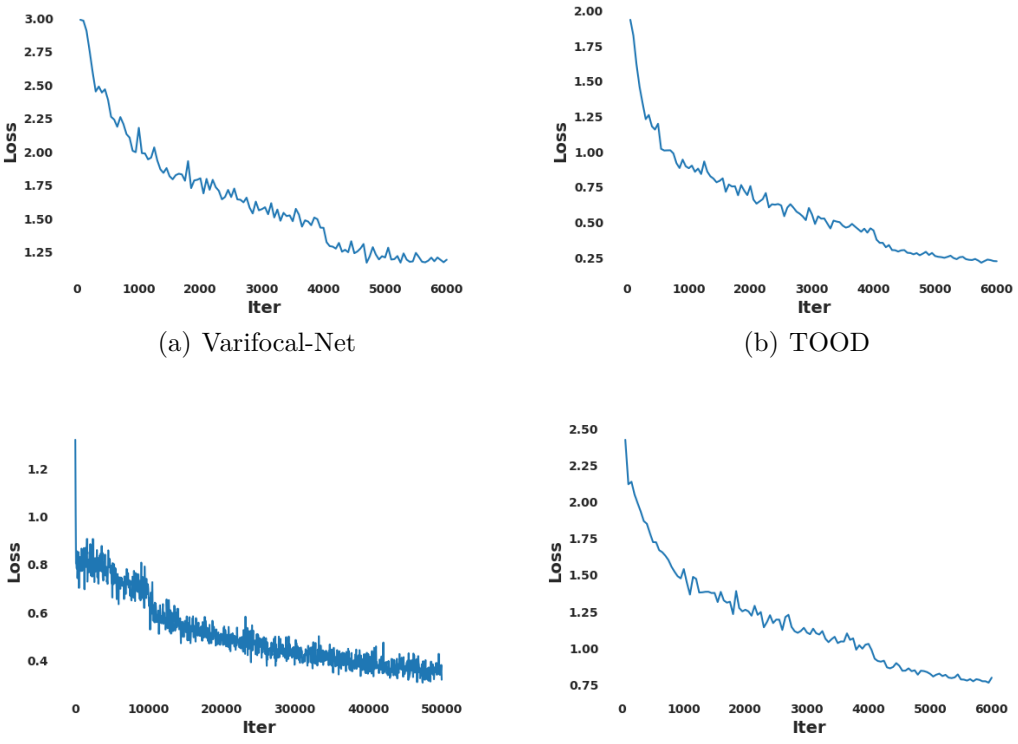
Para o processo de treinamento, os *backbones* dos quatro métodos foram inicializados com pesos pré-treinados da ImageNet (`http://www.image-net.org`). A configuração experimental utilizada para cada método foi estimada empiricamente seguindo os trabalhos originais (e.g [59], [57], [22], [23]) e encontra-se descrita na Tabela 6.

Tabela 6 – Descrição das configurações dos métodos.

Método	Épocas	Learning Rate	Gradient	Optimizer
ATSS	24	0.0001	0.9	SGD
Varifocal-Net	24	0.0001	0.9	SGD
TOOD	24	0.0001	0.9	SGD
Deformable DETR	200	0.0001	0.9	AdamW

A configuração de cada método foi estimada individualmente. A curva de perda por época, demonstrada na Figura 14 abaixo, demonstra que o erro tende a estabilizar após o número de épocas estabelecido para cada um dos métodos experimentados.

Figura 14 – Erro durante o treinamento dos métodos.



(a) Varifocal-Net

(b) TOOD

(c) Deformable DETR

(d) ATSS

Os métodos experimentados foram desenvolvidos utilizando o *framework*

disponibilizado pela MMDetection (https://github.com/open-mmlab/mmdetection). A máquina utilizada para os procedimentos de treino e testes possui sistema operacional Ubuntu 18.04 e é equipada com processador Intel Xeon(E) E3-1270 @3.80 GHz CPU, com 64 GB de memória RAM. O computador possui GPU do modelo Titan V, contendo 5120 CUDA (*Compute Unidet Device Architecture*) e 12 GB de memória gráfica.

4.1.5 Avaliação dos Métodos

A performance quantitativa dos métodos foi avaliada através de precisão média (AP). Esta métrica é calculada através da área sob a curva no gráfico de precision versus *recall*. Para obtenção destes parâmetros calcula-se a intersecção sob a união (IoU), que é área sobreposta entre a região predita pelo método e o *bounding box* do *ground-truth*, dividido pela área de união entre estes. Caso a intersecção fique acima de um limiar mínimo, a predição é considerada verdadeira positiva (TP), caso contrário a predição é classificada como um falso positivo (FP). Quando nenhuma detecção é predita pelo método, esta é classificada como um falso negativo (FN). Em nossos experimentos, utilizamos limiar de IoU de 0.5. Usando as métricas descritas acima, precisão e *recall* são calculados utilizando as expressões descritas nas Equações (4.1) e (4.2) abaixo.

$$P = \frac{TP}{TP + FP} \tag{4.1}$$

$$R = \frac{TP}{TP + FN} \tag{4.2}$$

4.2 Resultados

Esta seção descreve os resultados obtidos com os métodos experimentados. Os resultados serão analisados a partir da óptica quantitativa e de custos computacionais.

4.2.1 Análise Quantitativa

A partir da perspectiva quantitativa, os métodos foram analisados utilizando limiar de IoU de 0.5 (AP_{50}). Definido isto, analisamos a precisão média (AP) que cada classe de poste obteve para os métodos, conforme Tabela 7. Os resultados demonstram que a precisão média das classes varia conforme a metodologia utilizada pelas arquiteturas experimentadas. De forma geral, as classes Poste Circular e Poste 5 Pontas demonstraram os melhores resultados, sobressaindo em relação às outras em grande parte devido ao *background* em que se encontram, visto que esses objetos costumam ser instalados em regiões de campo aberto e com poucos objetos ao redor.

Tabela 7 – Precisão média por classe para os métodos.

Método	**T**	**Circ.**	**3 P.**	**4 P.**	**5 P.**	**Ilu. Dupla**	**Média**
VRF-Net	82.2	89.1	82.3	90.4	97.0	83.8	87.4
Def. DETR	83.7	87.4	78.4	91.2	97.4	83.3	86.9
ATSS	83.4	97.3	84.5	88.3	95.0	72.1	86.8
TOOD	81.2	90.9	75.9	86.3	96.9	73.9	84.3

Através dos resultados, nota-se que as quatro arquiteturas experimentadas possuíram desempenho próximo, com a VRF-Net sendo apenas 0,6% e 0,5% mais precisa em comparação com ATSS e Deformable DETR, e 3,1% mais precisa em relação a TOOD.

4.2.2 Custos Computacionais

Utilizamos a mesma estação de trabalho descrita na Seção 4.1.4 para avaliar o custo computacional dos métodos. Para isto, obtivemos o tempo de processamento das 285 imagens de teste para cada um dos métodos. Considerando que cada imagem cobre uma área de 655m², também obtivemos a métrica de área processada por segundo para cada método. Os resultados são demonstrados na Tabela 8.

Tabela 8 – Custo Computacional dos Métodos.

Método	**Tempo Médio (s)**	**Velocidade de Varredura (m^2/s)**
ATSS	0.142	4.585
Varifocal-Net	0.208	3.149
Deformable DETR	0.167	3.930
TOOD	0.151	4.323

Os resultados demonstram que ATSS obteve o melhor custo computacional entre os quatro métodos. TOOD e Deformable DETR obtiveram custo computacional próximos, sendo apenas 0.009 e 0.015 segundo respectivamente mais lentos em comparação ao ATSS. Neste quesito, VRF-Net não foi capaz de desempenhar de forma equiparada aos outros três métodos, sendo a mais custosa.

A utilização do mesmo *backbone* (R50) para os métodos ATSS e TOOD é evidenciado no custo computacional semelhante de ambas as arquiteturas. Com os resultados do método ATSS, podemos realizar a comparação entre os custos computacionais obtidos para o método treinado somente para a tarefa de localização em relação ao método treinado para classificação de postes. Na tarefa descrita no caítulo anterior, o método levou cerca de 0.048 segundos por imagem no processo de detecção de postes. Ao compararmos os experimentos do Capítulo 3 e Capítulo 4, observamos que a adição do processo de classificação aumentou o tempo de processamento em cerca de 0.094 segundos.

A velocidade de escaneamento obtida corrobora com os resultados demonstrados por Zhang [35], comprovando a grande vantagem no uso de imagens áreas quando comparadas ao mapeamento terrestre, possibilitando a detecção e classificação de postes em grandes áreas em apenas alguns segundos. Todos os métodos seriam capazes de percorrer mais de 3000 m^2 por segundo.

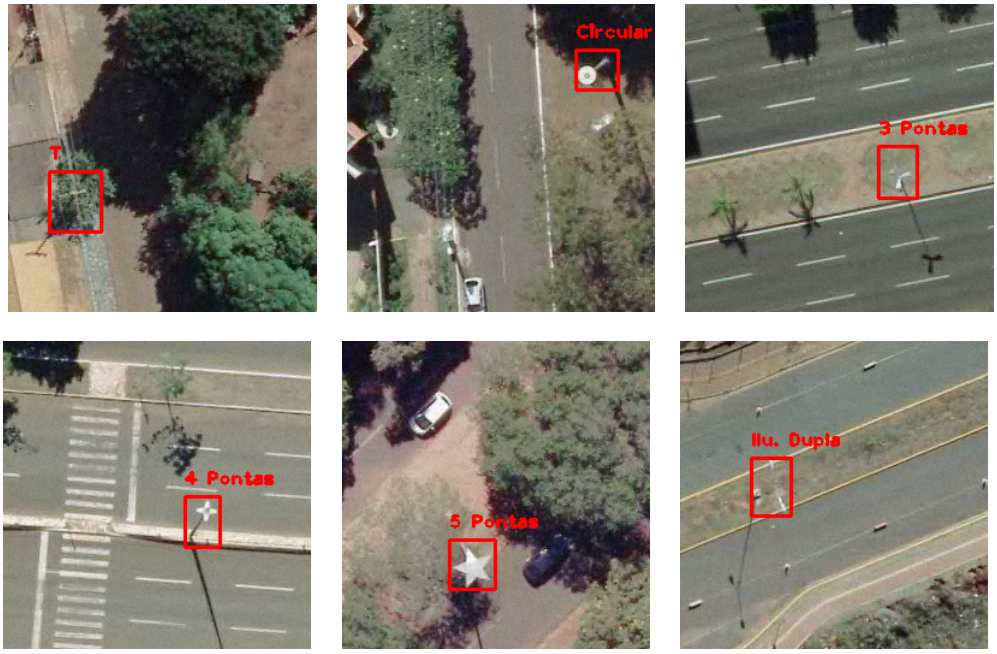
4.3 Discussão

Ao analisar o dataset de nosso experimento, observa-se que a classificação de postes em ortoimagens aéreas é uma tarefa desafiadora. A detecção por si só já é uma tarefa complexa, com a oclusão de postes por árvores e sombras, além

da variedade de terrenos que estes objetos se encontram. A classificação se torna ainda mais desafiadora pela alta incidência em imagens aéreas de objetos com características semelhantes aos postes e as características comuns entre as classes.

Um ponto interessante que nosso dataset permite avaliar é a generalização do aprendizado do método. Como evidenciado na Figura 15, determinadas classes possuem características divergentes entre si, enquanto outras são semelhantes. A dificuldade se encontra que as características majoritárias que definem um objeto se encontram no topo do poste, em uma região ínfima do objeto e que em determinados cenários pode estar obstruído ou sombreado por outro objeto. Todo o restante do objeto, basicamente a base do poste, possui praticamente nenhuma diferença entre as classes.

Figura 15 – Exemplos de nosso dataset demonstrando os desafios da classificação de diferentes tipos de postes. Nota-se nos exemplos que as principais características estão parcialmente cobertas por árvores.

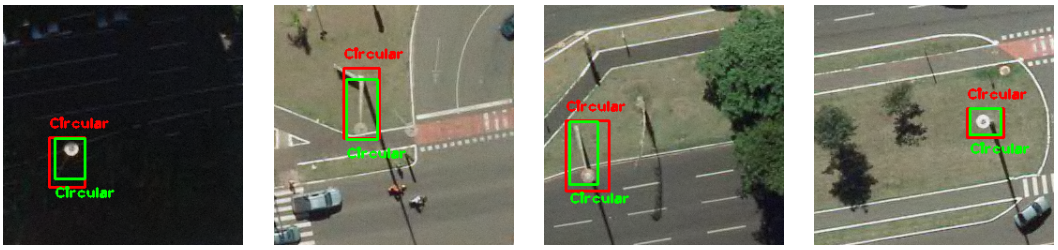


Através dos resultados quantitativos, observa-se que as redes ATSS, VRF-Net e Deformable DETR obtiveram resultados semelhantes em nossos experimentos. A arquitetura TOOD obteve performance levemente inferior quando compa-

rado com os outros três métodos. Nos parágrafos abaixo, analisaremos o desempenho individualmente de cada um dos métodos experimentados, justificando sua eficiência em nosso conjunto de imagens com base nos recursos propostos por cada uma.

A arquitetura ATSS obteve desempenho satisfatório em nossos experimentos, obtendo grande eficácia na classe de Poste 5 Pontas e Poste Circular, destacando-se principalmente nesta última em relação aos demais métodos. Na classe Poste Circular, ATSS foi 9,9% mais preciso que Deformable DETR, além de 8,2% e 6,4% mais preciso que VRF-Net e TOOD, respectivamente. A Figura 16 abaixo apresenta alguns exemplares da classe Poste Circular detectados pelo método ATSS.

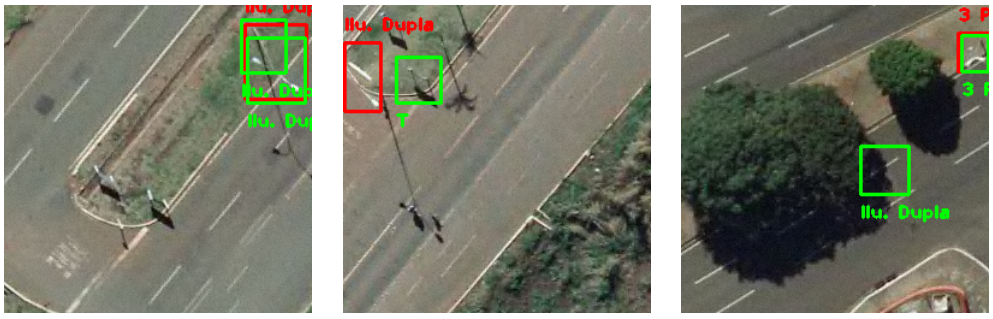
Figura 16 – Postes Circulares detectados pelo ATSS. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



O método não repetiu a mesma eficiência na classificação da classe Iluminação Dupla. Para este objeto, foram observados algumas causas que minaram com a precisão do método (Figura 17), sendo estas: falsos negativos, geração de múltiplos *bounding box* para uma mesma detecção e a atribuição desta classe a objetos do fundo da imagem com aspectos semelhantes ao objeto de interesse, gerando assim um falso positivo.

Em um contexto geral, o método ATSS obteve boa performance na classificação de todas as classes de nosso dataset. Tal eficiência corrobora com os resultados apresentados no capítulo 3. Nestes experimentos, o método ATSS obteve boa performance na localização de postes de diferentes modelos e tamanhos, sendo generalista na localização dos mais variados tipos deste objeto.

Figura 17 – Detecções errôneas da classe Iluminação Dupla pelo método ATSS. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



Assim como demonstrado no capítulo anterior, acreditamos que a proposta de IoU individual para cada *ground truth* do ATSS é uma ferramenta poderosa também para a classificação de cada tipo de poste. Em nossa aplicação, observa-se que o objeto de interesse corresponde a uma porção ínfima do *bounding box*. Além de impedir que *bounding boxes* que interseccionam apenas o *background* sejam considerados como amostras positivas, a utilização de um limiar de IoU individual possibilitou ao método o aprendizado de características particulares de cada classe e impediu que a arquitetura aprendesse atributos desnecessários presentes no fundo da imagem, gerando assim boa precisão na classificação dos postes.

O VRF-Net se destacou com a maior precisão média (mAP) entre todos os métodos experimentados. Apesar de possuir a maior AP individualmente em apenas uma classe, obteve uma eficiência constante em todas as outras classes, se equiparando aos melhores desempenhos para todos os tipos de objetos. Seu desempenho se manteve próximo do Deformable DETR e ATSS, sendo apenas 0,5% e 0,6% respectivamente mais preciso do que os métodos comparados. O desempenho próximo entre ATSS e VF-Net já era esperado e evidencia a similaridade entre as duas arquiteturas, visto que muitos dos elementos presentes no método Varifocal foram inspirados na rede ATSS [23].

Apesar de seu melhor desempenho ter sido para o Poste 5 Pontas, o método VF-Net se destacou dos demais na classe Iluminação Dupla, superando neste tipo de objeto o Deformable DETR em 0,5%, o ATSS em 11,7% e em 8,9% o TOOD.

Na Figura 18 são demonstrados alguns exemplares do Poste Iluminação Dupla detectados pelo método VRF-Net.

Figura 18 – Postes Iluminação Dupla detectados pelo Varifocal-Net. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



O pior desempenho do VRF-Net ocorreu para os objetos do tipo Poste T. Apesar disso, ainda objeto eficiência na média entre os métodos experimentados para esta classe. Na Figura 19 abaixo são demonstrados alguns exemplos em que o VRF-Net falhou em detectar a classe citada.

Figura 19 – Detecções errôneas da classe tipo T pelo método VRF-Net. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



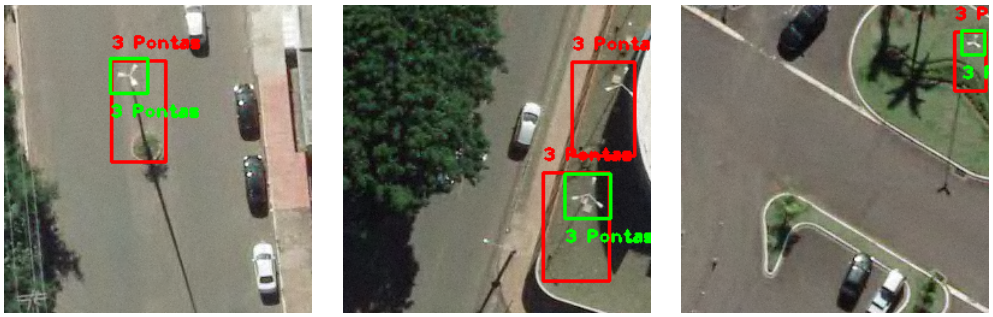
Os objetos presentes em nosso dataset são de difícil localização e por vezes geram detecções com scores baixos. Acreditamos que o recurso de IACS proposto pela VF-Net auxilia neste contexto. Ao gerar uma representação conjunta de presença de objeto e acurácia de localização, o método VF-Net previne que uma

localização com boa acurácia seja removida da detecção final por possuir uma pontuação baixa. Além disso, os postes são facilmente confundidos com vários outros objetos presentes em uma ortoimagem aérea. O IACS também previne que um objeto semelhante ao objeto de interesse seja considerado como uma amostra positiva simplesmente por possuir uma probabilidade alta de classificação.

Atribuímos a boa generalização do método para todas as classes a proposta de *Star Shaped Bounding Box*. A convolução deformável e os nove pontos fixos de amostragem possibilitaram ao VF-Net a captura de informações contextuais da região do retângulo envolvente em que o poste se encontra. Acreditamos que isto possibilitou ao método aprender também outras características que pudessem diferenciar as classes entre si mas que por vezes podem não estar presentes dentro do *bounding box*, como a base do poste, disposição dos fios de energia ao redor, terreno em que o poste se encontra instalado, entre outros.

Dentre os métodos experimentados, o Deformable DETR obteve a segunda melhor precisão na média entre todas as classes. O método baseado em atenção foi apenas 0,1% mais preciso em relação ao ATSS e 2,6% mais preciso em relação ao TOOD. O método obteve maiores dificuldades na detecção da classe Poste 3 Pontas, obtendo precisão média de 78,4%. O problema mais observado para este tipo de objeto foi relacionado às dimensões do retângulo envolvente predito. Em diversos exemplos é possível observar o retângulo envolvendo somente a parte superior do poste e suprimindo a base do poste na detecção. Na Figura 20 são expostos alguns exemplos das dificuldades encontradas pelo método na classificação do Poste 3 Pontas.

Figura 20 – Detecções errôneas de Postes 3 Pontas pelo método Deformable DETR. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



O método possuiu os melhores resultados nas classes Postes 5 Pontas e Postes 4 Pontas, alcançando o melhor desempenho em ambas as classes entre todos os métodos. Na Figura 21 são apresentados alguns exemplares das classes Postes 5 Pontas e Postes 4 Pontas detectados pelo método Deformable DETR.

Figura 21 – Postes 4 e 5 Pontas detectados pelo método Deformable-DETR. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



Através da imagem acima podemos elencar dois aspectos fundamentais que dificultam a classificam destas classes: ambas as classes possuem características

e formatos similares entre si e a própria classe Poste 4 Pontas possui diferentes formatos e tamanhos em seu topo. Atribui-se o melhor desempenho nestes dois cenários ao módulo de *deformable attention* proposto pela arquitetura. O mecanismo de amostragem por pontos em torno de uma referência aliado ao mecanismo de atenção, possibilitou ao método a aprendizagem de características fundamentais de cada uma destas classes, auxiliando na diferenciação entre objetos similares de diferentes classes bem como generalizando objetos levemente dissemelhantes mas pertencentes à mesma classe.

Além disso, observamos que o Deformable DETR foi eficiente em resolver um problema de sua precursora DETR, que era a detecção de objetos em escala reduzida. A adição de módulos de atenção multi-escala demonstrou-se um fator de extrema relevância no sucesso da arquitetura Deformable DETR em nossos experimentos, possibilitando a localização e classificação com maior eficiência dos postes presentes em nosso dataset. O número de iterações para convergência do método foi cerca de 8 vezes maior em comparação aos outros experimentados, porém este comportamento já era esperado devido aos *encoders*, *decoders* e módulos de atenção presentes na arquitetura.

O refinamento de *bounding box* proposto por Zhu também se mostrou uma ferramenta poderosa para o desempenho do Deformable DETR. Durante testes secundários sem o refinamento dos retângulos envolventes, o método não foi capaz de convergir durante os treinamentos, demonstrando portanto a importância deste recurso para o bom desempenho obtido pelo método.

O TOOD obteve um desempenho satisfatório em nosso experimentos, alcançando precisão média de 84,3% entre as classes presentes no dataset. Apesar de possuir a menor mAP entre todos os métodos, foi capaz de se igualar a alguns dos melhores desempenhos individuais por classe. As classes que este método obteve o melhor desempenho foram Poste 5 Pontas e Poste Circular, alcançando 96,9% e 90,9% de precisão respectivamente. Na Figura 23 são demonstrados alguns exemplares de ambas as classes detectadas pelo TOOD. Observa-se que o método foi capaz de classificar estas classes a partir de diferentes ângulos de vista.

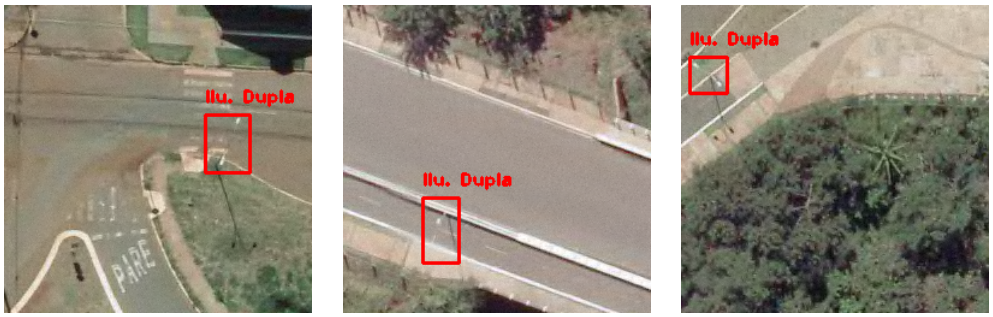
Figura 22 – Postes 5 Pontas e Circular detectados pelo método TOOD. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



Como também observado na arquitetura ATSS, um detector de estágio único assim como TOOD, o método também não obteve um desempenho comparável aos métodos Deformable DETR e VF-Net para a classe Iluminação Dupla, alcançando precisão de 74,5%. Diferentemente do observado para o método ATSS, o motivo do baixo desempenho do TOOD está ligado ao número elevado de falsos negativos gerados. Na Figura 23 são apresentados exemplares não detectados pelo método.

As propostas de alinhamento entre as tarefas de classificação e localização das caixas de âncora se demonstraram efetivas para o contexto de nosso dataset. Acreditamos que a proposta de *Task Alignment Learning* e *Task Alignment Predictor* possibilitaram ao método uma maior generalização durante as detecções. A interação entre as tarefas de classificação e localização preveniram o método da geração de falsos positivos em objetos de características semelhantes aos objetos de interesse.

Figura 23 – Detecções errôneas da classe Iluminação Dupla pelo método TOOD. As caixas delimitadoras vermelhas representam o *ground-truth* enquanto a caixa verde representa a detecção do método.



Os métodos TOOD e VF-Net trazem boas inovações para o estado da arte de detecção e classificação de objetos. Acreditamos que a intenção de promover um melhor alinhamento entre as métricas de classificação, localização e presença de objeto representaram uma grande parcela do bom desempenho obtido por estas duas arquiteturas em nosso dataset. Atribuímos o melhor desempenho do VF-Net ao recurso de Star-Shaped Convolution. Em nosso dataset, a convolução deformável aliado aos pontos de amostragem propostos pelo método demonstraram-se mais eficientes no processo de generalização e aprendizagem do contexto em que os objetos se encontram em comparação aos recursos de TAL e TAP propostos pelo TOOD.

Na comparação entre os detectores de um estágio, ATSS se sobressaiu em relação ao TOOD, contrário ao relatado por Feng et. al. [22] onde o método se sobressaiu em 3,1% no dataset COCO em comparação ao ATSS. Acreditamos que isso se deve ao fato de que os recursos propostos pela ATSS se aplicam melhor ao contexto encontrado em nosso dataset. Como já relatado, as características dos objetos de interesse correspondem a uma parte pequena do retângulo envolvente, de forma que o recurso de *ground-truth* individual potencializa o aprendizado do método de características relevantes para sua detecção e classificação. Por outro lado, os recursos de alinhamento entre classificação e localização propostos por TOOD podem ser mais relevantes em datasets onde exista a presença de vários objetos de interesse de diferentes classes em uma mesma imagem, contexto este do

dataset COCO e não presente em nosso conjunto de imagens.

4.4 Considerações Finais

Neste experimento, propusemos a avaliação e comparação da performance dos métodos do estado da arte ATSS, Deformable DETR, Varifocal-Net (VF-Net) e *Task-aligned One-Stage Object Detection* (TOOD) na tarefa de detecção e classificação de postes em ortoimagens aéreas.

A tarefa foi realizada satisfatoriamente por todos os quatro métodos, porém com o método VF-Net obtendo a melhor performance entre todas as arquiteturas experimentadas, sendo 0,5% mais precisa que a rede Deformable DETR, 0,6% em relação ao método ATSS e 3,1% em comparação com o método TOOD. Em termos de custo computacional, observou-se que as arquiteturas de um estágio, ATSS e TOOD, obtiveram os melhores desempenhos.

Os resultados indicam que os quatro métodos experimentados são fortes candidatos para o desenvolvimento de uma ferramenta operacional para localizar e classificar postes em ortoimagens aéreas. Nossa aplicação obteve melhores resultados em comparação a outras propostas encontradas, tais como o mapeamento de postes através de imagens no nível da rua [35].

5 Conclusão

O mapeamento geográfico de ativos é um processo essencial para a saúde econômica das empresas do setor elétrico. Este tipo de informação é chave para o planejamento de novos investimentos e avaliação da estrutura da empresa. Além disso, durante a revisão tarifária periódica as companhias elétricas são obrigadas a prestar contas de seus ativos para a ANEEL, incluindo informações das coordenadas geográficas de onde estes se encontram instalados.

No capítulo 3, avaliamos a performance de métodos do estado da arte na detecção de postes em ortoimagens aéreas. Comparamos a performance do método ATSS, um método recente e ainda não testado no contexto de sensoriamento remoto, em relação aos métodos Faster R-CNN e Retina Net. O ATSS obteve o melhor desempenho entre os três experimentados, alcançando 91,3% de precisão média em IoU50, sendo cerca de 4% mais preciso do que os métodos Faster R-CNN e RetinaNet.

No capítulo 4, acrescentamos a proposta a tarefa de classificar seis classes de postes presentes nas ortoimagens aéreas de nosso dataset. Avaliamos a performance do ATSS em relação aos métodos Varifocal-Net, Deformable DETR, e TOOD, pertencentes ao estado da arte e ainda também não aplicados no contexto de sensoriamento remoto. Nos experimentos deste capítulo, os quatro métodos experimentados obtiveram desempenho satisfatório na classificação das seis classes de postes rotuladas. Nesta proposta, a rede Varifocal-Net se sobressaiu com uma precisão média de 87,4% . O método foi cerca de 0,5% mais preciso que a arquitetura Deformable DETR, e 0,6% e 3,1% em relação aos métodos ATSS e TOOD.

O ATSS foi aplicado tanto no cenário de detecção quanto de classificação. A aplicação do método nos dois contextos se deve ao fato de que o ATSS é o método mais bem estabelecido na literatura entre as arquiteturas recentes e não experimentadas ainda em sensoriamento remoto. A adição da tarefa de classificação de postes acarretou em um decremento de 6,8% na precisão média do método ATSS.

Em termos de custo computacional, a proposta de classificação de postes foi cerca de 0,094 segundos mais custosa em comparação a tarefa de detecção. As inovações propostas pelo ATSS se aplicam muito bem ao contexto deste experimento. Em ambos os experimentos dos capítulos 3 e 4, o mecanismo de IoU individual auxiliou o método no aprendizado de características pertencentes de fato ao objeto de interesse, eliminando o aprendizado de características de *background*. O capítulo 4 reforça a importância deste recurso, demonstrando o aprendizado generalista obtido pelo método ATSS na tarefa de classificação de postes.

Os métodos Varifocal-Net, TOOD e Deformable DETR foram publicados no ano de 2021, posterior aos experimentos descritos no capítulo 3. Devido ao bom desempenho obtidos por estes métodos na tarefa de classificação do capítulo 4, acreditamos irrelevante testa-los no contexto de detecção proposto no capítulo 3. Os resultados obtidos no processo de classificação nos permitiu avaliar e comparar a eficiência dos recursos propostos por cada método quando aplicados no contexto de nosso dataset.

Apesar dos métodos terem classificados um número satisfatório de modelos, a qualidade das imagens utilizadas para este experimento inviabilizou o treinamento dos métodos para classificação de algumas outras classes de postes presentes na cidade de Campo Grande. A adição de novas imagens ao treinamento pode viabilizar a inclusão destas outras classes, tornando a tarefa de classificação ainda mais completa.

Até então métodos do estado da arte na área de detecção de objetos nunca haviam sido avaliados em tal tarefa. Desta maneira, através dos experimentos deste trabalho demonstramos a viabilidade da criação de uma ferramenta que preencha as lacunas encontradas no contexto de gestão de ativos do setor elétrico.

A performance obtida pelos métodos em nossos experimentos atestam a suficiência da tecnologia atual na criação de uma uma ferramenta eficiente de detecção automática de postes através de ortoimagens aéreas. O capítulo 3 demonstra que os métodos do estado da arte são capazes de localizar os postes mesmo em imagens distantes do solo. O capítulo 4 reforça esta idéia, acrescentando que a classificação do tipo do poste também é possível através do mapeamento aéreo.

As ortoimagens aéreas se mostraram um grande aliado no desenvolvimento desta ferramenta. O mapeamento aéreo possibilitou a varredura de grandes áreas em questão de segundos, algo que não seria possível utilizando o mapeamento do ponto de vista da rua.

Como trabalhos futuros sugerimos a aplicação de técnicas de pós processamento nas detecções dos postes, visando sanar as ocorrências em que múltiplas caixas envolventes são geradas para um mesmo objeto. A adoção de regras para validação da detecção, e.g., a não intersecção entre os objetos, mitigaria as falhas observadas no capítulo 3 e 4, elevando assim a precisão média dos métodos.

O dataset utilizado neste experimento cobre apenas uma região da nossa área de estudo. Sugerimos a aplicação dos métodos em um dataset que abranja toda a cidade de Campo Grande, gerando dados que possibilitem a quantificação da presença de cada tipo de poste por toda a área da cidade. Este tipo de experimento possibilita a análise dos métodos em um contexto mais próximo do que seria a utilização de uma ferramenta operacional de mapeamento de postes no dia a dia.

Referências

1 FIGUEIREDO, A. G. A importância da gestão dosativostécnicos para o negócio de distribuição de energia elétrica no brasil. *Revista Online IPOG ESPECIALIZE*, Jul 2019. Citado na página 1.

2 CRUZ, E. C. da. *Análise de procedimentos metodológicos para georreferenciamento e cadastramento de atributos de redes elétricas urbanas em atendimento às normas da ANEEL*. Dissertação (Mestrado), 2013. Citado 3 vezes nas páginas 1, 2 e 3.

3 PEREIRA, F. E. L. *Determinação do Intervalo de Manuntenção Programada da Proteção de Linhas de Transmissão Considerando-se Penalidades Associadas à Indisponibilidade.* Tese (Doutorado), 2008. Citado na página 1.

4 GISSONI, R. T. *Aplicação de Recursos de Gestão Produtiva Total (TPM) para Gestão de Ativos em Empresas de Distribuição de Energia Elétrica*. Dissertação (Mestrado), 2016. Citado na página 1.

5 VISCONTI, I.; ROSADO, m. Cálculo de perdas na média e baixa tensão numa rede de distribuição real, a partir do modelo de dados bdgd da aneel. In: . [S.l.: s.n.], 2019. Citado na página 1.

6 CANALENERGIA, d. A. *Aplicações inteligentes marcam nova fase de gestão de ativos no setor elétrico*. Disponível em: <https://canalenergia.com.br/noticias-/53114378/aplicacoes-inteligentes-marcam-nova-fase-de-gestao-de-ativos-no-setor-eletrico>. Citado na página 2.

7 LEONARDI, I. *SIG Regulatório no setor elétrico*. 2013. Disponível em: <https://mundogeo.com/2012/04/15/geodireito-4/>. Citado na página 2.

8 ANEEL. *Resolução Normativa 367/09*. [S.l.], 2009. Disponível em: <https:-//www.lexml.gov.br/urn/urn:lex:br:agencia.nacional.energia.eletrica:resolucao-.normativa:2009-06-26;367>. Citado na página 2.

9 ANEEL. *MANUAL DE CONTROLE PATRIMONIAL DO SETOR ELÉTRICO*. [S.l.], 2015. Disponível em: <http://www.aneel.gov.br/livros-/-/asset_publisher/eZ674TKh9oF0/content/manual-de-controle-patrimonial-do-setor-eletrico-mcpse/656835?inheritRedirect=false>. Citado na página 2.

10 ANEEL. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST Módulo 10 – Sistema de Informação Geográfica Regulatório*. 2. ed. [S.l.], 2021. Disponível em: <https://www.aneel.gov.br-/modulo-10>. Citado 2 vezes nas páginas 2 e 3.

11 ANEEL. *NOTA TÉCNICA Nº 07/2020–SFF/ANEEL*. [S.l.], 2020. Disponível em: <http://www2.aneel.gov.br/cedoc/ren2015686.pdf>. Citado na página 3.

12 SUBMóDULO 2.3 Base de Remuneração Regulatória. [S.l.]. Citado na página 3.

13 SOARES, S. M. *Revisão Tarifária e o Equilíbrio Econômico-Financeiro das Distribuidoras de Energia Elétrica.* Dissertação (Mestrado) — Fundação Getúlio Vargas - Escola Brasileira de Administração Pública e de Empresas / Mestrado Executivo em Gestão Empresarial - MEXS, 2019. Citado na página 3.

14 ZHU, L. et al. Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network. *Sensors*, v. 20, n. 6, 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/6/1576>. Citado 2 vezes nas páginas 3 e 15.

15 Cheng, W.; Song, Z. Power pole detection based on graph cut. In: *2008 Congress on Image and Signal Processing*. [S.l.: s.n.], 2008. v. 3, p. 720–724. Citado 2 vezes nas páginas 3 e 10.

16 Cetin, B.; Bikdash, M.; McInerney, M. Automated electric utility pole detection from aerial images. In: *IEEE Southeastcon 2009*. [S.l.: s.n.], 2009. p. 44–49. Citado 2 vezes nas páginas 3 e 11.

17 Sharma, H. et al. Image analysis-based automatic utility pole detection for remote surveillance. In: *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. [S.l.: s.n.], 2015. p. 1–7. Citado 2 vezes nas páginas 3 e 14.

18 ZHANG, S. et al. *Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection*. 2019. Citado 4 vezes nas páginas 5, 16, 20 e 21.

19 LIN, T.-Y. et al. *Focal Loss for Dense Object Detection*. 2017. Citado 4 vezes nas páginas 5, 17, 19 e 21.

20 REN, S. et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2015. Citado 4 vezes nas páginas 5, 17, 19 e 20.

21 ZHU, X. et al. Deformable DETR: deformable transformers for end-to-end object detection. *CoRR*, abs/2010.04159, 2020. Disponível em: <https://arxiv-.org/abs/2010.04159>. Citado na página 5.

22 FENG, C. et al. *TOOD: Task-aligned One-stage Object Detection*. arXiv, 2021. Disponível em: <https://arxiv.org/abs/2108.07755>. Citado 4 vezes nas páginas 5, 40, 41 e 54.

23 ZHANG, H. et al. *VarifocalNet: An IoU-aware Dense Object Detector*. arXiv, 2020. Disponível em: <https://arxiv.org/abs/2008.13367>. Citado 4 vezes nas páginas 5, 39, 41 e 48.

24 PRESS, S.; AUSTIN, D. Approaches to pole detection using ranged laser data. 01 2004. Citado na página 7.

25 LEHTOMäKI, M. et al. Detection of vertical pole-like objects in a road environment using vehicle-based laser scanning data. *Remote Sensing*, v. 2, 03 2010. Citado 2 vezes nas páginas 7 e 14.

26 HUANG, J.; YOU, S. Pole-like object detection and classification from urban point clouds. *Proceedings - IEEE International Conference on Robotics and Automation*, v. 2015, p. 3032–3038, 06 2015. Citado 2 vezes nas páginas 7 e 8.

27 JAVANMARDI, M.; SONG, Z.; QI, X. Automated traffic sign and light pole detection in mobile lidar. *IET Intelligent Transport Systems*, v. 13, 12 2018. Citado na página 8.

28 BARRANCO-GUTIÉRREZ, A. I.; MARTÍNEZ-DÍAZ, S.; GÓMEZ-TORRES, J. L. An approach for utility pole recognition in real conditions. In: HUANG, F.; SUGIMOTO, A. (Ed.). *Image and Video Technology – PSIVT 2013 Workshops*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. p. 113–121. ISBN 978-3-642-53926-8. Citado na página 8.

29 Liao, W. et al. Pole detection for autonomous gripping of biped climbing robots. In: *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. [S.l.: s.n.], 2019. p. 834–839. Citado na página 9.

30 WATANABE, J.-I. Electric pole detection using deep network based object detector. In: . [S.l.: s.n.], 2018. p. 26. Citado na página 9.

31 Blaga, B.; Nedevschi, S. A method for automatic pole detection from urban video scenes using stereo vision. In: *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*. [S.l.: s.n.], 2018. p. 293–300. Citado na página 9.

32 ZHU, Z. et al. Automatic utility pole inclination angle measurement using unmanned aerial vehicle and deep learning. *Proceedings of the Institute of Industrial and Systems Engineers Annual conference 2019*. Disponível em: <https://par.nsf.gov/biblio/10114064>. Citado na página 9.

33 TANG, Q. et al. Fine-grained distribution grid mapping using street view imagery. In: *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 10 e 15.

34 CABELLO, F. et al. Automatic detection of utility poles using the bag of visual words method for different feature extractors. In: *Computer Analysis of Images and Patterns. CAIP 2017*. [S.l.: s.n.], 2017. p. 116–126. ISBN 978-3-319-64697-8. Citado 2 vezes nas páginas 10 e 14.

35 ZHANG, W. et al. Using deep learning to identify utility poles with crossarms and estimate their locations from google street view images. *Sensors*, v. 18, p. 2484, 08 2018. Citado 6 vezes nas páginas 10, 15, 29, 33, 45 e 55.

36 ALAM ZANBO ZHU, B. E. T. J. Z. . S. H. M. M. Automatic assessment and prediction of the resilience of utility poles using unmanned aerial vehicles and computer vision techniques. *International Journal of Disaster Risk Science*, 2020. Citado 4 vezes nas páginas 11, 14, 15 e 16.

37 CHEN, B.; MIAO, X. Distribution line pole detection and counting based on yolo using uav inspection line video. *Journal of Electrical Engineering Technology*, v. 15, 07 2019. Citado na página 11.

38 Liu, L. et al. Deep inspection: An electrical distribution pole parts study via deep neural networks. In: *2019 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2019. p. 4170–4174. Citado 2 vezes nas páginas 12 e 16.

39 JOUKOSKI, A. et al. The influence of cement type and admixture on life span of reinforced concrete utility poles subjected to the high salinity environment of Northeastern Brazil, studied by corrosion potential testing. *CerÃ*, scielo, v. 50, p. 12 – 20, 03 2004. ISSN 0366-6913. Disponível em: <http://www.scielo.br-/scielo.php?script=sci_arttextpid=S0366-69132004000100003nrm=iso>. Citado na página 14.

40 DAS, S. et al. Severity analysis of tree and utility pole crashes: Applying fast and frugal heuristics. *IATSS Research*, v. 44, n. 2, p. 85 – 93, 2020. ISSN 0386-1112. Disponível em: <http://www.sciencedirect.com/science/article/pii-/S0386111218301596>. Citado na página 14.

41 MENG, L. et al. Real-time detection of ground objects based on unmanned aerial vehicle remote sensing with deep learning: Application in excavator detection for pipeline safety. *Remote Sensing*, v. 12, n. 1, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/1/182>. Citado na página 15.

42 ZHANG, X. et al. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sensing*, v. 12, n. 3, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/3/417>. Citado na página 15.

43 YUAN, Q. et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, v. 241, p. 111716, 2020. ISSN 0034-4257. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0034425720300857>. Citado na página 15.

44 CHAUDHURI, U. et al. Cmir-net : A deep learning based model for cross-modal retrieval in remote sensing. *Pattern Recognition Letters*, v. 131, p. 456 – 462, 2020. ISSN 0167-8655. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167865520300453>. Citado na página 15.

45 OSCO, L. P. et al. A convolutional neural network approach for counting and geolocating citrus-trees in uav multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 160, p. 97 – 106, 2020. ISSN 0924-2716. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0924271619302989>. Citado na página 15.

46 TORRES, D. L. et al. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution uav optical imagery. *Sensors*, v. 20, n. 2, 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/2/563>. Citado na página 15.

47 CASTRO, W. et al. Deep learning applied to phenotyping of biomass in forages with uav-based rgb imagery. *Sensors*, v. 20, n. 17, 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/17/4802>. Citado na página 15.

48 LI, K. et al. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 159, p. 296 – 307, 2020. ISSN 0924-2716. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0924271619302825>. Citado 3 vezes nas páginas 15, 16 e 17.

49 LI, L. et al. Quantifying understory and overstory vegetation cover using uav-based rgb imagery in forest plantation. *Remote Sensing*, v. 12, p. 298, 01 2020. Citado na página 15.

50 SANTOS, A. A. d. et al. Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs. *Sensors*, v. 19, n. 16, 2019. ISSN 1424-8220. Citado 2 vezes nas páginas 16 e 17.

51 SANTOS, A. et al. Storm-drain and manhole detection using the retinanet method. *Sensors*, MDPI AG, v. 20, n. 16, p. 4450, Aug 2020. ISSN 1424-8220. Disponível em: <http://dx.doi.org/10.3390/s20164450>. Citado 2 vezes nas páginas 16 e 17.

52 TIAN, Z. et al. FCOS: Fully convolutional one-stage object detection. In: *Proc. Int. Conf. Computer Vision (ICCV)*. [S.l.: s.n.], 2019. Citado na página 20.

53 CHEN, K. et al. *MMDetection: Open MMLab Detection Toolbox and Benchmark*. 2019. Citado na página 20.

54 LIN, T.-Y. et al. *Feature Pyramid Networks for Object Detection*. 2016. Citado na página 21.

55 Gabrielle Tavares. *Árvores do futuro: Campo Grande é uma das cidades mais arborizadas do mundo*. 2021. https://correiodoestado.com.br/cidades/campo-grande-e-uma-das-cidades-mais-arborizadas-do-mundo/390129. Citado na página 34.

56 CARION, N. et al. *End-to-End Object Detection with Transformers*. arXiv, 2020. Disponível em: <https://arxiv.org/abs/2005.12872>. Citado na página 38.

57 ZHU, X. et al. *Deformable DETR: Deformable Transformers for End-to-End Object Detection*. arXiv, 2020. Disponível em: <https://arxiv.org/abs/2010-.04159>. Citado 2 vezes nas páginas 39 e 41.

58 TEED, Z.; DENG, J. *RAFT: Recurrent All-Pairs Field Transforms for Optical Flow*. arXiv, 2020. Disponível em: <https://arxiv.org/abs/2003.12039>. Citado na página 39.

59 ZHANG, S. et al. *Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection*. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1912.02424>. Citado 2 vezes nas páginas 40 e 41.