



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE
PRODUÇÃO
CURSO DE ENGENHARIA DE PRODUÇÃO

GUSTAVO CAMARGO ROCHA LIMA

**APLICAÇÃO DE VISÃO COMPUTACIONAL NA GESTÃO DE RISCOS DA
INDÚSTRIA DE ÓLEO E GÁS**

Recife

2025

GUSTAVO CAMARGO ROCHA LIMA

**APLICAÇÃO DE VISÃO COMPUTACIONAL NA GESTÃO DE RISCOS DA
INDÚSTRIA DE ÓLEO E GÁS**

Trabalho de Conclusão de Curso apresentado
ao Curso de Engenharia de Produção da
Universidade Federal de Pernambuco, como
requisito parcial para obtenção do título de
Bacharel em Engenharia de Produção.

Orientador (a): Márcio José das Chagas Moura

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Lima, Gustavo Camargo Rocha.

Aplicação de visão computacional na gestão de riscos da indústria de Óleo e Gás / Gustavo Camargo Rocha Lima. - Recife, 2025.

74 p. : il., tab.

Orientador(a): Márcio José das Chagas Moura

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia de Produção - Bacharelado, 2025.

Inclui referências.

1. Gestão de riscos. 2. Visão computacional. 3. Indústria de Óleo e Gás. 4. Aprendizagem de máquina. 5. Equipamento de proteção individual. 6. Segurança no trabalho. I. Moura, Márcio José das Chagas. (Orientação). II. Título.

620 CDD (22.ed.)

**APLICAÇÃO DE VISÃO COMPUTACIONAL
NA GESTÃO DE RISCOS DA INDÚSTRIA DE
ÓLEO E GÁS**

Trabalho de Conclusão de Curso
apresentado ao Curso de Engenharia de
Produção da Universidade Federal de
Pernambuco, como requisito parcial para
obtenção do título de Bacharel em
Engenharia de Produção.

Aprovado em: 08/07/2025

BANCA EXAMINADORA

Prof. Dr. Márcio José das Chagas Moura (Orientador)
Universidade Federal de Pernambuco

Profa. Dra. Maisa Mendonça Silva (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Lucas Borges Leal da Silva (Examinador Interno)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Ficam aqui os meus agradecimentos, em primeiro lugar, à minha mãe, que me deu suporte durante toda a minha formação; aos meus amigos do colégio, em especial os “Kazokus”, que acompanharam toda essa jornada de perto. Aos amigos que conheci durante a minha primeira graduação em Engenharia de Controle e Automação, à equipe Maracatronics e a todo o time de futebol de robôs, que moldou muito da minha trajetória acadêmica — Júlia, Pedro, Pinho, Dani, Thiago — que estiveram comigo em todos os campeonatos que participamos e organizamos, seja na LARC, na OBR ou na RoboCup. Eram os famosos “paus para toda obra”.

Aos professores e colegas com quem tive a oportunidade de trabalhar no Voxar Labs, que foi a porta de entrada para meus estudos em visão computacional; e, em seguida, ao CEERMA, o segundo laboratório que me acolheu durante a graduação e incentivou todas as minhas iniciativas de trabalho e empreendedorismo.

Agradeço à July, que me orientou e acompanhou todo o processo de desenvolvimento, nos altos e baixos, sempre com a resposta certa para cada momento. Aos professores Márcio e VT, que foram duas grandes referências de incentivo e que enxergam a ciência como engrenagem para o desenvolvimento da sociedade — por isso reconhecem a importância das startups que surgem na academia.

Agradeço aos meus companheiros da InSpace, Pedro e Jorge — sem eles, com certeza eu estaria seguindo outros rumos na vida. Mas o sangue que eles dão pela nossa visão de futuro me incentiva sempre a continuar nesse caminho.

Por fim, agradeço ao apoio financeiro do Programa de Recursos Humanos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis — PRH-ANP-UFPE-Fapesp 38.1 — que acreditou no meu trabalho e contribuiu para mais uma qualificação profissional.

RESUMO

As indústrias de transformação se caracterizam pela complexidade de suas operações e pelo elevado risco ocupacional. Estudos extensivos sobre acidentes industriais apontam que os desvios operacionais causados por fatores humanos figuram entre as principais causas desses eventos. Nesse contexto, uma indústria do setor de óleo e gás (O&G) propôs uma série de desafios relacionados à segurança do trabalho, considerando a dificuldade das equipes em monitorar todos os desvios operacionais. A partir desse cenário, surgiu a motivação para investigar como realizar uma gestão preventiva mais eficaz da segurança no trabalho. Dentre eles, os desvios relacionados ao uso de equipamentos de proteção individuais foram destaque, que neste trabalho foram estudados os usos de capacete, luva e vestimenta de proteção. Além disso, a segurança é um tema comum a diversas indústrias, e para demonstrar a adaptabilidade da solução, também houve interesse por parte de duas outras empresas (uma no setor de vidro e outra da indústria alimentícia) que contribuíram para a execução de testes de rastreamento de pessoas e veículos. Diante disso, o presente trabalho buscou desenvolver um sistema de monitoramento com visão computacional (VC), capaz de identificar, em tempo real, o uso correto de EPIs, além de rastrear a movimentação de pessoas e veículos em ambientes industriais. Para tal, foi implementado um modelo baseado no *Faster R-CNN*, utilizando dados combinados de *datasets* públicos e de coletas realizadas em campo. Os resultados obtidos demonstram que o sistema é eficaz na detecção de desvios operacionais, alcançando mais de 70% de *mean Average Precision* (mAP) em todos os testes e atingindo um máximo de 94% de mAP na detecção de capacetes. O sistema mostrou-se capaz de operar em tempo real e adaptável a diferentes contextos industriais, tendo sido validado em cenários de três setores distintos. Essa solução oferece às equipes de segurança o potencial de realizar um monitoramento contínuo nas principais áreas de atividade das plantas industriais, facilitando a identificação de causas recorrentes de desvios operacionais. Isso possibilita a implementação de ações corretivas imediatas e até mesmo de treinamentos direcionados para fortalecer a cultura de segurança nas empresas. Com sua capacidade de adaptação, a tecnologia proposta revela o potencial da visão computacional como uma ferramenta versátil para enfrentar desafios complexos e diversos no ambiente industrial.

Palavras-chave: Gestão de riscos; Visão computacional; Indústria de Óleo e Gás; Aprendizagem de máquina; Equipamento de proteção individual; Segurança no trabalho.

ABSTRACT

Manufacturing industries are characterized by the complexity of their operations and a high level of occupational risk. Extensive studies on industrial accidents indicate that operational deviations caused by human factors are among the leading causes of such events. In this context, a company in the oil and gas (O&G) sector proposed a series of challenges related to occupational safety, given the difficulty teams face in monitoring all operational deviations. This scenario motivated the investigation of how to achieve more effective preventive management of workplace safety. Among the challenges, deviations related to the use of personal protective equipment (PPE) stood out, specifically the use of hardhats, gloves, and safety vest, which were the focus of this study. Additionally, safety is a cross-cutting issue in various industries, and to demonstrate the adaptability of the solution, two other companies, one in the glass sector and another in the food industry, also took part in the project by contributing to testing scenarios involving people and vehicle tracking. Accordingly, this work aimed to develop a computer vision (CV) monitoring system capable of identifying, in real time, the correct use of PPE and tracking the movement of people and vehicles in industrial environments. To achieve this, a model based on Faster R-CNN was implemented, using a combination of public datasets and field-collected data. The results demonstrate the system's effectiveness in detecting operational deviations, achieving over 70% mean Average Precision (mAP) in all tests and up to 94% mAP in helmet detection. The system proved capable of operating in real time and adapting to different industrial contexts, having been validated in scenarios from three distinct sectors. This solution provides safety teams with the ability to perform continuous monitoring in key operational areas of industrial plants, facilitating the identification of recurring causes of operational deviations. It enables the implementation of immediate corrective actions and even targeted training programs to strengthen safety culture within companies. With its adaptive capability, the proposed technology highlights the potential of computer vision as a versatile tool for addressing complex and diverse challenges in industrial environments.

Keywords: Risk Management; Computer Vision; Oil and Gas Industry; Machine Learning; Personal Protective Equipment; Occupational Safety.

LISTA DE ILUSTRAÇÕES

Figura 1: Artigos publicados entre 2010 e 2024 no contexto de segurança na indústria de O&G.	19
Figura 2: Representação de uma imagem colorida em dados computacionais.	25
Figura 3: Padronização da orientação dos pixels.	27
Figura 4: Esquerda - Erro de orientação da imagem e anotação; Direita - Orientação correta.	27
Figura 5: Um paralelo do processo de aprendizagem de máquina e a biologia humana.	29
Figura 6: Ilustração de um modelo MLP genérico.	30
Figura 7: Max-Pooling com tamanho de janela (2x2).	32
Figura 8: Fluxo de dados no LeNet. A entrada é um número escrito à mão, a saída é a probabilidade em 10 possíveis resultados.	33
Figura 9: Ilustração da Busca Seletiva.	36
Figura 10: Modelo simplificado da R-CNN.	36
Figura 11: Modelo simplificado do Fast R-CNN.	37
Figura 12: Modelo simplificado do <i>Faster R-CNN</i>	37
Figura 13: Ilustração do RPN.	38
Figura 14: Representação da saída do Keypoint R-CNN.	39
Figura 15: Tipos de modificação em imagens.	41
Figura 16: Fluxograma de desenvolvimento.	43
Figura 17: Fotos de algumas câmeras instaladas dentro da indústria petroquímica parceira.	44
Figura 18: (a) exemplo de oclusão na imagem. (b) exemplo de anotação dos objetos de interesse	45
Figura 19: Imagens da indústria parceira, anotadas para posterior adição no dataset original (Brilho e saturação alterados para não exposição de logomarcas).	45
Figura 20: Fluxograma de operação do sistema.	49
Figura 21: Sequência de frames próximos com identificações erradas.	50
Figura 22: Resultados de data augmentation. (a) Imagem com ruído. (b) Imagem com saturação aumentada. (c) imagem invertida e com ruído. (d) Imagem invertida, com ruído e com coloração em escala de cinza.	52
Figura 23: Gráfico mAP e de perda ao longo das 280 épocas do modelo YOLOv12.	53
Figura 24: Gráfico de mAP ao longo de 50 épocas para o modelo Faster R-CNN - ResNet50 FPN.	54

LISTA DE ILUSTRAÇÕES

Figura 25: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo Faster R-CNN - ResNet50 FPN.	55
Figura 26: Gráfico de mAP ao longo de 50 épocas para o modelo Faster R-CNN - MobileNetv3 Large 320 FPN.	56
Figura 27: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo Faster R-CNN - MobileNetv3 Large 320 FPN.	56
Figura 28: Gráfico de mAP ao longo de 50 épocas para o modelo Faster R-CNN - MobileNetv3 Large FPN.	57
Figura 29: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo Faster R-CNN - MobileNetv3 Large FPN.	58
Figura 30: Teste de detecção de EPIs em condições padrão de iluminação e distância.....	59
Figura 31: Teste de detecção de EPIs em ambiente poluído.	60
Figura 32: Teste de detecção de EPIs em distância e oclusão.....	61
Figura 33: Testes de detecção de pose, destacado pelos círculos das juntas conectados por linhas pretas.	62
Figura 34: Detecção de veículos industriais em ambientes externo e interno.....	64
Figura 35: Teste de interação de veículo-pessoa. Em cima da imagem de alerta tem a descrição do desvio.....	65
Figura 36: Testes de veículo-sinalização/homem-sinalização; à esquerda, um veículo passando por cima da faixa, no meio uma pessoa andando corretamente e em seguida fora da faixa de pedestres. Em cima das imagens de alerta tem a descrição do desvio.....	65
Figura 37: Teste de uso correto de EPI. Na esquerda a detecção sem correlação com os keypoints do corpo. Na direita, a integração com keypoints está ativa, sendo perceptível pelos círculos coloridos nos pontos de interesse.....	66

LISTA DE TABELAS

Tabela 1 - Número de publicações de artigos de IA na indústria de O&G nos últimos 5 anos.	18
Tabela 2 - Número de publicações de artigos de VC na indústria de O&G nos últimos 5 anos.	18
Tabela 3 - Exemplos de modelos CNN	33
Tabela 4 - Balanceamento inicial do dataset, já contabilizando data augmentation.	51
Tabela 5 - Resultado da avaliação dos modelos em um mesmo dataset de testes.....	58
Tabela 6 - Matriz de confusão para detecção de EPIs.	62
Tabela 7 - Métricas de avaliação por classe de EPI.	63

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVO	14
1.2.1	<i>Objetivo Geral</i>	<i>14</i>
1.2.2	<i>Objetivo Específicos</i>	<i>15</i>
1.2	ORGANIZAÇÃO DO TRABALHO	15
1.3	CATEGORIZAÇÃO DA METODOLOGIA	15
2	REVISÃO DA LITERATURA	18
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	ANÁLISE DE RISCO.....	24
3.2	PROCESSAMENTO DE IMAGEM	25
3.2.1	<i>Tratamento de imagens.....</i>	<i>26</i>
3.3	APRENDIZAGEM DE MÁQUINA.....	28
3.3.1	<i>Multi Layer Perceptron (MLP).....</i>	<i>30</i>
3.3.2	<i>Convolutional Neural Network (CNN)</i>	<i>31</i>
3.3.3	<i>You Only Look Once (YOLO)</i>	<i>34</i>
3.3.4	<i>Region Based Convolutional Neural Network (R-CNN)</i>	<i>35</i>
3.3.5	<i>Keypoint R-CNN</i>	<i>38</i>
3.3.6	<i>Métricas de avaliação.....</i>	<i>39</i>
3.4	DATA AUGMENTATION	40
4	METODOLOGIA	42
4.1	DESCRIÇÃO DO PROBLEMA.....	43
4.2	COLETA DE DADOS E ANOTAÇÃO DO DATASET	44
4.3	DATA AUGMENTATION	46
4.4	TREINAMENTO	46

SUMÁRIO

4.5	LÓGICA DA APLICAÇÃO	48
5	RESULTADOS.....	51
5.1	COLETA DE DADOS E <i>DATA AUGMENTATION</i>	51
5.2	TREINAMENTO	52
5.3	MODELOS DE DETECÇÃO.....	59
5.4	DETECÇÃO DE DESVIO OPERACIONAL	64
5.5	CONTRIBUIÇÕES	67
6	CONCLUSÃO	68
	REFERÊNCIAS.....	70

1 INTRODUÇÃO

O petróleo e o gás natural ainda são responsáveis por cerca de 70% do fornecimento de energia global. Além disso, são matérias-primas na produção de plásticos, cosméticos, tintas, solventes e outros produtos derivados. Portanto, esses elementos são estratégicos para decisões econômicas e interesses políticos (VASUDEVAN; ZOLGHADRI; MAKAREM, 2023).

Com o crescente interesse das nações na transição energética (DELOITTE, 2022), muita especulação tem surgido sobre o futuro do setor de óleo e gás (O&G). Apesar disso, sendo um dos maiores pilares da matriz energética mundial, a demanda por petróleo continua em ascensão (DELOITTE, 2022). O estudo realizado pela Deloitte (2024) aponta que esse setor deve gerar 800 bilhões de dólares em fluxo de caixa livre, o que indica o potencial para investir em novas tecnologias e negócios. O mesmo estudo apresenta as cinco principais tendências do setor, sendo uma delas a adoção da inteligência artificial (IA) para soluções inovadoras e criação de valor para a indústria. Nesse contexto, várias soluções baseadas em IA vêm sendo propostas com o objetivo de reduzir perdas e aumentar a eficiência produtiva.

A indústria de O&G é um dos setores com maior nível de risco (AYDIN; SEKER; ŞEN, 2022), com alto potencial de causar danos aos colaboradores, ao meio ambiente e ao patrimônio da própria indústria (MACÊDO, et al., 2022; MAIOR et al., 2022). Esse risco decorre, principalmente, das características intrínsecas dos materiais processados, armazenados e transportados por essa indústria, como a inflamabilidade, toxicidade e explosividade. Soma-se a isso, o alto grau de complexidade das operações nas etapas de exploração e refino, principalmente nos ambientes *offshore*, onde a disponibilidade de suporte é limitada (MARIANO; COUTINHO, 2022; RANGEL, 2018).

Segundo Nwankwo et al. (2022), 90% dos acidentes catastróficos no setor têm o fator humano como principal causa. Nesse contexto, a hierarquia de controles de risco ocupacional ressalta a importância dos Equipamentos de Proteção Individual (EPIs), que representam a última barreira de proteção do colaborador contra os perigos presentes no ambiente operacional (NIOSH, 2024).

É importante ressaltar que, além dos impactos já mencionados, tais incidentes têm um impacto econômico considerável. Segundo o último dado levantado pelo Observatório de Segurança e Saúde no Trabalho do Brasil, no setor de O&G brasileiro, em 2022 foram concedidas mais de 148 mil Benefícios Previdenciários Acidentários. Entre os anos 2000 e

2022, foram concedidos mais de 5 milhões de benefícios (SMARTLABBR, 2022). A esses números somam-se outros custos, frequentemente subestimados, como aposentadorias por invalidez, pensões por óbito, prejuízos ambientais e danos patrimoniais à própria indústria. Além disso, os acidentes desencadeiam longos processos de investigação e de disputas legais, resultando em ônus adicionais para empresas, trabalhadores e a sociedade em geral. Tais impactos evidenciam a urgência de medidas preventivas mais eficazes.

Apesar dos esforços em estabelecer diretrizes e procedimentos de segurança, a manutenção de um desempenho consistente em termos de segurança na indústria de O&G permanece um desafio. Abordagens convencionais muitas vezes enfrentam dificuldades para prever e controlar eficazmente o comportamento humano (THEOPHILUS *et al.*, 2017). Em muitos casos, as equipes de Saúde e Segurança do Trabalho (SST) apenas conseguem identificar desvios de conduta ou falhas após sua ocorrência, pois a detecção antecipada depende de um monitoramento constante que, quando realizado exclusivamente por indivíduos, torna-se inviável. Isso também é consequência da escala operacional de uma indústria: segundo a NBR 04 (MINISTÉRIO DO TRABALHO E EMPREGO, 2024), para cada turno em um ambiente com 5.000 colaboradores são necessários 10 técnicos em segurança do trabalho para supervisionar a operação. Mesmo que a indústria contrate mais técnicos para o monitoramento constante, essa solução acaba não sendo financeiramente sustentável, o que evidencia uma oportunidade para desenvolvimento de tecnologias que consigam suprir essa demanda de forma mais eficiente.

Diante desse cenário, a aplicação de técnicas de visão computacional (VC) surge como uma solução promissora para a melhoria da segurança no trabalho. A relevância dessa abordagem é evidenciada por iniciativas concretas do próprio setor de O&G, como a chamada de inovação aberta realizada por uma indústria petroquímica, que selecionou o projeto da InSpace¹, startup fundada na Universidade Federal de Pernambuco. Um dos desafios consiste em identificar o uso correto de EPIs durante operações na planta. É importante ressaltar que a simples detecção de um objeto na imagem não garante seu uso adequado; a ferramenta deve ser capaz de reconhecer, por exemplo, que o capacete está sendo utilizado corretamente pelo colaborador, e não apenas presente na cena. Esse projeto serviu de motivação para o presente trabalho

Este trabalho propõe um sistema de detecção e monitoramento de atividades em tempo real, com foco na segurança do trabalho. O objetivo principal é integrar diferentes

¹ O autor deste manuscrito é sócio-fundador e CEO da InSpace

modelos de visão computacional para detecção de situações de risco e auxiliar a tomada de medidas preventivas. Os modelos serão responsáveis pela detecção de: (i) EPIs, (ii) Pessoas e (iii) Veículos. Para o treinamento do modelo de detecção, foi necessário utilizar uma base de dados contendo imagens com os objetos de interesse e foram utilizadas também imagens obtidas na própria indústria petroquímica parceira. No entanto, por motivos de confidencialidade e para preservar a identidade dos colaboradores, a divulgação de imagens provenientes dessa planta será limitada.

O objetivo principal deste trabalho é a melhoria da segurança em ambientes industriais, especialmente no setor de O&G, a partir da análise de riscos operacionais e da otimização de processos de monitoramento. Além disso, para demonstrar a adaptabilidade da solução proposta para diferentes contextos, o modelo desenvolvido foi aplicado em indústrias diferentes, mais especificamente em uma indústria alimentícia e em uma indústria de vidro.

A formação na área de Engenharia de Produção proporciona uma visão sistêmica que integra aspectos técnicos, humanos e organizacionais, essenciais para propor soluções voltadas à eficiência, à confiabilidade e à sustentabilidade das operações produtivas. Ao empregar tecnologias emergentes para apoiar a gestão da segurança, este trabalho reforça o papel estratégico do engenheiro de produção na identificação e mitigação de riscos, contribuindo para a melhoria contínua dos sistemas industriais (CAMPOS, 2019).

Espera-se que o sistema desenvolvido possa auxiliar as equipes de SST a obter respostas mais eficientes a atos e condições inseguras, melhorando os índices de prevenção de acidentes.

1.1 OBJETIVO

1.2.1 Objetivo Geral

O objetivo geral deste trabalho consiste em avaliar a capacidade de uso de modelos de VC para o monitoramento de atividades na indústria, com ênfase na verificação do uso correto de EPIs, no rastreamento de veículos e no controle da circulação de pessoas em áreas de risco. Além disso, a pesquisa busca garantir que as soluções propostas sejam compatíveis com sistemas de videomonitoramento em tempo real e comercialmente viáveis, de modo a atender às demandas práticas das equipes de SST.

1.2.2 Objetivo Específicos

- Desenvolvimento de um modelo de VC para a detecção de EPIs, pessoas e veículos em ambientes industriais.
- Desenvolvimento de algoritmos com a lógica para identificar desvios operacionais: distância segura entre pessoas e veículos e uso correto de EPI.
- Avaliar a eficiência do modelo *Faster R-CNN* para detecção em tempo real, garantindo viabilidade comercial.
- Integração dos modelos de detecção em um sistema unificado, capaz de emitir alertas de risco.
- Avaliação da eficácia do sistema em diferentes cenários operacionais.

1.2 ORGANIZAÇÃO DO TRABALHO

O Capítulo 2 apresenta uma revisão da literatura sobre as tendências do uso de aplicações de aprendizagem de máquina na indústria, as principais áreas de investimento e o posicionamento da segurança no trabalho nesse contexto. O Capítulo 3 introduz conceitos de análise de risco e aborda os aspectos essenciais para a construção de uma solução baseada em VC para detecção de objetos, de acordo com os objetivos de segurança ocupacional, como o treinamento supervisionado, redes neurais convolucionais e modelos específicos, incluindo YOLO e *Faster-RCNN*. Também serão discutidas as métricas de avaliação utilizadas na problemática de detecção de objetos e apresentadas outras informações pertinentes ao projeto.

A metodologia proposta é detalhada no Capítulo 4, incluindo o fluxo de desenvolvimento, a base de dados utilizada, o treinamento dos modelos de detecção e a aplicação desses modelos. O Capítulo 5 apresenta os resultados obtidos, com suas análises quantitativas e qualitativas, avaliando a eficácia e a usabilidade dessa solução em contextos operacionais reais na indústria. Por fim, o Capítulo 6 trata das conclusões e limitações do trabalho, e sugestões para pesquisas futuras.

1.3 CATEGORIZAÇÃO DA METODOLOGIA

Este estudo enquadra-se como uma pesquisa aplicada, uma vez que busca resolver um problema prático identificado em ambientes industriais: a dificuldade no monitoramento e na detecção precoce de desvios operacionais relacionados à segurança do trabalho. Além disso, trata-se de uma pesquisa exploratória, pois tem como objetivo verificar a aplicabilidade da tecnologia para a resolução do problema no cenário proposto, obtendo uma compreensão aprofundada de um contexto específico.

A abordagem adotada pelo estudo é quantitativa, visto que utiliza indicadores objetivos para a avaliação do desempenho dos modelos implementados. A análise será conduzida com base em métricas como precisão e taxas de falsos positivos/negativos. Ademais, as regras operacionais de segurança, que definem as conformidades e inconformidades, serão determinadas de acordo com os cenários específicos observados em campo.

Com base nessas premissas, os constructos teóricos para o estudo de caso foram definidos a partir dos objetivos específicos da pesquisa:

1. Constructo: Conformidade com EPIs.

Indicadores: Taxa de detecção correta de EPIs.

2. Constructo: Rastreamento de pessoas e veículos.

Indicadores: Precisão na detecção de pessoas e veículos; taxa de falsos positivos e negativos.

3. Constructo: Capacidade de monitoramento por VC.

Indicadores: Desempenho do sistema em diferentes condições operacionais (por exemplo, iluminação e ângulos de câmera).

A metodologia envolve a realização de um estudo de caso, no qual os modelos serão testados com dados coletados em cenários industriais reais. A ferramenta será aplicada de acordo com as normas de segurança do respectivo local. Também será elaborado um protocolo experimental padronizado, que orientará a replicação dos experimentos realizados.

O método científico adotado é o fenomenológico, pois propõe descrever e interpretar os fenômenos diretamente nos locais de estudo, considerando as particularidades do ambiente e a experiência prática com a aplicação da tecnologia (CAUCHICK-MIGUEL *et al.*, 2012).

A técnica de pesquisa utilizada baseia-se na documentação direta, em que os dados serão levantados no local onde os eventos ocorrem. A estratégia metodológica combina pesquisa de laboratório, para a validação inicial dos modelos em ambiente controlado, com

pesquisa de campo, voltada à avaliação do desempenho da tecnologia em cenários reais (CAUCHICK-MIGUEL *et al.*, 2012).

2 REVISÃO DA LITERATURA

O uso de técnicas de IA na indústria de O&G vem seguindo uma tendência de crescimento nas últimas décadas. Em uma busca inicial, realizada com o objetivo de obter uma visão geral sobre o tema, na plataforma *Web of Science*, foram encontrados, entre 2010 e 2024, um total de 8.321 artigos publicados nessa temática, sendo os últimos cinco anos os mais representativos nesse avanço (Tabela 1). Para a pesquisa, foram utilizadas as palavras-chave traduzidas para o inglês: “Óleo e gás/petróleo/petroquímica”, “Indústria” e “Aprendizagem de Máquina/Inteligência artificial” e seus respectivos acrônimos.

Tabela 1 - Número de publicações de artigos de IA na indústria de O&G nos últimos 5 anos.

Ano	Publicações
2024	1.290
2023	1.157
2022	1.116
2021	803
2020	725

Fonte: O autor (2025).

Quando adicionadas as palavras-chaves “Visão computacional” ou “Processamento de imagem” no lugar de “aprendizagem de Máquina/Inteligência artificial”, para encontrar trabalhos que utilizam VC como abordagem para resolução de desafios da indústria, o número de artigos publicados cai para 1.613. Ao analisar os últimos cinco anos, percebe-se que, assim como o macrotema geral de IA, que está em crescimento, o uso de técnicas de VC também apresenta aumento, porém com uma taxa menos acentuada (Tabela 2). Esse dado pode indicar uma possível dificuldade na implementação dessas técnicas para o estudo das práticas industriais.

Tabela 2 - Número de publicações de artigos de VC na indústria de O&G nos últimos 5 anos.

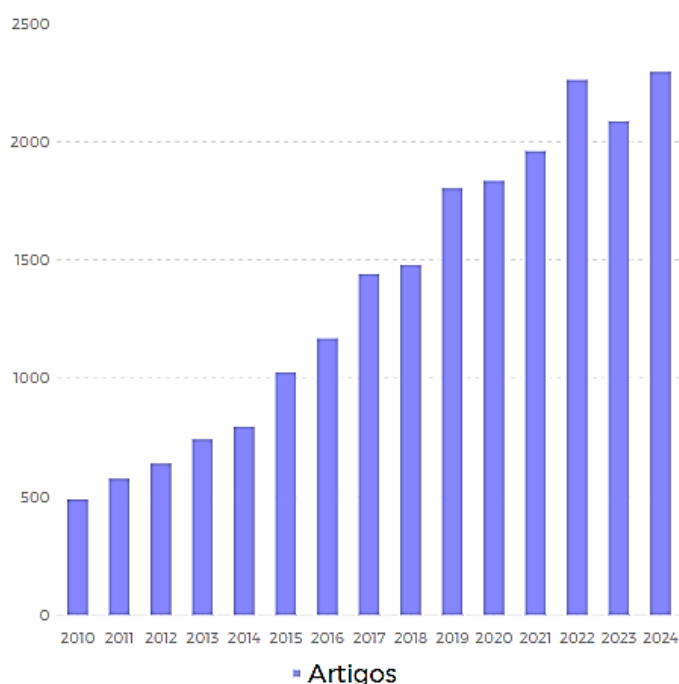
Ano	Publicações
2024	238
2023	193
2022	248
2021	175
2020	145

Fonte: O autor (2025).

A partir desse conhecimento, para direcionar a busca ao contexto deste trabalho, foram selecionadas palavras-chaves que foram combinadas e traduzidas para o inglês a fim de realizar a pesquisa na base de dados internacional *Web of Science*. As palavras-chave selecionadas foram: (i) “Visão Computacional”, (ii) “Fatores Humanos”, (iii) “Segurança”, (iv) “Óleo e Gás”, (v) “Equipamento de Proteção Individuais”, (vi) “Risco”, (vii) “Processamento de Imagem”, além de seus respectivos acrônimos.

Na primeira iteração da busca, o objetivo foi delimitar o tamanho do campo de estudo abordado, sem buscar uma solução específica. Utilizando as palavras-chaves (ii), (iii), (iv), (v) e (vi), e aplicando filtros para artigos publicados entre os anos de 2010 e 2024 foram encontrados 20.590 trabalhos. Nessa busca foi observado um número crescente de publicações ao longo dos anos (Figura 1).

Figura 1: Artigos publicados entre 2010 e 2024 no contexto de segurança na indústria de O&G.



Fonte: O autor (2025).

Ao especificar mais a busca, utilizando apenas as palavras-chave relacionadas a fatores humanos (ii) e EPIs (v) na indústria, o número de trabalhos encontrados foi reduzido para 310 artigos. Ao avaliar essas pesquisas, percebe-se que a maioria faz uso de frameworks de análise de dados históricos para compreender as principais causas de acidentes e orientar novas práticas gerenciais de segurança (ROBB; MILLER, 2012; YUAN et al., 2019; ZHANG; QU; CHEN, 2022), ou para aprimorar processos existentes, como checklists voltados à identificação de causas de erros humanos em inspeções

periódicas (YANG; KWON, 2022). Ou seja, ainda existe uma lacuna no desenvolvimento de soluções capazes de intervir em tempo real na operação, com o objetivo de prevenir acidentes decorrentes de desvios operacionais relacionados a fatores humanos.

No trabalho de Al-Mhdawi et al. (2024), ao realizar uma Análise de Modos de Falha e seus Efeitos (*Failure Mode and Effect Analysis* – FMEA) baseada na Teoria Fuzzy, os autores foram capazes de analisar, de forma quantitativa, os níveis de significância dos riscos na indústria de O&G. O estudo identificou 41 fatores de risco, classificados em sete categorias: gerenciais, técnicos e de qualidade, financeiros e econômicos, de saúde, segurança, ambientais, jurídicos e relacionados às partes interessadas. A avaliação final concluiu que a não conformidade com as normas de uso de EPIs é o fator de risco mais significativo entre todas as categorias analisadas no setor.

Outra vertente para a mitigação do risco ocupacional na indústria recai sobre a redução da exposição do colaborador a ambientes perigosos. Essa estratégia pode ser viabilizada, principalmente, por meio da automação e da robótica, que possibilitam substituir a presença humana em áreas de risco (BENSACI *et al.*, 2023). Em outros contextos, é possível reduzir o número de pessoas expostas simultaneamente por meio da digitalização de processos ou do uso de novas ferramentas que permitam a fusão de funções em uma planta industrial, diminuindo a necessidade de operadores em um mesmo processo (WANASINGHE *et al.*, 2024). Entretanto, nem sempre é viável eliminar a presença de trabalhadores nessas áreas, seja por limitações técnicas ou econômicas.

Ao aplicar os filtros relacionados ao uso de VC (i) e (vii) na indústria de O&G, foram encontrados 1.435 trabalhos. Grande parte dessas aplicações está voltada à automação de processos produtivos ou ao uso em robôs (SILVA et al., 2022; WANG et al., 2024; YESSENBAYEV; KOZHIRBAYEV; SHINTEMIROV, 2022). Quando combinadas todas as palavras-chave (i), (ii), (iii), (iv), (v), (vi), (vii), relacionando fatores humanos e risco na indústria com o uso de processamento de imagem, o número de trabalhos encontrados reduz-se para 164. Nota-se, nesses estudos, o uso promissor da VC para a identificação de falhas estruturais e corrosão (CHEN et al., 2024; HASAN et al., 2024). No entanto, entre todos os trabalhos identificados, apenas três mencionaram o uso de visão computacional no título ou no resumo em associação com fatores humanos.

Ramos et al. (2022) investigaram a detecção de sonolência em operadores de centros de controle, onde atividades monótonas e de alto nível de estresse podem levar à fadiga. O

estudo utilizou o monitoramento de sinais vitais para medir o nível de estresse, aliado à análise de imagens de câmeras para rastrear movimentos faciais e detectar sonolência. Embora relevante para a segurança operacional, esse cenário é bastante atípico no contexto de monitoramento com VC em ambientes industriais, pois trata-se de um ambiente controlado, com apenas um colaborador observado, centralizado e de frente para a câmera, permanecendo quase estático, o que facilita a análise. Essa configuração não reflete a maioria das operações industriais, onde há maior complexidade e variabilidade nas imagens.

Casas et al. (2024) propôs identificar possíveis riscos em operações de tubulações subterrâneas de óleo, por meio do rastreamento de veículos nas proximidades utilizando imagens aéreas de drones para treinar um modelo YOLO capaz de identificar sete tipos de veículos. Apesar dos resultados promissores, os autores destacam a necessidade de aumentar a robustez do modelo, considerando as restrições impostas por fatores climáticos nas imagens de drone, além da intenção de incorporar novos modelos para enriquecer a análise de risco.

Embora esses estudos sejam relevantes para a área de segurança, poucos abordam diretamente o problema do uso de EPIs, considerado um dos fatores mais importantes para a proteção humana.

Neste contexto o trabalho de Hodge (2024) explorou estudos de caso sobre o uso de VC nas operações da indústria de O&G, incluindo a identificação de capacetes. Em um dos casos, foi apresentado um sistema robusto para monitoramento da atividade de carregamento de caminhões, com um modelo de VC treinado para detectar a conexão de cabos, o bombeamento de combustível e a operação de risers, além de verificar se o operador utilizava os EPIs adequados. Contudo, os resultados indicaram que apenas o capacete era identificado, sem garantir que ele estava sendo usado corretamente ou apenas presente na cena. Além disso, o autor não detalhou o modelo utilizado nem os critérios de treinamento para os eventos-chave da operação, levantando dúvidas sobre a robustez do sistema frente a variações de posicionamento de câmeras ou cenários operacionais.

Considerando essas limitações, a metodologia proposta neste trabalho visa abordar um número maior de EPIs, verificar o uso correto desses equipamentos e integrar modelos adicionais para detectar pessoas e veículos, permitindo a identificação de eventos complexos com maior assertividade.

Apesar do número ainda reduzido de aplicações na indústria de O&G, observa-se em outros setores o uso crescente de VC para segurança relacionada a fatores humanos. Na indústria da construção civil, por exemplo, a detecção de EPIs com técnicas de VC tem sido utilizada para reduzir custos e agilizar a intervenção das equipes gestoras (FERDOUS; AHSAN, 2022). A maioria desses estudos emprega o modelo YOLO, amplamente utilizado para a detecção em tempo real de objetos em imagens, por sua capacidade de gerar “caixas delimitadoras” que identificam a localização e classificação de objetos (LI et al., 2024; WU, 2023).

No entanto, alguns desafios permanecem. Para o processamento em tempo real, é necessário elevado poder computacional. No trabalho de Li et al. (2024), os autores alcançaram taxas elevadas de quadros por segundo (FPS) utilizando uma GPU RTX 3090, porém com apenas uma classe de objeto monitorada (capacete). Em cenários industriais, diversos elementos precisam ser acompanhados simultaneamente para uma avaliação abrangente do risco ocupacional. Ferdous e Ahsan (2022) treinaram um modelo YOLO para detectar oito classes de objetos (quatro cores diferentes de capacetes, um tipo de vestimenta, óculos de proteção, além da cabeça e corpo dos trabalhadores). A maior complexidade do modelo demandou maior capacidade computacional, sendo os testes realizados com uma GPU Tesla K80, com variação de FPS entre 6 e 13, dependendo da versão do modelo utilizada. Esses dados indicam a necessidade de ponderar o *trade-off* entre desempenho e custo computacional para aplicações em tempo real na indústria.

Além disso, apenas detectar a presença de um objeto na imagem não é suficiente para garantir a conformidade operacional. Por exemplo, é necessário verificar se o capacete está sendo usado corretamente na cabeça do trabalhador e não apenas segurado na mão. Nos estudos apresentados anteriormente, os modelos apenas detectam a existência do objeto na cena e não o seu uso efetivamente.

Nesse sentido, algumas pesquisas têm explorado sistemas inteligentes capazes de monitorar operações em máquinas de menor escala (AHN *et al.*, 2023), rastreando partes do corpo (mãos, pés, cabeça) e verificando se estão dentro de áreas seguras durante a operação. Outras abordagens analisam a interação entre veículos e pessoas em ambientes industriais maiores (PEREZ et al., 2014; ZHU et al., 2022), priorizando a eficiência e a acurácia da detecção para calcular distâncias com precisão e prevenir interações perigosas. Esse tipo de sistema permite que gestores de segurança recebam dados antecipados sobre

situações de risco e gerem alertas imediatos para evitar acidentes.

Com base nessas observações, este trabalho propõe a implementação de um sistema para a detecção múltipla de objetos de interesse, como EPIs, pessoas e veículos. O sistema será capaz de cruzar essas informações para verificar, por exemplo, se veículos circulam próximos a pessoas, se sinalizações são respeitadas e se os EPIs estão sendo utilizados corretamente, identificando desvios operacionais que possam comprometer a segurança. Além disso, este estudo busca avaliar o desempenho do modelo *Faster R-CNN* para múltiplas detecções com baixo custo computacional, possibilitando aplicações em tempo real e viabilizando o uso em operações industriais reais.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 ANÁLISE DE RISCO

A análise de risco é um método amplamente utilizado, especialmente na indústria de O&G, para identificar, avaliar e mitigar perigos que possam resultar, entre outras consequências, em acidentes ou doenças ocupacionais. A norma ISO 31000 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018) define risco como o “efeito da incerteza sobre os objetivos”. No contexto da segurança do trabalho, o risco refere-se à possibilidade de ocorrência de eventos que comprometam a saúde e a integridade física dos trabalhadores, também conhecido como segurança ocupacional.

Nesse sentido, a análise de risco está diretamente relacionada à identificação de condições perigosas no ambiente de trabalho, como a exposição a agentes nocivos, inadequações ergonômicas, uso inadequado de equipamentos e comportamentos inseguros. Essa abordagem é regulamentada por normas brasileiras, como a NR-01 (BRASIL, 2019) e a NR-09 (BRASIL, 2025), que orientam o empregador a adotar um Programa de Gerenciamento de Riscos (PGR) e um Programa de Prevenção de Riscos Ambientais (PPRA), respectivamente. Tais diretrizes reforçam a importância da hierarquia de controle de riscos, priorizando a eliminação e a substituição de perigos antes do uso de EPIs, sendo estes a última barreira de proteção para o colaborador.

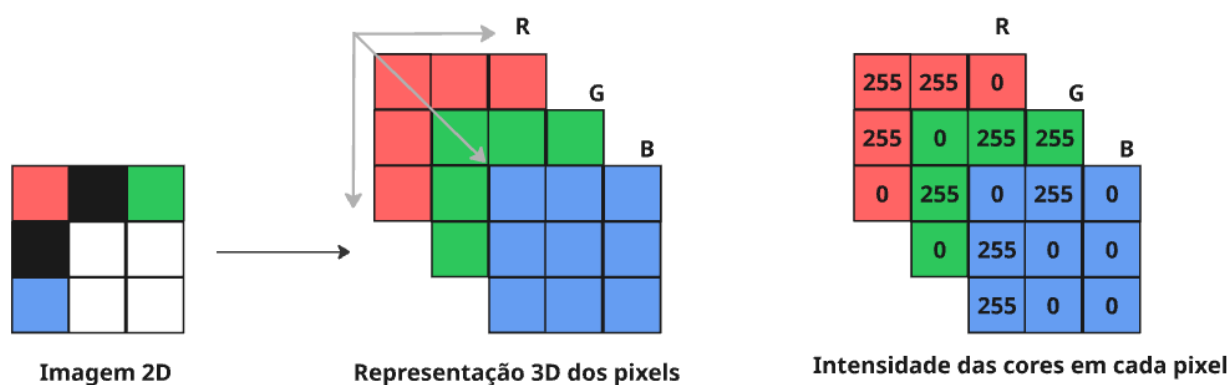
É importante destacar que a segurança ocupacional, além de sua relevância própria, integra o conceito mais amplo de segurança operacional, que abrange a integridade dos processos e sistemas produtivos como um todo. No livro *Managing the Risks of Organizational Accidents* (REASON, 2016), o autor associa essa abordagem à prevenção de falhas técnicas, organizacionais ou humanas que possam comprometer a continuidade das operações. Em sistemas complexos, como plantas industriais, a segurança operacional envolve a gestão da confiabilidade, do desempenho e da interação entre pessoas, tecnologias e processos.

Assim, a segurança operacional configura-se como um fator-chave não apenas para garantir a integridade dos colaboradores, mas também para assegurar a continuidade da operação como um todo. Um acidente pode levar à paralisação de uma linha produtiva, ocasionando falhas críticas no sistema. Nesse contexto, garantir o uso correto de EPIs é um dos pilares da segurança operacional (NIOSH, 2024), assim como a mitigação de desvios operacionais provocados por fatores humanos. Por isso, neste trabalho propõe-se o uso de visão computacional (VC) como uma abordagem para enfrentar esses desafios.

3.2 PROCESSAMENTO DE IMAGEM

Para os algoritmos de visão computacional (VC), uma imagem nada mais é do que uma matriz bidimensional de pixels, independentemente de ser monocromática ou colorida (ZHANG *et al.*, 2021). O que varia entre os tipos de imagem é a quantidade dos chamados “canais” de informação. No caso da imagem colorida tradicional, por exemplo, ela é composta por três canais: *Red* (vermelho), *Green* (verde) e *Blue* (azul), simplificados para “imagem RGB” (Figura 2). Isso significa que, em cada pixel da matriz, está inserido um vetor com três elementos contendo informações numéricas sobre cada uma dessas intensidades de cor, variando de 0 até 255. Com base nessa combinação de dados, o computador é capaz de formar as imagens que visualizamos, apresentando uma ampla variedade de cores resultante da combinação dos três canais.

Figura 2: Representação de uma imagem colorida em dados computacionais.



Fonte: O autor (2025).

De forma simplificada, ao compreender do que as imagens são compostas em termos computacionais, pode-se afirmar que todas as chamadas técnicas de visão computacional (VC) consistem em algoritmos desenvolvidos para ler e interpretar esse tipo de conjunto de dados, com o objetivo de extrair informações de maneira semelhante à visão humana. Esses algoritmos são baseados em aprendizado de máquina e redes neurais.

Um fator essencial para a construção de modelos de aprendizado de máquina e redes neurais é a elaboração de um conjunto de dados (*dataset*) representativo. A obtenção de imagens para a construção de um dataset pode ser um desafio, seja por limitações de tempo, acesso a dados, quantidade de imagens disponíveis na internet ou restrições de recursos em

geral. Ao preparar um conjunto de dados para treinamento, os modelos recebem imagens individuais como entrada. No entanto, quando há acesso a vídeos no contexto de interesse, é possível transformá-los em imagens estáticas. Vídeos nada mais são do que uma sequência de múltiplas imagens (*frames*) reproduzidas em alta velocidade, o que dá origem à expressão “quadros por segundo”. O conceito de extração de imagens a partir de vídeos consiste em isolar essas imagens e salvá-las separadamente.

Por se tratar de um volume elevado de imagens contidas em um vídeo, é fundamental selecionar quadros isolados com um intervalo de tempo adequado, a fim de evitar o salvamento de imagens muito semelhantes. Por exemplo, um método simples de extração consiste em, manualmente ou por meio de um algoritmo, reproduzir o vídeo e salvar uma imagem a cada 10 segundos. Assim, em um vídeo de 1 hora, seria possível extrair 360 imagens com essa abordagem.

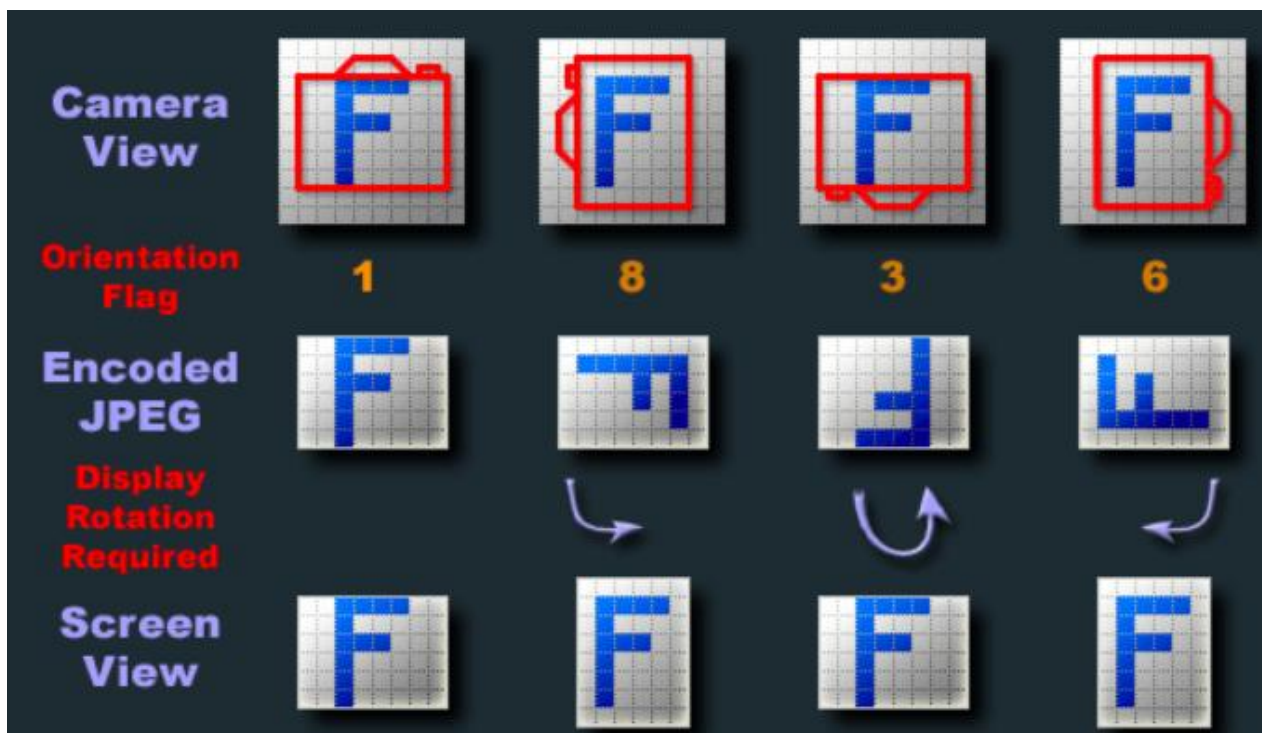
3.2.1 Tratamento de imagens

Na área de estudos em aprendizado de máquina, a etapa de pré-processamento de dados é tão importante quanto a coleta adequada destes. O objetivo é extrair o máximo desempenho do modelo. As técnicas de tratamento de imagens aplicadas neste estudo foram:

- **Ajuste de orientação:**

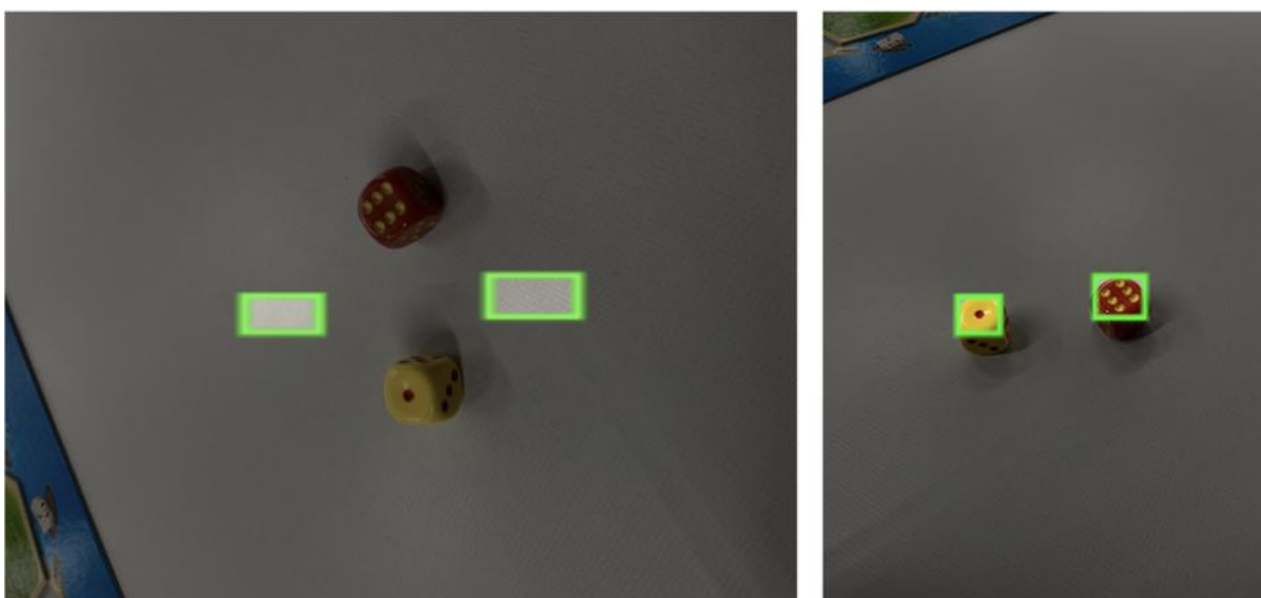
Quando uma imagem é capturada por um dispositivo, ela contém metadados que indicam como deve ser exibida em relação à forma como os pixels são armazenados no disco rígido do equipamento (Figura 3). Caso a aplicação responsável pela exibição das imagens não interprete corretamente esses metadados, podem ocorrer erros. Ou seja, se algumas imagens forem capturadas na orientação (x,y) e outras em (y,x) o modelo poderá ser alimentado com informações inconsistentes, conforme observado na Figura 4 (DWYER, 2020). Ao realizar a correção dessa orientação, o modelo torna-se mais robusto, sendo capaz de processar imagens provenientes de diferentes dispositivos e orientações, o que amplia sua aplicabilidade em diversos contextos.

Figura 3: Padronização da orientação dos pixels.



Fonte: JDHAO'S (2019).

Figura 4: Esquerda - Erro de orientação da imagem e anotação; Direita - Orientação correta.



Fonte: DWYER (2020).

- **Dimensão das imagens:**

A depender do dispositivo utilizado, as capturas podem ser feitas em diversas proporções, como 720p, 1080p ou 4K. Em termos computacionais, quanto maior a imagem, maior o custo de processamento e, conseqüentemente, mais pesado o modelo se torna para ser treinado. Por isso, os modelos de visão computacional (VC) utilizam proporções menores de imagem para facilitar o treinamento (REDMON *et al.*, 2015). Além disso, muitas arquiteturas de aprendizado de máquina exigem que os dados de entrada estejam todos no mesmo formato (GIRSHICK, 2015).

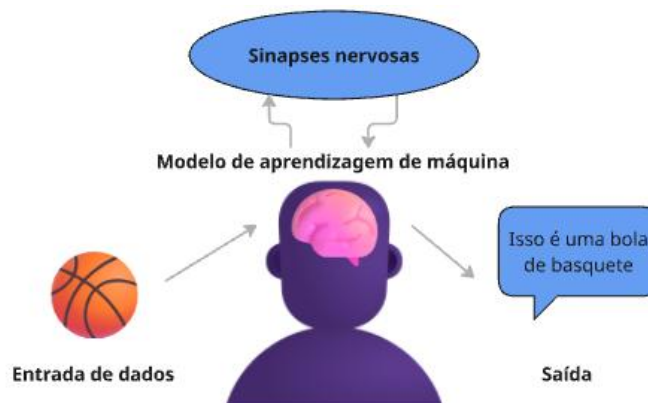
- **Ajuste de contraste das imagens:**

Considerando que o objetivo do modelo é identificar diferentes tipos de objetos em cenários não controlados, um ajuste eficaz é o aumento do contraste das imagens. Isso pode ser realizado por meio da equalização de histograma, uma técnica que permite o aumento global do contraste em imagens (BASSIOU; KOTROPOULOS, 2007). Com esse procedimento, os objetos passam a apresentar contornos mais bem definidos, o que auxilia nas etapas de detecção.

3.3 APRENDIZAGEM DE MÁQUINA

Como muitas tecnologias já desenvolvidas, a aprendizagem de máquina assim como o termo “visão computacional”, deriva de conceitos da biologia. Nesse caso, foi inspirada na própria capacidade biológica de aprendizado que os primeiros modelos de aprendizagem foram desenvolvidos (MCCULLOCH; PITTS, 1943). Antes de apresentar os modelos e suas funções, uma ilustração simplificada desse sistema é apresentada na Figura 5. A base conceitual dessa abordagem deriva da estrutura e do comportamento dos neurônios biológicos, unidades fundamentais do sistema nervoso responsáveis pela recepção, processamento e transmissão de informações (HÜGLE *et al.*, 2020).

Figura 5: Um paralelo do processo de aprendizagem de máquina e a biologia humana.



Fonte: O autor (2025).

Essa lógica inspirou o modelo do neurônio artificial proposto por McCulloch e Pitts (1943), que representava matematicamente essa dinâmica por meio de entradas ponderadas, uma função de soma e uma função de ativação responsável por determinar a saída do sistema. A partir desse modelo inicial, surgiram arquiteturas mais complexas, como o *Perceptron* (ROSENBLATT, 1958), cuja ideia central é que, assim como o cérebro humano aprende ajustando a força das conexões sinápticas entre neurônios com base na experiência, os algoritmos de aprendizado de máquina ajustam os pesos de suas conexões internas para minimizar erros e melhorar a precisão das previsões (DOMARATZKI; KIDANE, 2022).

Esse treinamento pode ser realizado de forma “supervisionada” ou “não supervisionada” (AN *et al.*, 2023). A principal característica de um treinamento supervisionado é a utilização de um conjunto de dados rotulados, ou seja, dados de entrada previamente associados a um resultado desejado. Nessa abordagem, a máquina ajusta seus pesos avaliando, a cada etapa do treinamento, a diferença entre a saída gerada e o resultado esperado pela rotulagem prévia. Para os fins deste trabalho, foram utilizadas técnicas supervisionadas.

O processo de rotular os dados é conhecido como “anotação” e, ao longo deste trabalho, o termo anotação de *dataset* refere-se tanto à atividade de rotular quanto aos dados já rotulados utilizados para treinamento. No contexto de (VC) para detecção de objetos, os rótulos representam cada objeto presente em uma imagem. Isso significa que uma única imagem pode conter diferentes rótulos, uma vez que o interesse está nos objetos dentro da cena e não na imagem como um todo. Esses rótulos são definidos por caixas delimitadoras (*bounding boxes*), que correspondem a retângulos envolvendo cada objeto de interesse e

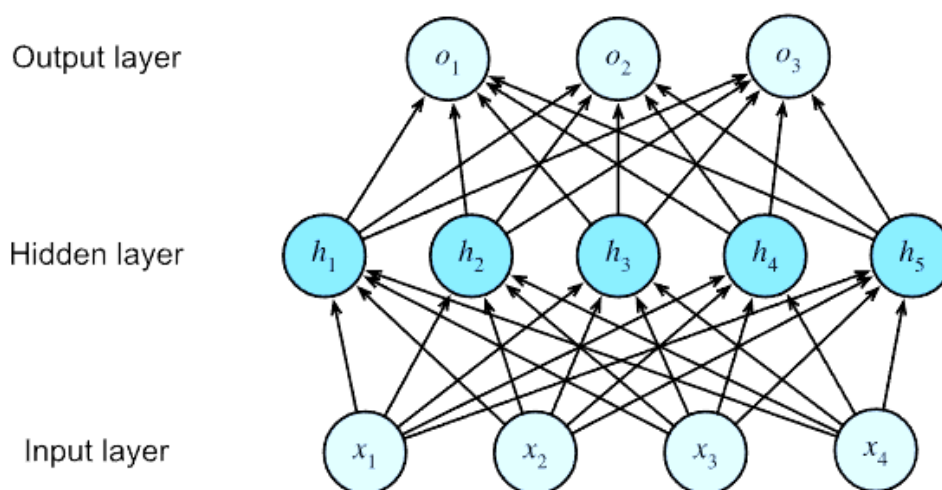
armazenam a informação de localização em coordenadas de pixels.

Neste trabalho, foi utilizada a ferramenta RoboFlow², que oferece acesso a diversos *datasets* públicos, além da possibilidade de criar e gerenciar um *dataset* próprio. A plataforma permite realizar a anotação dos dados para aprendizado supervisionado, obter estatísticas sobre o balanceamento de cada classe e aplicar transformações nas imagens para gerar novos dados (*data augmentation*). Caso necessário, a ferramenta também possibilita o treinamento na nuvem própria ou, alternativamente, a exportação do *dataset* e das anotações para treinamento local.

3.3.1 Multi Layer Perceptron (MLP)

Após a invenção do *perceptron*, considerado o modelo mais simples de uma rede neural artificial capaz de simular o funcionamento de um neurônio biológico, no final da década de 1950, suas aplicabilidades ainda eram limitadas. O modelo se restringia a problemas lineares, mas a capacidade de evolução do campo de estudo da aprendizagem de máquina já se mostrava notável. A partir desse cenário, surgiu o conceito dos “Perceptrons Multicamadas” (MLP), que consiste em interligar múltiplos *perceptrons* em um único modelo (Figura 6).

Figura 6: Ilustração de um modelo MLP genérico.



Fonte: ZHANG et al. (2021).

Uma MLP é composta por pelo menos três camadas: uma camada de entrada, uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída. Cada neurônio de uma camada

² [Roboflow: Computer vision tools for developers and enterprises](https://roboflow.com/)

é conectado a todos os neurônios da camada seguinte por meio de pesos sinápticos ajustáveis.

No livro *Dive Into Deep Learning* (ZHANG et al., 2021) temos que MLP é essencialmente uma função composta parametrizada, que aproxima uma função de mapeamento $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ onde n é o número de entradas e m o número de saídas. O objetivo da rede é aprender essa função com base em dados. Em cada camada cálculos fundamentados em princípios da álgebra linear são realizados que podem ser expressos pelas equações 1 e 2:

$$z_{(1)} = W_{(1)}X + b_{(1)} \quad (1)$$

$$a_{(1)} = \phi(z_{(1)}) \quad (2)$$

onde $X \in \mathbb{R}^n$ é uma entrada vetorial, $W_{(1)} \in \mathbb{R}^{h \times n}$ é a matriz de pesos da primeira camada oculta (com h neurônios), $b_{(1)} \in \mathbb{R}^h$ é o vetor de vieses (bias), $\phi()$ é uma função de ativação não-linear, como ReLU, sigmoide, tanh, entre outras, $a_{(1)}$ é a saída da camada após a ativação. Este processo é repetido em cada camada subsequente do modelo, sendo generalizado pelas equações 3 e 4:

$$z_{(k)} = W_{(k)}a_{k-1} + b_{(k)} \quad (3)$$

$$a_{(k)} = \phi(z_{(k)}) \quad (4)$$

Por fim, ao chegar na última camada, uma função de ativação final é ativada a depender do objetivo do modelo. No caso deste estudo a função softmax é utilizada para realizar a classificação final do objeto, dado que o problema contém múltiplas classes. A saída da rede é obtida através da equação 5:

$$y = \psi(W_{(K)}a_{(K-1)} + b_{(K)}) \quad (5)$$

sendo ψ uma função de ativação final.

O MLP é uma técnica que só aceita entradas unidimensionais, então quando se deseja passar uma imagem dentro de um modelo MLP, é necessário realizar um achatamento desses dados para um vetor de números ao invés de uma matriz bidimensional. Apesar de ser possível obter bons resultados através dessa metodologia, ela é uma clara limitação para uma análise de dados que originalmente não é unidimensional. Foi com essa problemática que surgiu a técnica apresentada a seguir, as “Redes Neurais Convolucionais” (CNN) (ZHANG et al., 2021).

3.3.2 Convolutional Neural Network (CNN)

Um dos primeiros trabalhos com CNN foi o de (JACKEL; DENKER, 1995), que propôs a rede LeNet, usada inicialmente para reconhecer dígitos manuscritos. Mas foi em

2012, com a arquitetura AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), que as CNNs ganharam destaque por vencerem competições internacionais de reconhecimento de imagem com grande margem de acerto.

Como o nome já sugere, a principal característica dessa técnica é a aplicação das chamadas convoluções que é definida pela matemática como a sobreposição entre duas funções $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, em que uma delas é espelhada ou translada em x (RUDIN, 1991), como apresenta a equação 8:

$$(f * g)(x) = \int f(z)g(x - z)dz \quad (8)$$

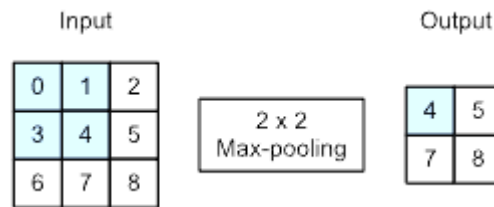
Nas redes neurais, as convoluções operam, de forma discreta e bidimensionais, entre filtros (ou *kernels*) treináveis e as entradas, resultando em mapas de ativação (ou *feature maps*). Considerando K o filtro (*kernel*) e I a imagem de entrada, a equação 8 se torna a 9:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) \cdot K(i - m, j - n) \quad (9)$$

Sendo $S(i, j)$ o mapa de ativação da respectiva saída da convolução. A ideia é que o *kernel* sendo uma pequena matriz 2D irá “deslizar” (tecnicamente chamado de *stride*) sobre a imagem multiplicando localmente os valores e somando os produtos. Isso faz com que o kernel seja capaz de extrair padrões locais como bordas e texturas.

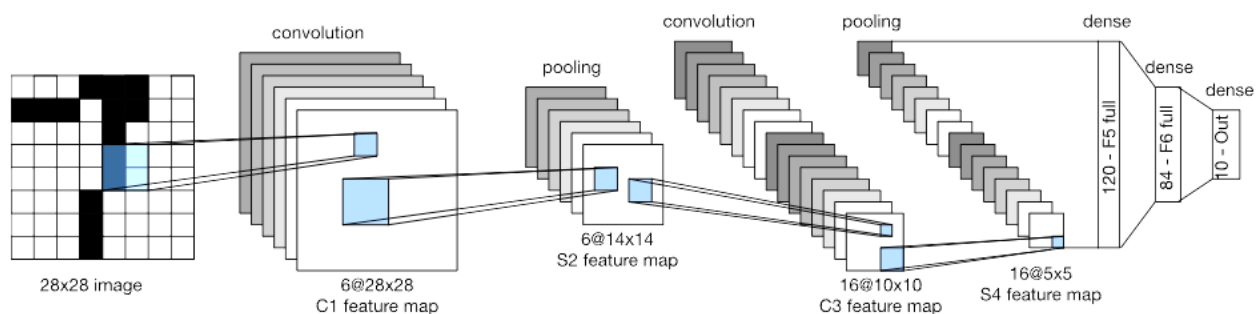
Em combinação com as convoluções, outra técnica presente nas CNN's são os *poolings*. Sua principal função é diminuir a dimensionalidade dos dados, isso é realizado substituindo agrupamentos de pixels por um valor que os represente. Um exemplo é o “MaxPool” que irá guardar o maior valor dentro de um conjunto de pixels, como é demonstrado na Figura 7. O fluxo completo de um modelo CNN pode ser visto na Figura 8.

Figura 7: Max-Pooling com tamanho de janela (2x2).



Fonte: ZHANG et al. (2021).

Figura 8: Fluxo de dados no LeNet. A entrada é um número escrito à mão, a saída é a probabilidade em 10 possíveis resultados.



Fonte: ZHANG et al. (2021).

Existem diversos modelos de VC que utilizam como base CNNs, na Tabela 3 são apresentados alguns citados neste trabalho, para uma maior contextualização e entendimento.

Tabela 3 - Exemplos de modelos CNN

Nome	Principal característica	Referência
VGG16	Rede neural convolucional profunda e homogênea. Demonstra que a profundidade da rede é um fator crucial para performance em tarefas de classificação de imagens em larga escala.	(SIMONYAN; ZISSERMAN, 2014)
ResNet	Introdução das “conexões residuais” ou “atalhos” (<i>skip connections</i>). A ResNet50 é uma versão reduzida da ResNet101 que possui 50 camadas intermediárias ao contrário das 101 que exige muito mais poder computacional para rodar.	(HE <i>et al.</i> , 2016)
MobileNet	Família de arquiteturas projetadas especificamente para dispositivos móveis e embarcados. Utiliza convoluções separáveis em profundidade (<i>depthwise separable convolutions</i>).	(HOWARD <i>et al.</i> , 2017)
Feature Pyramid Network	Lida com o desafio de detecção de	(LIN <i>et al.</i> , 2017)

Nome	Principal característica	Referência
(FPN)	objetos de diferentes tamanhos. Combina características de alta resolução com características de baixa resolução, resultando em uma pirâmide de características permitindo a detecção de objetos em ampla gama de escalas.	

Fonte: O autor (2025).

Cada arquitetura de modelo possui vantagens e desvantagens, sendo constantemente modificada de acordo com as necessidades de suas aplicações. Diante disso, nas próximas seções serão apresentados os modelos estados da arte para detecção e rastreamento de objetos, cada um com suas características específicas, como o YOLO e o R-CNN.

3.3.3 *You Only Look Once* (YOLO)

O modelo YOLO adota uma estratégia de detecção unificada, tratando o problema de detecção de objetos como uma única tarefa de regressão, indo diretamente da imagem para as *bounding boxes* e as classe. O YOLO divide a imagem de entrada em uma grade (*grid*) e, para cada célula, a rede fornece as *bounding boxes*, seus níveis de confiança e as probabilidades de classe. Essa abordagem reduz significativamente o tempo de processamento, tornando o YOLO adequado para aplicações em tempo real, como monitoramento por vídeo, direção autônoma e inspeção industrial (REDMON *et al.*, 2015).

A arquitetura original do YOLO consiste em uma CNN profunda, na qual as camadas finais são totalmente conectadas para produzir os vetores de predição. O modelo é treinado com um conjunto de imagens rotuladas com *bounding boxes*, utilizando uma função de perda que combina os erros nas coordenadas das caixas, nas probabilidades de classe e na confiança quanto à presença de objetos.

Uma das principais vantagens do YOLO é o seu equilíbrio entre velocidade e precisão. Por exemplo, em contextos nos quais é necessário detectar múltiplos objetos em tempo real (como drones, câmeras de segurança, robôs ou dispositivos móveis), as versões mais recentes do YOLO apresentam desempenho superior a métodos mais precisos, porém mais lentos, como o *Faster R-CNN*, que será abordado no próximo tópico.

A principal desvantagem do YOLO para este projeto refere-se à sua licença comercial. Como apresentado anteriormente, este trabalho visa resolver um problema industrial por meio da startup InSpace, o que implica uma futura licença comercial. Atualmente, a licença do YOLO para empresas de pequeno porte é de US\$ 5.000,00 por ano, tornando a solução financeiramente menos viável.

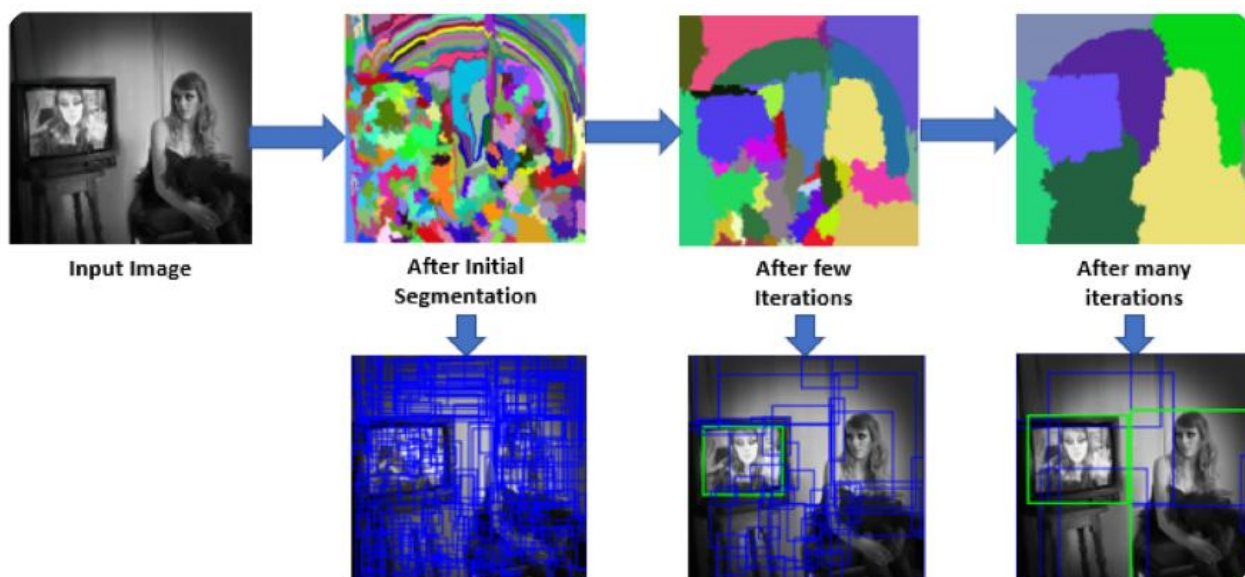
3.3.4 *Region Based Convolutional Neural Network (R-CNN)*

As R-CNNs representam um avanço significativo na área de visão computacional, especialmente na detecção e classificação de objetos em imagens. Diferenciando-se das CNNs tradicionais, que processam a imagem como um todo, as R-CNNs direcionam sua atenção para áreas específicas com maior probabilidade de conter objetos de interesse (GIRSHICK, 2015).

O processo de detecção de objetos com R-CNNs é dividido em algumas etapas. Inicialmente, ocorre a extração de regiões, na qual são identificadas áreas na imagem que possivelmente contêm objetos de interesse por meio do método “Busca Seletiva” (UIJLINGS *et al.*, 2013). Para essa seleção, são empregados algoritmos gulosos para segmentação da imagem, baseados em características como textura, cor e tamanho da região (Figura 9).

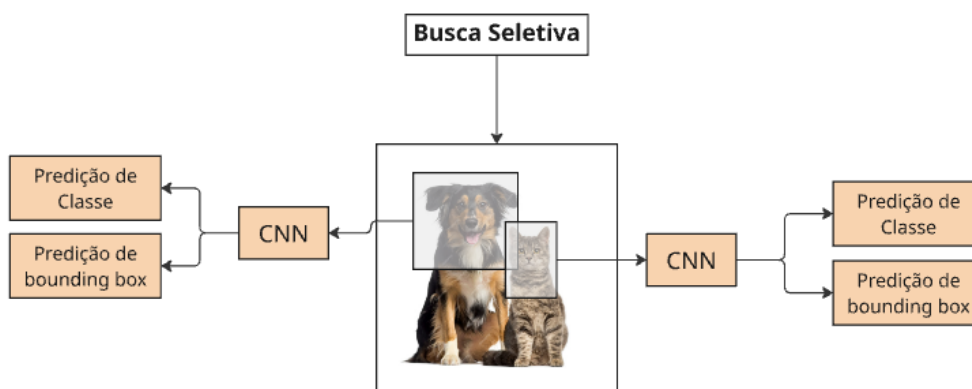
Em seguida cada região é avaliada por uma rede neural pré-treinada, que extrai *features* (características) e as submete a múltiplas SVMs para a determinação da classe. Paralelamente, essas informações são utilizadas por um regressor linear para estimar a *bounding box* daquela região. Um modelo simplificado da R-CNN pode ser observado na Figura 10.

Figura 9: Ilustração da Busca Seletiva.



Fonte: VINA (2024).

Figura 10: Modelo simplificado da R-CNN.

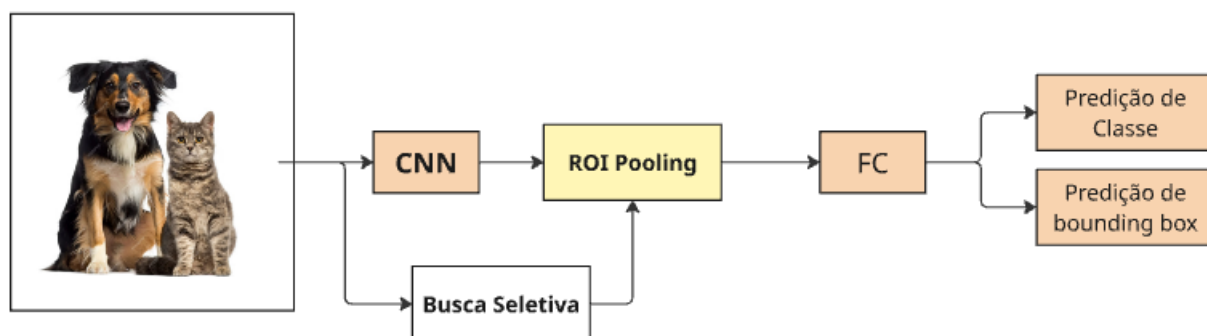


Fonte: O autor (2025).

Pode-se imaginar como essa técnica se torna custosa computacionalmente, já que cada região estimada precisa ser avaliada individualmente por uma CNN. Considerando a possibilidade de diversas regiões com sobreposição entre si, observa-se também o problema de retrabalho para a rede.

Por esse motivo, dois anos depois surgiu o *Fast R-CNN*, que propôs o uso de uma única CNN para processar a imagem completa, em vez de aplicar a rede em cada região de interesse (Figura 11). A partir do mapa de features gerado, realiza-se uma operação denominada *RoI Pooling*, que combina o mapa com as regiões de interesse, reduzindo significativamente o custo computacional (GIRSHICK, 2015).

Figura 11: Modelo simplificado do Fast R-CNN.

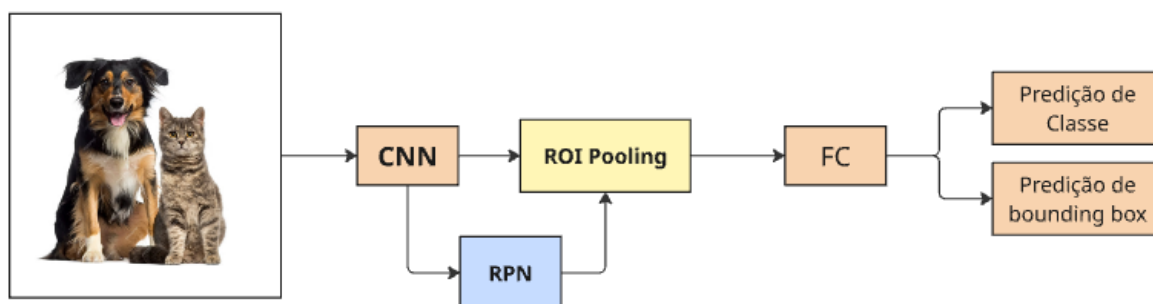


Fonte: O autor (2025).

Porém, a geração das regiões de interesse por algoritmos gulosos ainda representava uma limitação. O *Faster R-CNN* foi proposto com o objetivo de substituir essa etapa por uma Region Proposal Network (RPN).

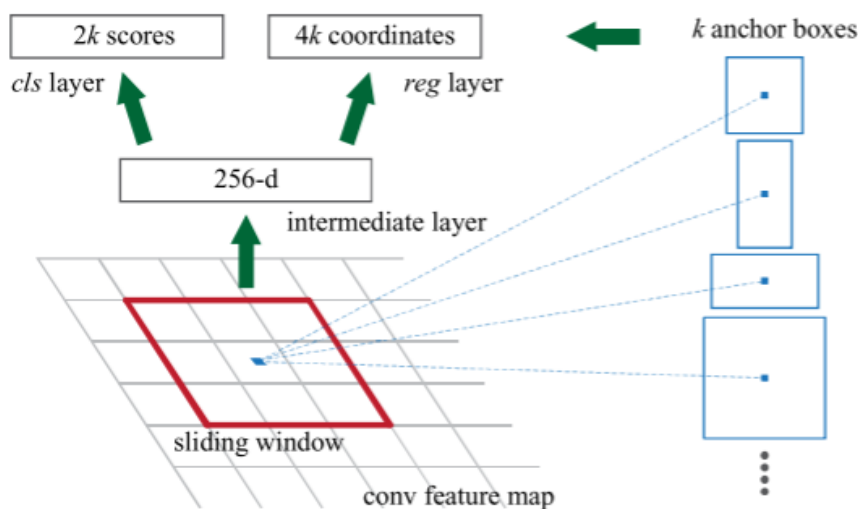
De forma simplificada, a RPN identifica, no *feature map* gerado pela CNN, as áreas com maior probabilidade de conter um objeto de interesse, em vez de elementos do plano de fundo. Isso é realizado por meio da técnica de *sliding window*, que desloca caixas chamadas âncoras, inicializadas em diferentes posições do feature map. Essas caixas possuem tamanhos variados e, à medida que são avaliadas, uma pequena rede convolucional calcula um *object score*, indicando a probabilidade de a caixa conter um objeto relevante ou apenas fazer parte do fundo da imagem (REN *et al.*, 2015).

As ilustrações do modelo *Faster R-CNN* e da RPN podem ser vistas, respectivamente, nas Figura 12 e Figura 13.

Figura 12: Modelo simplificado do *Faster R-CNN*.

Fonte: O autor (2025).

Figura 13: Ilustração do RPN.



Fonte: REN et al. (2015).

Apesar de ser possível encontrar, na literatura, usos isolados das tecnologias apresentadas para diferentes problemas na área de segurança, não foram encontrados trabalhos que visem integrar múltiplos sistemas. Neste trabalho, portanto, o objetivo foi criar uma aplicação eficiente e adaptável, utilizando o modelo *Faster R-CNN* (REN et al., 2015), capaz de operar em cenários complexos, levando em consideração múltiplos fatores, como a detecção de EPIs, a detecção e o rastreamento de pessoas e veículos. Uma IA será responsável por analisar esses elementos e identificar desvios operacionais para emitir alertas de risco.

3.3.5 Keypoint R-CNN

O último modelo que será apresentado tem como objetivo a detecção e o rastreamento de partes do corpo humano. O *Keypoint R-CNN* é uma extensão do framework *Mask R-CNN* (HE et al., 2017), desenvolvido com o objetivo de realizar a detecção de pontos-chave (*keypoints*), como articulações humanas, em imagens. O *Mask R-CNN* se originou a partir da adição de uma ramificação de segmentação de máscaras ao já apresentado *Faster R-CNN*. Por sua vez, o *Keypoint R-CNN* acrescenta uma nova ramificação à arquitetura, voltada especificamente para a regressão das coordenadas dos *keypoints* corporais. O modelo mantém a estrutura básica de duas etapas: primeiro, propõe regiões de interesse (RoIs) por meio da Rede Regional de Propostas (RPN); em seguida, realiza classificações, refinamento de caixas delimitadoras e predição dos *keypoints* em paralelo.

Esse modelo está disponível na biblioteca de mesmo nome no PyTorch. Ele é utilizado neste trabalho para identificar as partes de interesse do corpo humano, permitindo uma avaliação mais acurada da segurança ocupacional. Como apresentado na Figura 14, esse modelo retorna 17 pontos do corpo humano, representados por cada círculo colorido na imagem.

Figura 14: Representação da saída do Keypoint R-CNN.



Fonte: HE et al. (2017).

3.3.6 Métricas de avaliação

Existem diversas métricas de avaliação de algoritmos, principalmente de modelos de aprendizagem de máquina. No entanto, a depender da função do modelo, diferentes métricas podem ser mais adequadas e representar melhor seu desempenho. No caso da VC, não é diferente: para cada finalidade, alguns conjuntos de métricas são escolhidos. Neste trabalho, temos como objetivo duas tarefas principais: a detecção de objetos e a classificação.

Para detecção, o $mAP@0.5$ (*mean Average Precision*) é a métrica mais utilizada. Ela consiste na média das curvas de precisão-recall de todas as classes do modelo (JEGHAM *et al.*, 2024). O termo “@0.5” indica que essa métrica está integrada a IoU (*Intersection over Union*), que mede a sobreposição entre a caixa delimitadora predita e a caixa real do objeto identificado. Nesse caso, se a interseção for inferior a 50%, o algoritmo interpreta como uma

detecção não válida e não a contabiliza no cálculo. Também existe o $mAP@0.5:0.95$, uma métrica mais rigorosa que avalia o modelo em vários intervalos de limiares de sobreposição e calcula sua média, mantendo o mesmo princípio de funcionamento.

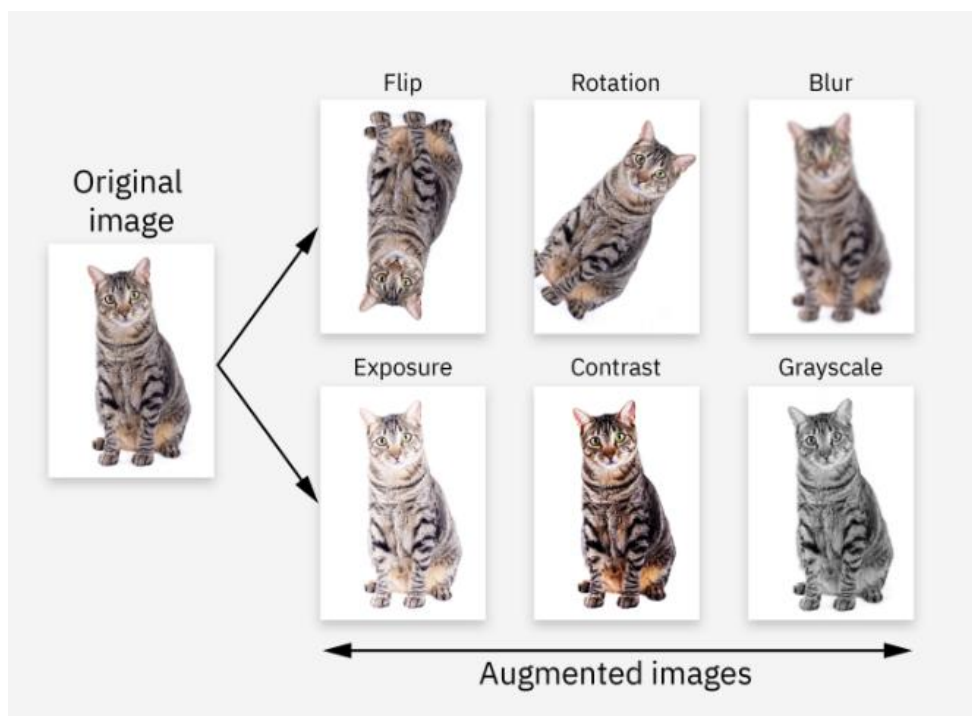
Já no caso da classificação, precisão e recall são duas das métricas mais relevantes. Como ambas se relacionam com o mAP na detecção de objetos, sua utilização contribui para a coerência da avaliação (ZHU *et al.*, 2022). A precisão mede a proporção de verdadeiros positivos entre todos os positivos preditos, enquanto o recall mede a proporção de verdadeiros positivos detectados entre todos os positivos reais.

3.4 DATA AUGMENTATION

Ao considerar a solução para aplicação em cenários adversos, onde nem sempre as condições de monitoramento vão ser sempre as mesmas para diferentes localidades, uma boa prática é usar a técnica de *Data Augmentation* para adicionar ao banco de dados imagens com características distintas.

Na VC isso pode ter dois objetivos: (1) Rebalancear classes, buscando evitar que uma ou mais estejam muito mais representadas, ou muito menos representadas, do que as demais; (2) Gerar imagens com características distintas, simulando diferentes cenários de entrada de dados, como, por exemplo: A modificação da exposição a luz, simulando diferentes luminosidades; A rotação do objeto, simulando diferentes orientações de captura de imagem; Adição de borramento nas imagens, simulando uma falha de foco da câmera (YANG *et al.*, 2022), como demonstrado na Figura 15. Todos esses métodos contribuem para aumentar a robustez da técnica na identificação de objetos em contextos adversos, tornando o modelo mais generalista e apto a operar com imagens não vistas durante o treinamento.

Figura 15: Tipos de modificação em imagens.



Fonte: MUREL e KAVLAKOGLU (2024).

4 METODOLOGIA

Neste trabalho, foi desenvolvido um modelo baseado em aprendizagem de máquina e VC, aplicado à problemática da segurança do trabalho em uma indústria, com o objetivo de realizar a detecção de pessoas, veículos e EPIs. Para isso, foram utilizadas combinações de técnicas de rastreamento voltadas à verificação da conformidade com os padrões operacionais da indústria, em tempo real. Parte dessa metodologia foi publicada e apresentada no evento *Rio Oil and Gas* (LIMA *et al.*, 2024) e na Reunião Anual de Avaliação do PRH-ANP em 2024, mais especificamente, as etapas de rastreamento e verificação do uso correto de EPIs, ainda sem aplicação na indústria petroquímica. O fluxo geral de desenvolvimento do modelo pode ser observado na Figura 16.

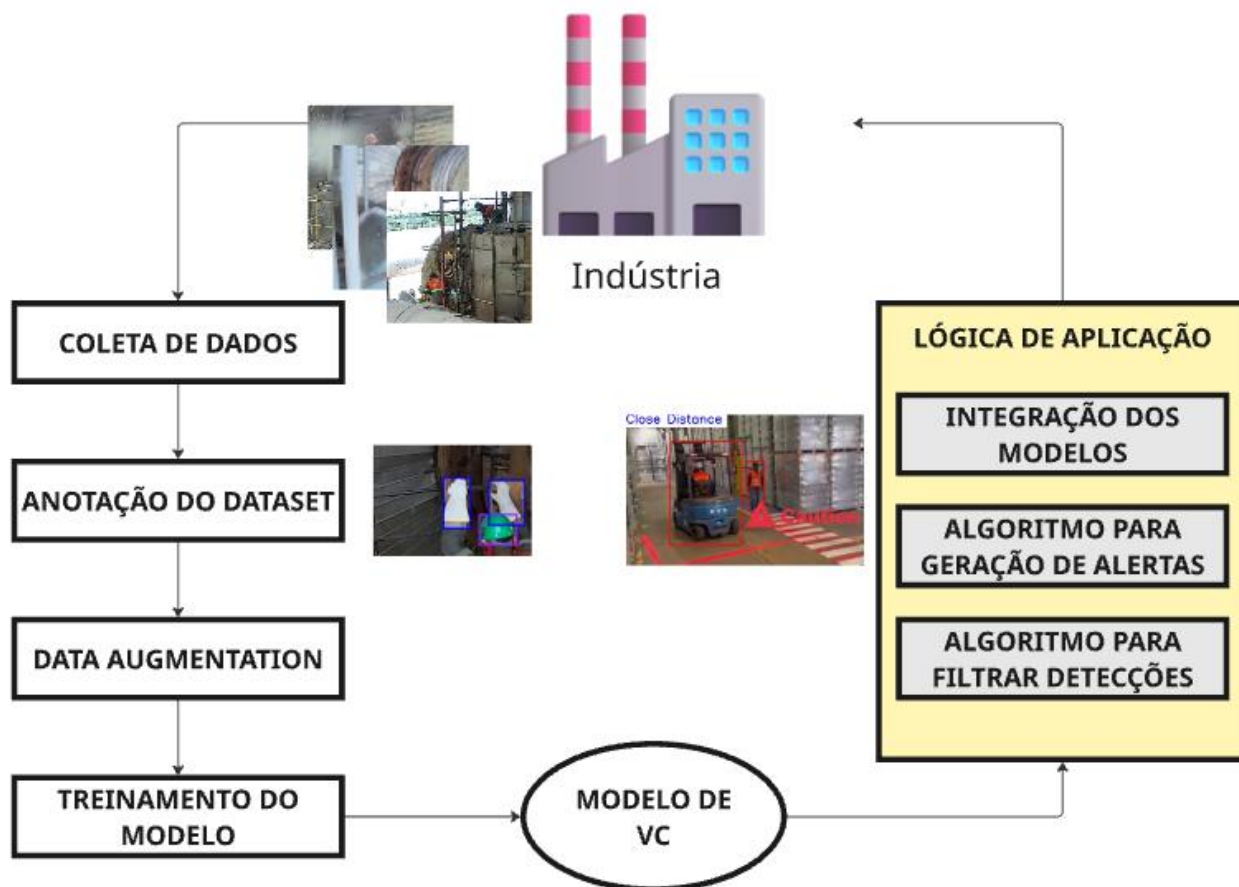
Na etapa de coleta de dados, foi utilizada uma combinação entre a busca por bases públicas e dados obtidos diretamente da indústria parceira, com o intuito de compor o *dataset* de estudo. Para os novos dados adicionados, foi necessário realizar a anotação das classes em cada imagem, a fim de possibilitar o uso de aprendizado supervisionado. Visando tornar o modelo mais robusto, aplicaram-se técnicas de *data augmentation* para gerar imagens de treinamento com ruídos, variações de exposição à luz, entre outros efeitos, de modo a expor o modelo a condições variadas.

Com os dados prontos, o treinamento representa a última etapa antes da aplicação prática. Nessa fase, as performances de diferentes modelos (YOLOv12, ResNet50 FPN, MobileNetv3 Large 320 FPN e MobileNetv3 Large FPN) foram comparadas, e o modelo mais apropriado foi selecionado.

Este trabalho focou na utilização de técnicas com viabilidade comercial, o que influenciou diretamente a escolha dos modelos utilizados como base para o desenvolvimento do sistema: o *Faster R-CNN* e o *Keypoint R-CNN*, ambos com licenças comerciais permissivas.

O modelo final treinado foi integrado a um sistema programado para detectar desvios operacionais e testado em filmagens realizadas em campo, com o objetivo de validar a tecnologia. A seguir, cada etapa será descrita em detalhes, apresentando os recursos e processos utilizados no desenvolvimento da solução.

Figura 16: Fluxograma de desenvolvimento.



Fonte: O autor (2025)

4.1 DESCRIÇÃO DO PROBLEMA

Para orientar as etapas iniciais do desenvolvimento do sistema, foi fundamental compreender os principais desafios enfrentados na gestão de segurança. A partir do contato com a indústria petroquímica, foi possível identificar os tipos mais comuns de desvios operacionais, permitindo um direcionamento mais preciso na definição das regras do sistema.

Entre os comportamentos observados, destacam-se a negligência no uso de EPIs, como a remoção temporária de equipamentos de proteção devido a desconfortos, e o descumprimento da sinalização, como pedestres caminhando fora das faixas sob a justificativa de baixo fluxo de veículos. Também foram relatadas situações de veículos circulando em velocidades acima do permitido, bem como a manipulação de equipamentos fora da ordem operacional, representando riscos significativos à integridade das operações.

A necessidade de controle eficaz se estende não apenas aos funcionários da própria empresa, cujo treinamento tende a ser mais estruturado, mas, sobretudo, aos trabalhadores

terceirizados, sobre os quais a organização muitas vezes não possui clareza quanto ao nível de capacitação técnica e comportamental. Essa lacuna representa um risco relevante à segurança da planta. Com base nessas observações, foi possível conduzir de forma mais assertiva as etapas seguintes, como a coleta de dados e a definição das regras operacionais a serem monitoradas pelo sistema.

4.2 COLETA DE DADOS E ANOTAÇÃO DO DATASET

O problema proposto é composto por três grandes classes de objetos: pessoas, veículos e EPIs. Para servir de base para a solução, foi utilizado o *dataset* público *Construction Site Safety Computer Vision Project*³ que contém 717 imagens e 5.570 anotações distribuídas em 25 classes de objetos, as quais contemplam pessoas, diversos veículos e alguns EPIs.

Vale ressaltar que o número de anotações é significativamente superior ao número de imagens, pois uma única imagem pode conter várias anotações. Ainda que se trate de uma base voltada à indústria da construção civil, muitos dos equipamentos presentes são comuns a diversas outras indústrias.

Contudo, é importante adequar o conjunto de dados ao contexto específico do estudo. Por esse motivo, foi obtido acesso a imagens de uma indústria petroquímica parceira. Esses dados foram coletados por meio de câmeras de segurança previamente instaladas (Figura 17).

Figura 17: Fotos de algumas câmeras instaladas dentro da indústria petroquímica parceira.



Fonte: O autor (2025).

Antes de realizar as anotações do conjunto de dados, foi necessário filtrar as imagens.

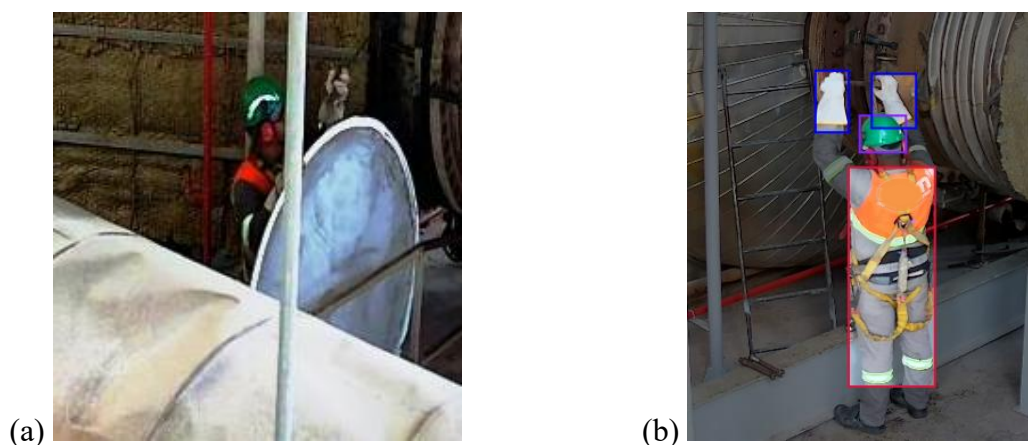
³ [Construction Site Safety Dataset > Overview](#)

Como os dados foram obtidos de filmagens, foi executado um processo de separação de quadros, o que gerou dois problemas principais: (i) sequência de imagens muito semelhantes e (ii) imagens com borramento devido ao movimento. Para melhorar a eficiência do *dataset*, foram removidas as imagens excessivamente repetidas, bem como aquelas de pior qualidade, já que a definição dos objetos estaria comprometida.

Em seguida, foi feita a anotação do *dataset*. Para padronizar esse processo, algumas regras foram estipuladas:

- Evitar a seleção de objetos com alta oclusão (acima de 50%) (Figura 18 (a)).
- Ajustar a caixa de seleção de forma a maximizar a captura do objeto de interesse e minimizar a inclusão de outros elementos (Figura 18 (b)).
- Focar nas principais classes de EPIs: Vestimenta de proteção, Capacete e Luvas (Figura 18 (b)).

Figura 18: (a) exemplo de oclusão na imagem. (b) exemplo de anotação dos objetos de interesse



Fonte: O autor (2025).

Exemplos do resultado da anotação podem ser vistos na Figura 19.

Figura 19: Imagens da indústria parceira, anotadas para posterior adição no *dataset* original (Brilho e saturação

alterados para não exposição de logomarcas).



Fonte: O autor (2025).

Para resolver o problema da verificação do uso correto dos EPIs, faz-se necessário rastrear os *keypoints* do corpo humano. Para isso, foi utilizada uma implementação nativa do PyTorch chamada *Keypoint R-CNN*⁴. Por se tratar de um modelo amplamente consolidado, não houve necessidade de buscar um novo *dataset* para essa tarefa.

4.3 DATA AUGMENTATION

Neste trabalho, as técnicas aplicadas estão listadas a seguir:

- *Flip* (Horizontal e Vertical).
- Escala de cinza aplicada em 25% das imagens.
- Saturação aplicada em algumas imagens entre -25% e +25%.
- Borramento (Blur) até 2,5 pixels.
- Ruído em até 1,5% dos pixels.

Todos esses métodos estão disponíveis na plataforma *RoboFlow*. Esse procedimento foi necessário para aumentar a quantidade de imagens no conjunto de treinamento e tornar o modelo mais robusto a imagens com interferências ou baixa qualidade.

4.4 TREINAMENTO

Neste trabalho, na divisão do conjunto de dados, foi dada prioridade aos subconjuntos de treinamento e validação, com uma proporção de 80% para os dados de treinamento, 15% para validação e 5% para teste, dentro das 784 imagens do *dataset*. Essa decisão foi tomada

⁴ [Keypoint R-CNN — Torchvision main documentation](#)

considerando que o modelo final será avaliado em um conjunto de dados externo, o que reduz a relevância do subconjunto de teste nesta etapa. A priorização do treinamento se justifica pela escassez de dados; portanto, é essencial maximizar o aproveitamento das amostras disponíveis.

Após a preparação do *dataset*, quatro modelos foram treinados com o objetivo de comparar seu desempenho. O primeiro treinamento foi realizado na plataforma *RoboFlow*, que permite o treinamento em nuvem de alguns modelos pré-configurados, dentre eles o YOLOv12, a versão mais recente do popular modelo de VC amplamente utilizado em diversos estudos (JEGHAM *et al.*, 2024). No entanto, essa plataforma não permite a personalização dos parâmetros de treinamento nem a configuração de critérios de *early stopping*.

O modelo YOLOv12 foi treinado por 280 épocas no mesmo conjunto de dados, com o intuito de servir como referência comparativa frente ao *Faster R-CNN*, considerado neste trabalho. Vale ressaltar, conforme mencionado anteriormente, que o uso do YOLOv12 não foi adotado como solução final devido à sua licença comercial restritiva.

Os outros três modelos considerados são baseados no *Faster R-CNN*, e foram treinados a partir de um procedimento disponibilizado em código aberto no GitHub⁵. Esse código é compatível com *datasets* no formato Pascal VOC (EVERINGHAM *et al.*, 2015) e permite a substituição do *backbone* da rede, arquitetura CNN responsável pela extração de features, originalmente baseada na VGG16 (SIMONYAN; ZISSERMAN, 2014). Neste trabalho, foram avaliados três *backbones* disponíveis no módulo torchvision da biblioteca PyTorch, todos pré-treinados no COCO Dataset (LIN *et al.*, 2014). São eles:

- ResNet50 FPN.
- MobileNetv3 Large 320 FPN.
- MobileNetv3 Large FPN.

Esses modelos foram treinados por 50 épocas, com batch size de 16, taxa de aprendizado do otimizador de 0,001 e imagens de entrada no tamanho 640×640, utilizando processamento com dispositivo CUDA. O treinamento e os testes foram realizados em um computador desktop equipado com uma placa de vídeo NVIDIA RTX 4080 com 16 GB de VRAM, processador Intel i5-10400F de 2,90 GHz e 16 GB de RAM.

Tanto o treinamento quanto os testes foram executados com esses modelos, e aquele que apresentou melhor desempenho foi selecionado para os testes em imagens captadas nos

⁵ <https://github.com/sovits-123/faster-rcnn-pytorch-training-pipeline>

ambientes industriais parceiros. Embora seja esperado que o YOLOv12 apresente melhores resultados, por ser uma técnica mais recente, opta-se pela utilização do modelo com desempenho mais próximo ao dele, desde que viável do ponto de vista comercial.

4.5 LÓGICA DA APLICAÇÃO

Com as ferramentas de detecção definidas, este tópico apresenta a lógica de operação do sistema, visando sua adequação ao contexto operacional.

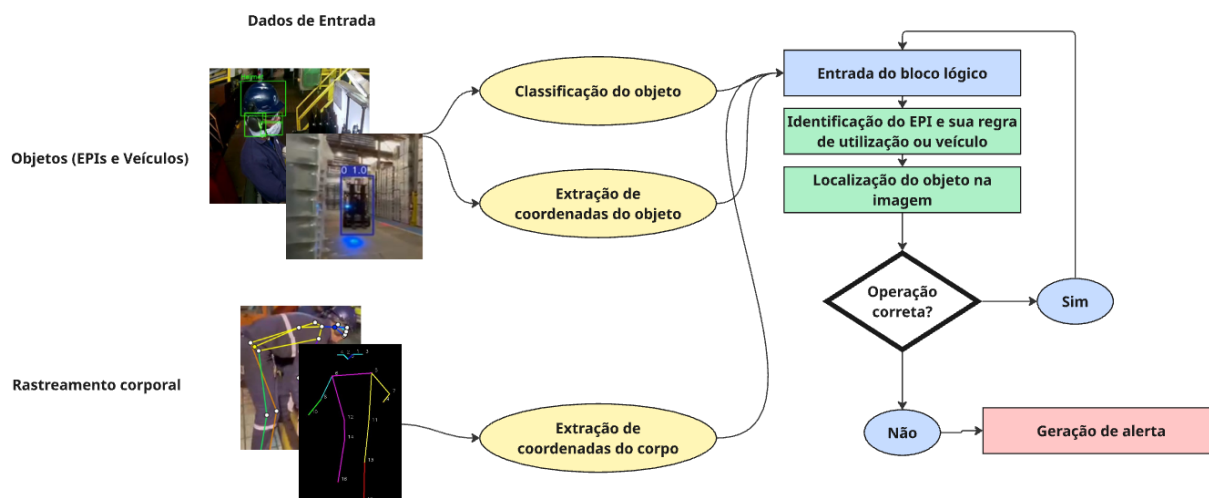
Inicialmente, é importante contextualizar o ambiente de funcionamento: trata-se de uma área industrial monitorada por câmeras fixadas em pontos específicos. Cada região pode possuir regras de operação distintas. Por regras de operação, entendem-se as normas de segurança aplicáveis ao local, como: tipos de EPIs obrigatórios, circulação permitida de veículos, delimitação de zonas seguras, faixas de pedestres, entre outros. As imagens são coletadas em tempo real, e os alertas devem ser processados de forma imediata. Portanto, é fundamental que os algoritmos de detecção apresentem alto desempenho.

Para que o sistema execute essa tarefa com eficácia, dois procedimentos são necessários. O primeiro é a integração dos modelos de detecção de objetos com o modelo de detecção de pose humana. Essa etapa é responsável por analisar as regras de segurança, verificando o posicionamento de pessoas e veículos, bem como o uso correto dos EPIs. O esquema dessa arquitetura pode ser observado na Figura 20.

As situações de não conformidade são previamente definidas com base nas normas de segurança, e sua verificação, no código, ocorre por meio da análise de distâncias entre elementos de interesse. Por exemplo, no caso da verificação do uso correto do capacete, calcula-se a distância entre o objeto “Capacete” e o *keypoint* da cabeça da pessoa. Caso essa distância ultrapasse um limiar predefinido, entende-se que o capacete não está posicionado corretamente.

Outro exemplo envolve a circulação de veículos: caso um veículo transite muito próximo de uma pessoa, a partir do cálculo da distância entre as coordenadas envolvidas, o sistema classifica a situação como segura ou de risco, conforme os parâmetros definidos.

Figura 20: Fluxograma de operação do sistema.



Fonte: O autor (2025).

O segundo procedimento aborda um problema recorrente em aplicações de VC em tempo real: a instabilidade das imagens e a alta taxa de inferência. Como discutido em seções anteriores, os modelos de VC são treinados utilizando imagens estáticas como dados de entrada, as quais são processadas individualmente. Nesse contexto, não há um problema aparente, e os modelos devem, de fato, seguir esse protocolo de treinamento. No entanto, esse processo gera um viés esperado: a maioria das imagens utilizadas apresenta boas condições de foco, nitidez e definição.

Embora a etapa de *data augmentation* aplique técnicas para simular diferentes tipos de interferência nas imagens, isso não garante total robustez do sistema em situações reais, especialmente em contextos com vídeos contínuos.

Em aplicações com vídeo, nas quais se espera uma taxa de quadros entre 24 e 60 por segundo, uma variedade de erros pode ocorrer em apenas um segundo de filmagem. Um dos mais comuns é o *motion blur*, ou borramento por movimento, causado pela captura de objetos em deslocamento por câmeras com baixa velocidade de obturação.

Mesmo que o modelo apresente bom desempenho, ele não necessariamente processa todos os quadros por segundo. Por exemplo, se forem analisadas apenas 12 imagens por segundo em um cenário com baixa taxa de quadros, é possível que, em apenas um segundo de vídeo, ocorram de 1 a 12 detecções distintas. Essas variações podem ocorrer tanto por limitações do modelo, que pode apresentar falhas mesmo com imagens de boa qualidade, quanto por deterioração momentânea das imagens.

Sem um tratamento adequado, essa instabilidade pode gerar alertas incorretos e

comprometer o funcionamento do sistema (Figura 21). Por esse motivo, torna-se fundamental a aplicação de uma solução simples, porém eficaz: a utilização de um filtro temporal, capaz de suavizar as variações entre quadros consecutivos e reduzir a incidência de erros ocasionais, resultando em um sistema mais estável e confiável.

Figura 21: Sequência de frames próximos com identificações erradas.



Fonte: O autor (2025).

O princípio de funcionamento do filtro temporal baseia-se na preservação da detecção mais recorrente em uma janela de tempo. Por exemplo, suponha-se que o filtro esteja configurado para analisar lotes (*batches*) de 10 quadros consecutivos. Se, nesse intervalo, forem registradas 5 detecções da classe X, 3 da classe Y e 2 da classe Z, infere-se que as ocorrências das classes Y e Z são menos confiáveis. Assim, o sistema considera como resultado final daquele lote a classe X, por apresentar maior frequência dentro da janela temporal analisada.

Dessa forma, em vez de o sistema processar as imagens de maneira isolada, a inferência passa a ser feita com base na predominância estatística dentro de um grupo de quadros, promovendo maior estabilidade nas decisões e reduzindo a incidência de falsos positivos.

5 RESULTADOS

Nesta seção, são apresentados os resultados obtidos nas etapas de treinamento e testes. Como os dados foram extraídos de vídeos, serão exibidos quadros específicos (*frames*) com a devida contextualização da operação, a fim de ilustrar os resultados de maneira clara e objetiva.

5.1 COLETA DE DADOS E *DATA AUGMENTATION*

Como parte do esforço para melhorar a distribuição das classes no *dataset*, a primeira etapa consistiu na coleta de novas imagens, o que resultou no aumento do número total de amostras de 717 para 784.

Após a adição desses dados, o novo conjunto de imagens voltado à detecção de EPIs, veículos e pessoas passou a conter 784 registros, cuja distribuição balanceada pode ser observada na Tabela 4.

Tabela 4 - Balanceamento inicial do *dataset*, já contabilizando *data augmentation*.

Classe	Número de anotações (antes)	Número de anotações (depois)	Aumento percentual
Luvras (<i>Gloves</i>)	296	405	36,8%
Capacete (<i>Hardhat</i>)	574	691	20,4%
Pessoas (<i>Person</i>)	1148	1228	6,9%
Vestimenta de Proteção (<i>Safety Vest</i>)	424	556	31,1%
Veículos (<i>Vehicle</i>)	122	122	0%

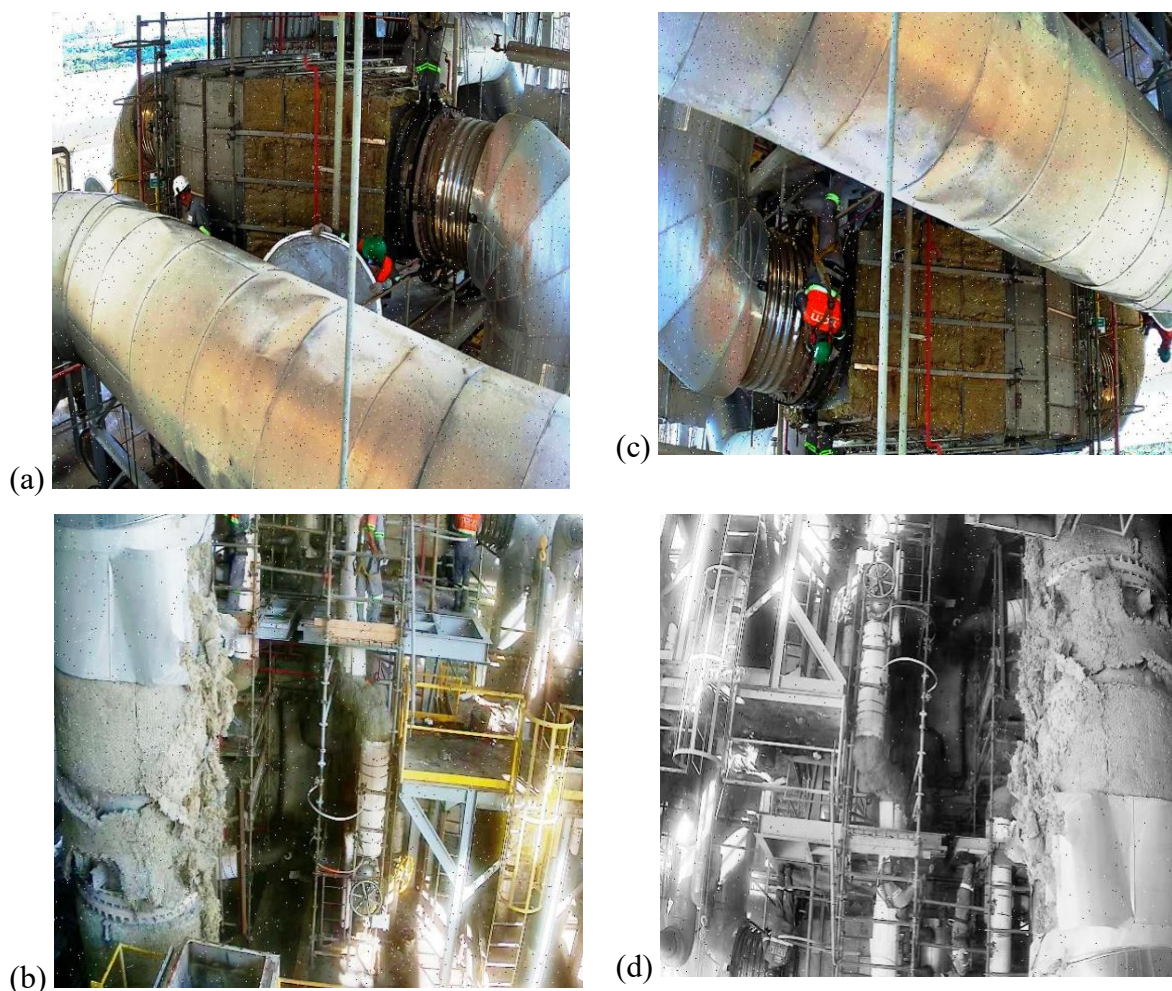
Fonte: O autor (2025).

Alcançar o balanceamento perfeito dos dados é sempre um desafio. No entanto, como pode ser observado pelo aumento percentual de cada classe, foi possível obter um crescimento significativo nas classes de EPIs que anteriormente eram menos representadas. A classe de veículos não foi ampliada devido à indisponibilidade de novos dados. Assim como nas demais categorias, trata-se de um processo contínuo de aprimoramento do *dataset*,

que deve ser atualizado à medida que novos dados se tornarem disponíveis.

Após a coleta e anotação dos dados adicionais, foram aplicadas técnicas de *data augmentation*. A Figura 22 apresenta exemplos de imagens geradas por esses algoritmos, as quais, ao serem incorporadas ao conjunto de treinamento, elevaram seu total de 627 para 1.290 imagens.

Figura 22: Resultados de *data augmentation*. (a) Imagem com ruído. (b) Imagem com saturação aumentada. (c) imagem invertida e com ruído. (d) Imagem invertida, com ruído e com coloração em escala de cinza.



Fonte: O autor (2025).

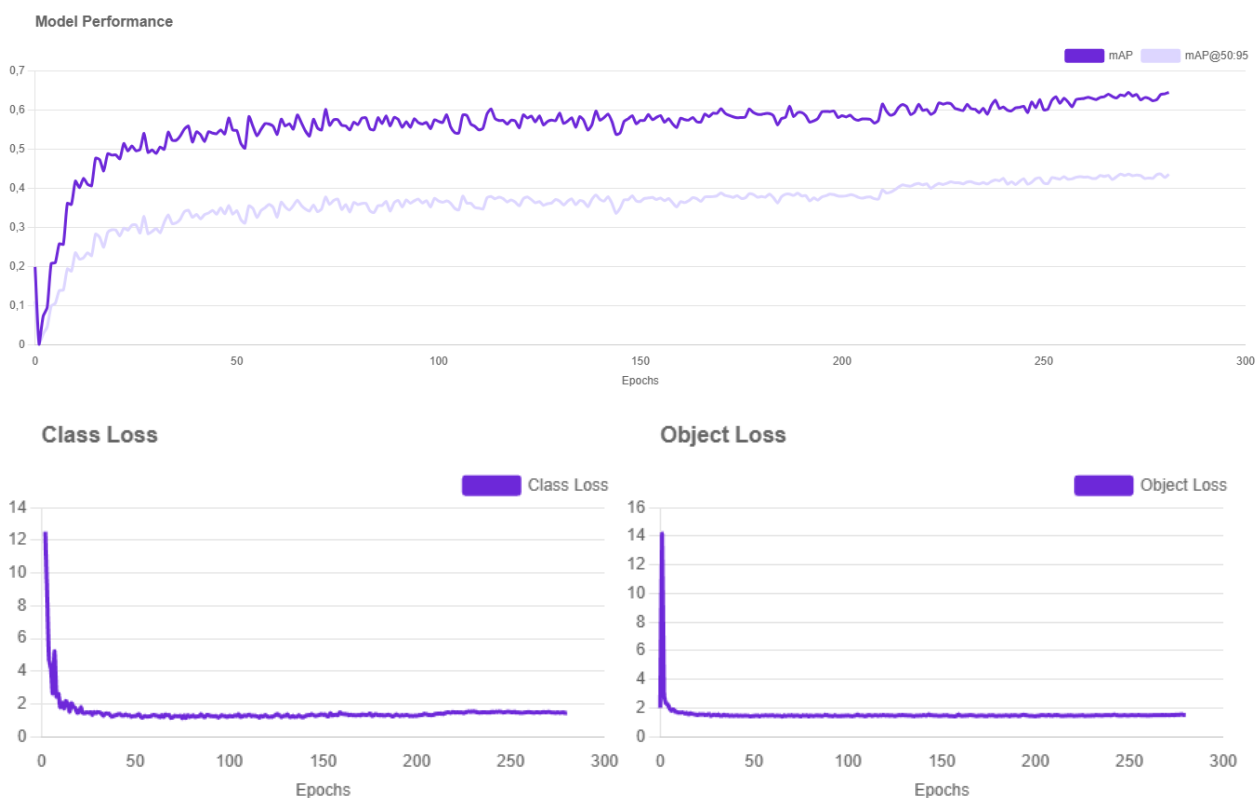
5.2 TREINAMENTO

É importante lembrar, como mencionado anteriormente, todos os modelos apresentados neste trabalho já tinham sido submetidos a um pré-treinamento no COCO *dataset*, o que resultou em curvas de aprendizagem mais rápidas. Essa característica facilita

o processo de readequação dos modelos a novas classes, permitindo sua reconfiguração com maior eficiência.

O modelo YOLO foi o único treinado por meio da plataforma RoboFlow e, devido às limitações da ferramenta quanto a ajustes finos no processo de treinamento, ele ultrapassou as 50 épocas, ao contrário dos demais modelos, cujos parâmetros foram definidos previamente para essa quantidade de iterações. Como o YOLO não foi incluído nos testes em campo desde o início, seus resultados são apresentados aqui apenas como uma base comparativa em relação aos outros modelos. Na Figura 23, observa-se uma evolução consistente nos indicadores de desempenho ao longo do treinamento.

Figura 23: Gráfico mAP e de perda ao longo das 280 épocas do modelo YOLOv12.



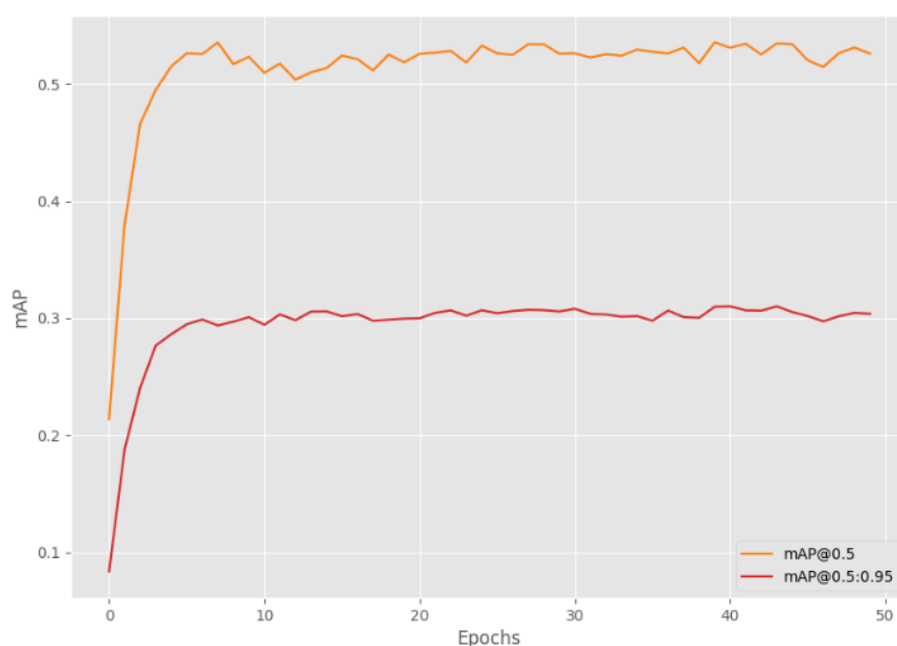
Fonte: O autor (2025).

A métrica mAP apresentou um crescimento rápido nas primeiras 50 épocas, seguido de uma estabilização progressiva, alcançando valores em torno de 0,6, o que indica um desempenho satisfatório na detecção dos objetos propostos. De maneira similar, o mAP@50:95 estabilizou-se em valores próximos de 0,35, refletindo uma capacidade razoável do modelo em lidar com diferentes níveis de sobreposição entre as predições e os objetos reais.

Em relação às funções de perda, tanto a *Class Loss* quanto a *Object Loss* apresentaram uma queda acentuada nas iterações iniciais, estabilizando-se posteriormente em patamares entre 1,5 e 2. Isso indica que o modelo conseguiu aprender de forma eficiente os padrões presentes nos dados, sem apresentar indícios de *overfitting* (quando o modelo performa bem apenas nos dados de treinamento). Esses resultados sugerem que o modelo atingiu uma boa convergência, embora ainda exista espaço para aprimoramentos.

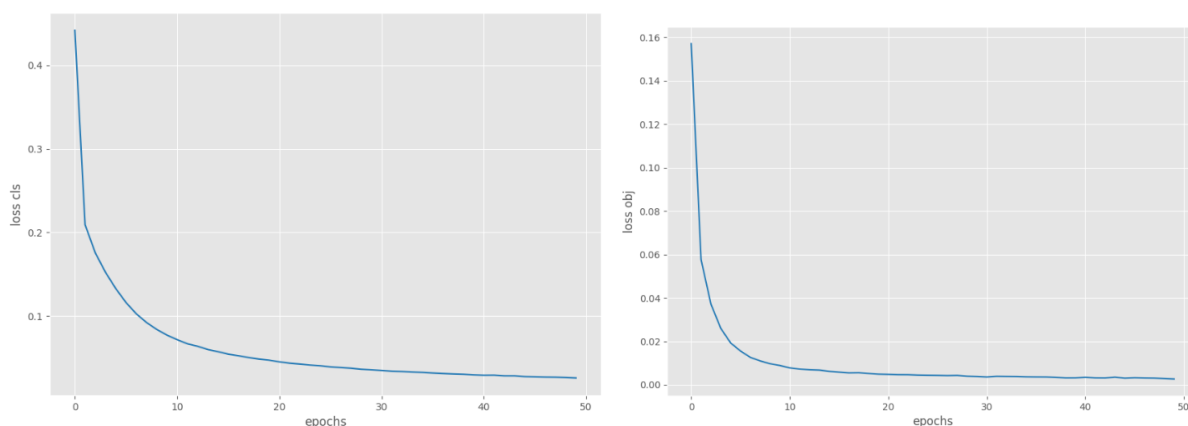
Na fase de treinamento do modelo Faster R-CNN com *backbone* ResNet50-FPN (Figura 24), observou-se uma evolução consistente dos indicadores de desempenho ao longo das 50 épocas analisadas.

Figura 24: Gráfico de mAP ao longo de 50 épocas para o modelo *Faster R-CNN* - ResNet50 FPN.



Fonte: O autor (2025).

Figura 25: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo *Faster R-CNN* - ResNet50 FPN.



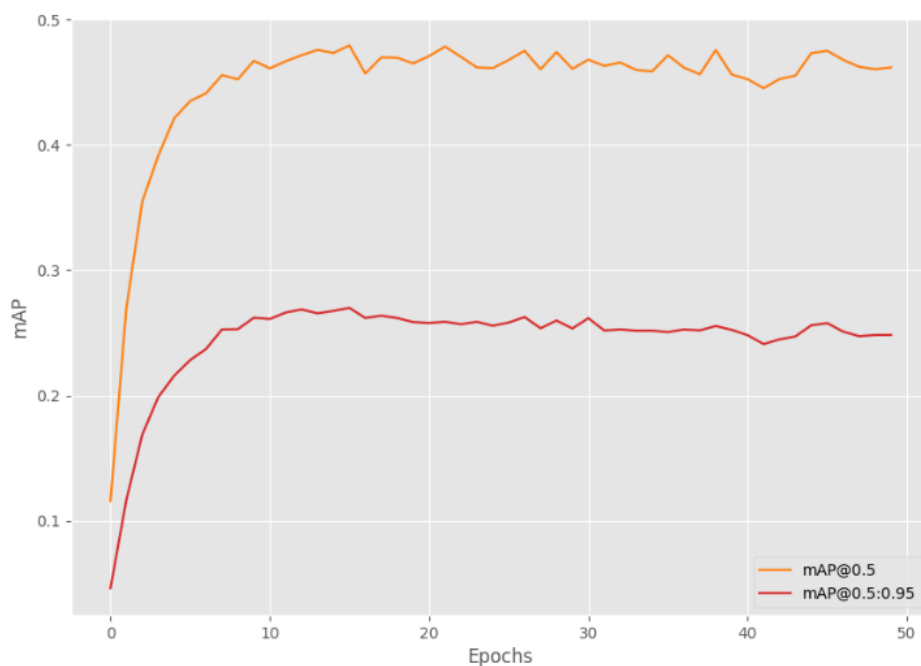
Fonte: O autor (2025).

O $\text{mAP}@0.5$ apresentou um crescimento acentuado nas primeiras 10 épocas, estabilizando-se em torno de 0,55, o que demonstra um bom desempenho na detecção de objetos com um limiar de IoU mais permissivo. Por outro lado, o $\text{mAP}@0.5:0.95$ apresentou valores mais baixos, como esperado, estabilizando-se em aproximadamente 0,30.

Em relação às funções de perda, tanto a *Class Loss* quanto a *Object Loss* exibiram uma redução expressiva nas iterações iniciais (Figura 25). Assim como no modelo anterior, esses resultados indicam uma boa capacidade de generalização, sem evidências de *overfitting* durante o período avaliado.

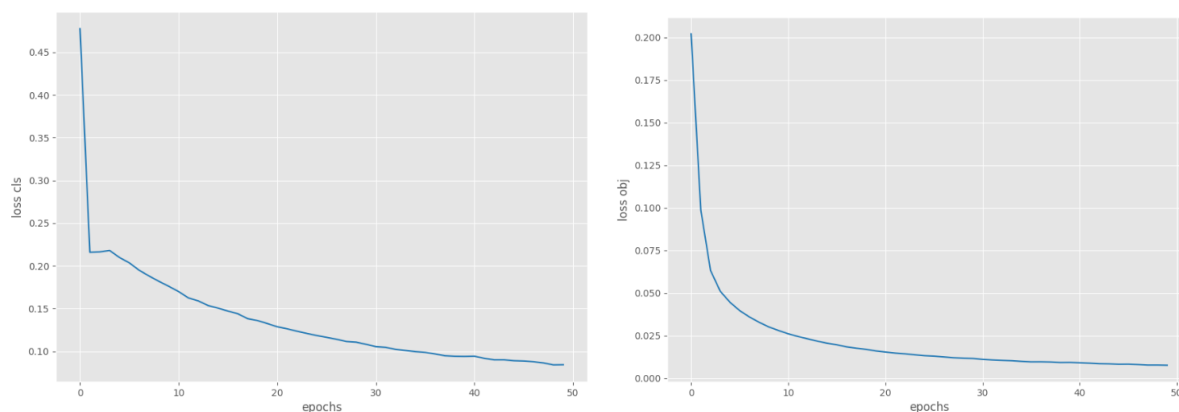
Na sequência, são apresentados os resultados dos modelos *Faster R-CNN* com arquitetura MobileNetV3, tanto com a redução da dimensão da imagem para 320 (Figura 26), quanto na dimensão padrão de 640 utilizada nos demais modelos (Figura 28).

Figura 26: Gráfico de mAP ao longo de 50 épocas para o modelo *Faster R-CNN* - MobileNetv3 Large 320 FPN.



Fonte: O autor (2025).

Figura 27: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo *Faster R-CNN* - MobileNetv3 Large 320 FPN.



Fonte: O autor (2025).

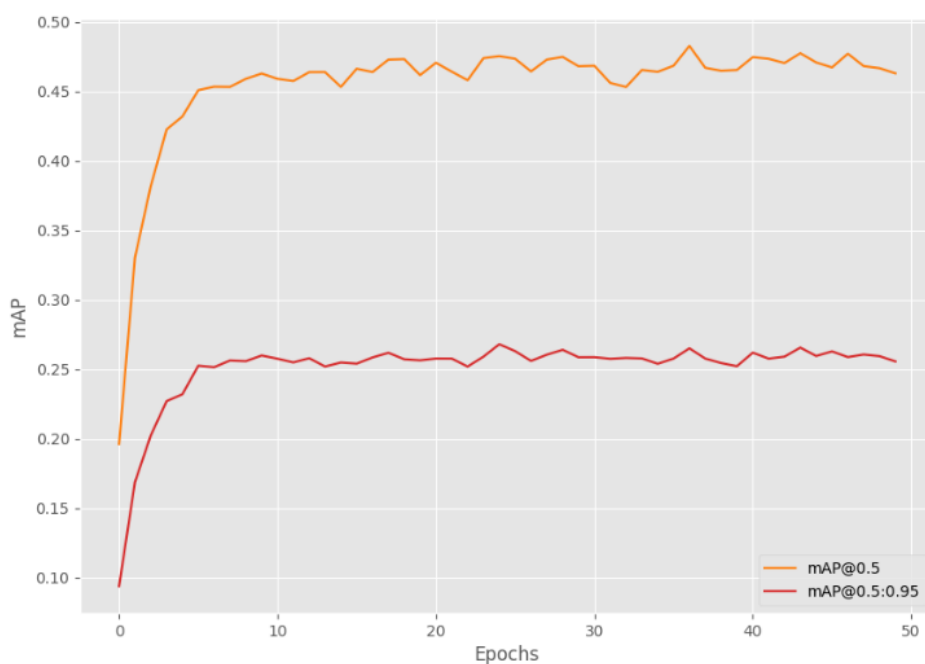
O treinamento apresentou um comportamento similar, indicando pouca variação de performance ao alterar a dimensão das imagens para este modelo. Em ambos os casos, as métricas de mAP@0.5 estabilizaram-se por volta da décima época, atingindo um patamar em torno de 0,47 — um valor ligeiramente inferior aos modelos anteriores, mas ainda considerado satisfatório, especialmente por se tratar de arquiteturas mais leves, projetadas

para rodar em dispositivos móveis.

Os gráficos de perda também exibem comportamentos semelhantes (Figura 29), com a única exceção sendo uma dificuldade aparente do modelo MobileNetV3 Large 320 FPN (Figura 27) em obter uma redução acentuada como os demais. No entanto, ao final do treinamento, ambos os modelos convergem para valores de perda similares, o que sugere a ausência de *overfitting*.

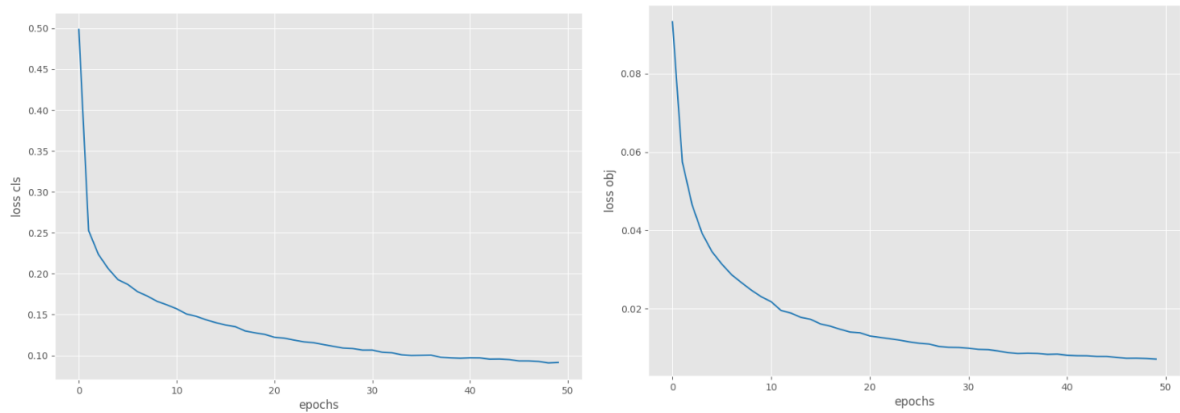
Dessa forma, os resultados indicam que, apesar do desempenho inferior em comparação a modelos de maior porte, o MobileNetV3 apresenta uma boa relação entre desempenho e custo computacional, mostrando-se uma alternativa viável para aplicações que exigem maior eficiência em tempo de execução ou que operem sob restrições de hardware.

Figura 28: Gráfico de mAP ao longo de 50 épocas para o modelo *Faster R-CNN* - MobileNetv3 Large FPN.



Fonte: O autor (2025).

Figura 29: Gráfico de perda da classe e objeto ao longo de 50 épocas para o modelo *Faster R-CNN* - MobileNetv3 Large FPN.



Fonte: O autor (2025).

Tabela 5 - Resultado da avaliação dos modelos em um mesmo *dataset* de testes.

Modelo	mAP@50	Recall médio
YOLOv12	63.5%	50.8%
<i>Faster R-CNN</i> ResNet50-FPN	48.8%	33.9%
<i>Faster R-CNN</i> MobileNetv3 Large 320 FPN	21.26%	13.9%
<i>Faster R-CNN</i> MobileNetv3 Large FPN	42.41%	26.5%

Fonte: O autor (2025).

Na

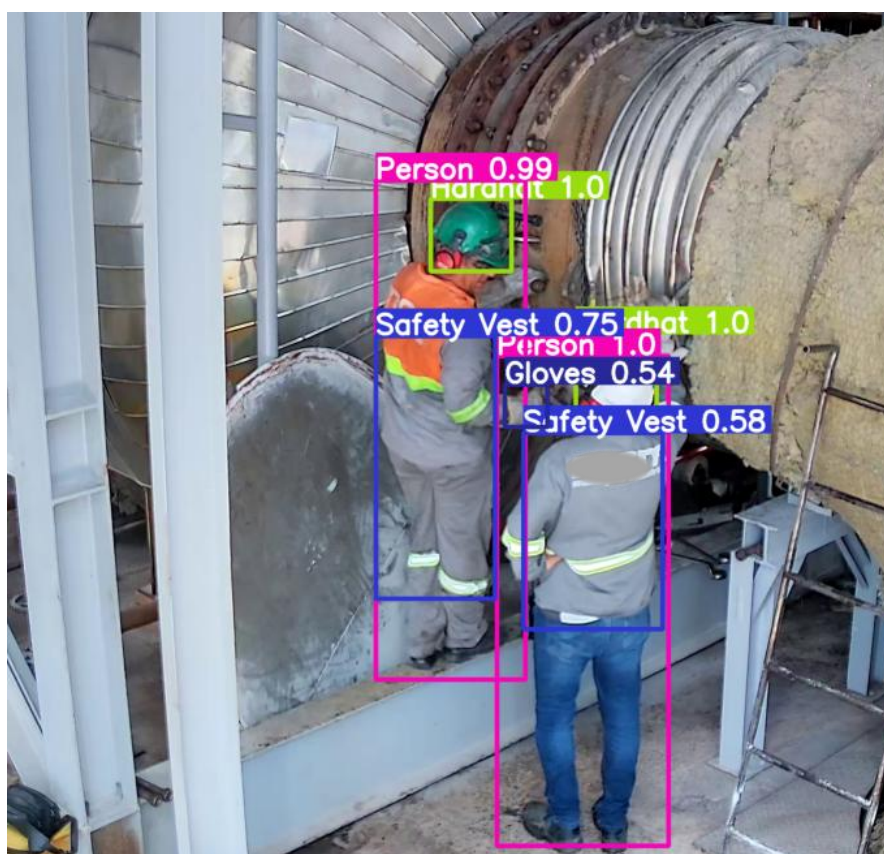
Tabela 5, são apresentados os resultados de cada um dos modelos avaliados em um *dataset* de testes. Como esperado, com base nos dados de treinamento, o modelo YOLOv12 obteve as melhores métricas de desempenho. Entre as arquiteturas do *Faster R-CNN*, a ResNet50 FPN, destacada em azul, apresentou resultados satisfatórios e foi selecionada para os testes em campo. Outro ponto que merece destaque é a inconsistência observada no modelo *Faster R-CNN* MobileNetV3 Large 320 FPN. Embora durante o treinamento ele tenha apresentado resultados semelhantes à sua versão alternativa com maior dimensão de imagem, no conjunto de testes houve uma redução significativa de performance. Esse comportamento pode indicar a ocorrência de *overfitting* neste modelo.

5.3 MODELOS DE DETECÇÃO

A primeira etapa da solução consiste na capacidade de detecção de objetos em cenários diversos. Vale lembrar que o sistema realiza o rastreamento de cinco elementos: Capacetes (*Hardhat*), colete/roupa de proteção (*Safety Vest*), luvas (*Gloves*), veículos (*Vehicles*) e pessoas (*Person*).

Na sequência de imagens, são apresentados testes para a detecção de EPIs e pessoas em distintos contextos operacionais. Na Figura 30, observa-se um ambiente em condições normais de iluminação e distância, no qual o modelo apresentou desempenho satisfatório, com exceção de alguns elementos, como as luvas, que tiveram um baixo grau de confiança.

Figura 30: Teste de detecção de EPIs em condições padrão de iluminação e distância.



Fonte: O autor (2025).

A Figura 31 apresenta um teste mais desafiador, realizado em um ambiente com poluição visual causada pela presença de fumaça. Nesse cenário, o modelo sofreu interferências significativas, resultando em uma redução na confiança das detecções e até na perda da identificação de alguns elementos, como *Safety Vest* e *Hardhat*, entre diferentes quadros. Esse resultado evidencia uma baixa eficiência das técnicas de *data augmentation*

aplicadas para lidar com interferências visuais desse tipo.

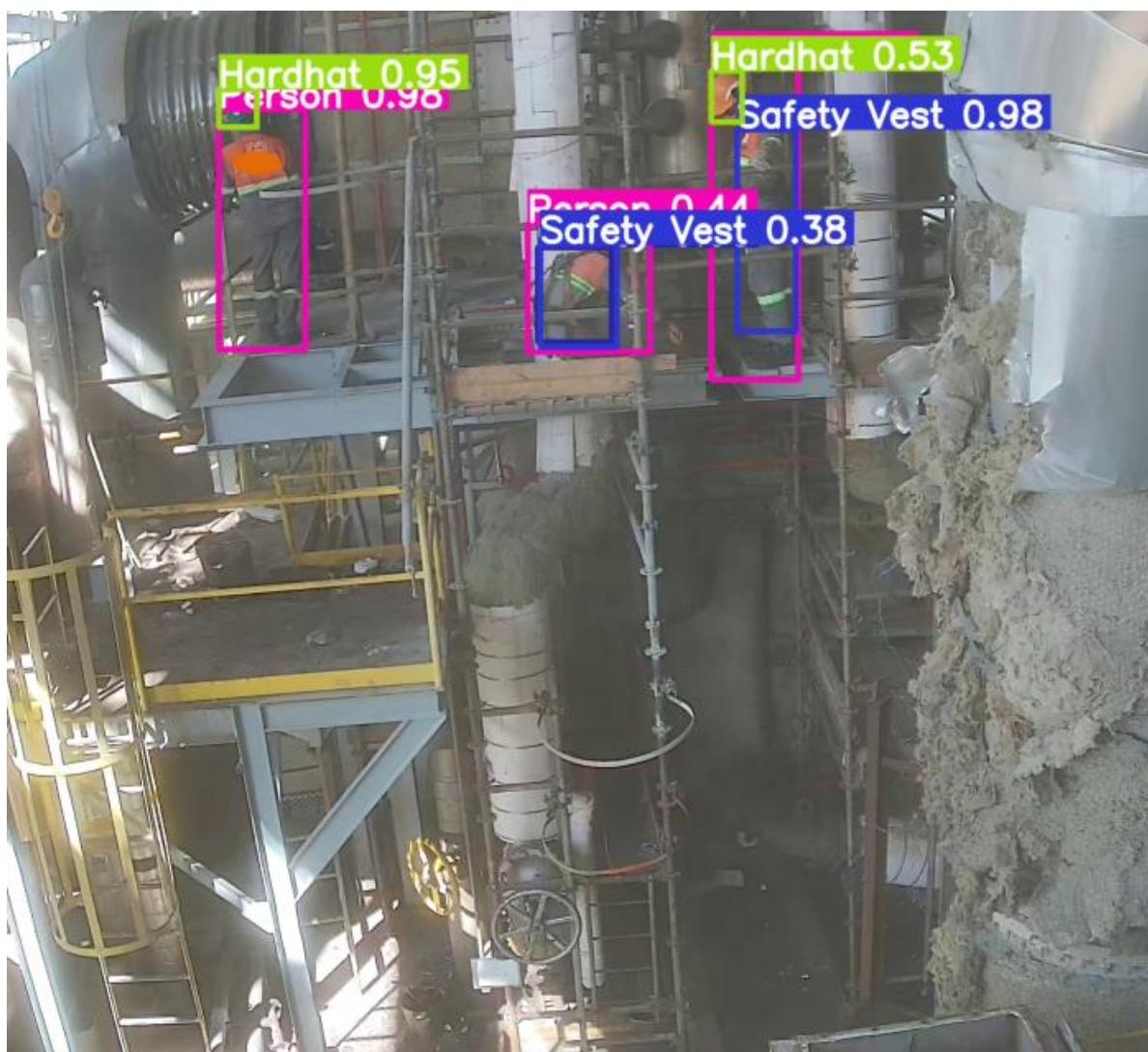
Figura 31: Teste de detecção de EPIs em ambiente poluído.



Fonte: O autor (2025).

Já na Figura 32, é apresentado um cenário mais complexo, com possibilidades de oclusão e maior distância entre a câmera e os objetos de detecção. Nesse teste, observa-se a dificuldade do modelo em identificar elementos menores, como as *Gloves*. A oclusão também impactou de maneira distinta a detecção do *Safety Vest* nos três indivíduos da imagem: à esquerda ocorreu ausência de detecção, no centro foi registrada uma detecção com baixa confiança devido à posição do operador, e à direita houve detecção completa.

Figura 32: Teste de detecção de EPIs em distância e oclusão.



Fonte: O autor (2025).

Em relação ao modelo de detecção de *keypoints* do corpo humano, a Figura 33 apresenta o modelo operando à distância, na imagem à esquerda, e em proximidade, na imagem à direita, ambos com resultados satisfatórios. No entanto, em outros testes foi possível perceber um alto nível de interferência quando o corpo está parcialmente ocluído, resultando na perda da orientação do rastreamento com relativa facilidade.

Figura 33: Testes de detecção de pose, destacado pelos círculos das juntas conectados por linhas pretas.



Fonte: O autor (2025).

Para avaliar o desempenho do modelo por classe de EPI, a Tabela 6 apresenta os resultados de assertividade. Observa-se um bom desempenho geral entre as classes, com baixa tendência a confusões (como detectar um capacete no lugar de uma luva). No entanto, de maneira geral, o modelo ainda apresenta limitações relacionadas a falsos negativos, ou seja, falhas na detecção da presença de um objeto na imagem.

Tabela 6 - Matriz de confusão para detecção de EPIs.

		Predição			
		Gloves	Hardhat	Safety Vest	Falso Negativo
Rótulos	Gloves	32	1	0	59
	Hardgat	0	117	0	102
	Safety Vest	0	0	60	74
	Falso Positivo	12	8	18	0

Fonte: O autor (2025).

Como mencionado anteriormente, a avaliação de modelos em vídeos apresenta desafios específicos, principalmente devido à ausência de controle sobre as condições da imagem de entrada. Dessa forma, falsos negativos podem ser ocasionados por oclusões parciais ou

movimentações rápidas dos objetos. Outro fator relevante é o desequilíbrio entre as classes do *dataset*. Como o treinamento foi realizado com imagens que contêm múltiplos elementos, torna-se difícil isolar as instâncias de cada classe e maximizar o aproveitamento de um *dataset* ainda limitado em tamanho.

Com relação aos falsos positivos, isto é, quando o modelo detecta um objeto inexistente na imagem, esses casos podem ser mitigados por meio da elevação do limiar de confiança exigido para a detecção. No entanto, considerando o volume total de inferências realizadas, trata-se de um número reduzido de ocorrências.

No contexto da segurança do trabalho, é importante destacar que falsos positivos representam um risco maior do que os falsos negativos. Ao aplicar os resultados do modelo em ambientes operacionais reais, é essencial garantir que o sistema não deixe de emitir alertas em situações de risco. Um falso positivo nesse cenário, como a detecção incorreta do uso de um EPI ausente, pode resultar na não geração de um alerta quando ele seria necessário, comprometendo a segurança do trabalhador. Nesse sentido, o fato de o modelo apresentar uma taxa baixa de falsos positivos, mesmo diante de uma incidência relativamente alta de falsos negativos, é considerado um bom desempenho para aplicações práticas.

A partir da avaliação individual de cada classe (Tabela 7), em conjunto com os dados de balanceamento do *dataset* (Tabela 4), observa-se que as classes com maior representação obtiveram melhores resultados. O item “*Hardhat*”, por exemplo, foi o EPI mais representado, com 691 instâncias e um valor de mAP de 0,94. Por outro lado, a classe “*Gloves*”, com 405 instâncias, apresentou um mAP de 0,73 nos testes realizados em campo.

Tabela 7 - Métricas de avaliação por classe de EPI.

Classe	mAP	Recall
<i>Gloves</i>	0.73	0.35
<i>Hardhat</i>	0.94	0.53
<i>Safety Vest</i>	0.77	0.45

Fonte: O autor (2025).

Em relação à detecção de veículos, os testes foram realizados em uma indústria do setor vidreiro. A

Figura 34 ilustra um experimento simples de detecção de veículos, tanto em áreas externas, a uma certa distância, quanto em ambientes internos, com maior proximidade. Por se tratar de objetos grandes e com boa distinção visual, os veículos não apresentaram

inconsistências significativas nas detecções.

Figura 34: Detecção de veículos industriais em ambientes externo e interno.



Fonte: O autor (2025).

5.4 DETECÇÃO DE DESVIO OPERACIONAL

Neste tópico, são apresentados três testes distintos: (1) interação homem-veículo, (2) interação veículo-sinalização / homem-sinalização e (3) verificação do uso correto de EPIs. Parte desses resultados foi previamente divulgada no evento *Rio Oil and Gas* (LIMA *et al.*, 2024) e na Reunião de Avaliação Anual do PRH-ANP em 2024.

Por motivos de confidencialidade os testes apresentados aqui são referentes às indústrias de vidro e alimentos. Vale a pena ressaltar que os modelos foram treinados com dados da petroquímica, logo é esperado que os resultados aqui apresentados sejam semelhantes ou inferiores aos obtidos na aplicação do modelo na própria petroquímica. Isso faz com que os resultados aqui mostrados sirvam para demonstrar a capacidade do sistema de adaptar-se em diferentes contextos industriais.

No primeiro teste, referente à interação entre homem e veículo (Figura 35), o sistema foi capaz de detectar com precisão a aproximação do veículo em relação a uma pessoa. Neste teste não foi levado em consideração as sinalizações de faixa de pedestre e afins, apenas a distância de um objeto para o outro. Por isso o sistema acusa um alarme mesmo o indivíduo estando na faixa de pedestres de forma segura.

Figura 35: Teste de interação de veículo-pessoa. Em cima da imagem de alerta tem a descrição do desvio.



Fonte: O autor (2025).

Nos testes relacionados à sinalização, a posição da faixa de pedestres foi previamente mapeada na imagem, com o objetivo de identificar objetos que interagissem com essa área. A Figura 36 apresenta uma sequência de quadros na qual é possível observar um veículo acessando indevidamente a faixa de pedestres, um operário caminhando dentro da faixa e, posteriormente, fora dela.

Figura 36: Testes de veículo-sinalização/homem-sinalização; à esquerda, um veículo passando por cima da faixa, no meio uma pessoa andando corretamente e em seguida fora da faixa de pedestres. Em cima das imagens de alerta tem a descrição do desvio.



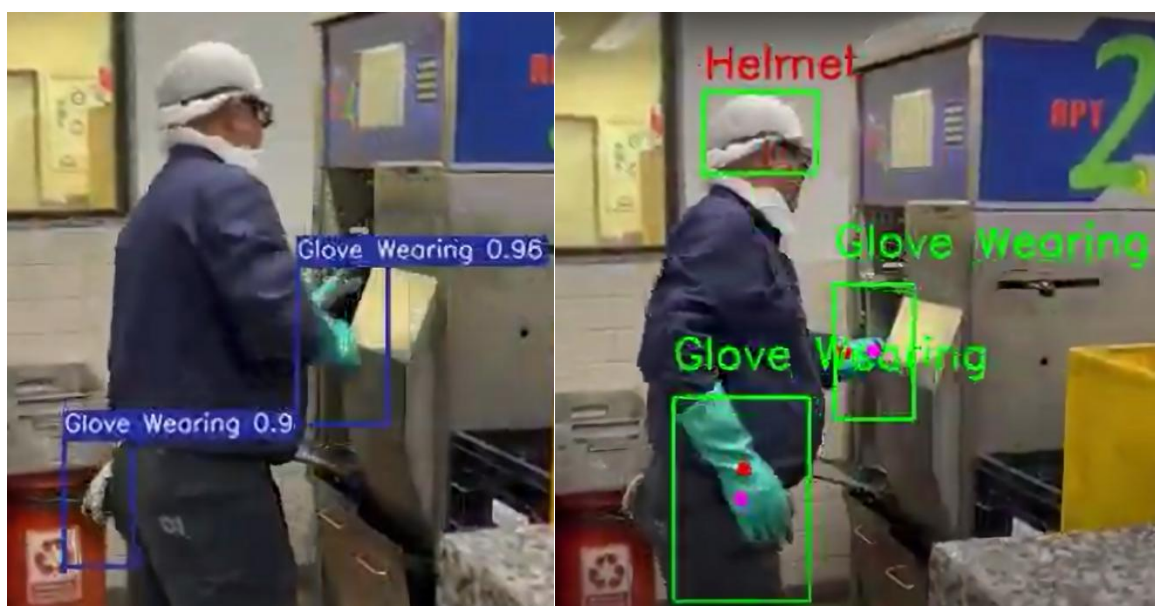
Fonte: O autor (2025).

No terceiro caso analisado, a verificação do uso correto de EPIs, os resultados também foram positivos. Assim como nos testes anteriores, o cruzamento das informações dos *keypoints* com a posição dos objetos mostrou-se eficaz. A Figura 37 ilustra um exemplo dessa metodologia: na imagem à esquerda, uma luva é detectada no bolso do indivíduo, o que não caracteriza uso adequado do equipamento. Já na imagem à direita, com a aplicação da lógica de posicionamento, a luva só é detectada se estiver corretamente posicionada na mão do

trabalhador.

Durante os testes em vídeo, o sistema alcançou uma média de 18 quadros por segundo (FPS), variando entre 13 e 22 FPS. Esse desempenho é considerado satisfatório para o tipo de aplicação proposta. A principal causa da oscilação observada está relacionada à quantidade de elementos presentes simultaneamente na imagem: quanto maior o número de objetos detectados e caixas delimitadoras exibidas, maior a sobrecarga de processamento e, consequentemente, menor a taxa de reprodução. É importante destacar que, em situações reais de operação, nem todos os elementos precisam ser exibidos continuamente, apenas aqueles envolvidos em desvios operacionais. Essa estratégia contribui para uma maior estabilidade na taxa de quadros.

Figura 37: Teste de uso correto de EPI. Na esquerda a detecção sem correlação com os *keypoints* do corpo. Na direita, a integração com *keypoints* está ativa, sendo perceptível pelos círculos coloridos nos pontos de interesse.



Fonte: O autor (2025).

A parte mais crítica deste trabalho consiste na detecção precisa dos objetos. Os testes demonstraram que, uma vez que os elementos de interesse são corretamente identificados, todas as regras operacionais são avaliadas com êxito e os alertas são emitidos de forma adequada. Com relação à detecção de EPIs, a taxa de acurácia superou 70%. Para a adoção ampla da ferramenta em ambientes industriais, ainda é necessário aprimorar os modelos de detecção, uma vez que o rigor operacional dessas indústrias exige altos níveis de confiabilidade, idealmente, acima de 90% de precisão.

Apesar de o sistema ainda não atuar de forma reativa, ou seja, não interrompe autonomamente uma operação em caso de falha, ele é capaz de notificar as equipes de segurança para intervenção imediata. Além disso, registra dados sobre o tipo e a quantidade de ocorrências ao longo das operações, possibilitando o uso dessas informações na definição de protocolos de treinamento para os colaboradores.

5.5 CONTRIBUIÇÕES

Os resultados demonstram o potencial da VC como ferramenta de apoio à segurança do trabalho em diferentes contextos industriais. A aplicação na indústria petroquímica permitiu avaliar o uso de EPIs em um ambiente de alto risco, onde pequenas falhas operacionais podem ter consequências severas. A detecção precisa de capacetes, luvas e coletes contribuiu para monitorar continuamente a conformidade dos operadores com os protocolos de segurança, atuando como um complemento às equipes de SST e viabilizando ações preventivas mais rápidas e embasadas. Além disso, o sistema gerou dados históricos de ocorrências que podem orientar o aprimoramento de treinamentos e políticas internas de segurança.

Na indústria de vidro, o sistema foi testado para detecção de veículos e sua interação com operadores. A detecção consistente mesmo em áreas externas ou com diferentes níveis de proximidade demonstrou a robustez do modelo em condições diversas de iluminação e escala. O cruzamento entre a posição dos veículos e a localização de trabalhadores permitiu detectar aproximações perigosas, alertando para situações de risco, como circulação fora de zonas seguras.

Já na indústria de alimentos, os testes se concentraram na detecção do uso correto de EPIs e no rastreamento da movimentação de operadores. O uso de *keypoints* para verificar se os equipamentos estavam de fato sendo utilizados corretamente (ex. luvas nas mãos e capacetes na cabeça) mostrou-se eficaz e trouxe ganhos relevantes na validação do uso real dos equipamentos. Isso é especialmente importante em ambientes com alta rotatividade de trabalhadores, incluindo terceirizados, onde a indústria não tem controle sobre o histórico de treinamento.

Apesar de indústrias distintas, o trabalho iniciado no setor de O&G pôde contribuir com aplicações em contextos diferentes, validando a capacidade do sistema de ser adaptado para os requisitos de cada indústria.

6 CONCLUSÃO

Neste trabalho, foram aplicadas técnicas de VC para a avaliação de atividades em imagens complexas, considerando o ambiente industrial estudado. Observou-se que, no contexto da segurança, há uma interseção significativa entre diferentes indústrias. Três tipos de EPIs, comuns às três plantas analisadas, foram testados com sucesso, enquanto os testes com veículos foram realizados em apenas uma das fábricas. Os resultados evidenciam que o princípio de funcionamento dos sistemas propostos é aplicável a diferentes contextos industriais, embora adaptações sejam necessárias para atender às especificidades de cada ambiente, como a inclusão de novos EPIs e classes de veículos no treinamento do modelo.

O modelo *Faster R-CNN*, com arquitetura ResNet-50-FPN, demonstrou ser adequado aos desafios propostos, assim como o modelo *Keypoint R-CNN* da biblioteca PyTorch, utilizado na análise de situações de risco. A solução alcançou uma média de 18 FPS, superando técnicas semelhantes da literatura e comprovando sua viabilidade para operação em tempo real. Durante os testes, também se constatou que fatores externos, como fumaça, distância da câmera e oclusões, afetam negativamente o desempenho dos modelos, uma limitação comum em aplicações de visão computacional, o que reforça a importância de adaptar a solução às particularidades de cada indústria.

Os resultados indicaram uma forte correlação entre o balanceamento do conjunto de dados e o desempenho por classe, sendo que as classes mais representadas no treinamento obtiveram maior acurácia. Assim, uma das principais oportunidades de melhoria consiste em ampliar a amostragem das classes menos representadas, a fim de elevar a performance geral do modelo. Além disso, embora tenham sido aplicadas técnicas de *data augmentation* para simular condições adversas, como baixa visibilidade, não foi possível verificar um ganho significativo de desempenho em imagens com interferência, como as que continham fumaça. Esse resultado sugere uma possível limitação da abordagem atual para lidar com esse tipo de ruído visual.

De forma geral, os resultados obtidos demonstram o potencial de soluções baseadas em monitoramento inteligente para apoiar a gestão da segurança e a mitigação de riscos operacionais em ambientes industriais. A estrutura da proposta, fundamentada em princípios da Engenharia de Produção, possibilitou o alinhamento entre viabilidade técnica e aplicação prática, com foco na melhoria do desempenho dos processos. Assim, este trabalho reforça a contribuição da área para o desenvolvimento de soluções inovadoras que integrem pessoas, tecnologia e gestão, promovendo ambientes industriais mais seguros, eficientes e resilientes.

Como próximos passos, será realizado o retreinamento do modelo utilizando exclusivamente dados provenientes de uma indústria-alvo, incorporando elementos específicos do seu contexto operacional, incluindo imagens sob condições adversas, como a presença de fumaça. Os resultados obtidos com esse novo modelo serão comparados com os do modelo atual, que foi treinado com uma base genérica ampliada por *data augmentation*, a fim de avaliar se o uso de dados dedicados proporciona ganhos de desempenho.

REFERÊNCIAS

- AHN, J. *et al.* SafeFac: Video-based smart safety monitoring for preventing industrial work accidents. **Expert Systems with Applications**, 1 abr. 2023. v. 215.
- AL-MHDAWI, M. K. S. *et al.* Implementation of a Combined Fuzzy Controller Model to Enhance Risk Assessment in Oil and Gas Construction Projects. **IEEE Access**, 2024. v. 12, p. 68319–68333.
- AN, Q. *et al.* A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. **Sensors**, 22 abr. 2023. v. 23, n. 9, p. 4178.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO Gestão de riscos-Diretrizes Risk management-Guidelines**. [S.l.]: [s.n.], 2018. Disponível em: <www.abnt.org.br>.
- AYDIN, N.; SEKER, S.; ŞEN, C. A new risk assessment framework for safety in oil and gas industry: Application of FMEA and BWM based picture fuzzy MABAC. **Journal of Petroleum Science and Engineering**, dez. 2022. v. 219, p. 111059.
- BASSIOU, N.; KOTROPOULOS, C. Color image histogram equalization by absolute discounting back-off. **Computer Vision and Image Understanding**, jul. 2007. v. 107, n. 1–2, p. 108–122.
- BENSACI, C. *et al.* Collision hazard modeling and analysis in a multi-mobile robots system transportation task with STPA and SPN. **Reliability Engineering & System Safety**, jun. 2023. v. 234, p. 109138.
- BRASIL. **Norma Regulamentadora nº 01 – Disposições Gerais e Gerenciamento de Riscos Ocupacionais (GRO)**. [S.l.]: [s.n.], 2019.
- _____. **Norma Regulamentadora nº 09 – Programa de Prevenção de Riscos Ambientais (PPRA)**. [S.l.]: [s.n.], 2025.
- CAMPOS, V. F. **TQC: Controle da Qualidade Total (no estilo japonês)**. 9. ed. Belo Horizonte: Falconi, 2019.
- CASAS, E. *et al.* An End-to-End Platform for Managing Third-Party Risks in Oil Pipelines. **IEEE Access**, 2024. v. 12, p. 77831–77851.
- CAUCHICK-MIGUEL, P. A. *et al.* **METODOLOGIA DE PESQUISA EM ENGENHARIA DE PRODUÇÃO E GESTÃO DE OPERAÇÕES**. 2. ed. Rio de Janeiro: Elsevier, 2012.
- CHEN, P. *et al.* A cascaded deep learning approach for detecting pipeline defects via pretrained YOLOv5 and ViT models based on MFL data. **Mechanical Systems and Signal**

Processing, jan. 2024. v. 206, p. 110919.

DELOITTE. **Energy Transition Trends Report 2022 A report in collaboration with Contents**. [S.l.]: [s.n.], 2022. Disponível em:

<<https://www.deloitte.com/br/pt/Industries/energy/research/gx-global-energy-transition-trends-2022.html>>. Acesso em: 2 dez. 2024.

_____. Full-PDF-report-2024-oil-and-gas-industry-outlook. 2024. Disponível em:

<<https://www.deloitte.com/br/pt/Industries/energy-chemicals/perspectives/perspectivas-oleo-gas-produtos-quimicos.html>>. Acesso em: 2 dez. 2024.

DOMARATZKI, M.; KIDANE, B. Deus ex machina? Demystifying rather than deifying machine learning. **The Journal of Thoracic and Cardiovascular Surgery**, mar. 2022. v. 163, n. 3, p. 1131- 1137.e4.

DWYER, B. When Should I Auto-Orient My Images? 8 mar. 2020. Disponível em:

<<https://blog.roboflow.com/exif-auto-orientation/>>. Acesso em: 29 jan. 2025.

EVERINGHAM, M. *et al.* The Pascal Visual Object Classes Challenge: A Retrospective.

International Journal of Computer Vision, 25 jan. 2015. v. 111, n. 1, p. 98–136.

FERDOUS, M.; AHSAN, S. M. M. PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. **PeerJ Computer Science**, 2022. v. 8.

GIRSHICK, R. Fast R-CNN. [S.l.]: IEEE, 2015. p. 1440–1448.

HASAN IMRAN, M. *et al.* A critical analysis of machine learning in ship, offshore, and oil & gas corrosion research, part I: Corrosion detection and classification. **Ocean Engineering**, dez. 2024. v. 313, p. 119600.

HE, K. *et al.* Deep Residual Learning for Image Recognition. [S.l.]: IEEE, 2016. p. 770–778.

_____. *et al.* Mask R-CNN. 20 mar. 2017. Disponível em:

<<http://arxiv.org/abs/1703.06870>>.

HODGE, P. Machine vision safeguard assurance. **Process Safety Progress**, 8 mar. 2024. v. 43, n. 1, p. 34–39.

HOWARD, A. G. *et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 16 abr. 2017. Disponível em: <<http://arxiv.org/abs/1704.04861>>.

HÜGLE, M. *et al.* Applied machine learning and artificial intelligence in rheumatology. **Rheumatology Advances in Practice**, 1 jan. 2020. v. 4, n. 1.

JACKEL, L.; DENKER, J. **Comparison of learning algorithms for handwritten digit recognition**. [S.l.]: [s.n.], 1995. Disponível em:

<<https://www.researchgate.net/publication/216792818>>.

- JDHAO'S. JPEG Image Orientation and Exif. 2019. Disponível em: <https://jdhao.github.io/2019/07/31/image_rotation_exif_info/>. Acesso em: 29 jan. 2025.
- JEGHAM, N. *et al.* YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. 31 out. 2024. Disponível em: <<http://arxiv.org/abs/2411.00201>>.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **ImageNet Classification with Deep Convolutional Neural Networks**. [S.l.]: [s.n.], 2012. Disponível em: <<http://code.google.com/p/cuda-convnet/>>.
- LI, Haibin *et al.* **YOLO-PL: Helmet wearing detection algorithm based on improved YOLOv4**. **Digital Signal Processing: A Review Journal**. Elsevier Inc.
- LIMA, G. C. R. *et al.* Computer Vision techniques applied in the Oil e Gas industry to enhance safety: detection of equipments, people and vehicles. [S.l.]: [s.n.], 2024. p. 2191–2205. Disponível em: <<https://biblioteca.ibp.org.br/pt-BR/search/49499>>.
- LIN, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. 1 maio. 2014. Disponível em: <<http://arxiv.org/abs/1405.0312>>.
- _____ *et al.* Feature Pyramid Networks for Object Detection. [S.l.]: IEEE, 2017. p. 936–944.
- MACÊDO, July Bias *et al.* Identification of risk features using text mining and BERT-based models: Application to an oil refinery. **Process Safety and Environmental Protection**, 1 fev. 2022. v. 158, p. 382–399.
- MAIOR, C. B. S. *et al.* Bayesian prior distribution based on generic data and experts' opinion: A case study in the O&G industry. **Journal of Petroleum Science and Engineering**, 1 mar. 2022. v. 210.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, dez. 1943. v. 5, n. 4, p. 115–133.
- MINISTÉRIO DO TRABALHO E EMPREGO. Norma Regulamentadora No. 4 (NR-4). **Norma Regulamentadora No. 4 (NR-4)**, 2024. Disponível em: <<https://www.gov.br/trabalho-e-emprego/pt-br/aceso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/comissao-tripartite-partitaria-permanente/normas-regulamentadora/normas-regulamentadoras-vigentes/norma-regulamentadora-no-4-nr-4>>. Acesso em: 13 jan. 2025.
- MUREL, J.; KAVLAKOGLU, E. O que é aumento de dados. 7 maio. 2024. Disponível em: <<https://www.ibm.com/br-pt/think/topics/data-augmentation>>. Acesso em: 15 jun. 2025.
- NIOSH. About Hierarchy of Controls. **About Hierarchy of Controls**, 2024. Disponível em: <<https://www.cdc.gov/niosh/hierarchy-of->

controls/about/index.html#:~:text=The%20hierarchy%20of%20controls%20identifies%20a%20preferred%20order,because%20they%20control%20exposures%20without%20significant%20human%20interaction.>. Acesso em: 13 jan. 2025.

NWANKWO, C. D. *et al.* Analysis of accidents caused by human factors in the oil and gas industry using the HFACS-OGI framework. **International Journal of Occupational Safety and Ergonomics**, 2022. v. 28, n. 3, p. 1642–1654.

PEREZ, C. A. *et al.* Automatic Safety Video Surveillance-Tracking System to Avoid Vehicle-Workers Interaction for Mining Applications. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2014. p. 23–27.

RAMOS, P. M. S. *et al.* Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals. **Process Safety and Environmental Protection**, ago. 2022. v. 164, p. 566–581.

REASON, J. **Managing the Risks of Organizational Accidents**. [S.l.]: Routledge, 2016.

REDMON, J. *et al.* You Only Look Once: Unified, Real-Time Object Detection. 8 jun. 2015. Disponível em: <<http://arxiv.org/abs/1506.02640>>.

REN, S. *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 4 jun. 2015. Disponível em: <<http://arxiv.org/abs/1506.01497>>.

ROBB, M.; MILLER, G. Human factors engineering in oil and gas - a review of industry guidance. **Work**, 2012. v. 41, p. 752–762.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, 1958. v. 65, n. 6, p. 386–408.

RUDIN, Walter. **Functional analysis**. [S.l.]: McGraw-Hill Education (India), 1991.

SILVA, Y. DA *et al.* Computer Vision Based Path Following for Autonomous Unmanned Aerial Systems in Unburied Pipeline Onshore Inspection. **Drones**, 14 dez. 2022. v. 6, n. 12, p. 410.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 4 set. 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>.

SMARTLABBR. Frequência de Afastamentos - INSS. **Afastamentos Previdenciários Acidentários**, 2022. Disponível em:

<<https://smartlabbr.org/sst/localidade/0?dimensao=frequenciaAfastamentos>>. Acesso em: 13 jan. 2025.

THEOPHILUS, Stephen Chukwubuike *et al.* Human factors analysis and classification system for the oil and gas industry (HFACS-OGI). **Reliability Engineering and System Safety**, 1 nov. 2017. v. 167, p. 168–176.

- UIJLINGS, J. R. R. *et al.* Selective search for object recognition. **International Journal of Computer Vision**, set. 2013. v. 104, n. 2, p. 154–171.
- VASUDEVAN, H.; ZOLGHADRI, S.; MAKAREM, M. A. Introduction to oil, gas, and petrochemical industries: importance to the current world. **Crises in Oil, Gas and Petrochemical Industries**. [S.l.]: Elsevier, 2023, p. 25–46.
- VINA, A. What is R-CNN? A quick overview. 2 nov. 2024. Disponível em: <<https://www.ultralitics.com/pt/blog/what-is-r-cnn-a-quick-overview>>. Acesso em: 15 jun. 2025.
- WANASINGHE, T. R. *et al.* Digitalization and the Future of Employment: A Case Study on the Canadian Offshore Oil and Gas Drilling Occupations. **IEEE Transactions on Automation Science and Engineering**, abr. 2024. v. 21, n. 2, p. 1661–1681.
- WANG, Wenbin *et al.* Solid–liquid separation state control system for shale shaker based on computer vision. **Petroleum Science and Technology**, 12 out. 2024. p. 1–26.
- WU, J. Design and Exploration of Real Time Monitoring System for Power Grid Operation Safety Based on YOLO Object Detection Algorithm. [S.l.]: Elsevier B.V., 2023. V. 228, p. 889–897.
- YANG, J.; KWON, Y. Human factor analysis and classification system for the oil, gas, and process industry. **Process Safety Progress**, 29 dez. 2022. v. 41, n. 4, p. 728–737.
- YANG, S. *et al.* Image Data Augmentation for Deep Learning: A Survey. 18 abr. 2022. Disponível em: <<http://arxiv.org/abs/2204.08610>>.
- YESSENBAYEV, Zh.; KOZHIRBAYEV, Zh.; SHINTEMIROV, A. Development of a computer vision module for autonomous vehicles. **Journal of Mathematics, Mechanics and Computer Science**, dez. 2022. v. 116, n. 4.
- YUAN, C. *et al.* Analysis method for causal factors in emergency processes of fire accidents for oil-gas storage and transportation based on ISM and MBN. **Journal of Loss Prevention in the Process Industries**, 1 nov. 2019. v. 62.
- ZHANG, A. *et al.* **Dive into Deep Learning**. [S.l.]: [s.n.], 2021.
- ZHANG, J.; QU, Q.; CHEN, X. Bo. Assessing the sustainable safety practices based on human behavior factors: an application to Chinese petrochemical industry. **Environmental Science and Pollution Research**, 1 jun. 2022. v. 29, n. 29, p. 44618–44637.
- ZHU, Z. *et al.* Deep learning for autonomous vehicle and pedestrian interaction safety. **Safety Science**, 1 jan. 2022. v. 145.