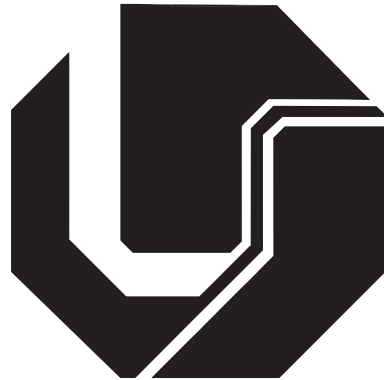


UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA



**Inserção Automática de Componentes
em Ambientes Virtuais de Treinamento para Subestações
de Energia utilizando Inteligência Artificial**

Leandro Sena Zuza

Uberlândia

2024

Leandro Sena Zuza

Inserção Automática de Componentes em Ambientes Virtuais de Treinamento para Subestações de Energia utilizando Inteligência Artificial

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para obtenção do Título de Mestre em Ciências.

Julho X, 2024.

Membros da Banca:

Prof. Alexandre Cardoso, Dr.
Orientador - UFU

Prof. Daniel S. Caetano, Dr.
Coorientador - UFU

Agradecimentos

[INSERIR NOVOS AGRADECIMENTOS]

“Então considere que as botas apertadas são uma das maiores venturas da Terra,
porque, fazendo doer os pés, dão azo ao prazer de as descalçar.”
(Machado de Assis, em Memórias Póstumas de Brás Cubas)

Abstract

The precise identification of equipment in images plays a crucial role in various operations related to power substations, facilitating not only maintenance but also the monitoring of these facilities. In this work, we present results on the efficiency of YOLOv8 in detecting equipment present in power substations from images obtained by Unmanned Aerial Vehicles (UAVs). We employ different optimization techniques to enhance detection efficiency, aiming to achieve more accurate and faster results. Additionally, this study aims to go beyond mere training with captured photos, seeking to identify the best-trained model to create a script capable of selecting, from a database of virtual reality models, the elements necessary for assembling a virtual power substation. Thus, we aim not only to improve the maintenance and monitoring processes of power substations in physical reality but also to streamline and enhance the generation of Virtual Training Environments for procedures related to these substations. With this advancement, we hope to not only optimize the use of detection technology in power substations but also to significantly contribute to the creation of realistic and efficient virtual environments for training in procedures related to the operation and maintenance of these facilities.

Keywords

Keywords - Power Substation; UAV; YOLOv8; Optimization; Virtual Training Environments

Resumo

A identificação precisa de equipamentos em imagens desempenha um papel crucial em várias operações relacionadas às subestações de energia, facilitando não apenas a manutenção, mas também o monitoramento dessas instalações. Neste trabalho, apresentamos resultados da eficiência da YOLOv8 na detecção de equipamentos presentes em subestações de energia, a partir de imagens obtidas por Veículos Aéreos Não Tripulados (VANTs). Utilizamos diferentes técnicas de otimização para aprimorar a eficiência na detecção, visando alcançar resultados mais precisos e rápidos. Além disso, este estudo visa ir além do mero treinamento com as fotos capturadas, buscando identificar o melhor modelo treinado para criar um script capaz de selecionar, a partir de uma base de modelos de realidade virtual, os elementos necessários para a montagem de uma subestação de energia virtual. Dessa forma, almejamos não apenas melhorar os processos de manutenção e monitoramento das subestações de energia na realidade física, mas também agilizar e aprimorar a geração de Ambientes Virtuais de Treinamento para procedimentos relacionados a essas subestações. Com este avanço, esperamos não só otimizar o uso de tecnologia de detecção em subestações de energia, mas também contribuir significativamente para a criação de ambientes virtuais realistas e eficientes para treinamento em procedimentos relacionados à operação e manutenção dessas instalações.

Palavras Chave

Subestação de Energia; VANTs; YOLOv8; Otimização; Ambientes Virtuais de Treinamento

Publicações

As publicações relacionadas à pesquisa e ao trabalho realizado são listadas a seguir:

1. Listar aqui

Sumário

1	INTRODUÇÃO	1
1.1	Motivação	1
1.2	Objetivos e Metas	3
1.3	Estrutura da Dissertação	3
2	FUNDAMENTAÇÃO TEÓRICA	4
2.1	Subestações de Energia	4
2.2	Redes Neurais Artificiais	4
2.2.1	You Look Only Once	9
2.2.2	Batch	10
2.2.3	Otimizadores	10
2.2.4	Precisão e Recall	10
2.3	Realidade Virtual	11
2.4	Considerações finais	13
	REFERÊNCIAS	14

Lista de ilustrações

Figura 1 – Representação de um NA (ADORNO, 2017)	5
Figura 2 – Topologia básica de uma rede neural (FLECK et al., 2016)	7
Figura 3 – Exemplificação de como uma imagem nada mais é que uma matriz, e como a filtragem busca padrões pré-estabelecidos dentro da imagem (LIMA; RUBIK; MORAIS, 2020)	8
Figura 4 – Esquema de funcionamento de uma RNC (VARGAS; PAES; VASCONCELOS, 2016)	8
Figura 5 – Operação de filtragem na camada convolucional. Nota-se que o filtro é uma multiplicação de matrizes (MUSIC, 2024)	9
Figura 6 – Cálculo de IoU. (PADILLA; NETTO; SILVA, 2020)	11
Figura 7 – Aumento de desempenho da precisão média, versão 5 para 8. (JOCHER; CHAURASIA; QIU, 2023)	12

Lista de tabelas

Lista de abreviaturas e siglas

VANT Veículo Aéreo Não Tripulado

RV Realidade Virtual

NA Neurônio Artificial

RNC Rede Neural Convolucional

1 Introdução

1.1 Motivação

As subestações de energia desempenham um papel fundamental no sistema de distribuição de energia elétrica no Brasil, permitindo a transferência eficiente e segura de eletricidade entre diferentes níveis de tensão. Elas são cruciais para garantir que a eletricidade gerada em usinas seja entregue aos consumidores com a qualidade e confiabilidade necessárias. Efetivamente, desempenham um papel crucial na estabilidade do sistema elétrico, facilitando a manutenção, controle e proteção da rede. Seu funcionamento envolve várias etapas, começando com a recepção da eletricidade gerada em usinas de energia. A eletricidade é então transformada em níveis de tensão adequados para distribuição por meio de transformadores. Nas subestações, também ocorrem operações de chaveamento, onde os dispositivos de comutação são usados para controlar o fluxo de eletricidade e direcioná-lo para as áreas desejadas da rede. As subestações também estão equipadas com sistemas de proteção que detectam e isolam falhas para evitar danos ao sistema elétrico e garantir a segurança dos equipamentos e dos operadores (RANDOLPH, 2013).

Contudo falhas de segurança durante as rotinas de um colaborador não são raras no setor elétrico, podendo causar danos a sua saúde, e em alguns casos levando a óbito. De acordo com o estudo de (LIMA; OLIVEIRA, 2021), acidentes de trabalho são muito comuns em ambientes como subestações de energia. Para avaliar as causas, foram analisados diversos processos trabalhistas durante um período de tempo contra empresas que prestam serviço neste setor. Foram concluídas as existências de várias as causas, mas uma se destaca: a falta de treinamento. No estudo citado, este fato é atrelado à terceirização dos serviços. Enquanto um funcionário direto da companhia de energia da região recebia 6 meses de treinamento, o funcionário terceirizado era treinado em um período médio de 30 a 40 dias. A qualidade do conteúdo destes treinamentos para o funcionário terceirizado também era muito mais superficial. Em uma análise dentro deste estudo, a respeito de um acidente fatal, foi verificado que um técnico foi acionado para resolução de um problema de rompimento de cabo. Devido a um descuido, um dos colaboradores tocou em um cabo energizado, sem se preocupar em verificar se todas as chaves estariam desligadas, acabando por o levar a óbito. Deste estudo, portanto, entendeu-se que a busca por custos mais baixos durante o treinamento, pessoas que se expõem ao risco, e o próprio funcionamento da transmissão de energia é colocado à prova.

Nesse cenário, intervenções tecnológicas seriam de grande valia para melhorar a condição de treinamento de operadores no sistema elétrico. São diversas as possibilidades de aplicações que podem atuar nesse sentido, desde sistemas avançados de monitoramento

até soluções de Realidade Virtual (RV), oferecendo oportunidades para aprimorar a gestão e operação das subestações. A incorporação dessas inovações, podem levar as empresas do setor elétrico a oferecer maior segurança aos seus colaboradores, reduzir custos operacionais e garantir um fornecimento de energia mais confiável para os consumidores finais (ZHOU; FU; YANG, 2016).

Para funções didáticas, como treinamentos, a RV se destaca como uma abordagem disruptiva em relação a métodos tradicionais, principalmente pelo alto nível de imersibilidade no contexto da aplicação, proporcionado ao usuário e ao alto resultado no aprendizado do conteúdo trabalhado. Contudo, para construir e preparar todo o ambiente para uma experiência imersiva em RV, faz-se necessário a elaboração de uma complexa estrutura que envolve desde a escolha do equipamento que será utilizado para a projeção ao usuário, como por exemplo, uma caverna de visualização, capacete de virtualização ou mesmo óculos de RV, até a criação, em softwares próprios para esse tipo de desenvolvimento, toda modelagem gráfica do ambiente até as interações que existentes na aplicação. Fatores como a capacidade gráfica e técnica são levadas em consideração nesta etapa, uma vez que aplicações com grande quantidade de interações e elementos, exigem do hardware que irá renderizar elevada capacidade de processamento. Se a demanda pela capacidade for alta, e não for suportada pelo hardware, será exigido do desenvolvedor redução na qualidade das texturas, assim como outros tratamentos para que toda a experiência durante a imersão não seja lenta ou mesmo careça de elementos que destitua a aplicação de imersibilidade (PALMEIRA et al., 2020).

Outro recurso que tem sido aplicado em várias áreas da ciência são as Redes Neurais Artificiais (RNA). Sua utilização tem sido atrelado a resolução de sistemas não-lineares em que nem todas as variáveis do problema são conhecidas, assim como problemas em que exista ruídos nos dados a serem tratados, ou seja, ideais para problemas do mundo real quando transportados para o mundo virtual. Ao simular o funcionamento do cérebro humano, replicando o aprendizado natural, as RNA exibem a capacidade de resolver problemas complexos, sendo, assim, ferramenta ótima a ser associada a um trabalho de pesquisa (OĞCU; DEMIREL; ZAIM, 2012).

Deste modo, motivado pela possibilidade de elaborar uma ferramenta que simplifique o desenvolvimento de um sistema em RV voltado para aplicações de treinamento de colaboradores em subestações de energia, este trabalho se propõe a construir uma ferramenta que faça a inserção automática de componentes em um ambiente de RV de uma subestação de energia. Toda a automação será alimentada por um modelo treinado a partir de uma RNA, alimentada por fotos capturadas por VANTs em duas subestações de energia diferentes.

1.2 Objetivos e Metas

O objetivo geral desta pesquisa é propor um sistema de inserção automática de componentes em ambientes virtuais de treinamento para subestações de energia a partir de imagens aéreas coletadas do local a ser mapeado virtualmente. Para alcançar esse objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Realizar uma revisão da literatura científica, para identificar quais os algoritmos utilizados no reconhecimento de padrões em imagens aéreas obtidas por VANTs;
- Estudar e avaliar quais são os hiperparâmetros do algoritmo de inteligência artificial a ser utilizado, que garanta maior eficiência no reconhecimento de componentes das subestações elétricas;
- Desenvolver uma automação capaz de receber uma imagem, reconhecer componente(es) da subestação elétrica inserir no ambiente virtual.

1.3 Estrutura da Dissertação

A presente dissertação é composta por sete capítulos, descritos da seguinte forma.

- No Capítulo 1 são apresentadas as motivações, os objetivos e a estruturação do trabalho;
- No Capítulo 2 são apresentados os principais fundamentos teóricos relacionados ao trabalho;
- No Capítulo 3 é apresentado o estado da arte da linha de pesquisa principal desse trabalho;
- Nos Capítulos 4 e 5 são apresentados materiais/métodos e detalhes de implementação;
- No Capítulo 6, são discutidos e apresentados os resultados obtidos nesse trabalho a partir do sistema desenvolvido;
- Por fim, no Capítulo 7, são apresentadas as conclusões e as perspectivas para trabalhos futuros.

2 Fundamentação Teórica

2.1 Subestações de Energia

A energia elétrica no Brasil é oriunda majoritariamente de fontes renováveis, em especial da geração hidroelétrica, correspondendo a 69% de toda malha produtora. Esta forma de gerar energia é limpa, de baixo custo e possui reduzida emissão de gases de efeito estufa (ANDRADE; SANTOS, 2015).

Após a geração, a energia passa pelo processo de transmissão, que será responsável por aplicar as devidas transformações, para que a eletricidade chegue em níveis seguros e desejados para o consumidor final. É nesta etapa que são usadas as subestações de energia. Trata-se, basicamente, de um sítio industrial, equipado com diversas máquinas, cujo objetivo é alterar e redirecionar as tensões elétricas, assim como manter íntegro e seguro todo esse processo de transferência. Para que isso ocorra, o processo inicial é realizado em altíssimas tensões, para que seja reduzida a perda durante o processo; assim como próximo aos consumidores finais, a tensão é reduzida para valores apropriados para uso doméstico (MUZY, 2012).

De maneira mais aprofundada, a energia é, portanto, obtida por geradores e transmitida para subestações primárias. Há aí a primeira aferição de qualidade dos níveis de transmissão. Após isso, a energia transformada é enviada para subestações de distribuição, alterada para níveis mais baixos para serem transmitidas para setores industriais e urbanos (JUNIOR et al., 2022).

2.2 Redes Neurais Artificiais

A RNA trata-se de um conjunto de técnicas que buscam simular o funcionamento do cérebro humano, a partir de algoritmos computacionais, a fim de resolver problemas específicos. Sua eficiência está relacionada com a quantidade de interações que as unidades de processamento que o compõem realizam entre si. Comparam-se as RNA à mente humana devido a sua capacidade de aprendizado, podendo generalizar funções a partir de alguns exemplos informados, e delas prever o comportamento de valores para os quais não foi fornecido a resposta esperada. A base do funcionamento da RNA está no conceito do neurônio artificial (NA), que se traduz como uma pequena unidade de processamento, capaz de receber um sinal simples, e a partir dele gerar uma resposta. De acordo com (FLECK et al., 2016), a primeira noção de NA surge em 1943, no trabalho Warren McCulloch e Walter Pitts, no artigo: “A Logical Calculus of the Ideas Immanent in Nervous Activity”.

Matematicamente, o NA recebe um valor de entrada, realizando com ele o produto desse valor a um outro denominado peso. O resultado é comparado com um diferenciador: se for maior, será dada a saída verdadeira; se menor, falso. Na figura 1, esse processo é demonstrado de maneira esquemática. A operação descrita, chamada também de threshold logic, i.e., lógica limiar, em tradução livre, mimetiza o funcionamento de um componente eletrônico chamado transistor, base do funcionamento dos processadores computacionais modernos, em que a passagem de corrente elétrica por ele é interrompida ou permitida com base no sinal de entrada (MCCULLOCH; PITTS, 1943).

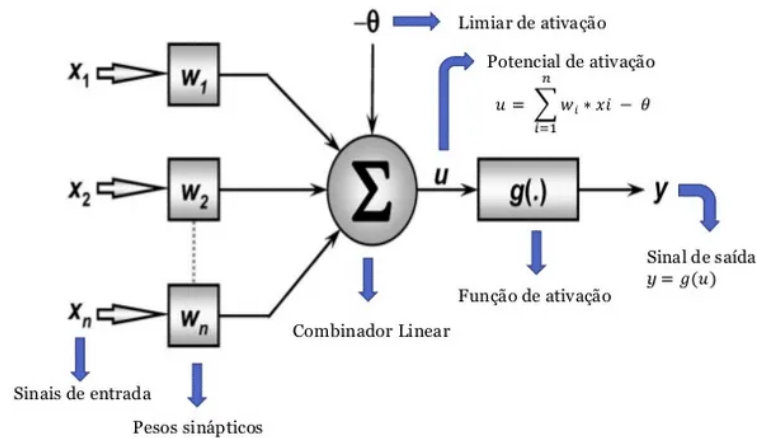


Figura 1 – Representação de um NA (ADORNO, 2017)

Aprofundando-se na álgebra relacionada do NA, pode-se visualizar o esquema do seu funcionamento na figura 1, e também acompanhar a definição de cada envolvido nesta operação (BRASIL, 2023):

- **Sinais de entrada:** Constituem-se os sinais, ou valores, externos ao modelo, muitas vezes submetidos a algum tratamento prévio, responsáveis por alimentar a rede.
- **Pesos sinápticos:** Também chamados apenas de pesos, ponderam os sinais de cada entrada da rede.
- **Combinador linear:** Operação aritmética envolvendo os sinais de entrada a fim de gerar um potencial de ativação.
- **Bias:** O bias é um valor adicional que é somado à combinação linear das entradas antes de passar pela função de ativação, visando ajustá-la para que os dados melhor se adaptem à rede.
- **Limiar de ativação:** Também denominado threshold, determina o nível adequado em que o resultado obtido pelo combinador linear pode acionar a ativação.
- **Potencial de ativação:** É o resultado decorrente da discrepância entre o valor gerado pelo combinador linear e o limiar de ativação. Se esse resultado for positivo,

indicando $u \geq 0$, o neurônio gera um potencial de excitação; caso contrário, o potencial será inibitório.

- **Função de ativação:** Sua função é restringir a saída de um neurônio dentro de um determinado intervalo de valores.
- **Sinal de saída:** Resultado final da saída, que pode ser utilizado como entrada para outros neurônios interligados sequencialmente.

Em suma, a RNA, enquanto um conjunto de NA, pode ser entendido como um processador robusto, distribuído de maneira paralela, com pequenas unidades de processamento. Sua semelhança ao cérebro humano, deve-se, portanto, à capacidade de aprendizado e aos sinais sinápticos, existentes entre os neurônios, responsáveis pelo processo de armazenamento de conhecimento (HAYKIN, 2001).

De acordo com (RAUBER, 2005), em 1958, após à concepção do NA, a primeira RNA a se notabilizar e a continuar sendo utilizada até o presente momento trata-se da Perceptron (ROSENBLATT, 1958). Este algoritmo possuía a capacidade de alterar os pesos dos neurônios, conforme o avanço do processamento da rede, de modo a resolver problemas resolvendo classificação linear. Seguido a ela, em 1960, houve a concepção da rede Adaline (WIDROW; HOFF et al., 1960). Diferentemente da perceptron, esta rede já se mostrava capaz de produzir resultados que iam além dos valores binários, sendo capaz de gerar respostas de valores presentes em todo o conjunto dos números reais. O cálculo da rede Adaline era baseado na regra Delta, que era capaz de aproximar os valores dos pesos gerados para aqueles valores com menor erro possível. Ambas as redes, contudo, apresentavam uma limitação quanto à modificação dos pesos em todas suas multicamadas. Para esse cenário, foi proposto o conceito de backpropagation, ou, em tradução livre, retropropagação do erro (RUMELHART; HINTON; WILLIAMS, 1986). A ideia da retropropagação se inicia com o feedforward, em que todos os dados de entrada de uma RNA são transmitidos do início à camada de saída, sem nenhuma alteração de peso. Desta forma, calcula-se o erro total dessa primeira passagem, comparando o valor resultante com o esperado. Essa métrica de erro torna-se o guia para o ajuste dos pesos entre as conexões de cada neurônio. Com ela, é feito o caminho de volta, e calculado um gradiente de erro, que será o fator decisivo para a atualização dos pesos durante todo o treinamento.

A topologia básica de uma rede neural é dividida em três níveis de camada: a camada de entrada, a camada oculta e a camada de saída, conforme mostrado na figura 2. A camada de entrada é, naturalmente, aquela que recebe os dados de entrada. Nela, cada nó representa uma característica ou atributo dos dados que estão sendo alimentados na rede neural. Por exemplo, em uma aplicação de reconhecimento de imagens, cada nó na camada de entrada pode representar um pixel da imagem. A camada oculta, que não se

restringe a apenas uma, podendo ser várias, representa um processo intermediário entre a camada de entrada e a camada de saída. Cada neurônio em uma camada oculta recebe entradas das camadas anteriores, realiza algum tipo de transformação não linear dessas entradas e passa o resultado para a próxima camada. A presença de múltiplas camadas ocultas permite que a rede aprenda representações complexas e abstratas dos dados. Por fim, a camada de saída representa a camada final, e ela é responsável por gerar as saídas desejadas. A estrutura da camada de saída depende do tipo de problema que está sendo resolvido. Por exemplo, em um problema de classificação, cada nó na camada de saída pode representar uma classe diferente, e a saída pode ser interpretada como a probabilidade de pertencer a cada classe (RAUBER, 2005).

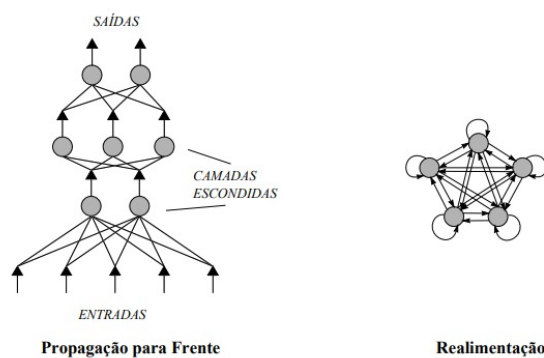


Figura 2 – Topologia básica de uma rede neural (FLECK et al., 2016)

Essencialmente, as redes neurais aprendem iterativamente ajustando os pesos de suas conexões através do processo de treinamento, onde são apresentados a um conjunto de dados de entrada e as correspondentes saídas desejadas. Com o tempo, a rede neural é capaz de aprender a mapear efetivamente os padrões nos dados de entrada para as saídas desejadas, tornando-se assim capaz de realizar tarefas como reconhecimento de padrões, classificação, regressão, entre outros (HAYKIN, 2001).

Contudo, para problemas envolvendo aprendizado por meio de imagens, a forma de lidar com as informações é diferente. De fato, imagens são dados, organizados em formato de matrizes de duas dimensões ou três dimensões (se for considerada a camada de cores), em que cada unidade de informação se chama pixel, conforme se nota no exemplo da figura 3. Para processar esses dados, um tipo de RNA destaca-se: a Rede Neural Convolutiva (RNC). Comparando ambas, nota-se que a RNA é composta por camadas intrinsecamente conectadas, em que cada neurônio de uma camada se conecta a todos os neurônios da camada seguinte. Isso faz com que ela seja adequada para dados vetorizados e para problemas de classificação e regressão sem uma estrutura espacial ou temporal específica. Por sua vez, a RNC recorre a dois tipos de camadas (figura 4): a convolutiva, cujo objetivo é extrair características locais, e a de pooling, responsável por manter a estrutura espacial essencial (VARGAS; PAES; VASCONCELOS, 2016).

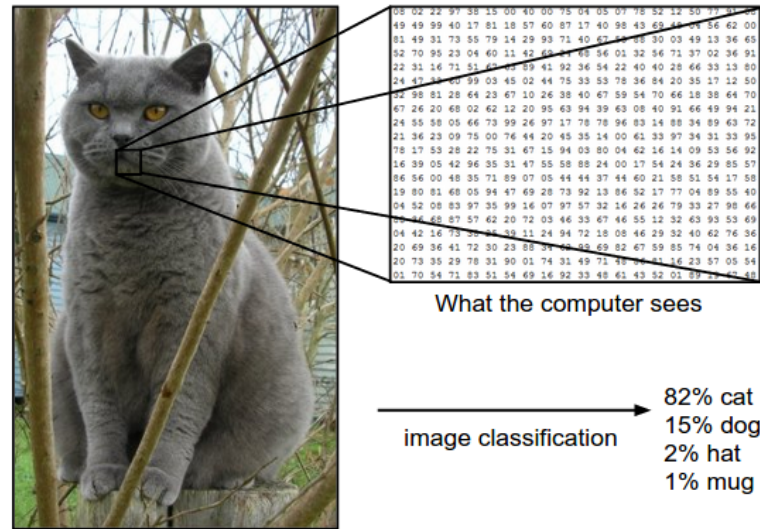


Figura 3 – Exemplificação de como uma imagem nada mais é que uma matriz, e como a filtragem busca padrões pré-estabelecidos dentro da imagem (LIMA; RUBIK; MORAIS, 2020)

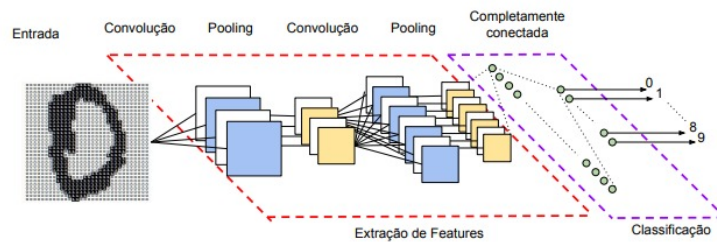


Figura 4 – Esquema de funcionamento de uma RNC (VARGAS; PAES; VASCONCELOS, 2016)

A camada de convolução, portanto, é responsável pela feature extractor, i.e., pela extração de características de interesse em uma imagem. Basicamente, são aplicados filtros, também chamados de kernel, à imagem a fim de buscar padrões. São comumente empregados filtros de detecção de bordas (edge detection), desfoque (blur), nitidez (sharpen). Conforme na figura 5, cada elemento do filtro é multiplicado pelo elemento de mesma posição na região em que o filtro está sendo aplicado naquele instante. Ao final dessas operações matriciais, adiciona-se os resultados dessas multiplicações para ter um único valor como saída, que será o pixel correspondente na imagem filtrada (RODRIGUES, 2023)

A camada de pooling é aplicada após cada camada convolucional, com o objetivo de reduzir as dimensões das imagens enquanto mantém a profundidade do volume. Esse processo promove a invariância da Rede Neural Convolutacional (RNC) às transformações geométricas, permitindo à rede reconhecer objetos na imagem independentemente de sua posição. Além disso, o uso da camada de pooling contribui para diminuir significativamente o custo computacional da rede. Existem diferentes tipos de pooling, como o max pooling,

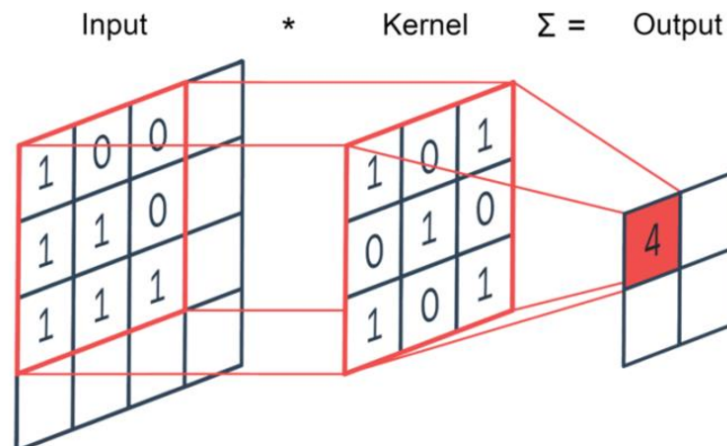


Figura 5 – Operação de filtragem na camada convolucional. Nota-se que o filtro é uma multiplicação de matrizes (MUSIC, 2024)

que seleciona o valor máximo em uma região específica da imagem, e o average pooling, que calcula a média dos valores nesta região. O max pooling é o mais comum, pois preserva as características mais salientes das imagens, enquanto reduz o ruído. A camada de pooling, ao reduzir o tamanho espacial dos mapas de características, também ajuda a evitar um problema comum no aprendizado de máquina, em que um modelo aprende não apenas os padrões relevantes dos dados de treinamento, mas também o ruído e as particularidades específicas desses dados. Chama-se este problema de overfitting. Evitando-o, a rede consegue construir um treinamento mais generalizável e que possa gerar resultados relevantes para dados diferentes do modelo treinado (LIMA; RUBIK; MORAIS, 2020).

2.2.1 You Look Only Once

YOLO (You Only Look Once) é um modelo de detecção de objetos em imagens e vídeos em tempo real, que se destaca por sua eficiência e precisão. Desenvolvido por Joseph Redmon, Santosh Divvala, Ross Girshick e Ali Farhadi, o YOLO aborda o problema de detecção de objetos como uma única tarefa de regressão, prevendo caixas delimitadoras e probabilidades de classe diretamente de imagens inteiras em uma única passagem pela rede neural. O funcionamento do YOLO começa com a divisão da imagem de entrada em uma grade, geralmente de dimensões como 7x7 ou 9x9. Cada célula dessa grade é responsável por prever um conjunto de caixas delimitadoras e as probabilidades das classes dos objetos contidos nessa célula. Para cada célula da grade, o modelo faz previsões sobre as caixas delimitadoras que contêm objetos, representadas por cinco valores: coordenadas (x, y) do centro da caixa, largura (w) e altura (h) da caixa, e a confiança de que a caixa contém um objeto. Além disso, são previstas as probabilidades de cada classe para cada caixa delimitadora. Após a predição das caixas delimitadoras, o YOLO utiliza um processo chamado Non-max Suppression (Supressão de Não-Máximo) para refinar as previsões e eliminar caixas sobrepostas ou redundantes, mantendo apenas as detecções

mais confiáveis. Esse processo envolve a supressão de caixas que têm uma sobreposição significativa e a escolha da caixa com a maior confiança entre as caixas sobrepostas. A saída do YOLO é uma lista de caixas delimitadoras, cada uma associada a uma classe prevista e sua confiança. Essas caixas delimitadoras representam os objetos detectados na imagem, fornecendo informações sobre sua localização e classificação. Em resumo, o YOLO oferece uma abordagem eficaz e eficiente para a detecção de objetos em tempo real, consolidando-se como uma ferramenta valiosa em diversas aplicações, desde sistemas de segurança até veículos autônomos (REDMON et al., 2016).

2.2.2 Batch

Um dos aspectos cruciais do funcionamento da YOLO é o conceito de "batch" (em tradução livre, "lote") durante o treinamento da rede neural. Ao agrupar várias imagens em lotes para processamento simultâneo, a YOLO aproveita a capacidade de processamento paralelo das GPUs, acelerando significativamente o treinamento. Durante a propagação direta, cada imagem no lote é processada pela rede neural para gerar previsões de detecção de objetos. Em seguida, a perda é calculada em relação às anotações verdadeiras, e os pesos da rede são atualizados para minimizar essa perda, usando algoritmos de otimização como o gradiente descendente. Esse processo é repetido para vários lotes de imagens até que a rede convirja para uma solução adequada. Assim, o uso eficiente de lotes na YOLO não apenas acelera o treinamento, mas também contribui para a robustez e eficácia dos modelos de detecção de objetos resultantes. (GOODFELLOW; BENGIO; COURVILLE, 2016)

2.2.3 Otimizadores

Falar sobre o Otimizadores.....!!!

2.2.4 Precisão e Recall

A fim de avaliar o desempenho de um treinamento na arquitetura YOLO, é preciso entender os resultados fornecidos pelo modelo. De acordo com (PADILLA; NETTO; SILVA, 2020), basicamente a YOLO utiliza a métrica chamada Average Precision (AP) ("Precisão Média", em tradução livre). Ela se baseia no conceito de IoU ("Intersection over the Union", Intersecção sobre a União, em tradução livre), que calcula uma razão entre a interseção da detecção feita pelo algoritmo com relação à marcação (bounding box) realizada em cima da área dividida pela união dessas duas áreas (Figura 1). Essa razão poderá ser comparada com um valor pré-estabelecido, o thresholds, que será referido por L. A partir desse valor, é possível que no processo de detecção retorne três diferentes resultados na pesquisa pela classe desejada. São eles: Verdadeiro Positivo (VP), Falso Positivo (FP) e Falso Negativo (FN). VP trata-se dos resultados considerados corretos que a rede neural retorna, que

seriam todos resultados que $IoU > L$. FP já seriam os resultados em que $IoU < L$, que são tidos como incorretos. FN, por sua vez, trata dos resultados totalmente fora do esperado.

$$\text{Recall} = \frac{VP}{VP + FN} \quad (1)$$

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

Com esses valores, pode-se calcular os resultados da saída que o treinamento da rede YOLO fornece. Em (1), tem-se o cálculo da Recall, que calcula a capacidade da rede de detectar todos os objetos relevantes em uma imagem. Já a Precisão (2), refere-se à capacidade da rede de encontrar apenas resultados relevantes.

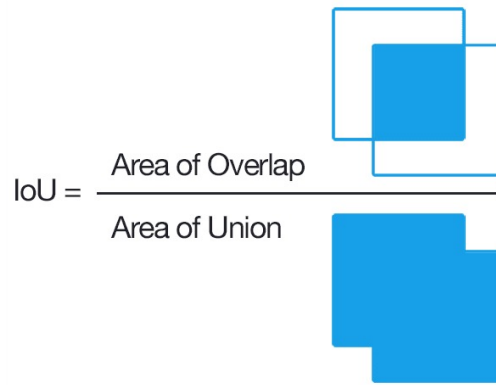


Figura 6 – Cálculo de IoU. (PADILLA; NETTO; SILVA, 2020)

A arquitetura YOLO disponibiliza um dataset, ou seja, um banco de imagens e weights, comum em todas as versões, chamado de COCO (“Common Objects in Context”, que em tradução livre seria “Classes Comuns de Objetos”) com classes pré-treinadas e imagens para realização de treinamentos. A partir dele, verificou-se por meio de testes a eficiência das quatro últimas versões da YOLO, a fim de identificar se nas mais recentes houve melhoras significativas em termos de performance e precisão. Na Figura 1, é apresentado o comparativo das versões. Nota-se que a v8, para um menor número de parâmetros que as demais, apresentou uma mAP50-95, maior que nas versões v5, v6 e v7. Além disso, com relação a velocidade de processamento, a v8 também se sobressai, com maior rapidez no processamento, ao processar maior quantidade de imagens para um mesmo intervalo de tempo (PADILLA; NETTO; SILVA, 2020).

2.3 Realidade Virtual

RV pode ser definida como um ambiente gerado a partir de um sistema computacional em que o usuário não apenas se sente dentro do contexto artificial, como também

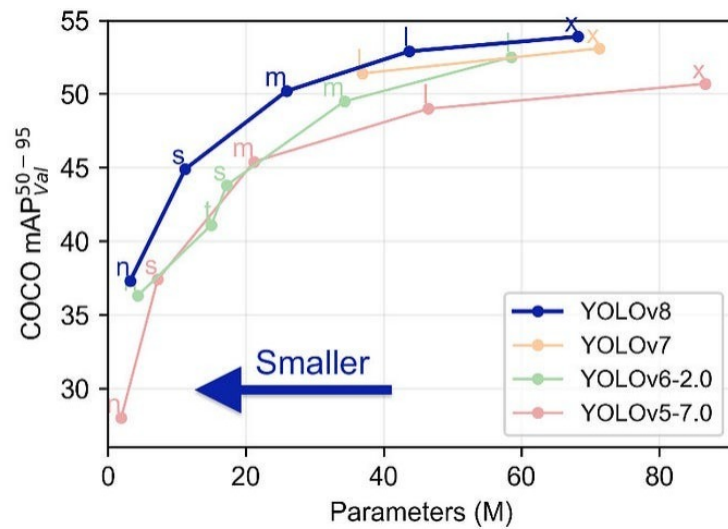


Figura 7 – Aumento de desempenho da precisão média, versão 5 para 8.
(JOCHER; CHAURASIA; QIU, 2023)

possibilita ao usuário a consciência de que pode navegar e interagir neste ambiente virtual. Trata-se, por isso, de uma interessante interface em um ser humano, um computador, capaz de gerar navegabilidade, interações e, principalmente, imersão em um ambiente sintético, gerado computacionalmente por meio de canais multisensoriais. Além disso, a RV pode ser classificada de duas formas diferentes: Imersiva e Não-imersiva. Na primeira, cria-se uma RV que isole o usuário do mundo real. Seus sentidos são bloqueados para recepção dos estímulos do seu entorno, para se direcionarem àqueles oriundos do mundo fictício. Para tanto, uma ampla variedade de equipamentos, como fones de ouvidos, luvas de dados, óculos de RV, entre outros, são empregados para os resultados esperados. Já a RV Não-Imersiva não se isola do mundo real. Ou seja, o usuário tem consciência de que está em um ambiente artificial. Da mesma forma, uma ampla variedade de equipamentos é empregada para gerar essa interação, incluindo dispositivos do cotidiano, como mouses, monitores de computador. Uma grande variedade de dispositivos convencionais e não-convencionais pode compor essa interação (CARDOSO et al., 2007).

É importante que os três conceitos da RV sempre sejam presentes em uma aplicação do tipo, sendo eles: interação, imersão e navegação (KALAWSKY, 1993).

O surgimento da RV data de 1963, em que foi apresentado uma aplicação de manipulação de objetos tridimensionais em um computador, denominada Sketchpad (SUTHERLAND, 1963). A aplicação conseguia reproduzir interatividade por meio de uma caneta óptica, que era utilizada para seleção de objetos projetados em uma tela. Neste trabalho, surgiram alguns dos principais termos da RV, como representação visual, dispositivos especiais, e interações em tempo real. Em 1968, o mesmo autor do trabalho anterior, Sutherland, publicou outro trabalho marcante para a história: A Head-Mounted Three Dimensional Display (SUTHERLAND, 1968). Nele, foi introduzido o conceito de

imersividade para uma RV; no caso, um capacete capaz de projetar fotos diretamente aos olhos do usuário, assim como rastrear o movimento da cabeça, afim de que este movimento fosse responsivo no ambiente virtual. Seguido a isso, uma série de outros equipamentos foram desenvolvidos a fim de sofisticar as soluções do ramo e propor novos modelos. A tendência sempre prevaleceu de surgir ferramentas mais simples e acessíveis ao usuário final (KIRNER; KIRNER, 2011).

Além das aplicações lúdicas, a RV atua de maneira séria e efetiva em diversas áreas técnicas. Como ferramenta para treinamento de operadores em ambientes de risco e de difícil simulação, torna-se uma alternativa viável para capacitação. O trabalho de (SILVA et al., 2012) demonstra bem essa ideia ao investigar técnicas de RV para que uma pessoa comum, sem treinamento anterior, possa, de maneira segura à uma subestação em funcionamento e à própria pessoa e todos ao seu redor, interagir com um transformador de energia, executando todos procedimentos que teria em um ambiente real, contudo de maneira virtual. Este treinamento já seria uma base interessante para que um profissional pudesse agilizar seu treinamento no mundo real ou mesmo absorvê-los. Em trabalhos de alta periculosidade tudo é crítico; então, qualquer intervenção que possa atenuar os riscos inerentes à natureza do trabalho, traz grandes avanços ao processo de profissionalização de técnicos e pessoas interessadas.

2.4 Considerações finais

Neste capítulo, foi apresentado o arcabouço teórico necessário para o entendimento da proposta dessa dissertação. A apresentação das redes neurais, e em específico, o modo que a arquitetura da YOLOv8 sobrepõe-se em termos de eficiência em relação às demais arquiteturas abordadas em outros trabalhos científicos, demonstra o direcionamento assertivo deste trabalho. Além disso, a apresentação da Realidade Virtual como uma disciplina inovadora e muito útil para servir à diversos propósitos dentro da indústria e ciência, corroboram para o entendimento da proposta desta dissertação.

Referências

- ADORNO, V. *Redes Neurais Artificiais*. 2017. Medium. Disponível em: <<https://medium.com/@avinicius.adorno/redes-neurais-artificiais-418a34ea1a39>>. Citado 2 vezes nas páginas 9 e 5.
- ANDRADE, A. de L.; SANTOS, M. A. D. Hydroelectric plants environmental viability: Strategic environmental assessment application in brazil. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 52, p. 1413–1423, 2015. Citado na página 4.
- BRASIL, D. L. B. *O Neurônio Biológico e Matemático*. 2023. Acessado em: 2 de junho de 2024. Disponível em: <<https://www.deeplearningbook.com.br/o-neuronio-biologico-e-matematico/#:~:text=Limiar%20de%20ativa%C3%A7%C3%A3o%20%7B%20CE%98%20%7D%3A,e%20o%20limiar%20de%20ativa%C3%A7%C3%A3o.>> Citado na página 5.
- CARDOSO, A. et al. Tecnologias e ferramentas para o desenvolvimento de sistemas de realidade virtual e aumentada. *Editora Universitária UFPE*, p. 1–19, 2007. Citado na página 12.
- FLECK, L. et al. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, v. 1, n. 13, p. 47–57, 2016. Citado 3 vezes nas páginas 9, 4 e 7.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 10.
- HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2001. Citado 2 vezes nas páginas 6 e 7.
- JOCHER, G.; CHAURASIA, A.; QIU, J. *YOLO by Ultralytics*. *Ultralytics*. 2023. Citado 2 vezes nas páginas 9 e 12.
- JUNIOR, C. d. L. B. et al. Uma proposta de sistema de autoria baseado em plantas baixas para projetar ambientes de realidade virtual: metodologia e estudo de caso aplicados a subestações de energia elétrica. Universidade Federal de Uberlândia, 2022. Citado na página 4.
- KALAWSKY, R. *The science of virtual reality and virtual environments*. [S.l.]: Addison-Wesley Longman Publishing Co., Inc., 1993. Citado na página 12.
- KIRNER, C.; KIRNER, T. G. Evolução e tendências da realidade virtual e da realidade aumentada. *Realidade Virtual e Aumentada: Aplicações e Tendências*. Cap, v. 1, p. 10–25, 2011. Citado na página 13.
- LIMA, M.; RUBIK, E.; MORAIS, R. *Introdução ao reconhecimento de imagens*. 2020. Acessado em: 2 de junho de 2024. Disponível em: <<https://lamfo-unb.github.io/2020/12/05/Captcha-Break/>>. Citado 2 vezes nas páginas 9 e 8.
- LIMA, M. E. A.; OLIVEIRA, R. C. Precarização e acidentes de trabalho: os riscos da terceirização no setor elétrico. *Revista Brasileira de Saúde Ocupacional*, SciELO Brasil, v. 46, p. e6, 2021. Citado na página 1.

- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943. Citado na página 5.
- MUSIC, R. N. C. of. *A Short History of Neural Synthesis*. 2024. Acessado em: 2 de junho de 2024. Disponível em: <<https://www.rncm.ac.uk/research/research-activity/research-centres-rncm/prism/prism-blog/a-short-history-of-neural-synthesis/>>. Citado na página 9.
- MUZY, G. L. C. d. O. Subestações elétricas. *Projeto Final (Graduação em Engenharia Elétrica)-UFRJ. Rio de Janeiro*, 2012. Citado na página 4.
- OĞCU, G.; DEMIREL, O. F.; ZAIM, S. Forecasting electricity consumption with neural networks and support vector regression. *Procedia-Social and Behavioral Sciences*, Elsevier, v. 58, p. 1576–1585, 2012. Citado na página 2.
- PADILLA, R.; NETTO, S. L.; SILVA, E. A. D. A survey on performance metrics for object-detection algorithms. In: IEEE. *2020 international conference on systems, signals and image processing (IWSSIP)*. [S.l.], 2020. p. 237–242. Citado 3 vezes nas páginas 9, 10 e 11.
- PALMEIRA, E. G. Q. et al. O uncanny valley das mãos virtuais em aplicações de realidade virtual imersiva: uma revisão sistemática da literatura. *Revista Ibérica de Sistemas e Tecnologias de Informação*, Associação Ibérica de Sistemas e Tecnologias de Informacao, n. E31, p. 497–512, 2020. Citado na página 2.
- RANDOLPH, J. Electric power substations engineering [book reviews]. *IEEE Power and Energy Magazine*, IEEE, v. 11, n. 3, p. 103–105, 2013. Citado na página 1.
- RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, v. 29, 2005. Citado 2 vezes nas páginas 6 e 7.
- 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 10.
- RODRIGUES, V. B. *Conhecendo a Visão do Computador: Redes Neurais Convolucionais*. 2023. Acessado em: 2 de junho de 2024. Disponível em: <<https://vitorborbarodrigues.medium.com/conhecendo-a-vis%C3%A3o-do-computador-redes-neurais-convolucionais-e1c2b14bf426>>. Citado na página 8.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 6.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Citado na página 6.
- SILVA, R. C. et al. Virtual substation um sistema de realidade virtual para treinamento de operadores de subestações elétricas. *Universidade Federal de Uberlândia*, 2012. Citado na página 13.

SUTHERLAND, I. E. Sketchpad: A man-machine graphical communication system. p. 329–346, 1963. Citado na página 12.

SUTHERLAND, I. E. A head-mounted three dimensional display. p. 757–764, 1968. Citado na página 12.

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. *Proceedings of the xxix conference on graphics, patterns and images*. [S.l.], 2016. v. 1, n. 4. Citado 3 vezes nas páginas 9, 7 e 8.

WIDROW, B.; HOFF, M. E. et al. Adaptive switching circuits. v. 4, n. 1, p. 96–104, 1960. Citado na página 6.

ZHOU, K.; FU, C.; YANG, S. Big data driven smart energy management: From big data to big insights. *Renewable and sustainable energy reviews*, Elsevier, v. 56, p. 215–225, 2016. Citado na página 2.