



Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Programa de Pós-Graduação em Engenharia Elétrica



REDES NEURONAIS CONVOLUTIVAS APLICADAS À DETECÇÃO DE EQUIPAMENTOS DE PLANTAS ELÉTRICAS DE ALTA POTÊNCIA

PROJETO DE PESQUISA I

CHRISTIANE RAULINO ALMEIDA MOLINA

ORIENTADORES:

RAIMUNDO CARLOS SILVÉRIO FREIRE

JUGURTA ROSA MONTALVÃO FILHO

Campina Grande - PB, julho de 2022



REDES NEURONAIS CONVOLUTIVAS APLICADAS À DETECÇÃO DE EQUIPAMENTOS DE PLANTAS ELÉTRICAS DE ALTA POTÊNCIA

Christiane Raulino Almeida Molina

Relatório apresentado ao Programa de Pós-graduação em Engenharia Elétrica - PPgEE, da Universidade Federal de Campina Grande, como parte dos requisitos necessários de avaliação do componente Projeto e Pesquisa I.

Orientadores:

Raimundo Carlos Silvério Freire

Jugurta Rosa Montalvão Filho

Campina Grande - PB, julho de 2022

Sumário

1	INTRODUÇÃO	2
1.0.1	Objetivos	3
1.0.2	Objetivos Específicos:	3
1.1	Estrutura do Projeto de Pesquisa	4
2	REVISÃO BIBLIOGRÁFICA	4
3	FUNDAMENTAÇÃO TEÓRICA	6
3.1	Rede MLP	6
3.1.1	Perceptron	6
3.1.2	Arquitetura da MLP	7
3.2	Rede Convolutiva	8
3.3	Treinamento e Teste	10
4	ABORDAGEM PROPOSTA	11
5	RESULTADOS	14
6	CONCLUSÃO	17
	Referências	19

1 INTRODUÇÃO

Monitorar e realizar manutenções em equipamentos de alta potência é uma tarefa bastante desafiadora em ambientes de produção e distribuição de energia elétrica. A dificuldade desta tarefa pode ser relacionada a alguns fatores: perdas na produção, uma vez que o equipamento precisa ser desligado para uma intervenção; o alto risco de acidentes de trabalho para operadores que realizam a tarefa; além de possíveis falhas inerentes ao processo manual de inspeção. Desta maneira, métodos de monitoramento automatizados [1, 2] compõem uma solução com o mínimo de intervenção e recursos humanos no espaço físico operante, criando uma malha energética inteligente.

Partindo desse contexto, imagens adquiridas de plantas elétricas de alta potência através de *drones* ou de captura manual remota geram dados que, ao serem analisados, podem trazer um preciso diagnóstico de problemas em curso, assim como a antecipação de problemas futuros, como é o caso da inspeção e manutenção de isoladores [3, 4], indicando a necessidade de manutenções preventivas ou substituição de equipamentos com vida útil limítrofe, por exemplo. Com efeito, a utilização de técnicas de visão computacional podem fornecer alternativas aos sistemas de monitoramento baseados em sinais elétricos ou mesmo trabalhar em conjunto com eles na construção de um monitoramento mais robusto. No entanto, as imagens adquiridas a partir de coletas não estruturadas podem ser de difícil manipulação e análise, especialmente quando coletadas de diferentes perspectivas ou através de diferentes equipamentos, situações nas quais o processo de segmentação e extração de características robusto torna-se indispensável para a boa utilização dos dados.

A meta final que motivou o estudo inicial contido neste projeto de pesquisa consiste em perceber não conformidades de operação em equipamentos presentes na área de alta potência em uma termelétrica. Através da imagem capturada, um equipamento elétrico pode ser isolado por segmentação e classificado de acordo com suas características. Todavia, algumas dificuldades como diferentes resoluções e perspectivas das imagens capturadas, além da possibilidade de utilização de imagens capturadas em diferentes espectros (visível e infravermelho, por exemplo [5]) costumam dificultar a tarefa de reconhecimento dos equipamentos de interesse.

Dessa forma, é proposto nesse projeto o início dos estudos para viabilizar o desenvolvimento de sistemas de monitoramento automatizado de plantas elétricas de alta tensão. Definida como primeira tarefa do sistema de visão, será investigado o uso de redes neurais convolutivas aplicados à detecção de 9 equipamentos presentes na planta elétrica. De acordo com o que é largamente aceito hoje, no âmbito de visão computacional, esse tipo de rede é capaz de classificar imagens de forma robusta e, portanto, uma boa ferramenta para ser tomada como foco de estudo deste projeto de pesquisa.

Primeiramente, é executada uma revisão bibliográfica, tal que seja estabelecido

um embasamento teórico capaz de proporcionar um domínio do estado da arte no que diz respeito às redes neurais convolutivas aplicadas à visão computacional, bem como os fundamentos teóricos necessários para a implementação. Por conseguinte, utilizando uma base pública como entrada, as características das redes convolutivas mais significativas são implementadas computacionalmente, assim como a rede de perceptrons em múltiplas camadas (*Multilayer perceptrons - MLP*), que é tomada como linha de base no contexto de redes neurais artificiais (RNA). Desta maneira, a implementação destas arquiteturas garante um bom entendimento das ferramentas utilizadas.

Além disso, em paralelo ao estudo teórico e experimentações, acontece a coleta das imagens dos equipamentos monitorados, realizada pelos operadores de uma termelétrica instalada no estado de Sergipe. Os dados são coletados a partir de câmeras funcionando nos espectros infravermelho, ultravioleta e também no espectro visível (RGB). Em seguida, são realizados experimentos utilizando os dados disponíveis para a aplicação descrita anteriormente. Tal experimentação proporciona a verificação da viabilidade do uso de redes neurais convolutivas ao problema de interesse, provocando o surgimento de questionamentos quanto às suas limitações e de propostas de melhorias voltadas à aplicação.

1.0.1 Objetivos

O objetivo geral dessa disciplina é adquirir domínio e investigar as características de redes neuronais convolutivas, aplicadas a métodos de visão computacional, mais especificamente em detecção de objetos, para monitoramento automatizado e/ou semi-automatizado de plantas elétricas de alta potência.

1.0.2 Objetivos Específicos:

- Estudar os últimos 10 anos de redes convolutivas e suas arquiteturas/métodos mais utilizados para detecção de objetos;
- Implementar computacionalmente, usando apenas programação matemática (i.e. sem usar "toolboxes" ou plataformas voltadas às redes convolutivas), um rede baseada na LeNet5, proposta por LeCun[6];
- Realizar um estudo comparativo entre método de detecção baseado em redes convolutiva e rede MLP, ambas aplicadas a uma base pública de imagens;
- Aplicar os métodos, implementados no item anterior, ao problema de interesse usando uma base de imagens capturadas em uma usina termelétrica instalada no estado de Sergipe (o acesso a esses dados é realizado através de colaboração com o projeto GImpSI do INESC P&D Brasil, já em andamento);

1.1 Estrutura do Projeto de Pesquisa

A sequência deste relatório de pesquisa está estruturada da seguinte maneira:

Capítulo 2 - é apresentada uma revisão bibliográfica no tocante a ao uso de redes neurais convolutivas aplicadas à tarefas de visão computacional;

Capítulo 3 - uma fundamentação teórica é apresentada acerca das redes neurais: perceptrons de múltiplas camadas e redes com camadas convolutivas;

Capítulo 4 - descreve-se a abordagem proposta, através das etapas desenvolvidas ao longo do projeto, a fim de atingir os objetivos da disciplina;

Capítulo 5 - apresentam-se a descrição dos experimentos realizados, resultados e discussão dos mesmos;

Capítulo 6 - são discutidas as conclusões e expectativas para trabalhos futuros.

A seguir é apresentada uma revisão bibliográfica em que são abordadas características relevantes quanto ao uso de redes convolucionais e trabalhos relacionados.

2 REVISÃO BIBLIOGRÁFICA

A tarefa de detecção de objetos no campo de visão computacional vem sendo desenvolvida, nos últimos 10 anos, com aplicação de métodos baseados em Redes Neurais Convolutivas (*Convolutional Neural Networks* - CNN ou ConvNets).

A motivação biológica do uso de campos receptivos locais [7] fez surgir as primeiras arquiteturas baseadas em operações convolucionais com o intuito de alcançar classificações mais robustas. Na década de 80, Fukushima propôs a Neocognitron [8] com estrutura similar ao modelo do sistema nervoso de visão proposto por Hubel e Wiesel. Em seguida, já na década de 90, com o objetivo de reconhecer dígitos manuscritos, Lecun implementou com sucesso a LeNet [6], onde a imagem é processada por por camadas convolucionais, também fazendo uso da ideia de sensibilidade local do sistema de visão. Sendo assim, ao invés de tratar a imagem inteira para entrada da rede neural, como uma MLP faria, as CNNs processam a imagem através de filtros (também chamados de *kernels*) treinados para capturar diferentes características da imagem (e.g. bordas, formas geométricas e sombras).

LeCun, em seu trabalho[6], obteve resultados superiores aos métodos clássicos da sua época, fazendo com que seu trabalho fosse um marco importante no desenvolvimento das redes convolutivas. Em sua arquitetura, a LeNet definiu três básicas características que

fundamentaram a estrutura padrão de uma ConvNet, que independem da sua variante. São elas: camadas convolucionais, camadas de subamostragem (em inglês, *pooling*) e camada totalmente conectada (do inglês, *fully connected*).

Durante alguns anos, as redes neurais convolucionais não foram muito utilizadas. Todavia, mais de 10 anos após Lecun, devido ao aumento da capacidade de memória e processamento computacional, tornou-se possível desenvolver CNNs mais profundas, mais difíceis de treinar, e com base de dados de treinamento muito grande, na ordem de milhões de imagens (*BigData*). Em 2012, a AlexNet [9] foi desenvolvida com 60 milhões de parâmetros e 650.000 neurônios, utilizou ReLu como função de ativação (evitando o problema de desaparecimento de gradiente) e técnicas como *dropout* e *data enhancement* para evitar o problema de sobreajuste. Tais incrementos na arquitetura e um treinamento com uma grande quantidade de dados, capaz de capturar fortes características das imagens, garantiu a esta rede um excelente desempenho na competição *ImageNet*. E foi a partir desse resultado que as ConvNets atingiram popularidade no campo de aprendizado de máquina.

Após esses dois trabalhos [6][9] importantes, que sedimentaram o sucesso do uso de CNN para classificação de imagens, diversas arquiteturas [10][11][12] surgiram com o intuito de elevar a eficiência da ferramenta e reduzir seu custo computacional. Por exemplo, a VGGNet [10] apresentou bons resultados utilizando um *kernel* menor que o utilizado em [9], bem como investigou o uso de mais camadas para melhorar a performance da rede.

Outro conceito importante, incorporado à tarefa de reconhecimento de imagens, foi combinar tarefas tradicionais de visão com as arquiteturas das ConvNets, gerando uma série de algoritmos (o original [13] e suas variantes melhoradas [14][15][16][17]) que utilizam uma etapa prévia à camada convolutiva, baseada em busca seletiva. Esse incremento foi realizado com o intuito de encontrar regiões de interesse (*regions of interest* - ROI) em uma imagem e atender a tarefa de detecção de objetos, onde suas localizações precisam ser determinadas e que cada elemento, encontrado na região, possa ter suas características extraídas e ser classificado. De outro modo, apesar da boa eficiência alcançada por essa família de algoritmos, o alto tempo de treinamento e de teste impossibilitou seu uso em aplicações de tempo real [18], fazendo surgir outros tipos de métodos de detecção [19][20][21][22].

Como visto nos parágrafo anteriores há diversos tipos de ConvNets bem estabelecidas, e suas aplicações estão presentes em diversas áreas de visão computacional, como auxílio em diagnóstico por análise de imagens [23], e na detecção de equipamentos em uma planta elétrica [24][25][26][27], para citar alguns exemplos. Portanto, de acordo com as considerações científicas apresentadas, para se realizar trabalhos inovadores na área de reconhecimento robusto de imagens, faz-se necessário conhecer profundamente as

Rede Neurais Convolutivas.

3 FUNDAMENTAÇÃO TEÓRICA

A fim de estabelecer uma fundamentação teórica para entendimento dos experimentos realizados durante o projeto de pesquisa, nesta seção serão apresentados os conceitos relevantes para esse trabalho acerca das redes neurais perceptrons de múltipla camada (em inglês, *Multilayer perceptron* - MLP) e das redes convolutivas, respectivamente.

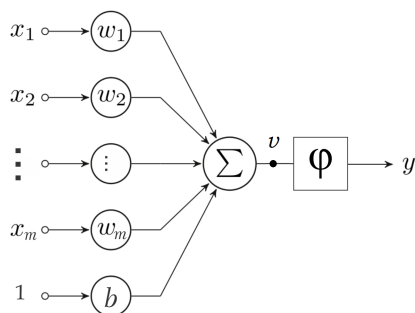
3.1 Rede MLP

Considerada um dos tipos de redes neurais mais clássico, a *Multilayer Perceptrons* é tipicamente formada por elementos básicos chamados perceptrons, e para um bom entendimento de seu funcionamento, primeiramente faz-se necessária a definição dessas unidades.

3.1.1 Perceptron

O perceptron é um modelo matemático simples, que realiza uma combinação entre a entrada e seus parâmetros (pesos sinápticos e viés), representada na figura 1.

Figura 1 – Perceptron



Em que a saída parcial v , a partir de uma entrada de dimensão m , pode ser escrita na forma:

$$v = \sum_{i=1}^m w_i x_i + b \quad (3.1)$$

Por uma perspectiva de classificação, esse neurônio matemático é capaz de definir um hiperplano ($\sum_{i=1}^m w_i x_i + b = 0$) como fronteira de decisão, separando o espaço m -dimensional (definido pelas variáveis da entrada) em em regiões.

Aplicada à saída parcial v , a função de ativação φ é utilizada para restringir a amplitude de saída y de um neurônio. Geralmente, as funções adotadas para restrição

normalizam a saída em intervalos fechados entre 0 e 1, ou entre -1 e 1. A saída do perceptron pode ser escrita, então, pela equação 3.2.

$$y = \varphi\left(\sum_{i=1}^m w_i x_i + b\right) \quad (3.2)$$

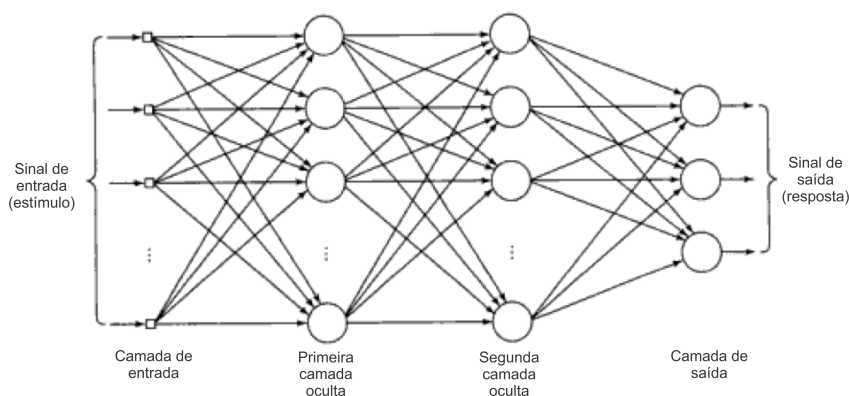
Há diversas funções de ativação capazes de normalizar a saída de um neurônio: sigmóide, tangente hiperbólica e unidade retificadora linear (ReLU), para citar alguns exemplos. Contudo, o melhor tipo de ativação utilizado no neurônio artificial depende do fenômeno a ser modelado. Em outras palavras, é escolhida a função de ativação que garante o melhor desempenho ao modelo. Por exemplo, é bastante comum utilizar funções ReLU como função de ativação para os perceptrons em imagens como dado de entrada.

Em resumo, o *perceptron* é a unidade básica de uma MLP e possui três parâmetros que o definem: vetor de pesos sinápticos (w_i , onde $i = 1, 2, \dots, m$), o viés (b) e o tipo da função de ativação (φ).

3.1.2 Arquitetura da MLP

Uma vez definida a unidade básica, uma MLP é uma rede com múltiplas camadas formadas por essas unidades, representadas por círculos na figura 2. Geralmente, sua arquitetura é composta por uma camada de entrada com dados a serem processados (sinal de estímulo), por camadas intermediárias (ou ocultas) e por uma camada de saída (resposta), que gera a informação útil de acordo com a tarefa a ser desenvolvida.

Figura 2 – Grafo arquitetural de um perceptron de múltiplas camadas com duas camadas ocultas [28]



No sentido da propagação da informação, os dados a serem processados são o sinal de entrada da primeira camada intermediária, e sua saída serve com entrada da segunda camada intermediária, seguindo o mesmo fluxo até chegar o final da rede. Logo, os neurônios de uma camada têm com entrada apenas os sinais de saída da camada anterior, estabelecendo um fluxo sequencial e caracterizando a rede de alimentação direta

(em inglês *feed forward*). O fluxo de operações para a arquitetura exibida em 2 está representado pelo algoritmo 1 em pseudocódigo.

Algorithm 1 Caminho direto da MLP de duas camadas intermediárias

```

 $X \leftarrow \text{Entrada}$ 
 $X \leftarrow [X \ 1]$  ▷ Inclusão do vies
 $A \leftarrow W_1 \cdot X$  ▷  $W_1$  Pesos sinápticos e vieses dos neurônios da primeira camada oculta
 $D \leftarrow \varphi(A)$ 
 $D \leftarrow [D \ 1]$  ▷ Inclusão do vies
 $B \leftarrow W_2 \cdot D$  ▷  $W_2$  Pesos sinápticos e vieses dos neurônios da segunda camada oculta
 $Y \leftarrow \varphi(B)$ 

```

Outro aspecto relevante observado na figura 2 é que cada neurônio de uma camada está conectado a todos os neurônios da camada anterior. Portanto, pode-se dizer que se trata de uma RNA totalmente conectada. Dessa maneira, se uma camada possui C_2 neurônios e está conectada a uma camada de entrada com C_1 , cada neurônio de C_2 possui $C_1 + 1$ (*vies*) graus de liberdades pra definir sua fronteira de classificação. Considerando todos os neurônios da camada, há $C_2 \cdot C_1 + C_2$ parâmetros livres para ajuste das fronteiras.

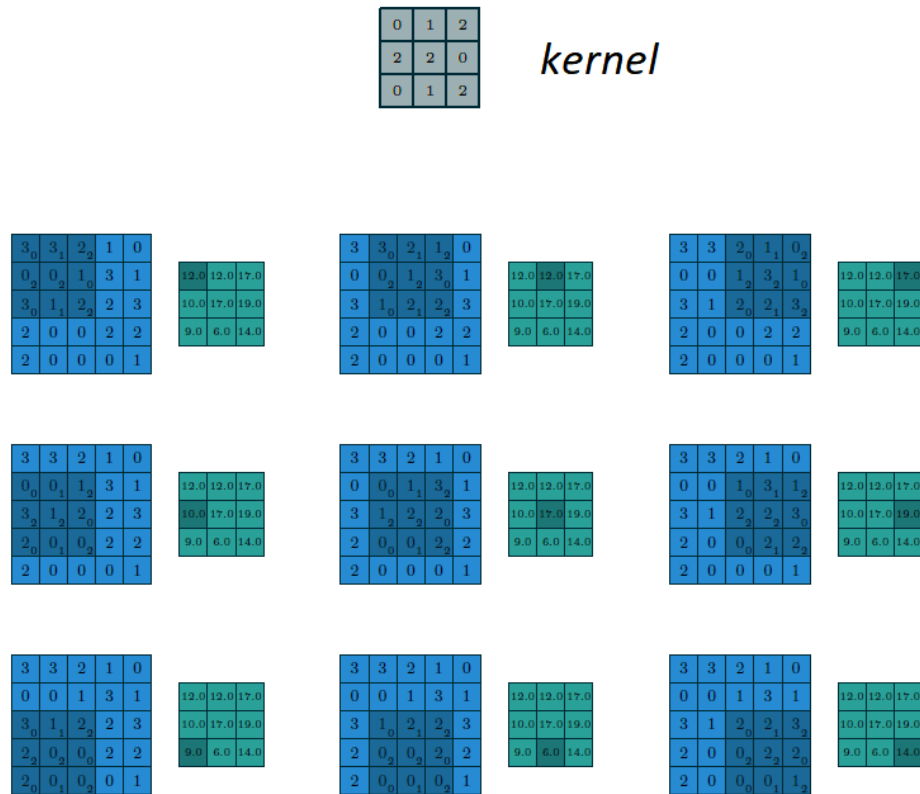
Com a arquitetura definida, o projeto da rede MLP, para um dado m -dimensional e uma saída de K classes, define primeiramente o número de camadas intermediárias (I) e o número de neurônios de cada camada (N_{C_i} , onde $i = 1, 2, \dots, I$). O número de camadas intermediárias e o número de neurônios de cada camada faz parte do projeto da rede, e são parâmetros a serem definidos para gerar um melhor desempenho na tarefa executada pela RNA. Em seguida os parâmetros livres de cada camada (pesos sinápticos e vieses dos neurônios) são ajustados, através de um algoritmo de aprendizado (treinamento), para cumprir o objetivo desejado (e.g, reconhecer um equipamento elétrico em uma imagem).

3.2 Rede Convolutiva

Redes convolutivas são um caso particular da MLP, e a principal particularidade desse tipo de rede está na substituição da camada totalmente conectada por uma camada de pesos compartilhados. A ideia de compartilhamento de pesos parte da motivação biológica em utilizar o conceito de campos receptivos locais [7], em que a camada convolutiva é capaz de reconhecer formas bidimensionais (estruturas) das imagens de forma invariante à translação, perspectiva e escala, assim como acontece no sistema de visão biológico.

A camada convolutiva é composta por um banco de filtros convolucionais (*kernels*), e para cada filtro é realizada a operação de convolução com a entrada 2-D, resultando em mapas de características como saída, onde cada filtro convolvido com a imagem gera um mapa. A convolução de um filtro com uma imagem de entrada é exibida na figura 3.

Figura 3 – Calculando os valores de saída da convolução discreta [29]



Para esta operação há 4 parâmetros a serem definidos:

- Número de filtros (NF)
- Dimensão dos filtros ($m \times n$)
- Avanço (S)
- Preenchimento com zero (P)

O número de filtros determina o número de mapas de características na saída da camada. Por exemplo, um camada convolucional com 3 filtros para uma entrada resulta em 3 mapas de características na saída. A dimensão do mapa de característica dependerá dos parâmetros adotados na camada e pode ser calculada por $(M - m)/S + 1$, para uma camada sem preenchimento com zero. O parâmetro S define o tamanho do avanço do filtro sobre a imagem e o preenchimento com zero amplia as bordas da imagem para que não haja perda de informação existente nas extremidades da entrada. Para a figura 3, o avanço utilizado foi $S = 1$ e não houve preenchimento com zero. Assim como na MLP, os

parâmetros listados acima também são definidos objetivando um melhor desempenho na tarefa executada pela RNA.

Duas importantes características desse tipo de camada podem ser evidenciadas: a redução do número dos parâmetros livres a serem ajustados e a extração robusta de formas bidimensionais da entrada. Onde a redução é devida ao compartilhamento dos pesos, um peso do filtro convolucional é compartilhado por diversos pixels da imagem de entrada. De forma visual na figura 3, o peso de posição $(1, 1)$ do *kernel* multiplica-se aos pixels da entrada de posições $(1, 1)$, $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 2)$, $(2, 3)$, $(3, 1)$, $(3, 2)$ e $(3, 3)$, diferentemente de uma MLP onde há um peso para cada entrada. Ademais, a extração robusta é obtida ao introduzir a operação de convolução com um pequeno kernel que ao deslizar sobre entrada extrai características invariantes a distorções.

Após uma camada convolutiva, uma rede convolutiva pode possuir uma camada de subamostragem, em que a dimensão do mapa de características é reduzida, com o efeito de reduzir a sensibilidade dos mapas a distorções na entrada. Como o objetivo desse projeto está focado na camada convolutiva, a subamostragem não é desenvolvida.

Uma vez definida o tipo de rede a ser utilizado, seja ela CNN ou MLP, a implementação da mesma é finalizada com duas últimas etapas: o treinamento e, logo em seguida, o teste.

3.3 Treinamento e Teste

Com o intuito de obter uma resposta classificando a entrada, o treinamento supervisionado consiste em apresentar à rede uma amostra de dados rotulados que representem o fenômeno real. É nesta etapa que ocorre o processo de aprendizagem e é importante ressaltar que o aprendizado torna-se mais eficiente à medida que é apresentado um conjunto numeroso de dados, que representa bem o universo do fenômeno. Uma boa maneira de garantir essa boa representação é apresentar bases de dados numerosas.

No contexto do processo de aprendizagem, há diversos métodos, sendo o método de aprendizagem por correção de erro mais comumente utilizado para treinar MLP e CNN, através do algoritmo de retropropagação de erro (em inglês, *error back-propagation*). Nesse caso, o erro é definido pela diferença entre a saída da rede e a classificação real (rótulo). A aprendizagem por retropropagação ocorre em duas etapas: o caminho para frente (algoritmo 1), onde os pesos sinápticos são fixos; e o caminho para trás, onde ocorre o ajuste dos pesos sinápticos de cada camada de acordo com uma regra de correção de erro. A regra mais utilizada consiste em definir uma função de custo composta do erro (J) e ajustar os parâmetros de maneira a minimizar essa função, através da descida do gradiente. Por exemplo, considerando o algoritmo 1, o ajuste de W_1 , pode ser descrito na equação 3.3.

$$W1(n+1) = W1(n) - \eta \frac{\partial J}{\partial W1} \quad (3.3)$$

Onde n é o n -ésimo ajuste e η é a taxa de aprendizado.

Como o J depende de todos parâmetros livres da rede, a derivada parcial $\frac{\partial J}{\partial w_{1ij}}$ pode ser determinada através da regra da cadeia (equação 3.4), evidenciando ideia de retropropagação do erro até a primeira camada.

$$\frac{\partial J}{\partial W1} = \frac{\partial J}{\partial erro} \frac{\partial erro}{\partial Y} \frac{\partial Y}{\partial B} \frac{\partial B}{\partial D} \frac{\partial D}{\partial A} \frac{\partial A}{\partial W1} \quad (3.4)$$

O ajuste dos parâmetros é realizado diversas vezes em múltiplas apresentações de um conjunto de dados de treinamento (totalizando uma época) e pode ser feito de modo sequencial, onde os parâmetros são atualizados a cada apresentação de uma amostra à rede. Por fim, finalizado o treinamento (ajuste dos parâmetros), na etapa de teste são apresentados dados que não foram usados previamente. A saída na rede é então comparada com o rótulo real da amostra, e caso os valores coincidam em boa parte dos testes, o aprendizado é validado, atestando a capacidade de generalização da RNA.

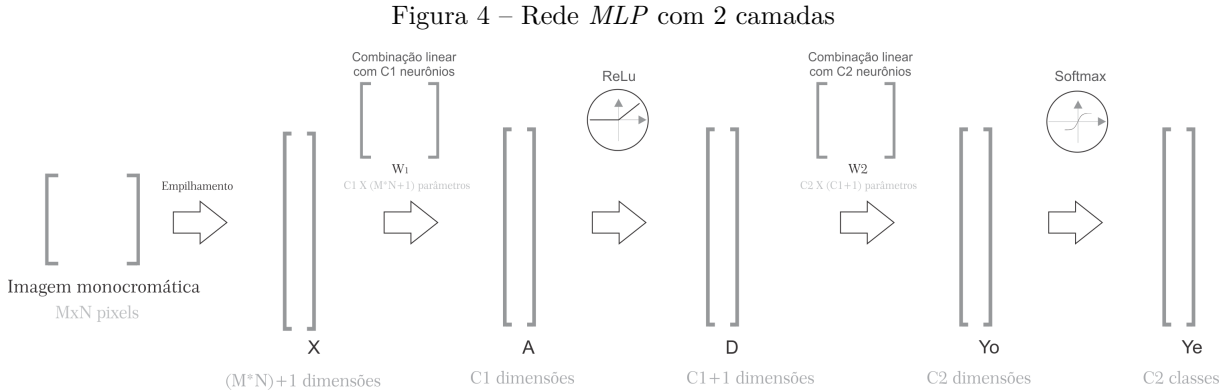
4 ABORDAGEM PROPOSTA

Com o intuito de atingir os objetivos desejados neste projeto, a ações de estudos e experimentos, executados durante o trabalho, são detalhados neste capítulo.

Inicialmente, foi executada uma revisão bibliográfica sobre métodos utilizados para segmentação e classificação de imagens de equipamentos em plantas elétricas, exposta no segundo capítulo deste relatório. A leitura realizada durante esta etapa deixou evidente o domínio das redes convolucionais, e de métodos derivados desta arquitetura, na área de visão computacional. Por conseguinte, apresentado no terceiro capítulo, realizou-se o estudo sobre as características das redes MLP e CNN.

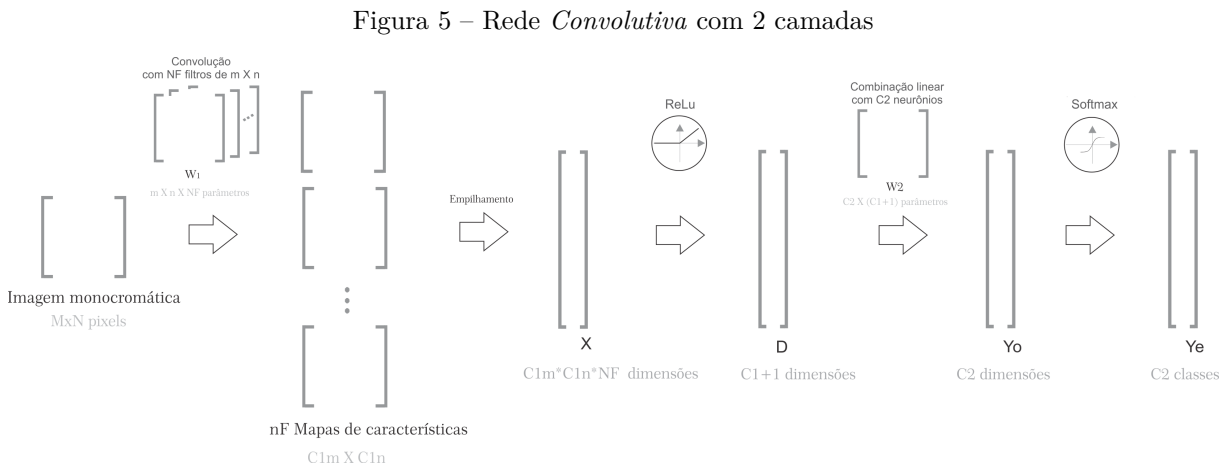
Após a etapa de estudo teórico, uma rede neural de perceptrons de múltiplas camadas e uma rede neural convolutiva foram implementadas. É importante ressaltar que, para fins didáticos, não foi utilizada nenhuma biblioteca de rede neural pública (como a *Keras*, por exemplo) durante o projeto das redes. Todas as operações matemáticas de caminho direto, bem como a retropropagação do erro para aprendizado, foram implementadas linha por linha. Consequentemente, um melhor conhecimento do efeito de cada camada da arquitetura das redes implementadas pode ser observado. A linguagem de programação escolhida foi *Matlab*, por ter uma sintaxe familiar à executora deste projeto, permitindo manter o foco voltado apenas ao funcionamento das redes e do mesmo modo que permitiu realizar o trabalho em tempo hábil.

As topologias implementadas para rede totalmente conectada e para a rede com camada convolutiva estão expostas na figura 4 e 5, respectivamente. E as operações realizadas no caminho direto dessas redes são representadas nos algoritmos 2 e 3, respectivamente.



Algorithm 2 Caminho direto da *MLP*

$X \leftarrow Empilhamento(Imagem)$
 $X \leftarrow [X \ 1]$ ▷ Inclusão do viés
 $A \leftarrow W_1 \cdot X$
 $D \leftarrow ReLu(A)$
 $D \leftarrow [D \ 1]$ ▷ Inclusão do viés
 $Y_o \leftarrow W_2 \cdot D$
 $Y_e \leftarrow Softmax(Y_o)$



A diferença existente entre as duas configurações que foram implementadas está na primeira camada da rede. Enquanto a rede totalmente conectada possui uma matriz completa W_1 com $C1 \cdot (M \cdot N + 1)$ parâmetros livres, a rede convolutiva possui cada filtro de W_1 formado por pesos compartilhados e um número menor de parâmetros livres, igual a $m \cdot n \cdot NF$.

Algorithm 3 Caminho direto da Conv

```

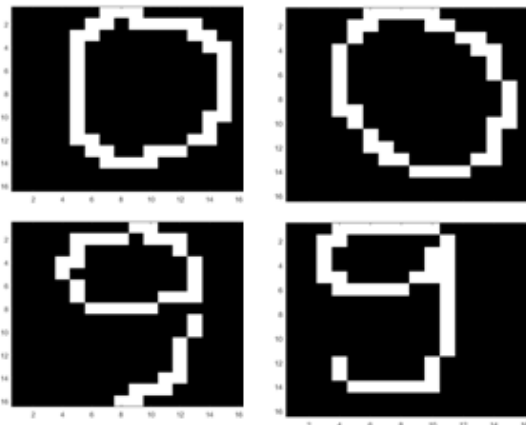
MapasdeCaracterísticas  $\leftarrow W_1 * Imagem$ 
A  $\leftarrow Empilhamento(MapasdeCaracterísticas)$ 
D  $\leftarrow ReLu(A)$ 
D  $\leftarrow ReLu(A)$ 
D  $\leftarrow [D \ 1]$ 
Yo  $\leftarrow W_2.D$ 
Ye  $\leftarrow Softmax(Y_o)$ 

```

▷ Inclusão do vés

Durante a etapa de implementação das RNAs, a base de dados pública "Números"[30] (dígitos manuscritos com dimensão 16X16 pixels) foi aplicada à entrada das redes. Amostras da base estão apresentadas na figura 6, composta por números de 0 a 9, totalizando 10 classes na saída. Criada por Lecun, essa base é uma miniatura da base MNIST. Ajustaram-se os parâmetros livres das matrizes de peso com o método descida do gradiente, descrito no capítulo 3. Com o intuito de garantir o aprendizado e a capacidade de generalização, utilizou-se 70% dos dados no treinamento e 30% no teste, durante 1000 épocas, ou até o até a acurácia no treinamento e teste estabilizar. O número de neurônios e o ajuste das não conformidades do algoritmo foram determinados baseados no desempenho padrão da rede, definido pelos resultados existentes na literatura. Depois da etapa da correta implementação das arquiteturas, diversos experimentos foram realizados com diferentes parametrizações a fim de estimar uma topologia com maior desempenho para a base "Números".

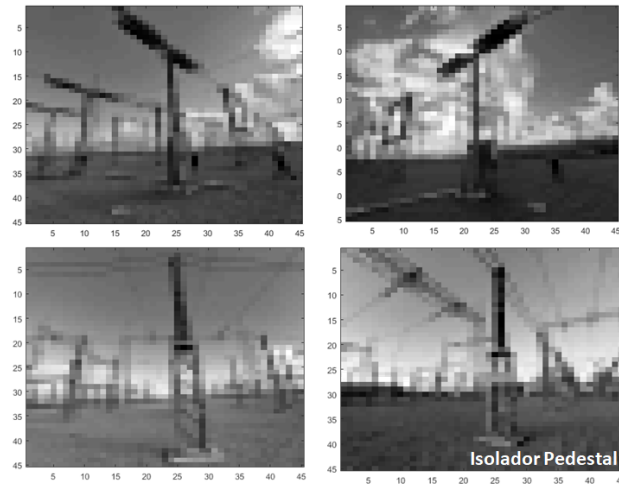
Figura 6 – Amostras da base Números



Por fim, visando a aplicação de interesse descrita no capítulo 1, 643 imagens monocromáticas de 45x45 pixels de 9 equipamentos (9 classes na saída) foram aplicadas como entrada para o desenvolvimento das RNAs descritas nas figuras 4 e 5. Amostras da base estão expostas na figura 7.

Vale ressaltar que a base da aplicação de interesse encontra-se em construção, e no momento da realização deste projeto possuía 643 amostras RGB de alta resolução,

Figura 7 – Amostras da base MiniBase GImpSI



sendo 382 amostras do equipamento 1, 59 do equipamento 2, 32 do equipamento 3 e 30 amostras para cada classe dos demais equipamentos. E que, para este trabalho, foram reduzidas a um canal (monocromáticas) e para dimensão 45x45 devido à capacidade de processamento limitada da linguagem adotada. Embora haja uma redução de informação, pode-se perceber que as estruturas visuais que definem as classes foram mantidas (figura 7).

Para treinamento e teste, adotaram-se os mesmos valores utilizados com a base anterior: 70-30% com até 1000 épocas no máximo. Contudo, a parametrização das redes para esse problema foi ajustada para o melhor resultado possível de acordo com a arquitetura adotada, já que se trata de uma tarefa inédita na literatura e não há desempenho de referência. Logo, tomou-se como referência para a parametrização a proporção do número de parâmetros livres entre as duas arquiteturas experimentadas com a base Números.

Os experimentos realizados para projeto das RNAs, bem como seus resultados e discussões, estão descritos no capítulo 5.

5 RESULTADOS

Os resultados obtidos nos experimentos realizados ao longo do projeto estão expostos neste capítulo, a seguir.

Fornecendo como entrada a base Números para a rede *MLP* desenvolvida (figura 4), e variando o número de neurônios na primeira camada, as acurácias no teste obtidas estão expostas na tabela 1.

É de referir que experimentos com maior número de neurônios não apresentaram um aumento da acurácia no teste. Com 200 neurônios, a acurácia obtida foi de ~90%,

Nº de Neurônios na primeira camada	Acurácia no teste
50	~89%
100	~90%
150	~93%

Tabela 1 – Acurácia no teste para *MLP* com Base Números

para citar um exemplo.

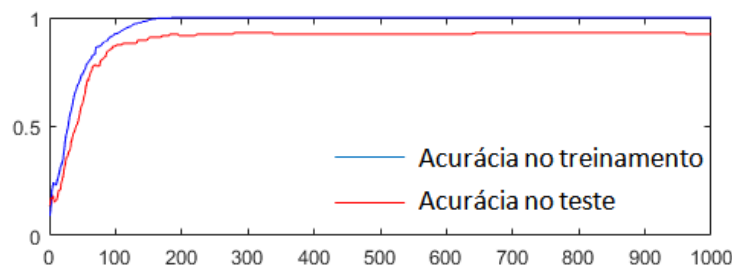
Para a mesma base de dados de entrada e utilizando a arquitetura da *ConvNet* (figura 5), os resultados obtidos estão expostos na tabela 2. Sendo que a parametrização dos filtros convolucionais foi escolhida respeitando a porcentagem de parâmetros livres da *MLP* definida em 4.

Dimensão dos filtros convolucionais	Acurácia no teste
3x3x4	~98%
4x4x4	~96%
5x5x6	~93%

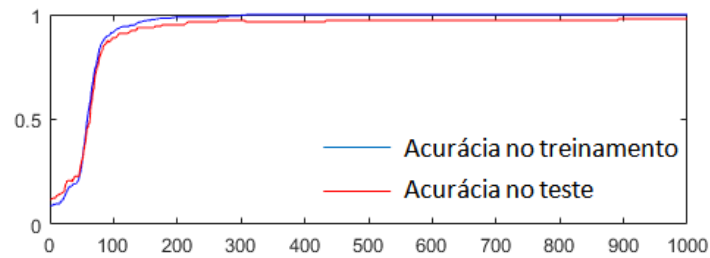
Tabela 2 – Acurácia no teste para *Convnet* com Base Números

Pelos desempenhos apresentados nas tabelas 2 e 1, pode-se verificar que a substituição da primeira camada totalmente conectada por uma camada convolutiva tornou a *Convnet* capaz de regularizar fronteiras com restrição de liberdade e de tirar proveito de um modelo analítico subjacente do que seja a imagem, implícito na escolha da estrutura a pesos compartilhados.

Os gráficos das acurácias (treinamento e teste) ao longo das épocas realizadas, do melhor desempenho de cada arquitetura, estão expostos nas figuras 8 e 9. Nota-se que a partir de 400 épocas o aprendizado já está estabilizado.

Figura 8 – Acurácia *versus* número de épocas na MLP- 150 neurônios na primeira camada

Similarmente aos experimentos com a base Números, agora aplicando a MiniBase GImpSI como entrada, as acurácias obtidas nas duas arquiteturas são observadas nas tabelas 3 e 4.

Figura 9 – Acurácia *versus* número de épocas na ConvNet- Filtro 3x3x4 na primeira camada

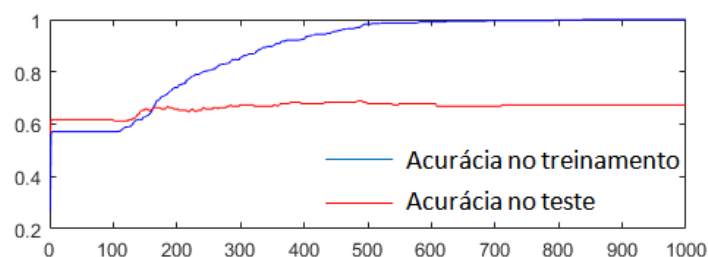
Nº de Neurônios na primeira camada	Acurácia no teste
50	~64%
150	~66%
200	~67%

Tabela 3 – Acurácia no teste para *MLP* com MiniBase GImpSI

Dimensão dos filtros convolucionais	Acurácia no teste
8x8x10	~68%
10x10x12	~72%
15x15x30	~68%

Tabela 4 – Acurácia no teste para *Convnet* com MiniBase GImpSI

Considerando que a Minibase GImpSI utilizada nos experimentos possui linha de base de 59% (382 amostras da classe 1 num total de 643 amostras). Então, uma acurácia no teste de ~72% não pode ser considerada como excelente, embora relevante. Essa característica pode ser observada nos gráficos expostos nas figuras 10 e 11.

Figura 10 – Acurácia *versus* número de épocas na MLP- 200 neurônios na primeira camada

Outro aspecto relevante dos resultados obtidos é que provavelmente nem mesmo a camada convolutiva conseguiu se ajustar a uma fronteira tão irregular (devido às liberdades de rotação e mudança de escala e fundo dos objetos-alvos) com apenas poucos exemplos de um fenômeno mais complexo. Complexidade que está relacionada aos graus de liberdade que as imagens da MiniBase permitem, como o fundo bastante heterogêneo e a mudança de perspectiva dos equipamentos a serem detectados. Todavia, mesmo

Figura 11 – Acurácia *versus* número de épocas na ConvNet- Filtro 10x10x12 na primeira camada

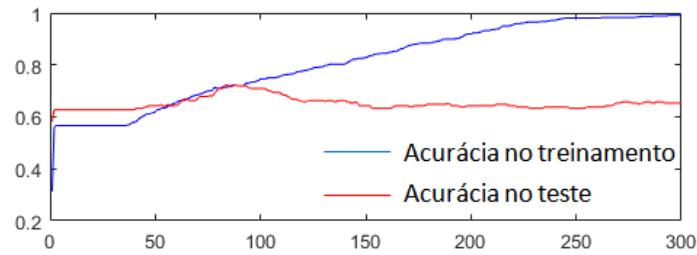
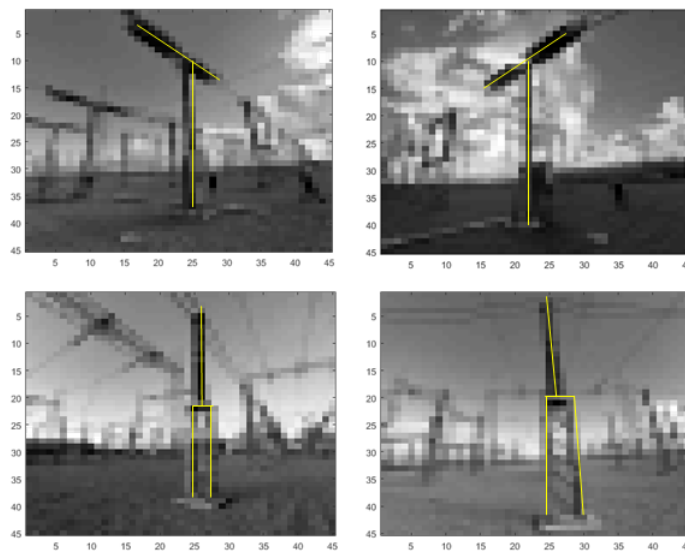


Figura 12 – Estruturas rígidas dos equipamentos na MiniBase GImpSI



com essas características variantes que dificultam o processo de detecção, o formato dos equipamentos presente nas amostras não se altera, ou seja, cada classe tem sua estrutura rígida, observada em destaque amarelo na figura 12. E, deste modo, essa rigidez pode ser explorada com o uso de camadas convolucionais.

6 CONCLUSÃO

A implementação linha por linha garantiu um bom conhecimento das etapas básicas que compõem as redes neurais convolutivas. Portanto, a escolha didática de não utilizar funções prontas foi assertiva quanto ao bom aprendizado da ferramenta na disciplina.

Para dar sequência ao trabalho desenvolvido, a fim de evoluir na análise da aplicação de interesse, faz-se necessário migrar a linguagem para uma com maior capacidade de processamento. Sendo assim, com a alteração da linguagem, é desejado implementar redes mais profundas, capazes de obter melhores resultados na detecção dos objetos de interesse. Além disso, a coleta das imagens dos equipamentos monitorados

ainda está sendo realizada pelos operadores da termelétrica. Dessa maneira, é esperado um aumento de amostras da base, a fim de garantir um melhor treinamento das redes, utilizando dados mais representativos do problema.

Por último, como proposta de novos experimentos, fazendo proveito das estruturas rígidas da aplicação de interesse, propõe-se a utilização de filtros convolucionais sintetizados capazes de representar pequenas partes estruturais dos equipamentos. Em outras palavras, a ideia parte da proposta de sintonizar os filtros de um maneira a forçar que seus pesos só possam extrair características dos equipamentos com estrutura rígida, ignorando as informações entre pixels presentes no fundo ou em outros elementos adicionais na imagem, por exemplo outro equipamento.

Referências

- [1] LEI, Y. et al. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, v. 63, n. 5, p. 3137–3147, 2016. Citado na página 2.
- [2] WANG, H.; ZHOU, B.; ZHANG, X. Research on the remote maintenance system architecture for the rapid development of smart substation in china. *IEEE Transactions on Power Delivery*, v. 33, n. 4, p. 1845–1852, 2018. Citado na página 2.
- [3] SALEM, A. A. et al. The leakage current components as a diagnostic tool to estimate contamination level on high voltage insulators. *IEEE Access*, v. 8, p. 92514–92528, 2020. Citado na página 2.
- [4] WERNECK, M. M. et al. Detection and monitoring of leakage currents in power transmission insulators. *IEEE Sensors Journal*, v. 15, n. 3, p. 1338–1346, 2015. Citado na página 2.
- [5] JIANG, Q. et al. A contour angle orientation for power equipment infrared and visible image registration. *IEEE Transactions on Power Delivery*, v. 36, n. 4, p. 2559–2569, 2021. Citado na página 2.
- [6] LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado 3 vezes nas páginas 3, 4 e 5.
- [7] HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, Wiley-Blackwell, v. 160, n. 1, p. 106, 1962. Citado 2 vezes nas páginas 4 e 8.
- [8] FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: *Competition and cooperation in neural nets*. [S.l.]: Springer, 1982. p. 267–285. Citado na página 4.
- [9] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: <<https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>>. Citado na página 5.
- [10] SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 5.
- [11] SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9. Citado na página 5.
- [12] HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 5.

- [13] GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 580–587. Citado na página 5.
- [14] GIRSHICK, R. Fast r-cnn. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 1440–1448. Citado na página 5.
- [15] REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 6, p. 1137–1149, 2017. Citado na página 5.
- [16] HE, K. et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2961–2969. Citado na página 5.
- [17] GKIOXARI, G.; MALIK, J.; JOHNSON, J. Mesh r-cnn. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2019. p. 9785–9795. Citado na página 5.
- [18] ARKIN, E. et al. A survey of object detection based on cnn and transformer. In: *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML)*. [S.l.: s.n.], 2021. p. 99–108. Citado na página 5.
- [19] REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788. Citado na página 5.
- [20] LIU, W. et al. Ssd: Single shot multibox detector. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 21–37. Citado na página 5.
- [21] LIN, T.-Y. et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2980–2988. Citado na página 5.
- [22] DUAN, K. et al. Centernet: Keypoint triplets for object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2019. p. 6569–6578. Citado na página 5.
- [23] KAYALIBAY, B.; JENSEN, G.; SMAGT, P. van der. Cnn-based segmentation of medical imaging data. *CoRR*, abs/1701.03056, 2017. Disponível em: <<http://arxiv.org/abs/1701.03056>>. Citado na página 5.
- [24] XIONG, X. et al. Identification of electrical equipment based on faster lstm-cnn network. In: *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*. [S.l.: s.n.], 2020. p. 1–6. Citado na página 5.
- [25] LEI, X.; SUI, Z. Intelligent fault detection of high voltage line based on the faster r-cnn. *Measurement*, Elsevier, v. 138, p. 379–385, 2019. Citado na página 5.
- [26] JIANLONG, G.; WEIXIA, F.; MANHUA, W. An improved faster r-cnn algorithm for electric equipment detection. In: *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*. [S.l.: s.n.], 2019. p. 138–141. Citado na página 5.

-
- [27] HAN, S. et al. Electrical equipment identification in infrared images based on roi-selected cnn method. *Electric Power Systems Research*, Elsevier, v. 188, p. 106534, 2020. Citado na página 5.
- [28] HAYKIN, S. *Redes Neurais - 2ed.* Bookman, 2001. ISBN 9788573077186. Disponível em: <<https://books.google.com.br/books?id=1Bp0X5qfyjUC>>. Citado na página 7.
- [29] DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016. Citado na página 9.
- [30] CUN, Y. L. et al. Handwritten zip code recognition with multilayer networks. In: IEEE. *[1990] Proceedings. 10th International Conference on Pattern Recognition*. [S.l.], 1990. v. 2, p. 35–40. Citado na página 13.