Estudo de Viabilidade da Classificação Visual de Danos Aparentes em Postes de Concreto Utilizando Visão Computacional

Leonardo Yuji Seo
Curso de Engenharia Elétrica
Universidade Federal de São Carlos,
Brasil
leonardoyuji@estudante.ufscar.br

Celso Ap.de França
Departamento de Engenharia Elétrica
Universidade Federal de São Carlos,
Brasil
celsofr@ufscar.br

Resumo **– Tradicionalmente, a avaliação das condições estruturais de ativos de distribuição elétrica é realizada por meio de inspeções de campo e análises visuais conduzidas por equipes técnicas, processo que demanda tempo e apresenta limitações para avaliação geral da rede elétrica. Nesse contexto, propõe-se a utilização de técnicas de reconhecimento de imagens associadas a modelos computacionais treinados, no caso MobileNetV2, com o intuito de classificar imagens de postes de concreto em bom ou mau estado de conservação, auxiliando na necessidade de manutenção. O projeto foi implementado na linguagem de programação Python, bibliotecas de código aberto voltadas ao processamento de imagens e modelos computacionais previamente treinados com o banco de dados ImageNet, aplicando transferência de aprendizado e mantendo congeladas as camadas responsáveis pela extração de características. Utilizando um conjunto próprio de 150 imagens, igualmente distribuídas entre as duas classes, as imagens foram redimensionadas, normalizadas e submetidas a técnicas de aumento artificial de dados. Na avaliação do conjunto de teste com modelo de imagens recortadas alcançou acurácia de 95% e 85% respectivamente, para a primeira e a segunda distribuição das imagens, mas ainda houve erros na identificação de postes em mau estado. Em posse dos resultados, conclui-se que a abordagem apresenta potencial como ferramenta auxiliar de inspeção, porém ainda não possui confiabilidade suficiente para realizar a classificação autônoma, sendo necessária a ampliação do banco de imagens e o aprimoramento de processamento de dados.**

Palavras-Chave: visão computacional; inspeção de estruturas; postes de distribuição; classificação de imagens.

Abstract — Traditionally, the structural condition of electrical distribution assets is assessed through field inspections and visual analyses conducted by technical teams, a process that is time-consuming and presents limitations for the overall evaluation of the electrical network. In this context, this work proposes the use of image recognition techniques associated with a trained computational model, specifically MobileNetV2, to classify images of concrete utility poles as being in good or poor condition, thereby supporting maintenance assessment. The project was implemented using the Python programming language, open-source libraries for image processing, and a computational model previously trained on the ImageNet dataset. Transfer learning was applied while keeping the layers responsible for feature extraction frozen. A dataset assembled by the author containing 150 images, equally distributed between the two classes, was used. The images were resized, normalized, and subjected to data augmentation techniques. During the evaluation of the test set, the model trained with cropped images achieved accuracies of 95% and 85% for the first and second image distributions, respectively. However, errors still occurred in the identification of poles in poor condition. Based on the results, it was concluded that the proposed approach has potential as an auxiliary inspection tool; however, it is not yet sufficiently reliable for autonomous classification. Therefore, expanding the image dataset and improving data preprocessing procedures are still necessary.

Keywords: computer vision; structural inspection; power distribution poles; image classification.

1. INTRODUÇÃO

O sistema elétrico brasileiro possui uma extensa rede de linhas aéreas [1], que representam mais de 98% da infraestrutura instalada, devido ao menor custo de instalação quando comparado à instalação de rede subterrânea. Nesse contexto, grande parte da distribuição de energia elétrica ocorre por meio de estruturas aéreas sustentadas por postes, os quais necessitam de inspeções periódicas para garantir a integridade estrutural e a segurança do sistema.

De acordo com documento técnico da AES Eletropaulo [2], disponível no portal da Enel, a avaliação das condições dos postes considera critérios como fissuras, lasqueamento, exposição de armaduras e degradação do concreto. Esses parâmetros técnicos orientam a decisão quanto à manutenção ou substituição do ativo.

A fiscalização da rede de distribuição, conforme descrito na literatura especializada [3] e [4], é realizada predominantemente por meio de visitas técnicas presenciais, que envolvem deslocamento de equipes qualificadas para análise visual dos ativos. Tal procedimento demanda tempo, recursos humanos especializados e custos operacionais, podendo limitar o alcance das inspeções diante da extensa rede de distribuição.

Embora existam normativas técnicas e estudos voltados à fiscalização e à gestão de ativos da rede de distribuição, observa-se que os processos descritos concentram-se predominantemente em inspeções presenciais e avaliações visuais realizadas por equipes especializadas. Nesse contexto, a principal contribuição deste trabalho consiste na aplicação de técnicas de visão computacional para a classificação do estado superficial de postes de distribuição, com base em critérios técnicos

normativos, propondo uma abordagem que possa atuar como ferramenta de apoio à decisão no processo de inspeção.

Dessa forma, este trabalho propõe o desenvolvimento de uma abordagem baseada em visão computacional capaz de identificar, a partir de imagens digitais, postes em mau estado de conservação, classificando-os conforme critérios técnicos previamente estabelecidos. A proposta não visa substituir as inspeções técnicas presenciais, mas atuar como ferramenta de apoio à decisão, assim, auxiliando na priorização de visitas técnicas e na otimização dos recursos operacionais.

O objetivo geral deste trabalho é desenvolver uma abordagem baseada em visão computacional capaz de classificar o estado de conservação de postes de distribuição de energia elétrica (enfoque nos postes construídos por concreto), a partir de imagens digitais, conforme critérios técnicos previamente estabelecidos; e como objetivos específicos: coletar e organizar um conjunto de imagens de postes de distribuição em diferentes estados de conservação, aplicar técnicas de visão computacional utilizando redes neurais convolucionais para classificação das imagens e avaliar o desempenho do modelo na identificação de postes em bom e mau estado.

2. TRABALHOS RELACIONADOS

Nos últimos anos, o avanço das técnicas de visão computacional, especialmente aquelas baseadas em redes neurais convolucionais (CNNs), tem permitido o desenvolvimento de soluções automatizadas em tarefas de reconhecimento de padrões em imagens [5]. Nesse contexto, diversos trabalhos utilizam CNNs para a detecção de fissuras e rachaduras em estruturas de concreto, proporcionando maior agilidade e consistência em comparação aos métodos tradicionais baseados em inspeção visual humana [6].

Entre essas abordagens, destacam-se modelos que combinam CNN com técnicas de processamento de imagem, permitindo não apenas a detecção, mas também a classificação e quantificação de fissuras. Tais métodos apresentam elevada acurácia, superior a 96% conforme [7]. No entanto, essas soluções tendem a possuir maior complexidade, exigindo redes desenvolvidas especificamente para a aplicação, além de grandes volumes de dados para treinamento.

Outra abordagem amplamente utilizada consiste no uso de técnicas de transfer learning com arquiteturas pré-treinadas, como VGG16 [8] e AlexNet [9]. Esses métodos permitem o treinamento de modelos mesmo com conjuntos de dados reduzidos, sendo particularmente úteis em cenários com limitação de imagens. Os resultados obtidos indicam acurácia elevada, podendo variar entre aproximadamente 92% e 99%, dependendo das condições do experimento.

Entretanto, apesar do bom desempenho, tais modelos ainda apresentam limitações, como a dependência de dados rotulados e dificuldades de generalização em cenários reais, nos quais há maior variabilidade nas condições das imagens, como iluminação, ruídos e presença de elementos de fundo.

De modo geral, observa-se que modelos baseados em CNN apresentam desempenho superior aos métodos tradicionais de processamento de imagem, especialmente em cenários com maior variabilidade nas condições das imagens. Entretanto, tais abordagens ainda apresentam limitações, como a dependência de grandes volumes de dados rotulados e dificuldades de generalização em cenários reais.

Nesse contexto, arquiteturas leves, como a MobileNetV2, têm se destacado como alternativas eficientes para aplicações práticas. Esses modelos apresentam menor custo computacional, sendo adequados para execução em dispositivos com recursos limitados, sem comprometer significativamente a acurácia. Quando combinados com técnicas de transfer learning, podem alcançar resultados elevados em tarefas de detecção de fissuras conforme [6].

Essa característica torna tais arquiteturas especialmente relevantes em cenários com restrições de hardware e conjuntos de dados reduzidos, aproximando-se das condições enfrentadas em aplicações reais.

3. FUNDAMENTAÇÃO TEÓRICA

3.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados para modelar a maneira como o cérebro realiza uma tarefa conforme [5] e [11]. Essas redes são compostas por camadas de neurônios artificiais, organizadas em camadas de entrada, camadas ocultas e camada de saída, com o objetivo de identificar padrões e relações em conjuntos de dados [10].

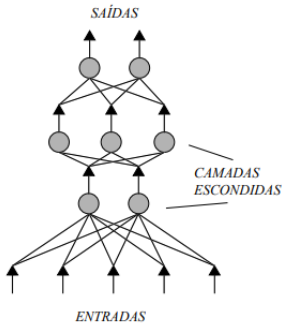
A partir dos dados fornecidos na camada de entrada, as informações são propagadas pelas camadas ocultas, onde cada neurônio realiza operações baseadas em combinações ponderadas entre os nós, associadas a pesos [10]. Essas operações permitem que a rede aprenda representação dos dados, ajustando seus parâmetros durante o processo de treinamento.

Por fim, a camada de saída é responsável por fornecer o resultado da rede neural, que pode apresentar uma classificação ou uma previsão com base nos padrões identificados durante o processo de treinamento.

De forma geral, as RNAs podem ser categorizadas em redes de propagação para frente (*feedforward*) e redes recorrentes, diferenciando-se pela presença ou não de realimentação no fluxo de informações [12].

A Figura 1 apresenta a estrutura de uma RNA do tipo *feedforward* totalmente conectada, na qual cada neurônio de uma camada está conectado a todos os neurônios da camada seguinte[12].

Figura 1 - Diagrama estrutural de uma rede neural artificial.



Fonte: [12].

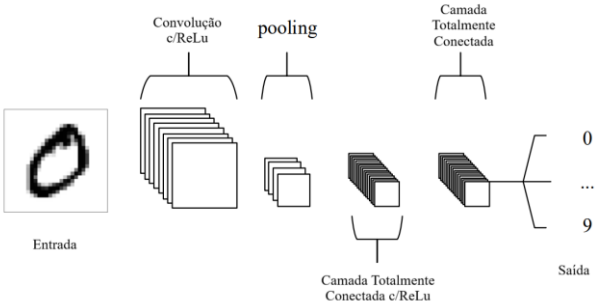
3.2 Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs) são uma especialização das RNAs, desenvolvidas especificamente para o processamento de dados estruturados em forma de grade, como imagens [5]. Essas redes são capazes de identificar padrões a partir dos valores presentes nos pixels da imagem, permitindo extrair características relevantes [10].

De forma geral, as CNNs são treinadas por meio de aprendizado supervisionado, utilizando algoritmos de retropropagação (*backpropagation*) para ajustar os parâmetros da rede. Sua arquitetura é composta por camadas de entrada, camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas (*fully-connected)*, responsáveis pela etapa final de classificação. As camadas convolucionais e de *pooling* correspondem às camadas ocultas da rede, sendo responsáveis pela extração e redução das características das imagens. Ao longo dessas camadas, funções de ativação, como a *Rectified Linear Unit* (ReLU), são aplicadas para introduzir não linearidade ao modelo, permitindo a aprendizagem de padrões mais complexos a partir dos dados de entrada de acordo com [5], [10] e [13].

A Figura 2 apresenta uma representação simplificada da arquitetura de uma CNN, apresentando o fluxo do processamento desde a entrada até a saída do modelo.

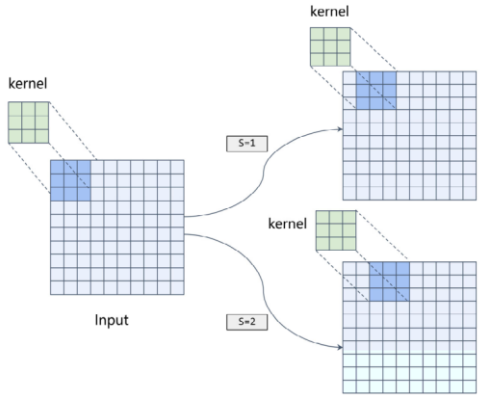
Figura 2 – Arquitetura de uma Rede Neural Convolucional.



Fonte: adaptado de [5].

A camada de convolução tem como objetivo extrair características relevantes da imagem aplicando-se operações de convolução, gerando como saída mapas de características (*feature maps*). Esse processo é realizado por meio de filtros, também chamados de *kernels*, que percorrem a imagem de entrada realizando operações matemáticas em regiões locais [5], permitindo a identificação de padrões como bordas, texturas e formas [10]. A Figura 3 apresenta o diagrama da camada convolucional e o uso do *kernel*.

Figura 3 - Camada convolucional

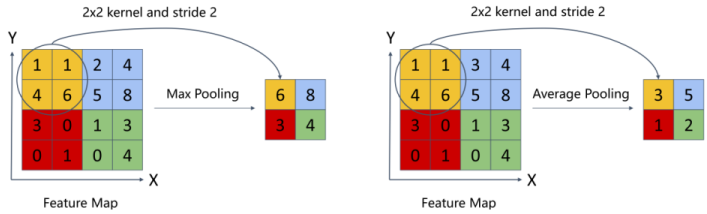


Fonte: [10].

A camada de pooling tem como objetivo reduzir a dimensionalidade dos mapas de características gerados pelas camadas convolucionais, diminuindo o número de parâmetros e o custo computacional do modelo [10]. Esse processo é realizado ao dividir o mapa em regiões locais e aplicar uma operação de agregação, sendo o *max pooling* o método mais comum, no qual o maior valor de cada região é selecionado. Geralmente utiliza-se um filtro de dimensão 2x2 com passo (stride) igual a 2, reduzindo a dimensão espacial do mapa [5]. Além disso, essa etapa contribui para a melhoria da capacidade de generalização da rede e para a redução de overfitting.

A Figura 4 apresenta o processo de *max pooling* (pooling máximo) e *average pooling* (pooling médio).

Figura 4 - Processo de pooling máximo e pooling médio.



(a) Max Pooling (b) Average Pooling

Fonte: [10].

3.3 Função de Ativação

Funções de ativação são componentes das redes neurais artificiais responsáveis por transformar a saída de um neurônio antes de enviá-la para a próxima camada. Sua principal função é introduzir não linearidade no modelo, permitindo que a rede aprenda padrões complexos e represente relações mais sofisticadas entre os dados de entrada e saída [14].

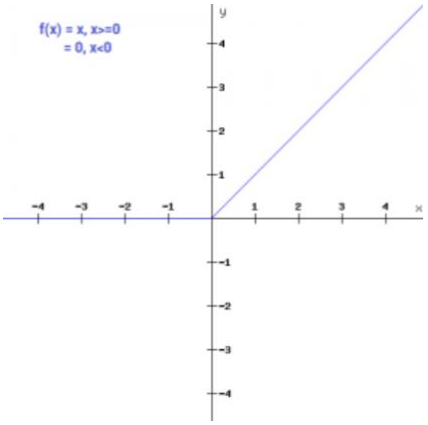
Sem funções de ativação, a rede se comportaria como um modelo linear simples, com capacidade limitada de aprendizado. Por isso, elas são essenciais para tarefas mais complexas, como reconhecimento de imagens, voz e texto. Além disso, funções de ativação diferenciáveis possibilitam o uso do algoritmo *backpropagation*, permitindo o ajuste dos pesos e a otimização

do modelo durante o treinamento [14].

Há diversas funções de ativação utilizadas na literatura, sendo a ReLU uma das mais empregadas. Para problemas de classificação binária, também são utilizadas funções como a função degrau binário e a função sigmóide. No entanto, a função degrau binário não é adequada para redes neurais modernas, pois não é diferenciável, o que impede a aplicação do algoritmo de *backpropagation* [14].

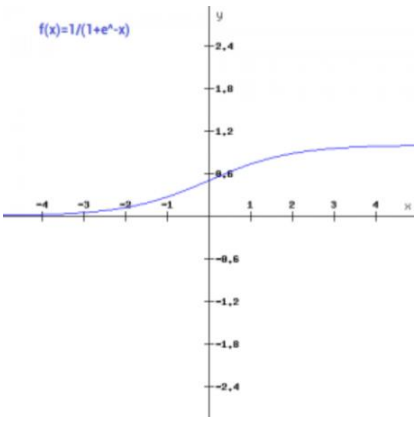
A função ReLU é amplamente utilizada nas camadas ocultas devido à sua eficiência computacional e à sua capacidade de aprender padrões complexos em dados de imagem. Já a função sigmóide é mais apropriada para a camada de saída em problemas de classificação binária, pois produz valores no intervalo entre 0 e 1, permitindo interpretar o resultado como a probabilidade de ocorrência de uma classe, como a presença ou ausência de rachaduras [14]. As Figuras 5 e 6 apresentam as funções de ativação ReLU e sigmóide.

Figura 5 - função de ativação ReLU.



Fonte: Adaptado de [14]

Figura 6 - Função de ativação Sigmóide.



Fonte: Adaptado de [14].

3.4 Transfer Learning

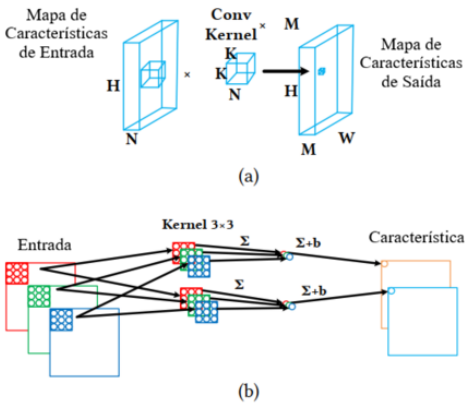
O *transfer learning* consiste na utilização de modelos previamente treinados em grandes bases de dados para a resolução de novos problemas, permitindo aproveitar características já aprendidas em tarefas similares para reduzir o tempo de treinamento e a necessidade de grande volume de dados. No contexto de redes neurais convolucionais, essa abordagem possibilita reutilizar características já aprendidas nas camadas iniciais da rede, como padrões presentes nas imagens, enquanto as camadas posteriores são responsáveis por se adaptar às características específicas da nova tarefa.

Uma estratégia comum envolve o congelamento (*freezing*) das camadas iniciais do modelo, responsáveis pela extração de características gerais, enquanto as camadas finais são ajustadas (*fine-tuning*) para adaptar o modelo ao problema proposto. Além disso, a camada de saída pode ser modificada conforme a tarefa, como em problemas de classificação binária. Dessa forma, o *transfer learning* permite o desenvolvimento de modelos mais eficientes, com menor custo computacional e melhor capacidade de generalização [6].

3.5 MobileNetV2

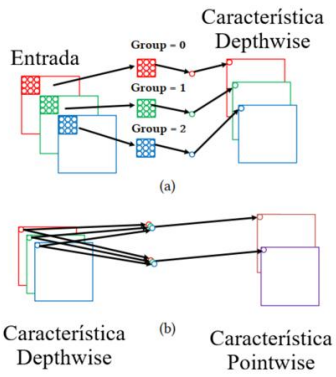
A MobileNetV2 é uma arquitetura de rede neural convolucional projetada para aplicações com restrições de processamento, como dispositivos móveis e sistemas embarcados. Sua principal característica é a eficiência computacional, obtida por meio do uso de convoluções separáveis em profundidade (*depthwise separable convolutions*) e da estrutura de blocos residuais invertidos (*inverted residuals*), que permitem reduzir o número de parâmetros do modelo sem comprometer significativamente seu desempenho [6]. As Figuras 7 e 8 exemplificam a convolução padrão e a *depthwise separable convolution*.

Figura 7: convolução padrão.



Fonte: adaptado de [15].

Figura 8 - *Depthwise Separable Convolution.*

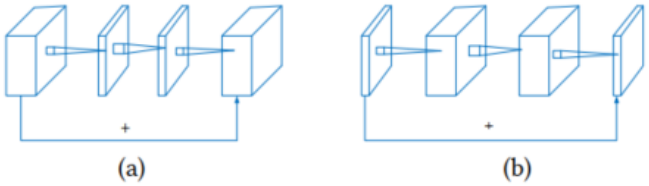


Fonte: adaptado de [15].

Na *depthwise convolution*, cada canal da imagem é processado separadamente, reduzindo o custo computacional da rede. Já a *pointwise convolution*, realizada com filtros 1×1,

combina as informações dos diferentes canais, permitindo a integração das características extraídas. A combinação dessas duas etapas forma a *Depthwise Separable Convolution*, utilizada na MobileNetV2 para tornar a arquitetura mais leve. A Figura 9 apresenta o bloco residual invertido.

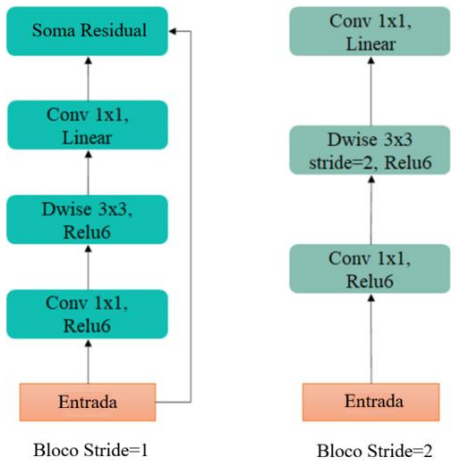
Figura 9 – Comparação entre bloco residual tradicional e bloco residual invertido.



Fonte: adaptado de [15].

Os blocos residuais invertidos (Inverted Residual Blocks) da MobileNetV2 realizam inicialmente a expansão dos canais por meio de convoluções 1×1, seguida da aplicação da depthwise convolution e da redução dimensional utilizando convoluções pointwise. Além disso, a arquitetura utiliza o conceito de *Linear Bottleneck*, no qual a camada final de redução dimensional emprega uma projeção linear sem função de ativação ReLU, buscando preservar informações relevantes durante o processamento e aumentar a eficiência computacional da rede. A Figura 10 apresenta a estrutura simplificada da arquitetura MobileNetV2.

Figura 10 - Estrutura Simplificada da arquitetura MobileNetV2.



Fonte: adaptado de [16].

Conforme apresentado na Figura 10, a MobileNetV2 utiliza blocos compostos por convolução 1x1 e convoluções separáveis em profundidade (*depthwise convolution*). Essa abordagem divide o processamento da imagem em etapas, tornando a rede mais leve e eficiente quando comparada às CNNs tradicionais.

Essas características tornam a MobileNetV2 adequada para tarefas de visão computacional em tempo real, especialmente em cenários com recursos limitados. Além disso, essa arquitetura pode ser utilizada em conjunto com técnicas de *transfer learning*, permitindo o reaproveitamento de modelos previamente treinados e facilitando sua adaptação para tarefas específicas, como a detecção de rachaduras [6].

3.6 Método de Avaliação

A matriz de confusão é utilizada para avaliar modelos de classificação, relacionando a classe real das amostras com a classe prevista pelo modelo. Em problemas de classificação binária, as classes podem ser representadas pelos valores 0 e 1, correspondentes, respectivamente, às classes negativa e positiva [10].

Tabela 1 - Matriz de Confusão.

	Previsto	
Real	0	1
0	TN	FP
1	FN	TP

Fonte: adaptado de [10].

O verdadeiro negativo (TN) ocorre quando uma amostra da classe 0 é corretamente classificada como 0, enquanto o falso positivo (FP) ocorre quando uma amostra da classe 0 é incorretamente classificada como 1. O falso negativo (FN) representa uma amostra da classe 1 classificada incorretamente como 0, e o verdadeiro positivo (TP) corresponde a uma amostra da classe 1 corretamente classificada como 1 [10].

A partir dos valores na matriz de confusão, podem ser calculadas as métricas que auxiliam na avaliação do desempenho do modelo, como acurácia, recall e F1-score [10]. A acurácia corresponde à proporção de amostras classificadas corretamente em relação ao total de amostras avaliadas, sendo calculada por:

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precisão indica a proporção de verdadeiros positivos entre todas as amostras classificadas pelo modelo como pertencentes à classe positiva:

$$Precisão = \frac{TP}{TP + FP} \quad (2)$$

Recall representa a proporção de amostras positivas que foram corretamente identificadas pelo modelo:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Por fim, o F1-score corresponde à média harmônica entre precisão e recall, servindo para avaliar o equilíbrio entre as duas métricas.

$$F1 - Score = 2 \cdot \frac{Precisão \cdot Recall}{Precisão + Recall} \quad (4)$$

Para este trabalho, a classe positiva corresponde aos postes em mau estado de conservação, enquanto a classe negativa representa os postes em bom estado. Entre as métricas analisadas, o *recall* da classe positiva assume maior relevância, pois indica a proporção de postes em mau estado corretamente identificados pelo modelo. Nesse contexto, os falsos negativos são mais críticos que os falsos positivos, uma vez que um poste

deteriorado classificado incorretamente como bom pode deixar de ser priorizado para inspeção e manutenção. Em contrapartida, um poste em bom estado classificado como ruim tende a resultar apenas em uma inspeção adicional.

4. METODOLOGIA

Para o desenvolvimento foi realizada a coleta de imagens de postes em diferentes estados de conservação, classificando-os em duas categorias: Bom e Ruim. A classe Bom corresponde a postes que não apresentam rachaduras ou danos estruturais visíveis, enquanto a classe Ruim inclui postes com fissuras ou rachaduras que indicam necessidade de manutenção, conforme documentação técnica de inspeção [2].

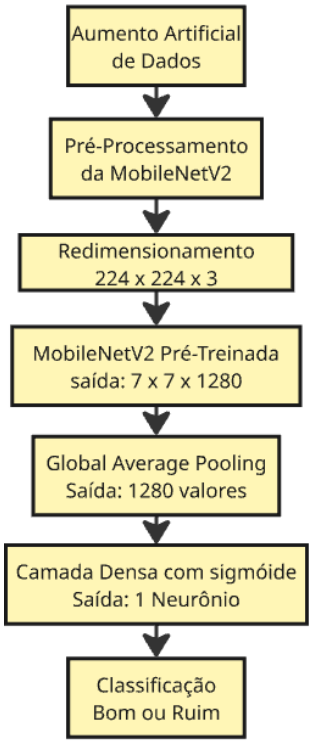
Ao todo, foram utilizadas 150 imagens obtidas por meio da plataforma Google Maps, sendo 75 classificadas pelo autor como Bom, por não apresentarem danos estruturais visíveis, e 75 como Ruim, por apresentarem sinais de deterioração, como fissuras, rachaduras ou exposição de armaduras. Para o treinamento e a avaliação do modelo, o conjunto de dados foi dividido em 110 imagens para treinamento, 20 para validação e 20 para teste. Cada subconjunto foi mantido balanceado entre as duas classes: o conjunto de treinamento foi composto por 55 imagens de cada classe, enquanto os conjuntos de validação e teste foram compostos por 10 imagens de cada classe.

Na etapa de pré-processamento, todas as imagens foram redimensionadas para 224 × 224 pixels, de modo a manter compatibilidade com o formato de entrada utilizado pela MobileNetV2 [16]. Em seguida, foi aplicada a função *preprocess_input*, específica dessa arquitetura, que transforma os valores dos pixels para o intervalo entre −1 e 1, adequando as imagens ao padrão de entrada esperado pelo modelo pré-treinado com pesos da base ImageNet [19].

Devido à quantidade reduzida de imagens disponíveis, foi utilizada a técnica de aumento de dados (*data augmentation*) apenas no conjunto de treinamento aumentando a diversidade do conjunto e reduzindo o risco de sobreajuste (*overfitting*) conforme [6] e [9]. Foram aplicadas rotação de até 20°, variação de zoom de até 20% e inversão horizontal. Esses valores foram limitados para evitar alterações muito grandes nas imagens, como postes excessivamente inclinados ou a retirada de partes importantes da estrutura. Mesmo assim, essas transformações podem alterar parcialmente o enquadramento original das imagens.

Para a construção do modelo, foi utilizada a arquitetura MobileNetV2, escolhida por sua eficiência computacional e adequação a aplicações em dispositivos com recursos limitados, de acordo com [6] e [10]. O modelo foi inicializado com pesos pré-treinados na base ImageNet, caracterizando o uso de *transfer learning* [6]. Nesse contexto, as camadas convolucionais da rede foram mantidas congeladas, ou seja, seus pesos não foram atualizados durante o treinamento, preservando características previamente aprendidas. A Figura 11 apresenta o fluxo da metodologia utilizada, incluindo as etapas aumento artificial de dados, pré-processamento, extração de características com a MobileNetV2 e classificação final das imagens.

Figura 11 - Estrutura da rede utilizada.



Fonte: autor.

Na Figura 11, as imagens entram na rede com dimensões de (224x224x3), sendo 224 pixels de altura, 224 pixels de largura e três canais de cor referentes ao padrão RGB. Na etapa de pré-processamento, os valores dos pixels são transformados para o intervalo entre −1 e 1. Então, as imagens são processadas pelas 154 camadas internas da arquitetura MobileNetV2, carregadas com pesos previamente treinados na base ImageNet. Essas camadas atuam na extração das características presentes nas imagens. Por fim, foram adicionadas duas camadas ao modelo: a camada *Global Average Pooling*, a qual reduz os mapas de características produzidos pela MobileNetV2 a um vetor com 1280 valores, e a camada densa, composta por um neurônio com função de ativação sigmóide, responsável pela classificação final entre as classes Bom e Ruim.

Diferentemente do pooling local, que opera sobre pequenas regiões da imagem, o *Global Average Pooling* calcula a média de todos os valores presentes em cada mapa de características, produzindo uma representação global das informações extraídas pela rede. Segundo [18], essa abordagem permite capturar características discriminativas independentemente de sua posição na imagem, além de reduzir o número de parâmetros do modelo e contribuir para a redução do sobreajuste.

A função sigmóide foi adotada por fornecer uma saída no intervalo entre 0 e 1, sendo adequada para problemas de classificação binária [14].

O modelo foi compilado utilizando o otimizador Adam, a função de perda *binary crossentropy* [17] e a métrica de acurácia. O otimizador Adam foi escolhido por sua capacidade de ajustar automaticamente a taxa de aprendizado durante o treinamento, contribuindo para uma convergência mais eficiente

do modelo conforme [6] e [10]. O treinamento foi configurado para um máximo de 30 épocas, utilizando interrupção antecipada com monitoramento da perda de validação, paciência de três épocas e restauração dos pesos correspondentes ao melhor resultado.

Após o treinamento e a avaliação do modelo com as imagens originais, foi realizado um segundo experimento com ajuste manual de enquadramento, buscando reduzir a presença de paisagens e objetos de fundo e manter o poste como elemento predominante da imagem. Foi adotado, como critério operacional para este conjunto de dados, um recorte lateral que resultasse em largura aproximada de 230 pixels (uma vez que para a MobileNetV2 as imagens deve ter 224 x 224), logo, nas imagens que apresentavam enquadramento distante e largura superior a 230 pixels, foi aplicado um recorte lateral, resultando em largura aproximada de 230 pixels, enquanto a altura original foi preservada. As imagens que já apresentavam enquadramento aproximado ou largura igual ou inferior a esse valor foram mantidas sem alterações adicionais, evitando recortes excessivos que destacam apenas uma seção do poste. Quando necessário, o enquadramento foi ajustado para manter o poste principal aproximadamente centralizado. O procedimento foi realizado por meio do editor de imagens nativo do Windows e aplicado segundo o mesmo critério às duas classes. Posteriormente, todas as imagens foram redimensionadas para 224 x 224 pixels antes de serem fornecidas ao modelo.

O modelo foi avaliado no conjunto de testes, formado por imagens não utilizadas no treinamento ou na validação. Foram calculadas a matriz de confusão, a acurácia e as métricas de precisão, recall e F1-score para a classe Ruim. Para a classificação, saídas iguais ou superiores a 0,5 foram atribuídas à classe Ruim, enquanto valores inferiores foram atribuídos à classe Bom. Esse limiar foi mantido nos experimentos para padronizar a avaliação dos modelos. Em uma aplicação prática, entretanto, poderia ser adotado um limiar menor, definido a partir de testes no conjunto de validação, com o objetivo de aumentar o *recall* e reduzir a quantidade de postes em mau estado classificados incorretamente como bons. Essa escolha aumentaria o número de postes em bom estado encaminhados para inspeção, mas reduziria o risco de postes deteriorados passarem despercebidos.

A Figura 12 apresenta a imagem de um poste em mau estado sem o recorte.

Figura 12 - Poste em mau estado.

Fonte: [20]

A Figura 13 apresenta o recorte da Figura 12, removendo características do fundo, as quais podem atrapalhar a rede.

Figura 13 - Poste em mau estado com recorte.

Fonte: adaptado de [20].

Para fins de comparação, foram realizados quatro experimentos: dois utilizando imagens sem recorte e dois utilizando imagens recortadas. O conjunto de dados foi composto por 150 imagens, sendo 75 de postes em bom estado e 75 de postes em mau estado, numeradas de 1 a 75 em cada classe.

Na primeira distribuição, as imagens numeradas de 1 a 55 foram destinadas ao treinamento, as imagens de 56 a 65 à validação e as imagens de 66 a 75 ao teste. Na segunda distribuição, as imagens de 21 a 75 foram utilizadas no treinamento, enquanto as imagens restantes, numeradas de 1 a 20, foram distribuídas aleatoriamente e em igual quantidade entre os conjuntos de validação e teste. As mesmas divisões foram adotadas para ambas as classes e para os experimentos com e sem recorte, permitindo a comparação dos resultados.

Os experimentos foram implementados no ambiente Google Colab, utilizando TensorFlow e Keras.

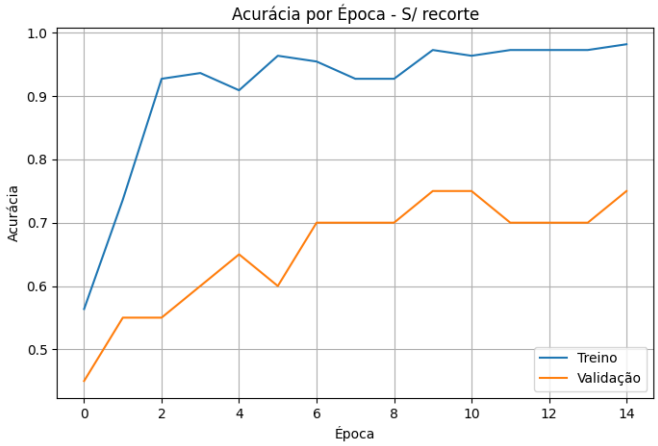
5. RESULTADOS E DISCUSSÕES

Para validar o desempenho da rede, foram feitos os gráficos de acurácia por época e perda por época para os dois casos (dados brutos e dados com recorte), assim como, a matriz de confusão da validação e do teste para determinar se a rede é apta ou não para classificar postes em bom ou mau estado.

5.1 Treinamento da Rede Sem Recorte das Imagens

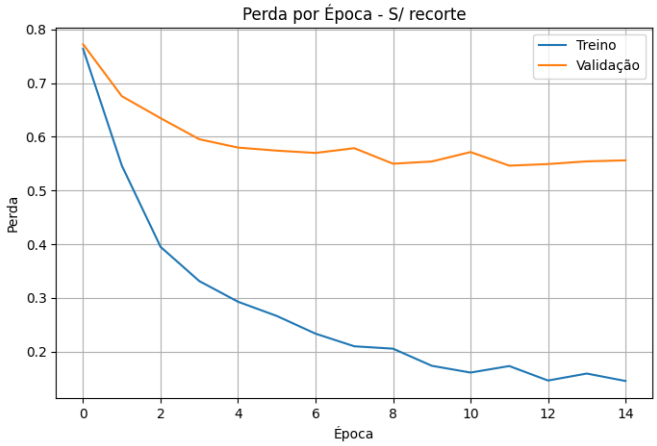
A Figura 14 e 15 apresentam o resultado obtido para o caso em que o treinamento foi feito sem modificação dos dados.

Figura 14 – Evolução da acurácia sem recorte.



Fonte: autor.

Figura 15 – Evolução da perda sem recorte.



Fonte: autor.

A Figura 14 mostra que a curva de acurácia do conjunto de treinamento cresceu ao longo das épocas, atingindo valor próximo de 98%. Porém, a acurácia do conjunto de validação apresentou oscilações entre aproximadamente 45% e 75%, sem acompanhar o crescimento observado no treinamento.

Com relação à Figura 15, a perda do conjunto de treinamento diminuiu continuamente, passando de aproximadamente 0,77 para 0,14. Por outro lado, a perda de validação apresentou redução nas primeiras épocas e, posteriormente, passou a oscilar em torno de 0,55, sem acompanhar a redução contínua da perda de treinamento.

A diferença observada entre as curvas de treinamento e validação indica que o modelo aprendeu as características das imagens utilizadas no treinamento, porém apresentou dificuldade para generalizar esse aprendizado para as imagens de validação. Esse comportamento indica a ocorrência de sobreajuste (*overfitting*), situação em que o modelo se adapta excessivamente ao conjunto de treinamento e apresenta desempenho inferior em imagens não utilizadas nessa etapa.

A matriz de confusão do conjunto de teste para o experimento sem recorte, apresentada na Tabela 2, mostra que, entre as dez imagens de postes em bom estado, sete foram classificadas corretamente e três foram classificadas incorretamente como pertencentes à classe Ruim. Para as imagens de postes em mau estado, todas as dez foram identificadas corretamente, não ocorrendo falsos negativos.

Tabela 2 - Matriz de confusão do conjunto de teste para o experimento sem recorte.

Classe Real	Previsto Bom	Previsto Ruim
Bom	7	3
Ruim	0	10

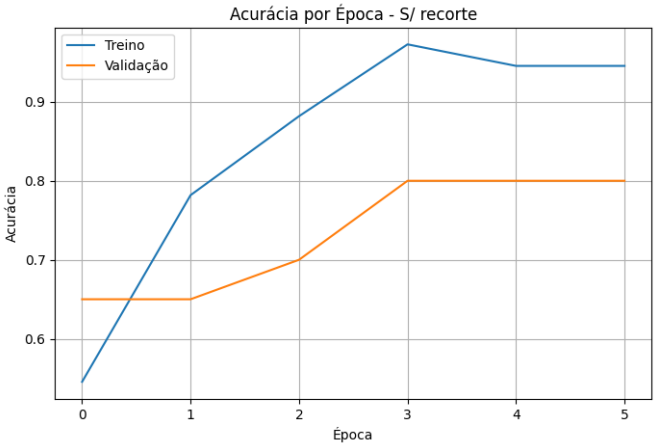
Fonte: autor.

Com isso, o modelo atingiu acurácia de 85% no conjunto de teste. Para a classe Ruim, foram obtidos precisão de 76,92%, recall de 100% e F1-score de 86,96%. O recall de 100% indica que todos os postes em mau estado presentes no conjunto de teste foram identificados pelo modelo, sendo um caso ideal em uma aplicação prática, uma vez que postes em mau estado não devem passar despercebidos devido ao perigo que apresentam. Entretanto, a precisão de 76,92% demonstra que três imagens de postes em bom estado foram classificadas incorretamente como ruins

Esses resultados devem ser interpretados com cautela, pois a acurácia obtida no conjunto de validação foi de 70%, inferior aos 85% observados no teste.

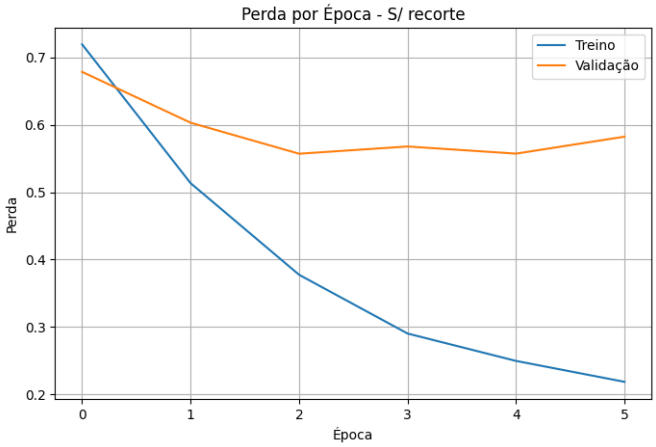
Após o primeiro teste para o caso sem recorte, as imagens foram distribuídas, como mencionado na metodologia, entre o treino, validação e teste para gerar um segundo teste e comparar o desempenho da rede. As Figuras 16 e 17 apresentam os resultados obtidos para a segunda distribuição para o caso sem recorte

Figura 16 – Evolução da acurácia com nova distribuição das imagens para o caso sem recorte.



Fonte: autor.

Figura 17 – Evolução da perda com nova distribuição das imagens para o caso sem recorte.



Fonte: autor.

A Figura 16 mostra que a curva do treinamento cresceu ao longo das épocas chegando a uma acurácia máxima de 97%, enquanto a acurácia da validação estabilizou em 80%. Enquanto na Figura 17 a perda do treinamento chegou a um mínimo de 0,21, enquanto a perda na validação oscilou entre 0,55 e 0,60. Novamente indicando que a rede está com dificuldade para generalizar.

Pela Tabela 3 do conjunto de teste, o modelo alcançou acurácia de 75%, correspondendo a 15 classificações corretas entre as 20 imagens avaliadas. A matriz indica que para classificação correta: oito postes foram classificados em bom

estado e 7 em mau estado. Para a classe Ruim, o modelo apresentou precisão de 77,78%, recall de 70% e F1-Score de 73,68%. Além disso, três postes em mau estado foram classificados incorretamente.

Tabela 3 - Matriz de confusão do conjunto de teste para o experimento sem recorte com nova distribuição das imagens.

Classe Real	Previsto Bom	Previsto Ruim
Bom	8	2
Ruim	3	7

Fonte: autor.

Novamente há uma discrepância entre a acurácia da validação e do conjunto de teste, sendo 70% a acurácia do conjunto de validação e 75% para o conjunto de teste. Além disso, como os conjuntos de validação e teste possuem apenas 20 imagens cada, uma única classificação corresponde a cinco pontos percentuais de acurácia, fazendo com que pequenas diferenças no número de acertos provoquem variações consideráveis nas métricas.

Nesta nova distribuição há a queda do *recall* de 100% para o primeiro teste para 70%, este é um caso que deve ser evitado, pois é melhor que postes em bom estado sejam classificados como mau estado do que o contrário.

5.2 Comparação conjunto de teste para o caso sem recorte

A seguir é apresentado a Tabela 4 com os dados obtidos para ambos os testes para o caso sem recorte.

Tabela 4 - Comparação dos testes para o caso sem recorte.

Métrica	Teste 1	Teste 2
Acurácia	85%	75%
Precisão Classe Ruim	76,92%	77,78%
Recall Classe Ruim	100%	70%
F1-Score Classe Ruim	86,96%	73,68%

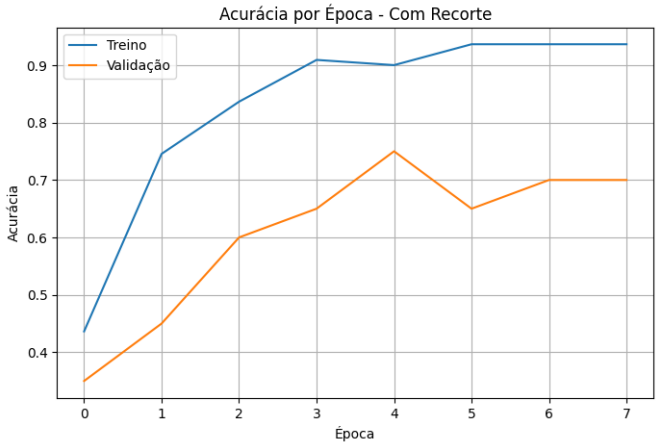
Fonte: autor.

A comparação entre os dois experimentos sem recorte demonstra que a distribuição das imagens entre os conjuntos influenciou o desempenho do modelo. No primeiro teste, foi obtida acurácia de 85%, enquanto o segundo apresentou 75%. A precisão da classe Ruim foi semelhante nos dois casos, com 76,92% no primeiro teste e 77,78% no segundo. Entretanto, o recall diminuiu de 100% para 70%, e o F1-score passou de 86,96% para 73,68%. Dessa forma, o primeiro teste apresentou melhor desempenho geral, principalmente na identificação dos postes em mau estado, pois classificou corretamente todas as imagens dessa classe, enquanto o segundo deixou de identificar três dos dez postes ruins. Esses resultados indicam que, devido ao conjunto reduzido de imagens, o modelo apresentou sensibilidade à distribuição dos dados entre treinamento, validação e teste.

5.3 Treinamento da Rede Com Recorte das Imagens

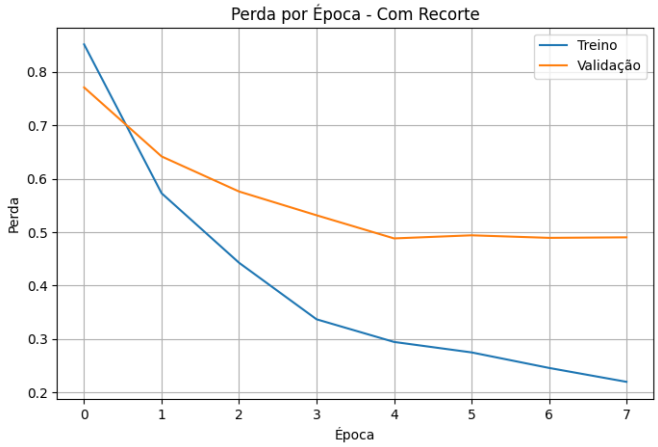
A Figura 18 e 19 apresentam o resultado obtido para o caso em que o treinamento foi feito com recorte das imagens na região do poste, a fim de evitar influência da paisagem e objetos de fundo.

Figura 18 – Evolução da acurácia com recorte.



Fonte: autor.

Figura 19 – Evolução da perda com recorte.



Fonte: autor.

Na Figura 18 a acurácia de treinamento aumentou rapidamente nas primeiras épocas e atingiu valores próximos de 94%. A acurácia de validação também apresentou crescimento inicial, alcançando 75%, mas passou a oscilar entre 65% e 70%, sem acompanhar a curva de treinamento

Pela Figura 19, a perda de treinamento atingiu valor próximo de 0,22 na última época, enquanto que a validação estabilizou em torno de 0,49.

Ainda existe diferença entre as curvas de treinamento e validação, indicando que o modelo apresentou dificuldade para generalizar completamente o aprendizado para as imagens de validação. Entretanto, quando comparado ao experimento sem recorte, houve melhora nos resultados de validação, indicando que a remoção de parte da paisagem e dos objetos de fundo contribuiu para que o modelo concentrasse sua análise na região do poste. Mesmo assim, ainda foram observados indícios de sobreajuste.

Tabela 5 - Matriz de confusão do conjunto de teste para o experimento com recorte.

Classe Real	Previsto Bom	Previsto Ruim
Bom	9	1
Ruim	0	10

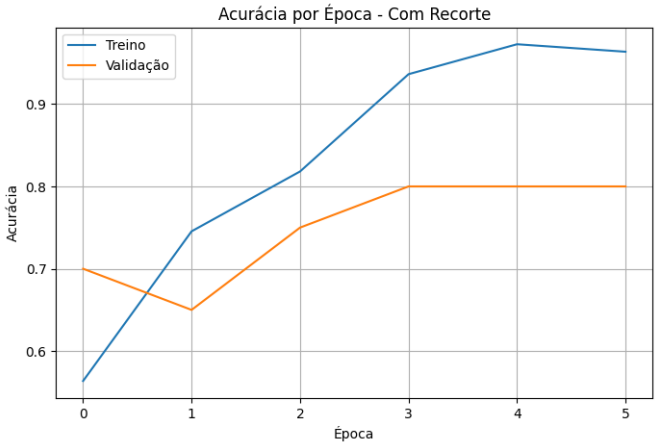
Fonte: autor.

A Tabela 5, referente ao conjunto de teste com recorte, mostra que, entre as dez imagens de postes em bom estado, nove foram classificadas corretamente e uma foi classificada como pertencente à classe Ruim. Para as imagens de postes em mau estado, todas as dez foram identificadas corretamente, sem ocorrência de falsos negativos.

Com isso, o modelo alcançou acurácia de 95% no conjunto de teste. Para a classe Ruim, foram obtidos precisão de 90,91%, recall de 100% e F1-score de 95,24%. O recall de 100% indica que todos os postes em mau estado presentes no conjunto de teste foram identificados corretamente. Entretanto, a precisão de 90,91% mostra que uma imagem de poste em bom estado foi classificada incorretamente como Ruim, caracterizando um falso positivo.

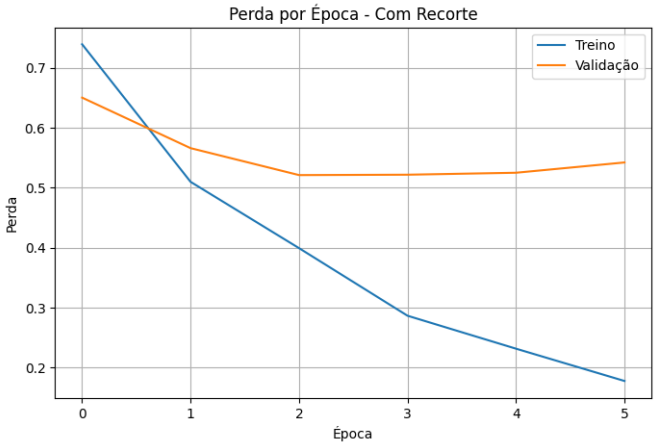
Para o segundo teste do caso com recorte, a distribuição das imagens foi feita conforme informado na seção de metodologia, e a seguir nas Figura 20 e 21 é apresentado os gráficos de acurácia e perda por época.

Figura 20 – Evolução da acurácia com a nova distribuição das imagens para o caso com recorte.



Fonte: autor.

Figura 21 – Evolução da perda com nova distribuição das imagens para o caso com recorte.



Fonte: autor.

Pela Figura 20 o treinamento evoluiu com o passar das épocas, atingindo um máximo de 97,27% de acurácia, enquanto a acurácia da validação estabilizou em 80%.

Na Figura 21 a perda cai constantemente até um mínimo de 0,17, mas a validação oscilou entre 0,52 e 0,65.

A Tabela 6 apresenta a matriz de confusão do segundo experimento para o caso com recorte.

Tabela 6 - Matriz de confusão do conjunto de teste para o experimento com recorte com nova distribuição das imagens.

Classe Real	Previsto Bom	Previsto Ruim
Bom	9	1
Ruim	2	8

Fonte: autor.

No segundo experimento com imagens recortadas e nova distribuição dos dados, o modelo teve 85% de acurácia, logo, 17 classificações corretas dentre as 20 imagens avaliadas. A matriz de confusão indica que nove postes em bom estado e oito postes em mau estado foram classificados corretamente. Entre os erros, um poste em bom estado foi classificado como ruim, enquanto dois postes em mau estado foram classificados como bons. Para a classe Ruim, o modelo apresentou precisão de 88,89%, recall de 80% e F1-score de 84,21%. Esses resultados indicam que, embora o modelo tenha apresentado bom desempenho geral, ainda ocorreram casos em que postes em mau estado não foram identificados corretamente.

5.4 Comparação conjunto de teste para o caso com recorte

A Tabela 7 a seguir apresenta os dados obtidos para ambos os testes para o caso com recorte.

Tabela 7 - Comparação dos testes para o caso com recorte.

Métrica	Teste 1	Teste 2
Acurácia	95%	85%
Precisão Classe Ruim	90,91%	88,89%
Recall Classe Ruim	100%	80%
F1-Score Classe Ruim	95,24%	84,21%

Fonte: autor.

A comparação entre os dois experimentos com recorte mostrou que a nova distribuição das imagens também influenciou o desempenho do modelo, tendo acurácia de 95% para o primeiro teste, enquanto teve 85% para o segundo teste. A precisão da classe Ruim apresentou pequena diferença de 2,02%. Entretanto o recall diminuiu de 100% para 80% e o F1-score passou de 95,24% para 84,21%. Dessa forma, o primeiro teste apresentou melhor desempenho geral, pois classificou corretamente todos os postes em mau estado, enquanto o segundo deixou de identificar dois dos dez postes ruins. Esses resultados reforçam que, devido ao número reduzido de imagens, o desempenho do modelo é sensível à distribuição das amostras.

Considerando os resultados dos quatro experimentos, o uso de imagens recortadas apresentou melhor desempenho em relação às imagens sem recorte. No teste 1 para a primeira distribuição, a acurácia aumentou de 85% (sem recorte) para 95% (com recorte), enquanto para teste 2 para a segunda distribuição, aumentou de 75% (sem recorte) para 85% (com recorte). Também havendo maiores valores de precisão, recall e F1-score para a classe Ruim nos experimentos com recorte.

Estes resultados indicam que a redução dos elementos de fundo permitiu ao modelo concentrar-se nas características visuais dos postes. Contudo, a diferença obtida entre as duas distribuições mostra que o modelo é sensível à escolha das imagens utilizadas no treinamento, validação e teste, possivelmente em razão do conjunto reduzido de dados. Portanto, o recorte contribuiu para melhorar a classificação, mas os resultados ainda não comprovam robustez para aplicação autônoma em situações reais.

6. CONCLUSÃO

Os resultados mostram que a MobileNetV2 foi capaz de extrair características úteis para diferenciar postes em bom e mau estado na maioria das imagens analisadas. Em ambas as distribuições avaliadas, o uso de imagens recortadas apresentou desempenho superior ao uso das imagens sem recorte. Na primeira distribuição, a acurácia aumentou de 85% para 95%, enquanto, na segunda, passou de 75% para 85%. Esses resultados indicam que a redução de elementos externos ao poste contribuiu para a classificação. Contudo, o modelo ainda não apresentou confiabilidade suficiente para realizar a classificação de forma autônoma em uma aplicação real, devido à ocorrência de erros, aos indícios de sobreajuste e à variação dos resultados entre as duas distribuições.

Entre as principais limitações do trabalho estão o tamanho reduzido do conjunto de dados e a pequena quantidade de imagens nos conjuntos de validação e teste, compostos por apenas 20 imagens cada. Nessas condições, cada classificação incorreta representa uma variação de cinco pontos percentuais na acurácia. A redução de dez pontos percentuais observada entre a primeira e a segunda distribuição, tanto nas imagens sem recorte quanto nas imagens recortadas, demonstra que o desempenho do modelo é sensível às imagens utilizadas em cada subconjunto.

Como trabalhos futuros, recomenda-se ampliar o banco de imagens, especialmente com diferentes tipos de danos, condições de iluminação, ângulos e distâncias de captura. Também podem ser investigadas técnicas mais refinadas de segmentação e pré-processamento, outras arquiteturas de redes neurais convolucionais e estratégias de treinamento voltadas à redução do sobreajuste. Com esses aprimoramentos, a abordagem poderá apresentar maior robustez e atuar como ferramenta de apoio à triagem e à priorização de inspeções em postes de distribuição.

7. REFERÊNCIAS

[1] G. N. da Silva, G. M. Locosselli, e L. S. Freitas, “Expansão das redes de distribuição elétrica subterrânea como instrumento de mitigação de impactos socioambientais no centro expandido de São Paulo” Instituto de Estudos Avançados da Universidade de São Paulo (IEA-USP), Mar. 2021. [Online]. Disponível em: https://www.iea.usp.br/pesquisa/projetos-institucionais/usp-cidades-globais/artigos-digitais/impacto-social-de-programas-de-pos-graduacao-no-brasil-com-interfaces-na-area-de-sustentabilidade-urbana

[2] ENEL Distribuição São Paulo, “Supervisão de Obras de Construção e Manutenção de Redes de Distribuição Aérea e de Iluminação Pública,” ID 4.038, ago. 2013.

[3] M. S. Urti, “O compartilhamento de postes de luz pelas distribuidoras de energia elétrica e as prestadoras de serviços de telecomunicações,” 2021.

[4] G. G. Borges, “Sistema de apoio à fiscalização das concessionárias de distribuição de energia elétrica,” Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, São Paulo, Brasil, 2005.

[5] O’SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. Disponível em: https://doi.org/10.48550/arXiv.1511.08458. Acesso em: 11 abr 2026.

[6] F. Yang, “Enhancing concrete crack image detection using MobileNetV2 and transfer learning,” Frontiers in Science and Engineering, vol. 4, no. 4, 2024.

[7] M. Flah, A. R. Suleiman, and M. L. Nehdi, “Classification and quantification of cracks in concrete structures using deep learning image-based techniques,” Cement and Concrete Composites.

[8] W. R. L. da Silva and D. S. de Lucena, “Concrete cracks detection based on deep learning image classification,” in Proceedings of the 18th International Conference on Experimental Mechanics (ICEM18), Brussels, Belgium, Jul. 2018.

[9] R.-S. Rajadurai and S.-T. Kang, “Automated vision-based crack detection on concrete surfaces using deep learning,” Applied Sciences, vol. 11, no. 11, p. 5229, 2021. doi: 10.3390/app11115229.

[10] F. Wu, “Method and research of concrete pavement crack detection based on convolution neural network,” Highlights

in Science, Engineering and Technology, vol. 56, p. 200, 2023.

[11] FLECK, Leandro; TAVARES, Maria Hermínia Ferreira; EYNG, Eduardo; HELMANN, Andrieli Cristina; ANDRADE, Minéia Aparecida de Moraes. Redes neurais artificiais: princípios básicos. Revista Eletrônica Científica Inovação e Tecnologia, Medianeira, v. 1, n. 13, p. 47–57, jan./jun. 2016.

[12] T. W. Rauber, Redes neurais artificiais. Vitória, ES, Brasil: Universidade Federal do Espírito Santo, Departamento de Informática, [n.d.].

[13] K. Dong, Y. Ruan, C. Zhou, and Y. Li, "Model for Image Classification," em 2020 2nd International Conference on Information Technology and Computer Application (ITCA), 2020.

[14] S. Sharma, S. Sharma and A. Athaiya, "Activation Functions in Neural Networks," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 12, pp. 310–316, 2020.

[15] Q. Xiang, G. Zhang, X. Wang, J. Lai, R. Li and Q. Hu," Fruit Image Classification Based on MobileNetV2 with Transfer Learning Technique," in Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (CSAE), Sanya, China, Oct. 2019, doi: 10.1145/3331453.3361658.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov e L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[17] R. Yu, Y. Wang, Z. Zou e L. Wang, "Convolutional neural networks with refined loss functions for the real-time crash risk analysis," Transportation Research Part C: Emerging Technologies, vol. 119, 2020, Art. no. 102740.

[18] A. Zafar et al., "A Comparison of Pooling Methods for Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 17, art. 8643, 2022, doi: 10.3390/app12178643.

[19] TENSORFLOW, "tf.keras.applications.mobilenet_v2.preprocess_input," TensorFlow API Documentation, 2024. [Online]. Disponível em: https://www.tensorflow.org/api_docs/python/tf/keras/applications/mobilenet_v2/preprocess_input. Acesso em: 20 jun. 2026.

[20] Portal Grande Tijuca, "Poste deteriorado tem risco de queda em frente ao Detran de Vila Isabel," 19 set. 2023. [Online]. Disponível em: https://grandetijuca.com.br/noticia/6148/poste-deteriorado-tem-risco-de-queda-em-frente-ao-detran-de-vila-isabel.html. Acesso em: 21 jun. 2026.