UNIVERSIDADE DE SÃO PAULO
PROGRAMA DE EDUCAÇÃO CONTINUADA – ESCOLA POLITÉCNICA DA USP

MATHEUS SCHABERLE GOVEIA

Sistema de visão com inteligência artificial embarcada para detecção de capacetes de segurança

SÃO PAULO
2024

MATHEUS SCHABERLE GOVEIA

Sistema de visão com inteligência artificial embarcada para detecção de capacetes de segurança

Trabalho de conclusão de curso apresentado à Escola Politécnica da Universidade de São Paulo no Programa de Educação Continuada (PECE) para a obtenção do título de Especialista em Inteligência Artificial.

Orientador (a): Profª. Drª. Larissa Driemeier

SÃO PAULO

2024

FOLHA DE AVALIAÇÃO

Nome: GOVEIA, Matheus Schaberle

Título: Sistema de visão com inteligência artificial embarcada para detecção de capacetes de segurança**.**

Trabalho de conclusão de curso apresentado à Escola Politécnica da Universidade de São Paulo no Programa de Educação Continuada (PECE) para obtenção do título de Especialista em Inteligência Artificial.

Aprovado em: / /

Banca Examinadora

Profº. Drº:_____

Instituição:_____

Julgamento:_____

Profº. Drº:_____

Instituição:_____

Julgamento:_____

AGRADECIMENTOS

Dedico essa trabalho a minha mãe que sempre me apoiou nos meu estudos e a minha noiva que apoiou durante todo o desenvolvimento do curso, sem seu suporte não teria conseguido chegar ao final.

“A verdadeira viagem de descobrimento não consiste em procurar novas paisagens, mas em ter novos olhos”.

(Marcel Proust)

LISTA DE ILUSTRAÇÕES

LISTA DE SIGLAS

EPI Equipamento de proteção Individual

EPC Equipamento de proteção coletiva

IA Inteligência artificial

NR Norma Regulamentadora

API *Application Programming Interface*

GPU *Graphics Processing Unit*

IP *Internet Protocol*

SUMÁRIO

RESUMO

O ambiente de trabalho pode ser perigoso para aqueles que nele atuam, para diminuir esse risco os trabalhadores devem usar corretamente seus equipamentos de segurança. É comum, entretanto encontrar trabalhadores os utilizando incorretamente e assim sujeitos a possíveis acidentes. Propõe-se, então a criação de um sistema de visão computacional capaz de identificar o uso correto dos equipamentos de segurança, através de um algoritmo de inteligência artificial utilizando o modelo YOLO, que foi treinado para ser capaz de detectar o uso correto de capacetes de segurança.

Palavras-chave: Inteligência artificial, sistema de visão, segurança no trabalho.

ABSTRACT

The work environment can be dangerous for those who operate within it, to decrease this risk, workers should correctly use their safety equipment. However, it is common to find workers using them incorrectly and thus subject to potential accidents. Therefore, the creation of a computer vision system capable of identifying the correct use of safety equipment is proposed, through an artificial intelligence algorithm using the YOLO model, which has been trained to detect the proper use of safety helmets.

Keywords: Artificial intelligence, vision system, workplace safety.

1. INTRODUÇÃO

O ambiente industrial, devido à sua natureza, apresenta riscos consideráveis para os operadores envolvidos. Para mitigar esses perigos, são estabelecidas diversas normas de segurança destinadas a regular um ambiente de trabalho seguro. Essas normas abrangem desde a disposição e espaçamento adequados das máquinas até a implementação de procedimentos de segurança e a utilização de equipamentos específicos para proteger os operadores e reduzir as possibilidades de acidentes.

Os equipamentos de proteção se dividem em dois tipos, os EPI (Equipamentos de Proteção Individuais), que devem ser utilizados pelos operadores, como luvas, óculos de proteção, capacete e roupas de proteção e os EPC (Equipamentos de Proteção Coletiva), que servem para prover uma proteção ao grupo, como sistema de ventilação e barreiras que impeçam o acesso a uma determinada área.

O Observatório de Segurança e Saúde do Trabalho representa uma entidade governamental responsável por compilar informações sobre acidentes de trabalho provenientes de diversas fontes, incluindo dados fornecidos por empresas e secretarias em todo o Brasil. Segundo o levantamento feito pelo observatório (1) no ano de 2022, houve 612,9 mil casos de acidentes no trabalho, dentre eles cerca de 2500 foram acidentes com óbito. Dentro desses casos há ainda acidentes com lesões permanentes que podem fazer com que o trabalhador afetado não possa mais exercer atividade remunerada, ou também fraturas e traumatismos que afastem o trabalhador por um período de tempo até a total recuperação e retorno as atividades de trabalho.

É frequente observar operadores retirando seus Equipamentos de Proteção Individual (EPI) enquanto executam suas tarefas, justificando desconfortos associados ao uso desses equipamentos. Essa prática aumenta o risco de lesões em caso de acidentes. Para lidar com essa questão, as empresas geralmente contam com técnicos de segurança do trabalho, encarregados de supervisionar a utilização adequada dos EPIs pelos colaboradores. No entanto, nem sempre é viável realizar esse monitoramento de forma contínua e assídua em todas as áreas ao longo de toda a jornada de trabalho da empresa. Conforme estipulado pela Norma Regulamentadora (NR) no. 4 (2), quando uma empresa é categorizada com o mais alto grau de risco, o grau de risco 4, é requerido que esta possua pelo menos 1 técnico de segurança do trabalho para cada 50 colaboradores e, a partir de 3501 colaboradores, a necessidade aumenta para 10 técnicos. Monitorar todas as pessoas na empresa ao mesmo tempo torna-se uma tarefa desafiadora, dada a escassez de técnicos em relação à quantidade de colaboradores.

Dentro das companhias dos mais diversos tipos é comum ter em suas instalações câmeras de segurança que têm como objetivo a segurança patrimonial, principalmente no período noturno, quando a circulação de pessoas é menor. Essas câmeras são geralmente distribuídas por todas as dependências com os propósitos de intimidar a presença de intrusos e auxiliar eventuais investigações em caso de extravios.

Assim, dado o reduzido contingente de técnicos de segurança nas empresas, a implementação de um sistema de visão computacional pode ser uma abordagem mais eficaz para a tarefa de monitoramento.

Os sistemas de visão computacional podem ser utilizados para resolver diversos problemas em setores distintos, como monitoramento de pacientes feito por Silva (3), monitoramento de robôs móveis para ambinetes educacionais feito por Rios e Netto (4), detecção de componentes em subestações elétricas feito por Oliveira (5) e até mesmo para controle de estoque numa loja de autopeças como feito por De Souza Oliveira (6).

O presente trabalho pretende utilizar as imagens capturadas por câmeras, sejam elas de segurança ou de controle de acesso de funcionários, para realizar a conferência correta do uso de capacetes pelos colaboradores. Dessa forma, o monitoramento contínuo do uso do EPI não apenas assegura a conformidade durante a execução das atividades, mas também minimiza substancialmente o risco de lesões em caso de acidentes no ambiente de trabalho.

2. REVISÃO BIBLIOGRÁFICA

Ao longo dos anos foram desenvolvidas diversas abordagens e metodologias para resolver esse mesmo problema da detecção de objetos. A técnica Haar Cascades desenvolvida por Viola e Jones, (7) em 2001 foi considerada a primeira abordagem a conseguir realizar detecções de faces na ordem de grandeza de milisegundos (0,067s), sendo aproximadamente 15 vezes mais rápida que os métodos anteriores e com um baixo número de falsos posistivos, sendo ainda utilizada nos dias de hoje como feito por Javed Mehedi Shamrat, et al. (8) para detecção de rostos.

Desenvolvido por Dalal e Triggs (9), alguns anos depois, em 2005, o *Histogram of Oriented Gradients* (HOG), considerado também um marco importante para essa área, pois foi um modelo capaz de fazer a detecção de pedestres e também de faces com sucesso acima de 90% em diversas bases de dados de rostos utilizadas na época.

A PASCAL VOC, base de dados considerada benchmark para detecção de objetos, que como descrita por Everingham et al. (10), trata-se de uma base de imagens já anotadas com 20 classes diversas de 4 tipos principais: pessoas, veículos, animais e objetos domésticos, posui aproximadamente 10000 imagens, todas retiradas do site flicker.com. Com a ascensão do Deep Learning, foram desenvolvidos métodos baseados em Redes Neurais Convolucionais, do inglês convolutional neural network (CNN). Dentre os principais, pode ser citado: R-CNN (do inglês, Region-based Convolutional Neural Networks), onde Girshick et al. (11), foi capaz de melhorar em 30 pontos percentuais a mAP do inglês *mean average precision*, em relação ao resultado obtido utilizando o HOG, obtendo um mAP de 53,3% um salto significativo para época em que o autor afirmava que a performance da detecção de objetos estava estagnada nos últimos anos.

A abordagem da SPP-net (Spatial Pyramid Pooling), desenvolvida por He et al. (12), trouxe uma notável inovação ao permitir o uso de CNNs sem a restrição de um tamanho de entrada fixo. A metodologia permite utilizar imagens de diversos tamanhos e proporções, eliminando a necessidade de empregar técnicas de corte de imagem anteriormente utilizadas. Essa flexibilidade na detecção de objetos resultou em um mAP de 59,2% na base de dados PASCAL VOC, demonstrando o avanço significativo proporcionado pela SPP-net.

Os obstáculos eram o treinamento das redes e o tempo de detecção dos objetos nas imagens. Então Girshick (13), apresenta a Fast R-CNN (do inglês, Fast Region-based Convolutional Neural Networks), sendo treinada 2.7 vezes mais rápida que a SPP-net, levando 9,5 horas vs. 25,5 horas, e detectando 7 vezes mais rápido, levando 0,32 fente a 2,3 segundos por imagem. Uma extensão desse trabalho levou Ren et al. (14) a propor a Faster R-CNN (do

inglês, Faster Region-based Convolutional Neural Networks) que além de obter um mAP de 73,2%, estado da arte para época, também era capaz de realiza detecções em 5fps, cerca de 0.2s por imagem, o que foi considerado detecção em tempo real, uma grande evolução considerando que a R-CNN levava em torno de 50s.

A YOLO (You Only Look Once), arquitetura proposta por Redmon et al. (15), surgiu em paralelo ao Faster R-CNN, destacando-se por sua eficiência ao realizar a detecção de objetos em tempo real, processando a imagem inteira de uma só vez, em oposição a abordagens que dividem a imagem em regiões para detecção. Notavelmente, a YOLO conseguiu alcançar uma taxa de detecção de até 155 frames por segundo na base de dados PASCAL VOC, destacando-se pela eficiência e rapidez em comparação com abordagens anteriores. As altas taxas de quadros por segundo são particularmente úteis em aplicações que exigem tempo de resposta rápido, como sistemas de vigilância e veículos autônomos.

Atualmente, o estado da arte das tarefas de visão computacional é representado pela DINOv2, desenvolvida por Oquab et al. (16). Este modelo, auto supervisionado, foi treinado com uma vasta base de dados composta por aproximadamente 142 milhões de imagens, criteriosamente selecionadas de um *dataset* de 1 bilhão de imagens. Sua arquitetura segue o padrão ViT (Vision Transformer), contendo cerca de 1 bilhão de parâmetros, mas oferece versões em modelos menores, adequando-se a ambientes com limitações de poder de processamento. Os autores testaram o modelo em diversas tarefas, incluindo classificação, segmentação, entendimento de vídeo, entre outras, obtendo consistentemente uma acurácia superior ou equivalente aos modelos considerados estado da arte na época de sua publicação.

3. MATERIAIS E MÉTODOS

As redes neurais artificais, do inglês *Convolution Neural Network* (CNN) são um tipo de arquitetura de rede neural artificial que é eficaz no reconhecimento de padrões, assim como descirta por O'Shea, Nash (17), é muito poderosa nas tarefas de reconehcimentos com imagens, sendo assim será a ferramenta utilizada para a resolução desse problema.

Com as CNNs será possivél reconhecer os padrões dos EPIs nas imagens que servirão de entrada para o modelo, não importando a orientação das fotos ou cenário de fundo já que elas conseguem isolar os EPIs desse contexto através da identificação de padrões. Para isso foi necessário definir um EPI que se desejava identificar e o escolhido foi o capacete.

Como os capacetes podem ter cores diferentes e possuem formas diferentes se estão sendo observados de frente ou por trás ou até mesmo por cima, era necessário um banco extenso e o ideal é que as imagens sejam as mais variadas possível, com fundos diferentes, cores dos capacetes diversos, possuirem pessoas de diferentes etnias, gênero, ângulos diferentes. Gerar esse tipo de banco levaria muito tempo, assim optou-se por utilizar a base de dados de Maranhão (17). Essa base de dados possui 5000 imagens, divididas em três classes diferentes: capacete, cabeça e pessoa e não possui restrição de direitos autorais, podem ser utilizadas livremente. Possui todas as imagens já devidamente segmentadas, desse modo não foi necessário se alongar com essa tarefa. A figura 1, exemplifica algumas imagens da base escolhida, onde destaca-se as classes nela presente.

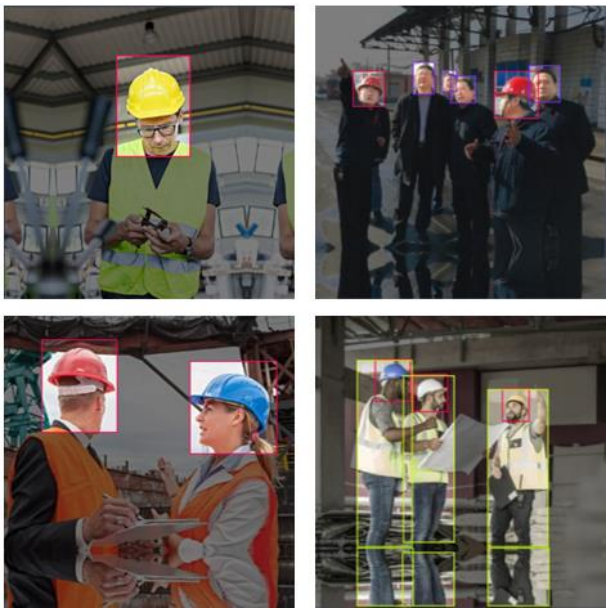

Figura 1 - Exemplo de imagens com classes capacete, cabeça e pessoa destacados, retirado da base de dados utilizada

A tarefa de detecção será feita pela YOLO que originalmente é uma rede neural composta por 24 camadas convolucionais seguidas de duas camidas totalmente conectadas, que seguem a seguinte lógica de funcionamento: a YOLO é baseada na ideia de segmentar uma imagem em imagens menores. Na figura 2 temos a imagem dividida em uma grade quadrada de dimensões S×S.

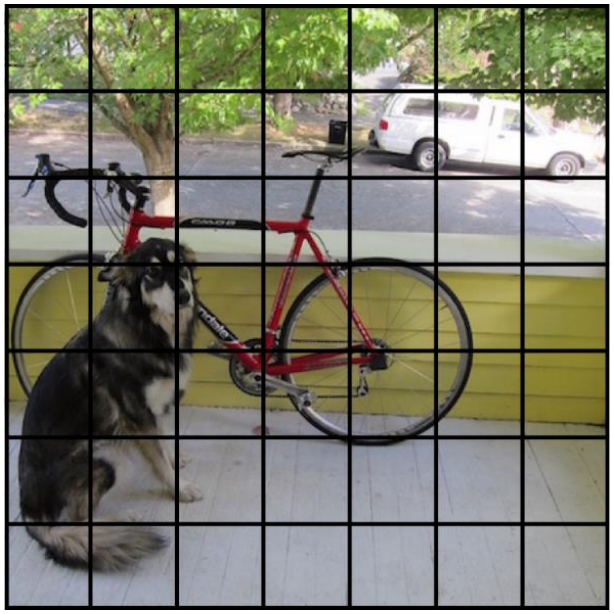

Figura 2 - Imagem segmentada – Redmon et al. (15)

A célula na qual o centro de um objeto, por exemplo, o centro do cachorro, reside, é a célula responsável por detectar esse objeto. Cada célula preverá B caixas delimitadoras e uma pontuação de confiança para cada caixa. O padrão para esta arquitetura é o modelo prever duas caixas delimitadoras. A pontuação de classificação será de **0,0** a **1,0**, sendo **0,0** o nível mais baixo de confiança e **1,0** o mais alto; se nenhum objeto existir naquela célula, as pontuações de confiança devem ser **0,0**, e se o modelo estiver completamente certo de sua previsão, a pontuação deve ser **1,0**. Esses níveis de confiança capturam a certeza do modelo de que existe um objeto naquela célula e que a caixa delimitadora é precisa. Cada uma dessas caixas delimitadoras é composta por 5 números: a posição x, a posição y, a largura, a altura e a confiança. As coordenadas **(x, y)** representam a localização do centro da caixa delimitadora prevista, e a largura e a altura são frações relativas ao tamanho total da imagem. A confiança representa a IOU entre a caixa delimitadora prevista e a caixa delimitadora real, referida como a caixa delimitadora ground truth. A IOU significa *Intersection Over Union*, Intersecção sob a União, e é a área da interseção das caixas previstas e das caixas corretas onde se encontra o

objeto dividida pela área da união das mesmas caixas previstas e caixas corretas. Na figura 3 é possível ver a imagem sendo segmentada.

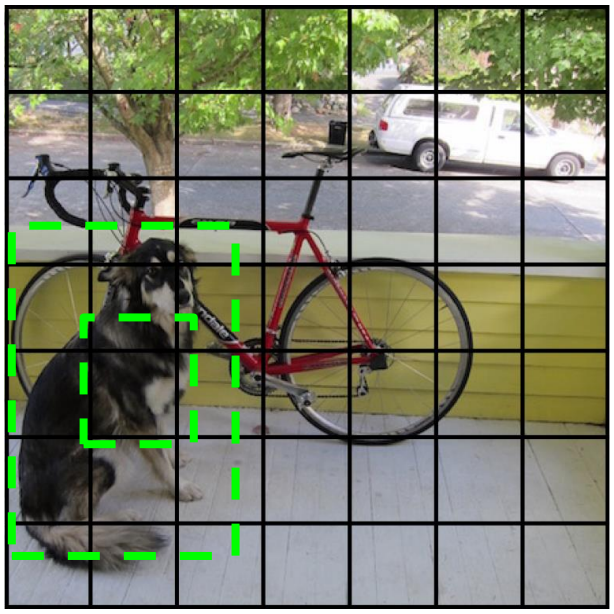

Figura 3 - Imagem segmentada com retângulos de detecção – Redmon et al. (15)

Além de produzir caixas delimitadoras e pontuações de confiança, cada célula prevê a classe do objeto. Essa previsão de classe é representada por um vetor one-hot de comprimento C, o número de classes no conjunto de dados. No entanto, é importante notar que, embora cada célula possa prever qualquer número de caixas delimitadoras e pontuações de confiança para essas caixas, ela prevê apenas uma classe. Isso é uma limitação do próprio algoritmo da YOLO, e se houver vários objetos de diferentes classes em uma célula de grade, o algoritmo falhará em classificar ambos corretamente. Assim, cada previsão de uma célula de grade terá a forma C + B * 5, onde C é o número de classes e B é o número de caixas delimitadoras previstas. B é multiplicado por 5 aqui porque inclui (x, y, w, h, confiança) para cada caixa. Como há S × S células de grade em cada imagem, a previsão geral do modelo é um tensor de forma S × S × (C + B ∗ 5).

É importante notar que o modelo prevê o centro da caixa delimitadora com larguras e alturas, em vez das posições dos cantos superior esquerdo e inferior direito. A classificação é

representada por um vetor one-hot e, neste exemplo trivial, existem 7 classes diferentes. A 5ª classe é a previsão e podemos ver que o modelo está bastante certo de sua previsão. Na figura 4 vemos a imagem de todas as caixas delimitadoras e previsões de classe que realmente seriam feitas e seu resultado final.

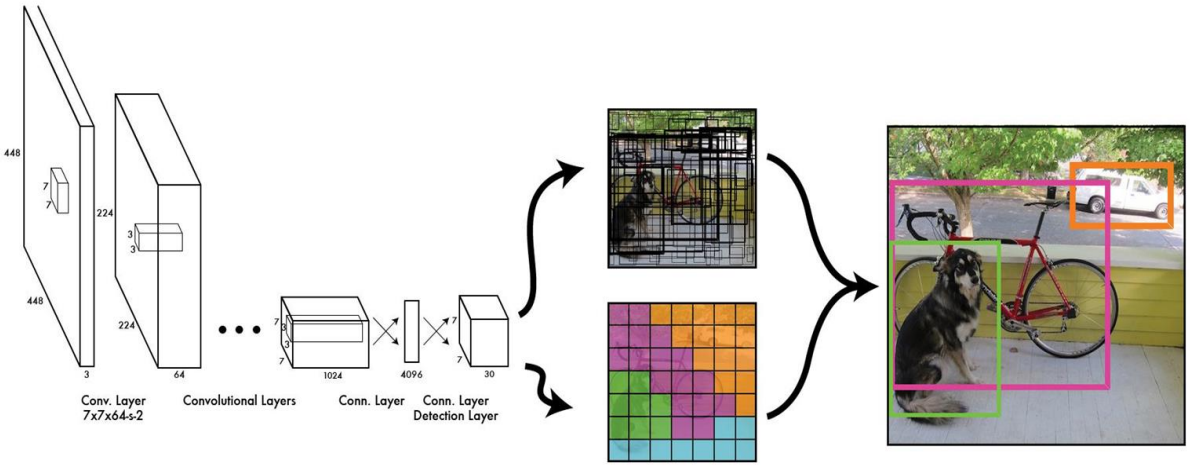

Figura 4 - Fluxo de detecção YOLO – Redmon et al. (15)

A rede YOLO vem sendo atualizada desde seu lançamento e atualmente encontra-se em sua versão 8, que pode ser encontrada diretamente ao se instalar a biblioteca *ultralytics*, sendo de implementação mais simples que as versões anteriores, tornando o código mais enxuto. A arquitetura da YOLOv8 é representa na figura 5.

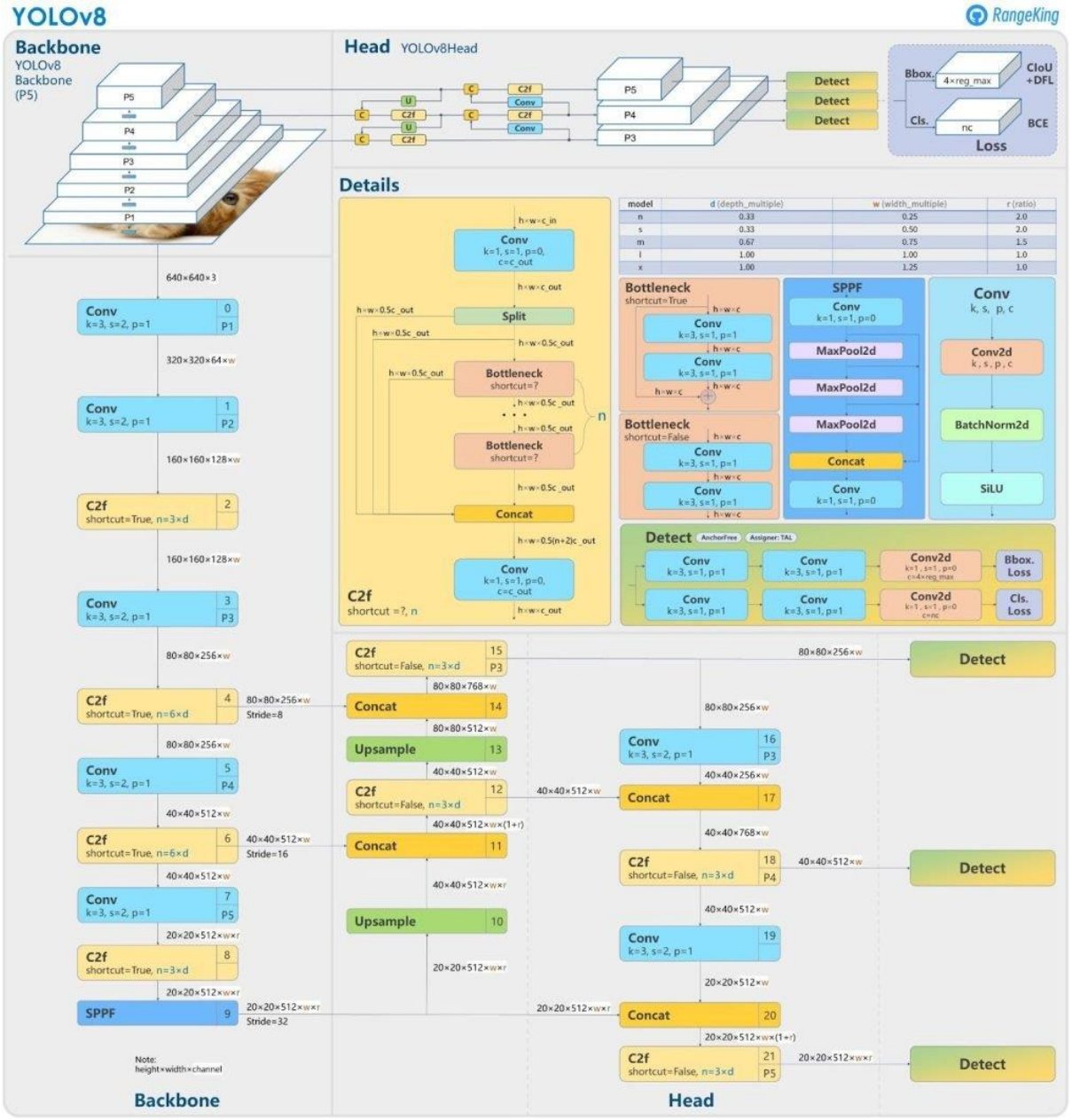

Figura 5 - Arquitetura YOLOv8 – Boesch (18)

Segundo Boesch (18) as principais características da YOLOv8 que diferem de sua versão anterior são:

- **Aumento de Dados em Mosaico:** Assim como a YOLOv4, a YOLOv8 utiliza aumento de dados em mosaico que mistura quatro imagens para fornecer ao modelo melhores informações de contexto. A mudança na YOLOv8 é que o aumento para nas últimas dez épocas de treinamento para melhorar o desempenho.

- **Detecção sem Âncora:** A YOLOv8 mudou para detecção sem âncora para melhorar a generalização. O problema com a detecção baseada em âncora é que as caixas de âncora predefinidas reduzem a velocidade de aprendizado para conjuntos de dados personalizados. Com a detecção sem âncora, o modelo prevê diretamente o ponto médio de um objeto e reduz o número de previsões de caixas delimitadoras. Isso ajuda a acelerar a Supressão Não-Máxima (NMS) - um passo de pré-processamento que descarta previsões incorretas.

- **Módulo C2f:** A espinha dorsal do modelo agora consiste em um módulo C2f em vez de um C3. A diferença entre os dois é que no C2f, o modelo concatena a saída de todos os módulos de gargalo. Em contraste, no C3, o modelo usa a saída do último módulo de gargalo. Um módulo de gargalo consiste em blocos residuais de gargalo que reduzem os custos computacionais em redes de aprendizado profundo. Isso acelera o processo de treinamento e melhora o fluxo de gradiente.

- **Cabeça Desacoplada:** O diagrama acima ilustra que a cabeça não executa mais classificação e regressão juntas. Em vez disso, ele realiza as tarefas separadamente, o que aumenta o desempenho do modelo.

- **Desalinhamento de Perda:** O desalinhamento de perda é possível, uma vez que a cabeça desacoplada separa as tarefas de classificação e regressão. Isso significa que o modelo pode localizar um objeto enquanto classifica outro. A solução é incluir um escore de alinhamento de tarefa com base no qual o modelo conhece uma amostra positiva e negativa. O escore de alinhamento de tarefa multiplica o escore de classificação com o escore de Interseção sobre União (IoU). O escore de IoU corresponde à precisão de uma previsão de caixa delimitadora. Com base no escore de alinhamento, o modelo seleciona as principais amostras positivas e calcula uma perda de classificação usando BCE e perda de regressão usando Complete IoU (CIoU) e Distributional Focal Loss (DFL). A perda de BCE simplesmente mede a diferença entre os rótulos reais e previstos. A perda de CIoU considera como a caixa delimitadora prevista é relativa à verdade em termos do ponto central e da proporção de aspecto. Em contraste, a perda focal distribucional otimiza a distribuição dos limites da caixa

delimitadora, concentrando-se mais em amostras que o modelo classifica erroneamente como falsos negativos.

Para ser possível utilizar a YOLO é necessária inicalmente retreiná-la, já que inicialmente ela possui a capacidade de realizar a detecção dos objetos do base de dados PASCAL VOC, mas não consegue realizar a detecção dos capacetes. O processo de retreinamento é feito de maneira supervisionada, onde todos os dados estam divididos em suas respectivas classes, também já possuem essas classes devidamente identificadas no formato PASCAL VOC, ou seja, com retêngulos delimitando a classe a qual aquela imagem pertence.

Para realizar o retreinamento utilizou-se a seguinte proporção: 70% para treinamento, 20% para validação e 10% para testes, as imagens foram embaralhadas aleatóriamente com o objetivo de tirar qualquer viés de ordenação da base de dados. Optou-se por treinar o modelo por 100 épocas, ao levantar a curva de precisão do modelo ao final do treinamento notou-se um comportamento não esperado ilustrado na figura 6.

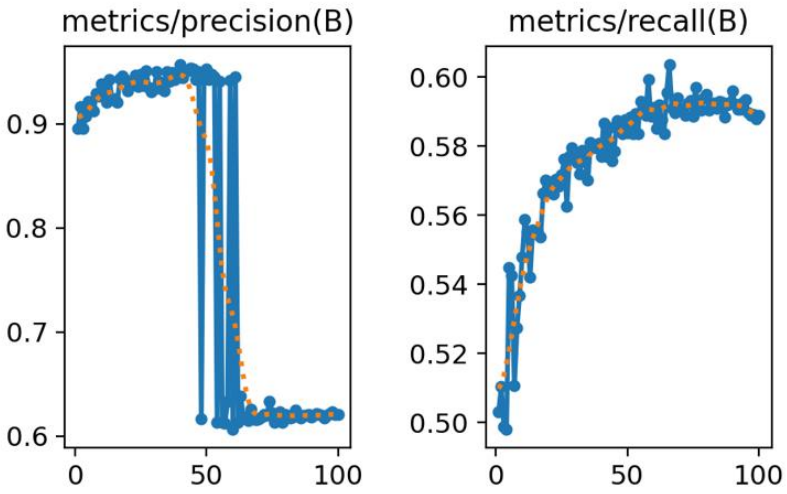

Figura 6 - Curva de precisão e recall após 100 épocas de treinamento

A curva de precisão teve uma queda abrupta próxima da época número 50, comportamento que não se refletiu na curva de recall que se manteve praticamente estável das épocas 50 a 100, como resultado disso o modelo identificou alguns objetos falso positivos. O algoritmo durante o treinamento salva os pesos da rede na última época e da época que obteve a maior precisão alcançada no treinamento, assim todos os resultados obtidos são referentes à época 44, a que teve a melhor precisão.

Na figura 7 temos a matriz de confusão gerada ao final do processo de treinamento, que ilustra o quanto as classes foram detectadas corretamente e quais não foram.

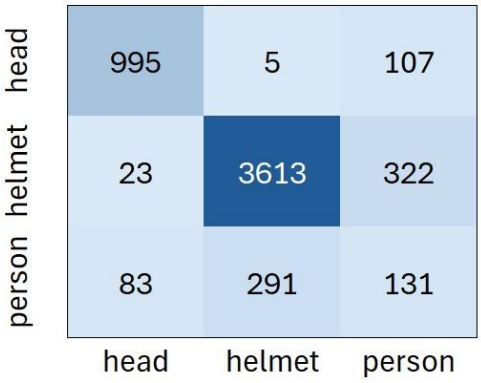

Figura 7 - Matriz de confusão do treinamento da rede

Observa-se que após o treinamento a rede foi capaz de idenficar corretamente 995 imagens da classe cabeças de um total de 1101, aproximadamente 90% de acertos, e 3613 imagens corretas de um total de 3909 da classe capacete, ou seja, aproximadamente 92% de acertos. Podemos ver que o modelo retreinado foi eficaz em identificar os capacetes, porém teve dificuldade na identificação das pessoas, classe onde o modelo mais errou as detecções.

Na figura 8, temos os exemplos de imagens onde o modelo não foi capaz de identificar as classes corretamente.

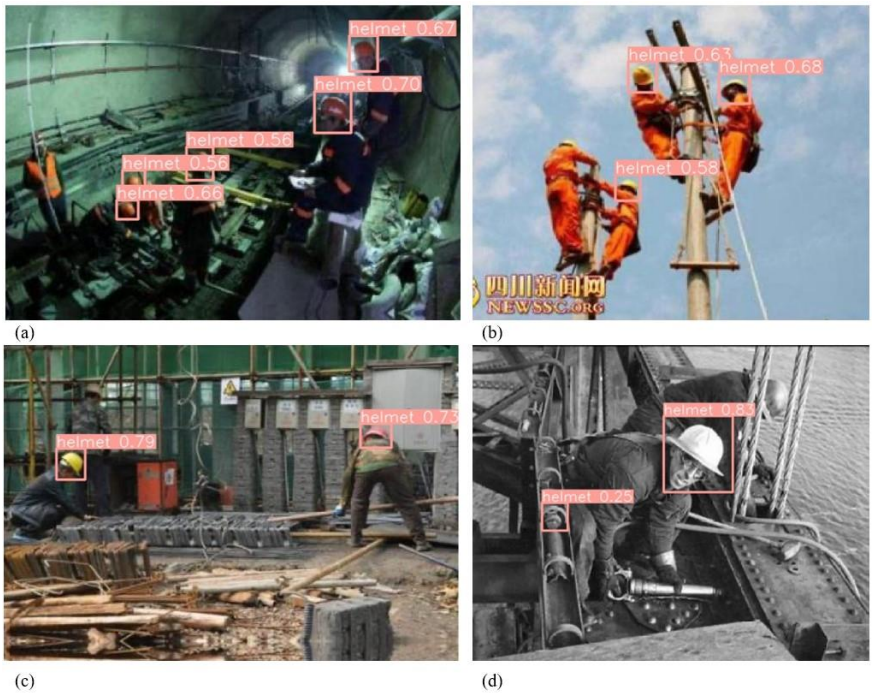

Figura 8 - Imagens não detectadas corretamente pelo algoritmo

Nota-se que nas imagens (a) e (d), o algoritmo deixou de identificar o capcete de um trabalhador, eles possuem objetos bloqueando o seu rosto, que dificultam sua identificação, na imagem (a) o operário segura uma barra de ferro na sua frente, e na imagem (d), o operário está atrás de cordas de aço. Já os outros capcetes na imagens (a) e (d) onde os capacetes se encontram sem nenhum objeto a sua frente, foram corretamente detectados. Já na figura (b), um único operador não teve o capacete detectado, devido a estar com a cabeça inclinada, deixando apenas uma pequena parte dele a mostra, que não foi suficiente para ser detectada, e na imagem (c), por mais que o capcete esteja a mostra, o mesmo se encontra numa parte mais escura da foto e como o operador está mais distante que os outros, o capacete também é menor que os outros detectados.

Para que o modelo, agora retreinado, possa realizar inferências com imagens capturadas em tempo real, utilizou-se uma câmera usb, para filmar continuamente um ambiente, e alimentar a entrada do modelo com imagens em tempo real, e então exibir o resultado da imagem em um servidor web para que seja possivel visualizar a inferência, na figura 9 temos o fluxograma que demonstra o funcionamento da solução.

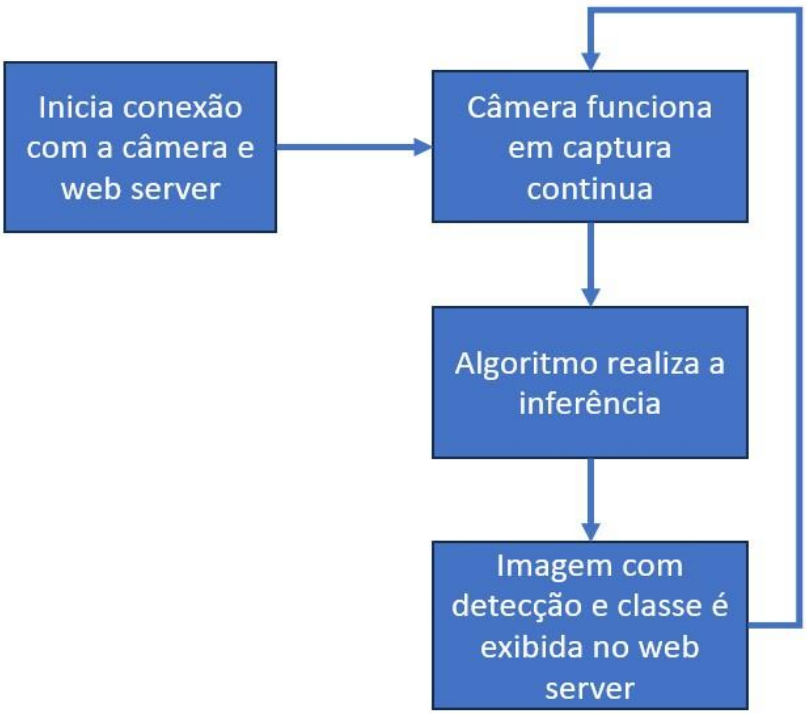

Figura 9 - Fluxograma de funcionamento do algoritmo

Ao executar o sistema é possível monitorar em tempo real a inferência obtida, na figura 10 é possível visualizar tanto a detecatação correta de uma pessoa sem utilizar o capacete quanto com a pessoal utilizando o capacete.

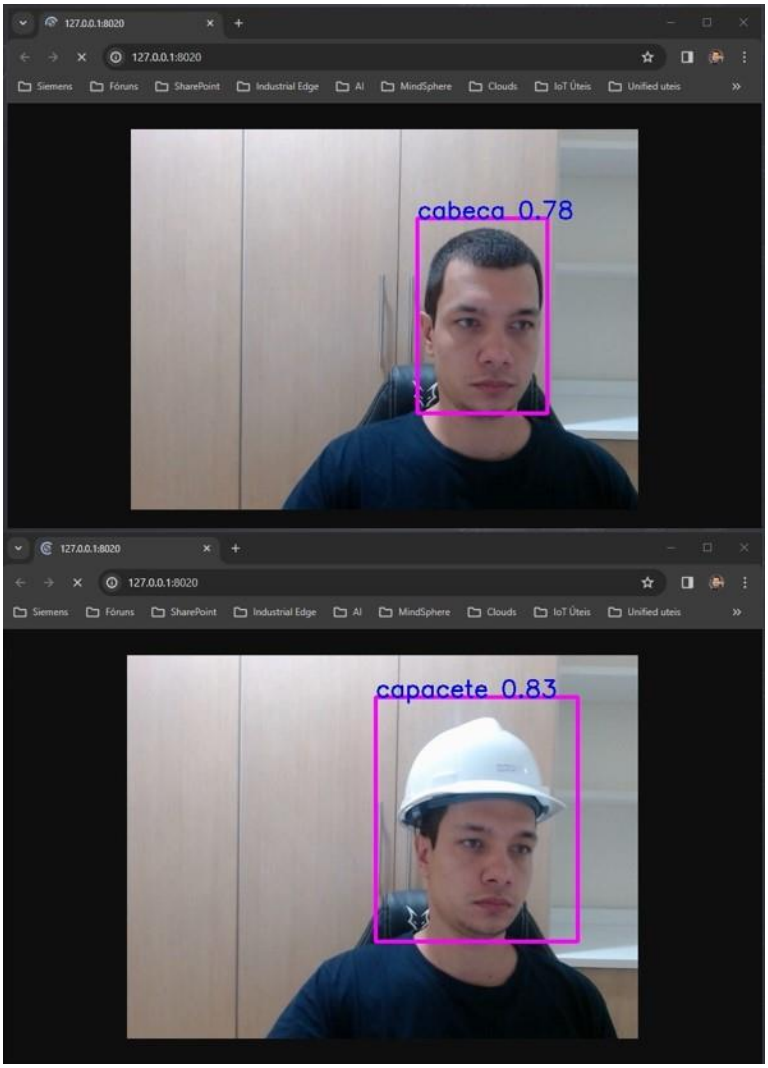

Figura 10 - Inferências realizadas em tempo real pelo algoritmo e exibidas no *webserver*

A detecção se mostrou com uma acurácia bem alta mesmo num ambiente fechado sem iluminação natural e com um fundo diferente dos normalmente encontrados em ambientes industriais como eram os das imagens de treinamento, mostrnado que de fato a solução é robusta para ser utilizada em diferentes ambientes e ainda sim ser bem-sucedida.

4. CONCLUSÃO

Um grande desafio foi o tempo de treinamento do modelo, já que a YOLO utiliza imagens como entradas levou-se algumas horas para a sua conclusão, vendo a queda da precisão do modelo após cerca de 50 épocas, poderia se treinar menos o modelo e economizar tempo de treinamento mantendo a mesma precisão.

O presente trabalho poderia ser ampliado para a detecção de mais EPIs como luvas, óculos de proteção, colete de identificação, protetores auriculares, tornando-se uma solução completa de identificação dos principais EPIs utilizados na indústria e se eles de fato estão vestidos pelos trabalhadores, e não apenas os segurando em suas mãos, seria necessário para isso outras bases de dados semelhantes e inclusive a mesma base também poderia ser utilizada caso esses outros EPIs fossem identificados nela, já que essa base possui um grande número de imagens de trabalhadores na indústria.

Por mais que o modelo tenha tido algumas dificuldades para detectar as imagens, nas quais haviam objetos a frente dos capacetes, o modelo se mostrou eficaz para realizar as inferências em cenários onde os capacetes não tem qualquer coisa a sua frente, havendo uma distribuição de câmeras de modo que seja possível coletar imagens diversas de um mesmo ponto pra evitar que o capacete seja bloqueado, o modelo pode ser utilizado para detectar o uso correto do capecete pelos operados durante a execução do seu trabalho.

5. REFERÊNCIAS

1. **Plataforma SmartLab de Trabalho Decente**. Disponível em: <https://smartlabbr.org/sst/localidade/0?dimensao=perfilCasosAcidentes>. Acesso em: 10 de jan. de 2024.

2. **NR 04 -SERVIÇOS ESPECIALIZADOS EM SEGURANÇA E EM MEDICINA DO TRABALHO**. [s.l: s.n.]. Disponível em: <https://www.gov.br/trabalho-e-emprego/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/comissao-tripartite-partitaria-permanente/arquivos/normas-regulamentadoras/nr-04-atualizada-2022-2-1.pdf>. Acesso em: 10 de jan. de 2024.

3. SILVA, Gustavo Henrique Souto da. **Um sistema de visão computacional para o monitoramento de parâmetros respiratórios de pacientes com esclerose lateral amiotrófica em ambiente hospitalar**. 2012. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte.

4. RIOS, Marcel; NETTO, José Francisco. Uma Abordagem Utilizando Visão Computacional para Monitoramento de Robôs Móveis em Ambientes de Tarefas na Robótica Educacional. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2016. p. 480.

5. OLIVEIRA, Bruno Alberto Soares; DE FARIA NETO, Abilio Pereira; FERNANDINO, Roberto Márcio Arruda; COSTA, Diego de Proença; GUIMARÃES, Frederico Gadelha. Deep Learning para Detecção de Componentes em Alimentadores de Subestações. **Simpósio Brasileiro de Telecomunicaçoes e Processamento de Sinais-SBrT**, v. 1, 2020.

6. DE SOUZA OLIVEIRA, Claudia Almerinda; BRITO, Soo Man Gimenes; PIRES, Ricardo. APLICAÇÃO DE UMA REDE NEURAL ARTIFICIAL YOLO PARA CONTROLE DE ESTOQUE POR IMAGENS EM UMA REVENDA DE AUTOPEÇAS. **REGRASP-Revista para Graduandos/IFSP-Câmpus São Paulo**, v. 8, n. 2, p. 21-43, 2023.

7. VIOLA, Paul; JONES, Michael. Rapid object detection using a boosted cascade of simple features. In: **Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001**. Ieee, 2001. p. I-I.

8. JAVED MEHEDI SHAMRAT, F.M.; MAJUMDER, Anup; ANTU, Probal Roy; BARMON, Saykot Kumar; NOWRIN, Itisha; RANJAN, Rumesh. Human face recognition applying haar cascade classifier. In: **Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021**. Springer Singapore, 2022. p. 143-157.

9. DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. In: **2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)**. Ieee, 2005. p. 886-893.

10. EVERINGHAM, Mark; GOOL, Luc Van; WILLIAMS, Christopher K. I.; WINN, John; ZISSERMAN, Andrew. The pascal visual object classes (voc) challenge. **International journal of computer vision**, v. 88, p. 303-338, 2010.

11. GIRSHICK, Ross; DONAHUE, Jeff; DARRELL, Trevor; MALIK, Jitendra. Rich feature hierarchies for accurate object detection and semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2014. p. 580-587.

12. HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Spatial pyramid pooling in deep convolutional networks for visual recognition. **IEEE transactions on pattern analysis and machine intelligence**, v. 37, n. 9, p. 1904-1916, 2015.

13. GIRSHICK, Ross. Fast r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. 2015. p. 1440-1448.

14. REN, Shaoqing; HE, Kaiming; GIRSHICK, Ross; SUN, Jian. Faster r-cnn: Towards real-time object detection with region proposal networks. **Advances in neural information processing systems**, v. 28, 2015.

15. REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 779-788.

16. OQUAB, Maxime; DARCET, Timothée; MOUTAKANNI, Théo; VO, Huy V.; SZAFRANIEC, Marc; KHALIDOV, Vasil; FERNANDEZ, Pierre; HAZIZA, Daniel; MASSA, Francisco; EL-NOUBY, Alaaeldin; ASSRAN, Mahmoud; BALLAS, Nicolas; GALUBA, Wojciech; HOWES, Russell; HUANG, Po-Yao; LI, Shang-Wen; MISRA, Ishan; RABBAT, Michael; SHARMA, Vasu; SYNNAEVE, Gabriel; XU, Hu; JEGOU, Hervé; MAIRAL, Julien; LABATUT, Patrick; JOULIN, Armand; BOJANOWSKI, Piotr. Dinov2: Learning robust visual features without supervision. **arXiv preprint arXiv:2304.07193**, 2023.

17. O'SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. **arXiv preprint arXiv:1511.08458**, 2015.

18. MARANHÃO, A. Safety Helmet Detection, 2020. Disponível em: <https://www.kaggle.com/datasets/andrewmvd/hard-hat-detection>. Acesso em novembro de 2023.

19. BOESCH, G. **A Guide to YOLOv8 in 2024**. Disponível em: <https://viso.ai/deep-learning/yolov8-guide/>. Acesso em: 15 de jan. de 2024.

6. APÊNDICE A – CÓDIGO PUBLICADO NO GITHUB

A seguir está link para acesso a página do repositório no github onde foi publicado o código do sistema apresentado neste trabalho: <https://github.com/mschaberle/tcc_helmet_detection>. Código publicado em 21 de março de 2024.