# VISÃO COMPUTACIONAL

## *Aplicada na Gestão da Tecnologia da Informação*

**Ricardo Lucas Pires**
ricardo.pires2@fatec.sp.gov.br

**Fábio Eder Cardoso**
fabio.cardoso6@fatec.sp.gov.br

## RESUMO

A Visão Computacional destaca-se como catalisadora na melhoria da Interação Homem-Computador (IHC) ao interpretar dados visuais. Inserida no âmbito da Tecnologia da Informação (TI), essa abordagem integra Inteligência Artificial (IA), Aprendizado de Máquina (AM) e Aprendizado Profundo (AP), impulsionando avanços na área. Técnicas específicas, como algoritmos de processamento de imagem, segmentação e classificação de objetos, desempenham papel crucial, expandindo as possibilidades na Gestão de TI. Os avanços em Redes Neurais Convolucionais (CNN) agregam refinamento e complexidade, destacando-se como ferramentas promissoras na otimização da eficiência e na tomada de decisões em TI. A evolução contínua nesse campo permite a implementação de algoritmos especializados, abrangendo desde o processamento de imagem até a classificação de objetos. A sinergia entre IA, AM, AP e CNN delineia um horizonte promissor para o desenvolvimento dessa disciplina, enfatizando seu impacto na otimização de processos, identificação de oportunidades e resolução de desafios na Gestão de TI.

**Palavras-chave:** Visão Computacional. Inteligência Artificial. Aprendizado de Máquina. Interação Homem-Computador. Tecnologia da Informação.

## ABSTRACT

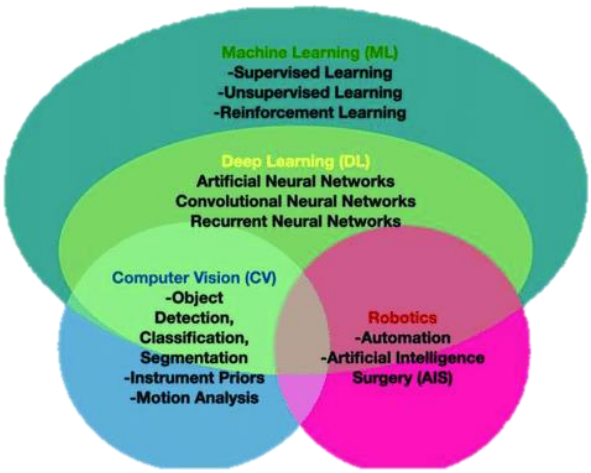
Computer Vision stands out as a catalyst for improving Human-Computer Interaction (HCI) by interpreting visual data. Embedded in the realm of Information Technology (IT), this approach integrates Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), propelling advancements in the field. Specific techniques, such as image processing algorithms, segmentation, and object classification, play a crucial role in expanding possibilities in IT Management. Advances in Convolutional Neural Networks (CNNs) add refinement and complexity, emerging as promising tools for optimizing efficiency and decision-making in IT. The continuous evolution in this field allows for the implementation of specialized algorithms, covering everything from image processing to object classification. The synergy among AI, ML, DL, and CNN delineates a promising horizon for the development of this discipline, emphasizing its impact on process optimization, identification of opportunities, and resolution of challenges in IT Management.

**Keywords:** Computer Vision, Artificial Intelligence, Machine Learning, Information Technology.

# 1 INTRODUÇÃO

A Visão Computacional, integrante da ciência da computação, destina-se a habilitar os computadores a interpretarem e compreenderem o mundo visual de maneira análoga aos seres humanos. Em conjunto com o Aprendizado de Máquina e a Inteligência Artificial, essas tecnologias estão promovendo transformações em diversos setores, impulsionando a automação, análise de dados e aprimorando a tomada de decisões.

**Figura 1** – Inteligência Artificial e seus campos.




**Fonte:** Andrew A (GUMBS et al., 2022)

Pesquisas recentes, exemplificadas pelos estudos de Wu *et al*. (2019), Liang e Seo (2022), e Sanyal (2023), como "*Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset.*"[1] e "*Automatic Detection of Construction Workers' Helmet Wear Based on Lightweight Deep Learning.*"[2] destacam a relevância da visão computacional na automação de processos e no aumento da precisão nas análises de dados. Essa tecnologia, além de sua aplicação evidente em setores como saúde, indústria, segurança e educação, apresenta um vasto potencial de contribuição em várias outras áreas.

A visão computacional, como subárea da inteligência artificial, é dedicada à análise e automação de tarefas de percepção visual realizadas por computadores, abrangendo uma ampla gama de desafios específicos. A notável capacidade de uma máquina ou computador em identificar

[1]Detecção automática de capacetes utilizados por pessoal de construção: Uma abordagem de aprendizado profundo e conjunto de dados de referência. (tradução nossa)
[2] Detecção Automática do Uso de Capacete por Trabalhadores da Construção com Base em Aprendizado Profundo Leve. (tradução nossa)

elementos em imagens complexas envolve processos intricados de cálculos matemáticos e treinamento de redes neurais, as quais aprendem a distinguir padrões, emulando o processo de aprendizagem humano.

No atual cenário de rápida evolução tecnológica, diversos modelos de redes neurais têm apresentado resultados notáveis e altamente precisos. No entanto, diante de possíveis deficiências nos procedimentos de fiscalização do uso de Equipamentos de Proteção Individual (EPIs), torna-se imperativo abordar implicações legais e de segurança decorrentes dessas falhas. A falta de conformidade com regulamentações de segurança no ambiente de trabalho pode acarretar riscos significativos tanto para os trabalhadores quanto para as empresas.

A segurança no ambiente de trabalho, uma preocupação crítica, requer a adoção rigorosa de práticas seguras, incluindo o uso adequado de EPIs, indispensáveis para reduzir acidentes e lesões. Nesse contexto, a visão computacional e a automação surgem como elementos fundamentais para aprimorar a segurança no local de trabalho, oferecendo ferramentas inovadoras para detectar a conformidade com o uso de EPIs e prevenir acidentes.

As redes neurais desempenham um papel crucial na automação de tarefas de percepção visual, destacando-se por sua relevância, especialmente quando aplicadas à fiscalização e garantia de segurança no ambiente de trabalho.

A proposta deste artigo concentra-se em analisar as tecnologias contemporâneas capazes de corrigir as deficiências no monitoramento do uso de EPIs em ambientes de trabalho. O projeto, inicialmente, envolve uma pesquisa bibliográfica detalhada das principais tecnologias de Inteligência Artificial que utilizam câmeras para controlar o acesso e o fluxo de trabalhadores em áreas onde o uso de EPIs é obrigatório.

O projeto foi implementado em fases distintas, iniciando com a investigação sobre segurança e os esforços aplicados na detecção de EPIs, desde os mais simples, como o reconhecimento de capacetes de segurança, até o avanço para redes neurais mais complexas. O foco principal foi garantir a precisão dos resultados, avaliando as funcionalidades e a capacidade de reconhecer diversos tipos de EPIs e cenários.

## Objetivo Geral

O projeto visa auxiliar a resolver um problema crônico de fiscalização inadequada no uso de equipamentos de proteção individual (EPIs), que resulta infelizmente em lesões e fatalidades de trabalhadores e custos para empregadores e Estado. No Brasil, a cada 48 segundos um trabalhador sofre acidente e um morre a cada 4h. (ROCHA, 2020)

O principal objetivo é implementar um sistema de fiscalização automática e inteligente usando visão computacional e inteligência artificial para garantir o uso correto de EPIs. Isso promoverá ambientes de trabalho mais seguros, reduzirá acidentes e custos associados, e terá um impacto positivo na sociedade e principalmente na saúde e bem-estar dos trabalhadores.

### Objetivos Específicos

1. Desenvolver e implementar um sistema de detecção de objetos com foco em Redes Neurais Convolucionais (CNN); 2. Avaliar e selecionar as melhores opções de hardware e software para a Visão Computacional; 3. Garantir a precisão e eficácia da rede neural na identificação de Equipamentos de Proteção Individual (EPIs).

Estes objetivos visam não apenas criar uma solução tecnológica eficiente, mas também integrá-la de maneira coesa e eficaz no contexto da segurança no trabalho.

# 2 REVISÃO DA LITERATURA

## 2.1 APRENDIZADO DE MÁQUINA

No contexto do aprendizado de máquina, o modelo não supervisionado é uma abordagem no qual os dados de treinamento não são previamente classificados ou rotulados. Nesse cenário, os algoritmos são alimentados apenas com os dados brutos e são capazes de identificar similaridades entre grupos de dados, agrupando-os em classes ou *clusters* com base em suas características comuns (MISHRA, 2017).

No entanto, é importante ressaltar que a aprendizagem não supervisionada não será o foco deste projeto, uma vez que o objetivo principal é a detecção de objetos específicos em imagens, o que se enquadra em uma categoria diferente de aprendizagem. Portanto, serão aprofundadas as abordagens relacionadas à detecção de objetos por meio de redes neurais convolucionais (CNN).

Na aprendizagem supervisionada, os dados de treinamento são acompanhados de rótulos (*Labels*) que indicam o tipo ou a categoria a que pertencem. Geralmente, esses dados são divididos em conjuntos de treinamento e teste para avaliação do modelo. Esse tipo de aprendizado pode ser subdividido em dois grandes campos: regressão e classificação (WILSON, 2019).

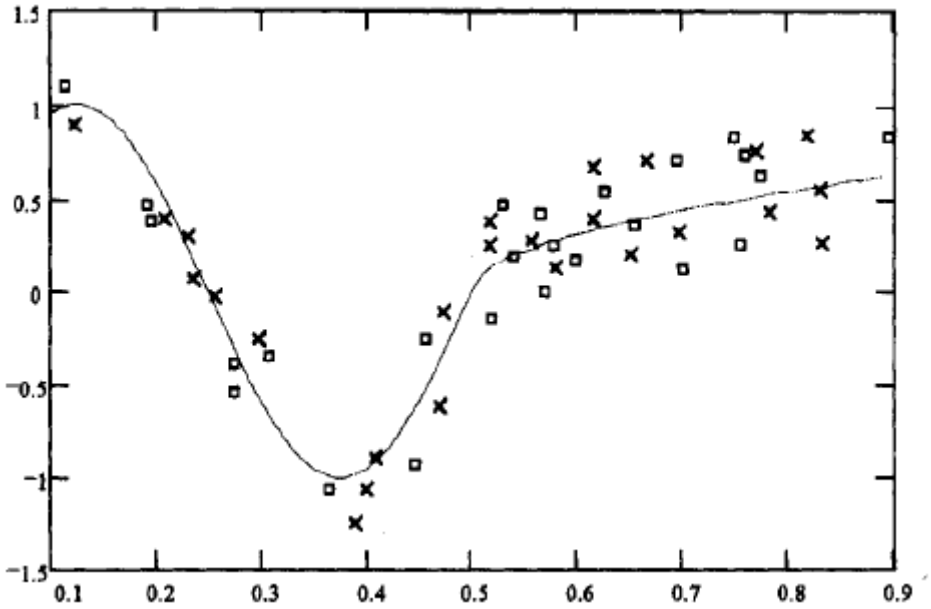
Na regressão, o objetivo é prever valores futuros com base em dados históricos. Já a classificação envolve agrupar os dados em classes ou categorias específicas. É importante notar que nem todos os parâmetros de um modelo são ajustáveis durante o treinamento. Alguns parâmetros, conhecidos como hiperparâmetros, precisam ser ajustados manualmente. No processo de ajuste de modelos, dois problemas principais podem surgir:

a) **Subajuste** (*Under-fitting*): Isso ocorre quando o modelo apresenta um alto erro tanto nos dados de treinamento quanto nos dados de teste. Nesse caso, as previsões não são precisas o suficiente

para ambas as amostras. A solução geralmente envolve a aquisição de mais dados ou a utilização de modelos mais complexos.

b) **Superajuste** (*Over-fitting*): No superajuste, o modelo funciona bem nos dados de treinamento, mas tem um alto erro ao lidar com novos exemplos. Isso ocorre quando o modelo é excessivamente complexo e perde a capacidade de generalização. A correção geralmente requer simplificar o modelo ou aplicar técnicas de regularização para evitar a complexidade excessiva.

**Figura 2** – Exemplo de *Overfitting*




**Fonte**: Tetko, Livingstone e Luik, 2020

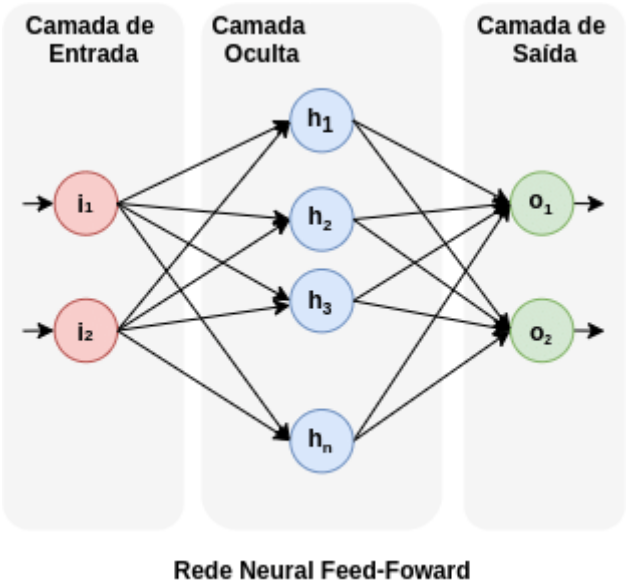
## 2.2 CLASSIFICAÇÃO

Em um modelo de classificação, o objetivo é atribuir um rótulo ou categoria a um dado de entrada, agrupando-o em uma classe específica. Esse tipo de modelo é amplamente utilizado para categorizar dados e é fundamental em diversas aplicações, incluindo reconhecimento de objetos em imagens, diagnóstico médico, detecção de fraudes e muito mais. O modelo de classificação é uma ferramenta essencial na área de aprendizado de máquina e desempenha um papel crucial em resolver uma variedade de problemas complexos. A regressão logística, apesar do nome, é um método amplamente utilizado para classificação binária. Essa abordagem é discutida em grande parte da literatura, como no trabalho de Bishop (2011).

## 2.3 REDES NEURAIS

Redes neurais artificiais são sistemas de computação inspirados na estrutura do cérebro humano, projetados para permitir que programas de computador identifiquem padrões de forma semelhante ao cérebro humano. Essas redes neurais são compostas por unidades individuais de processamento chamadas de neurônios artificiais, que são baseadas em funções matemáticas. Cada neurônio artificial recebe uma ou mais entradas, realiza operações matemáticas nessas entradas e produz uma saída. A figura 2 ilustra um exemplo de rede neural artificial.

**Figura 2**. Uma rede neural típica consiste em três camadas principais




**Fonte**: (BARBOSA et al., 2021)

1. **Camada de Entrada**: Recebe os dados de entrada, que geralmente são representados como matrizes de valores de intensidade de pixels em uma imagem.

2. **Camada Oculta**: Essas camadas processam as informações e identificam padrões nos dados de entrada.

3. **Camada de Saída**: Fornece a inferência final com base no processamento das camadas intermediárias.

As redes neurais são fundamentais em problemas de visão computacional, como a classificação de imagens, pois podem aprender a reconhecer padrões complexos a partir de grandes volumes de dados.

## 2.4 APRENDIZADO PROFUNDO

O aprendizado profundo, também conhecido como *Deep Learning*, é uma abordagem de aprendizado de máquina baseada em redes neurais artificiais que consistem em múltiplas Camadas Ocultas

(*Hidden Layers*). Essas redes são chamadas "profundas" devido ao grande número de camadas que podem conter.

Existem várias arquiteturas de aprendizado profundo, incluindo Redes Neurais Profundas (*Deep Neural Networks - DNN*), Redes de Crenças Profundas (*Deep Belief Networks - DBN*), Redes Neurais Recorrentes (*Recurrent Neural Networks - RNN*), e Redes Neurais Convolucionais (*Convolutional Neural Networks - CNN*). (Ren; Others, 2015).

Essas arquiteturas utilizam múltiplas camadas para extrair características e informações úteis das entradas. Por exemplo, em processamento de imagens, camadas inferiores podem identificar características simples, como bordas, enquanto camadas superiores podem identificar conceitos mais complexos, como dígitos, letras ou rostos.

Uma imagem digital é frequentemente representada como uma matriz, onde cada elemento da matriz representa a intensidade de um pixel. Em imagens coloridas, a representação padrão é o modelo RGB (*Red, Green, Blue*), que usa três canais para representar as cores vermelha, verde e azul.

Além disso, no processamento de imagens, a convolução desempenha um papel crucial. A convolução envolve a aplicação de uma matriz chamada *kernel* ou máscara a uma imagem para realizar várias operações, como desfoque, afiação, detecção de bordas e outras transformações que podem melhorar a extração de características e o processamento de imagens. Isso é fundamental em tarefas de visão computacional.

### 2.4.1 Redes Neurais Convolucionais (CNN)

Segundo (REN; OTHERS, 2015) as Redes Neurais Convolucionais (CNN) desempenham um papel crucial na análise de imagens e na extração de características relevantes, tais como bordas, texturas e padrões. Elas são a base arquitetônica primordial empregada na tarefa de processamento de imagens.

## 2.5 *FRAMEWORKS*

Vários modelos de detecção de objetos, incluindo o YOLO e o Darknet, fazem uso de camadas convolucionais para processar imagens e identificar características cruciais para a detecção de objetos. Portanto, a conexão entre o YOLO, o Darknet[3] e as CNNs residem no fato de que o YOLO é um modelo especializado em detecção de objetos que se aproveita da arquitetura

[3] https://github.com/pjreddie/darknet

das Redes Neurais Convolucionais para processar imagens. Com frequência, o YOLO é implementado com o auxílio do framework Darknet, que disponibiliza as ferramentas essenciais para o treinamento e uso eficaz do YOLO[4]

### 2.5.1 Darknet

Darknet é um *framework* de código aberto empregado para o desenvolvimento do YOLO. Ele oferece as ferramentas necessárias para treinar e implantar Redes Neurais Convolucionais, sobretudo para tarefas de visão computacional. Sua concepção visa eficiência e flexibilidade, o que o torna uma escolha popular na implementação de diversas arquiteturas de CNN em tarefas de detecção de objetos e outras aplicações de visão computacional.

### 2.5.2 YOLO (*You Only Look Once*)

O YOLO (*You Only Look Once*) "Você Só Olha Uma Vez" é um *framework* notável na detecção de objetos em imagens, fundamentado em princípios-chave. Ele realiza a detecção e classificação de objetos em uma única etapa, otimizando eficiência e velocidade. Para isso, divide a imagem em uma grade e prevê caixas delimitadoras e classes para cada célula dessa grade. Ademais, realiza previsões em múltiplas escalas, o que lhe permite detectar objetos de diferentes tamanhos na mesma imagem. O YOLO também utiliza uma função de perda personalizada, que penaliza erros na localização e classificação de objetos. Tudo isso é potencializado por redes neurais convolucionais profundas, que aprimoram a precisão e velocidade do processo. Versões melhoradas, como YOLOv1, YOLOv2, YOLOv3, continuam a evoluir e aprimorar o sistema.

O YOLO oferece várias vantagens, incluindo alta velocidade, tornando-o adequado para aplicações em tempo real, bem como uma precisão competitiva nas versões mais recentes. Além disso, possui a capacidade de detectar múltiplos objetos em uma única imagem. No entanto, o YOLO apresenta desvantagens notáveis, incluindo limitações na detecção de objetos muito pequenos ou que estejam muito próximos uns dos outros. Em suma, o YOLO é um *framework* de destaque na detecção de objetos em imagens, graças a suas características únicas e aplicações versáteis, embora com algumas limitações a serem consideradas.

[4] https://docs.ultralytics.com/

# 4 PROJETO

Para o presente trabalho foi utilizado um conjunto de dados (*Dataset*) criado com imagens de segurança de canteiro de obras. O conjunto de dados consiste em 200 amostras de imagens com rótulos no formato YoloV8. Essas imagens estão divididas em conjuntos de treinamento: 2605, validação: 114 e teste: 82. Cada pasta contém pastas de imagens e rótulos. Existem 10 classes para detectar no conjunto de dados: 'Capacete', 'Máscara', 'Sem-Capacete', 'Sem-Máscara', 'Sem-Colete de Segurança', 'Pessoa', 'Cone de Segurança', 'Colete de Segurança', 'maquinaria', 'veículo'.

## 4.1 ETAPAS

A seguir, é apresentada a tradução das etapas listadas:

1. Instale as ferramentas necessárias usando "Miniconda"; 2. Crie um ambiente virtual; 3. Reúna o conjunto de dados para treinamento; 4. Anote as imagens coletadas; 5. Configure o YOLOv8 (UltralyticsPlus); 6. Treine o modelo; 7. Avalie e teste o modelo treinado; 8. Implante o modelo na plataforma Hugging Face.

## 4.2 TECNOLOGIAS

Este projeto utiliza uma combinação de ferramentas de código aberto e proprietárias para alcançar seus objetivos:

***Git*** [1]- Controle de versão e colaboração.

**Miniconda**[2] - Gerenciamento de ambiente.

**Simple-Image-Download**[3] - Coleta de imagens.

**LabelImg**[4] - Anotação de conjuntos de dados.

**Ultralyticsplus**[5] - para YOLOv8 e implantação simplificada no Hugging Face.

**Hugging Face**[6] - Implantação de modelos.

## 4.3 HIERARQUIA DE ARQUIVOS

A pasta "data" contém o arquivo *yaml* necessário para o treinamento. Ela também contém 3 pastas: "*train*," "*valid*" e "*test*." Cada uma dessas pastas tem 2 subpastas: "*images*" (com arquivos .jpg) e

[1] https://git-scm.com/
[2] https://docs.conda.io/projects/miniconda/en/latest/#
[3] https://pypi.org/project/simple-image-download/
[4] https://pypi.org/project/labelImg/1.4.0/
[5] https://huggingface.co/ultralyticsplus/yolov8s
[6] https://huggingface.co/

"*labels*" (com anotações em .txt). A pasta "*results*" contém os resultados das previsões do modelo, o gráfico da matriz de confusão, visualizações dos lotes de treinamento e validação e curvas PR. A pasta "*models*" contém 2 modelos: "*yolov8n.pt*," que é o modelo pré-treinado no arquivo "*COCO128.yaml*," e "*best.pt*," que é o modelo *YoloV8n* treinado personalizado em nosso conjunto de dados. A pasta "*source_files*" contém vídeos e imagens para avaliação do nosso modelo treinado personalizado. A pasta "*output*" contém a saída produzida pelo nosso modelo de detecção de objetos personalizado após 100 épocas de treinamento.

## 4.4 CÓDIGO

Todo o projeto foi desenvolvido com tecnologias como, OpenCV[5], NumPy[6], TensorFlow[7], PyTorch[8], YOLO[9]. uso da linguagem de programação.

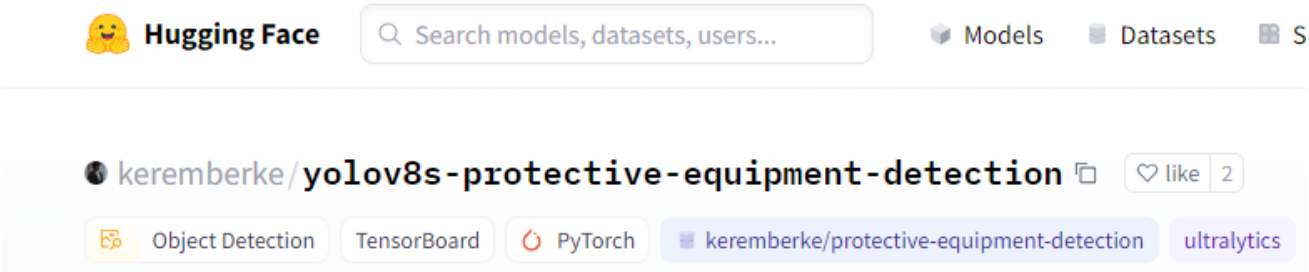
## 4.5 RESULTADOS

O código foi executado no *Google Colab*[10], com uma GPU T4. Foi instalada a biblioteca *Ultralytics*[11] para executar a detecção de objetos personalizada YoloV8 no conjunto de dados.

## 4.6 SAÍDAS

**Uso.** Para obter detalhes sobre treinamento e avaliação, é necessário consultar a pasta "Treinamento" (*Training*). Para entender o processo de implantação no *Hugging Face*, confira a pasta "Hugging Face".

**Figura 3** – Hugging Face




**Fonte**: Hugging Face (2023).

[5] https://opencv.org/
[6] https://numpy.org/
[7] https://www.tensorflow.org/
[8] https://pytorch.org/
[9] https://pjreddie.com/darknet/yolo/
[10] https://colab.google/
[11] https://www.ultralytics.com/

**Figura 4** – Detecção Capacete e Colete.


**Fonte**: (SANYAL, 2023)

**Figura 5** – Detecção em obras.


**Fonte**: (WU et al., 2019)

## 5 CONSIDERAÇÕES FINAIS

Na pesquisa, a detecção automática do uso de capacetes de segurança em locais de construção e a identificação das cores correspondentes foram abordadas com base no conjunto de dados YOLO.

O conjunto de dados YOLO é composto por um número significativo de imagens e instâncias, o que o torna robusto para treinamento e teste. Ele é dividido em classes e cada instância é anotada com uma etiqueta de classe e uma caixa delimitadora, facilitando a identificação de pessoas que não estão usando capacetes de segurança.

Essa pesquisa representa um avanço importante na automação da segurança em locais de construção, contribuindo para a redução de riscos ocupacionais. Utilizando o *hardware* já existente em pátios de obras como câmeras de segurança já existentes e aplicando *softwares* como YOLO os resultados alcançados têm o potencial de aprimorar a eficácia e eficiência da fiscalização de uso de equipamentos de proteção individual, melhorando a segurança dos trabalhadores em ambientes de construção. Portanto, essa pesquisa representa uma contribuição significativa para a área de segurança no trabalho em construção civil.

Como sugestão de trabalho futuro, recomenda-se primeiramente o treinamento do modelo por mais épocas, o que poderia aprimorar significativamente seu desempenho. Em seguida, seria valioso comparar o modelo atual com outros quatro modelos do YoloV8, para avaliar sua eficácia em diferentes configurações e cenários. Uma terceira sugestão envolve a criação de um sistema de rastreamento que identifique trabalhadores e salve as caixas delimitadoras daqueles que não estiverem usando o Equipamento de Proteção Individual (EPI) apropriado. Isso não apenas aumentaria a segurança no ambiente de trabalho, mas também forneceria dados valiosos para futuras análises. Por fim, propõe-se a implantação de um aplicativo de aprendizado de máquina que inclua um mecanismo de acionamento de alarme. Esse recurso poderia servir como um alerta imediato em situações em que o uso incorreto ou a falta do EPI é detectada, contribuindo para a prevenção de acidentes e reforçando as medidas de segurança.

# REFERÊNCIAS


BARBOSA, G. et al. **Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina.** Em: [s.l: s.n.]. p. 145–189.

GUMBS, A. A. et al. **The Advances in Computer Vision That Are Enabling More Autonomous Actions in Surgery: A Systematic Review of the Literature. Sensors**, v. 22, n. 13, p. 4918, 29 jun. 2022.

MISHRA, S. **Unsupervised Learning and Data Clustering**. Disponível em: <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>.

REN, S.; OTHERS. **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. CoRR**, v. abs/1506.01497, 2015.

ROCHA, R. **No Brasil, a cada 48 segundos um trabalhador sofre acidente e um morre a cada 4h**. Disponível em: <https://spbancarios.com.br/08/2018/no-brasil-cada-48-segundos-um-trabalhador-sofre-acidente-e-um-morre-cada-4h>.

SANYAL, S. **PPE Detection for Construction Site Safety using YoloV8**. , 8 out. 2023. Disponível em: <https://github.com/snehilsanyal/Construction-Site-Safety-PPE-Detection>. Acesso em: 8 out. 2023

WILSON, A. **A Brief Introduction to Supervised Learning**. Disponível em: <https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590>. Acesso em: 23 out. 2023.

WU, J. et al. Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. **Automation in Construction**, v. 106, p. 102894, 1 out. 2019.