FACULDADE UNIGUAÇU

IGUAZU SCIENCE

EDITORA UNIVERSITÁRIA UNIGUAÇU

v. 3, n. 7, jun. 2025
ISBN 978-65-83057-12-9

ARTIGO CIENTÍFICO ACESSO LIVRE

SISTEMA DE DETECÇÃO E CONTAGEM DE ALUNOS EM SALA DE AULA UTILIZANDO INTELIGÊNCIA ARTIFICIAL E PROCESSAMENTO DE VÍDEO

Mariana Molossi*; Alexssandro Ferreira Cordeiro; Guilherme Henrique Silvestri*; Bruno Luiz Schuster Rech****

* Acadêmico de Engenharia de Software, *marianamolossi@icloud.com; guilhermehsilvest@gmail.com*
**Mestre Tecnologias Computacionais para o Agronegócio – UTFPR Medianeira, *alexssandrofc@gmail.com*;
**Mestre Ciência da Computação – UNIOESTE, *brunolsrech@gmail.com*.

INFORMAÇÕES

Histórico de submissão:

Recebido em: 17 out. 2025
Aceite: 05 jun. 2025
Publicação *online*: jun. 2025

RESUMO

Este trabalho propõe o desenvolvimento de um sistema automatizado para detecção e contagem de alunos em salas de aula, utilizando técnicas de visão computacional e aprendizado profundo. A pesquisa aborda o problema da ineficiência e suscetibilidade a erros nos métodos tradicionais de controle de frequência, propondo uma solução baseada na arquitetura YOLOv11 (You Only Look Once, versão 11). O sistema utiliza câmeras posicionadas estrategicamente para capturar imagens em tempo real, que são processadas por algoritmos de detecção de objetos para identificar e contar o número de estudantes presentes. A metodologia emprega a biblioteca OpenCV para captura e pré-processamento das imagens, enquanto o modelo YOLOv11, treinado com o COCO Dataset, realiza a detecção. Os resultados demonstram que o sistema é capaz de operar com alta precisão mesmo em condições desafiadoras, como variações de iluminação e oclusões parciais. Conclui-se que a aplicação proposta oferece uma solução eficaz e não invasiva para o monitoramento de presença em ambientes educacionais, contribuindo para a melhoria da gestão acadêmica e segurança institucional.

Palavras-chave: Visão computacional; YOLOv11; detecção de pessoas; contagem automática; educação.

ABSTRACT

This study proposes the development of an automated system for detecting and counting students in classrooms, employing computer vision and deep learning techniques. The research addresses the inefficiencies and susceptibility to errors in traditional attendance control methods by introducing a solution based on the YOLOv11 (You Only Look Once, version 11) architecture. Strategically positioned cameras capture real-time images, which are processed by object detection algorithms to identify and count the number of students present. The methodology utilizes the OpenCV library for image capture and preprocessing, while the YOLOv11 model, trained with the COCO Dataset, performs the detection. Results demonstrate that the system operates with high accuracy even under challenging conditions, such as lighting variations and partial occlusions. It is concluded that the proposed application offers an effective and non-invasive solution for attendance monitoring in educational environments, contributing to improved academic management and institutional security.

Keywords: computer vision; YOLOv11; people detection; automatic counting; education.

Citação: MOLOSSI, Mariana; CORDEIRO, Alexssandro Ferreira; SILVESTRI, Guilherme Henrique; RECH, Bruno Luiz Schuster. Sistema de detecção e contagem de alunos em sala de aula utilizando inteligência artificial e processamento de vídeo. **Iguazu Science**, São Miguel do Iguaçu, v. 3, n. 7, p. 107-115, jun. 2025.

INTRODUÇÃO

A evolução tecnológica nas últimas décadas tem promovido transformações significativas em diversos

setores da sociedade, incluindo o setor educacional (Singh *et al*., 2021). Ferramentas baseadas em inteligência artificial, aprendizado de máquina e visão computacional vêm sendo amplamente exploradas para resolver problemas antes tratados de forma manual ou ineficiente (Russell *et al*., 2015). Dentre essas soluções, destaca-se a aplicação de tecnologias para detecção e contagem automática de pessoas, uma abordagem que vem ganhando cada vez mais relevância em ambientes internos, especialmente nas instituições de ensino (Kumar *et al.,* 2020).

A presença de alunos em sala de aula é um indicador fundamental para diversas análises pedagógicas, administrativas e até mesmo de segurança. Em contextos tradicionais, o controle de frequência é realizado manualmente, seja por meio de listas de chamada ou sistemas eletrônicos que exigem interação humana direta. No entanto, tais métodos estão sujeitos a falhas, como marcações indevidas, fraudes, omissões ou simplesmente perda de tempo em sala. Além disso, eles não oferecem uma visão contínua ou em tempo real da ocupação dos espaços educacionais (Faria *et al*., 2019).

Nesse sentido, a possibilidade de utilizar sistemas automáticos de contagem de pessoas por vídeo apresenta-se como uma alternativa moderna e eficaz. Com o uso de câmeras comuns e algoritmos de detecção de objetos, é possível monitorar a movimentação e a permanência de indivíduos em ambientes internos, promovendo ganhos em termos de eficiência operacional, segurança institucional e apoio à tomada de decisões. Por exemplo, a identificação de uma queda súbita na frequência pode sinalizar problemas como evasão escolar, ao passo que o excesso de alunos pode representar riscos relacionados à superlotação ou à inadequação do espaço físico (Medeiros *et al*., 2020).

Contudo, esse tipo de abordagem traz consigo desafios técnicos importantes. A qualidade e o posicionamento das câmeras, a iluminação do ambiente, a movimentação dos alunos e a possível obstrução entre corpos são fatores que influenciam diretamente a eficácia do sistema. Além disso, a necessidade de equilibrar precisão com desempenho computacional é uma preocupação recorrente, especialmente em aplicações que exigem análises em tempo real ou com recursos limitados (Zhang *et al*., 2021).

Diante desse cenário, o presente trabalho propõe o desenvolvimento de uma aplicação baseada em redes neurais convolucionais, utilizando a arquitetura YOLO (You Only Look Once), com o objetivo de detectar e contar automaticamente o número de alunos em uma sala de aula a partir da análise de imagens de vídeo. O sistema é projetado para operar de maneira otimizada, realizando o processamento em intervalos específicos, o que permite reduzir a carga computacional sem comprometer a acurácia dos resultados (Redmon *et al.*, 2016).

Com isso, espera-se contribuir para o avanço de soluções inteligentes no ambiente educacional, promovendo maior controle, segurança e eficiência por meio da integração entre tecnologia e gestão acadêmica.

FUNDAMENTAÇÃO TEÓRICA

2.1 Visão Computacional e Análise de Imagens

A visão computacional é um campo da inteligência artificial que visa permitir que computadores “enxerguem” e interpretem imagens ou vídeos de maneira similar à percepção humana. Essa área envolve tarefas como detecção de objetos, reconhecimento de padrões, rastreamento de movimento e segmentação de imagens. Com o avanço de técnicas de aprendizado de máquina, a análise de imagens ganhou um papel central em diversas aplicações, como vigilância, automação industrial e monitoramento de ambientes (Gonzalez; Woods, 2018).

No contexto educacional, a visão computacional possibilita a automação de processos como controle de presença, análise de comportamento e segurança. A detecção de pessoas em tempo real pode ser utilizada não apenas para fins administrativos, mas também para suporte à segurança institucional e planejamento de recursos.

2.2 Redes Neurais Convolucionais (CNN)

As Redes Neurais Convolucionais (CNNs – Convolutional Neural Networks) constituem um tipo de arquitetura de rede neural artificial projetada especialmente para o processamento de dados com estrutura em grade, como imagens bidimensionais. Diferentemente das redes neurais tradicionais, as CNNs são compostas por camadas convolucionais que aplicam filtros sobre regiões locais da imagem, extraindo características relevantes de forma automática e hierárquica. Essa abordagem permite capturar desde padrões simples, como bordas e contornos, até estruturas mais complexas, como formas e objetos (LeCun *et al*., 2015).

As CNNs são amplamente empregadas em tarefas de visão computacional, especialmente em problemas de detecção, reconhecimento e classificação de objetos. Essas redes apresentam alto desempenho mesmo em cenários com grande variabilidade de formas, tamanhos e posições, o que as torna ideais para aplicações em áreas como segurança pública, medicina diagnóstica, monitoramento urbano e, mais recentemente, em contextos educacionais (Goodfellow; Bengio; Courville, 2016).

No contexto escolar, as CNNs vêm sendo utilizadas para implementar sistemas de contagem automática de estudantes em sala de aula. Essa aplicação tem por objetivo otimizar o processo de controle de presença,

oferecer dados estatísticos em tempo real e apoiar a gestão educacional por meio da automação. O funcionamento geral do sistema envolve a captura de imagens ou vídeos por câmeras instaladas estrategicamente, seguida do pré-processamento dessas imagens e da detecção de pessoas com o uso de modelos baseados em CNNs, como as arquiteturas da família YOLO (You Only Look Once).

Esses modelos realizam a detecção de indivíduos ao identificar suas localizações por meio de caixas delimitadoras (bounding boxes). A partir das detecções, o sistema realiza a contagem dos estudantes presentes, gerando relatórios e alertas de forma automática. Esse tipo de solução contribui significativamente para a melhoria da eficiência operacional das instituições de ensino, além de possibilitar a análise de dados relacionados à frequência e à ocupação das salas de aula (Redmon *et al.*, 2016).

2.3 Detecção de Objetos com YOLO

A arquitetura YOLO (You Only Look Once), desenvolvida por Redmon *et al.* (2016), revolucionou a detecção de objetos ao permitir que a identificação e a classificação fossem feitas em tempo real, com uma única passagem pela imagem. Diferente de abordagens tradicionais como R-CNN, que exigem múltiplas etapas de processamento, o YOLO divide a imagem em grades e, para cada uma delas, estima diretamente as classes e posições dos objetos.

2.4 COCO Dataset

O Common Objects in Context (COCO) é um dos conjuntos de dados mais amplamente utilizados no desenvolvimento e avaliação de modelos de visão computacional, especialmente em tarefas de detecção de objetos, segmentação semântica e geração de legendas automáticas para imagens. Criado por Lin *et al.* (2014), o COCO possui mais de 330 mil imagens, das quais aproximadamente 200 mil são anotadas, totalizando mais de 2,5 milhões de instâncias de objetos rotuladas, distribuídas em 80 categorias distintas, como pessoas, veículos, utensílios domésticos, animais, entre outros.

A principal característica que diferencia o COCO de outros datasets é a complexidade contextual das imagens: os objetos aparecem em posições variadas, com oclusões, em diferentes escalas e em contextos reais e desafiadores. Essa diversidade torna o COCO um recurso essencial para o treinamento de modelos robustos, capazes de generalizar bem para situações do mundo real, onde os objetos não se apresentam em cenários limpos e isolados, como em laboratórios ou conjuntos de dados artificiais.

Modelos de detecção de objetos de alto desempenho, como os da família YOLO (You Only Look Once), são frequentemente treinados e avaliados com base no COCO, o que contribui significativamente para o aumento da precisão e da capacidade de generalização desses modelos. A comparação padronizada dos resultados obtidos no COCO também permite que pesquisadores avaliem o progresso da área de forma objetiva, utilizando métricas como mAP (mean Average Precision).

Assim, o COCO consolidou-se como uma referência fundamental na área de visão computacional, fornecendo uma base sólida para avanços em aplicações práticas que vão desde a segurança pública até a contagem de pessoas em ambientes educacionais e comerciais.

2.5 Aplicações da Contagem de Pessoas

A contagem automática de pessoas é um tema de pesquisa relevante em áreas como transporte público, eventos, varejo e segurança. A implementação em instituições de ensino ainda está em estágio inicial, mas apresenta grande potencial. Segundo Chen *et al.* (2020), sistemas automáticos de contagem podem auxiliar na gestão de espaços, no controle de acesso e no monitoramento de padrões de presença e movimentação.

Em ambientes educacionais, a contagem de alunos pode otimizar o uso de salas, identificar evasão escolar, detectar anomalias como superlotação e ainda contribuir para a resposta rápida em situações emergenciais. Com o suporte de tecnologias como visão computacional e deep learning, essas análises podem ser feitas de forma não invasiva, com alta precisão e baixo custo operacional.

2.6 Limitações e Desafios

Embora promissoras, as tecnologias de detecção e contagem de pessoas ainda enfrentam alguns desafios. A oclusão (quando uma pessoa bloqueia parcialmente outra na imagem), a iluminação inadequada, a baixa resolução das câmeras e os diferentes ângulos de captação são fatores que podem afetar negativamente os resultados. Além disso, o uso contínuo de algoritmos de detecção em tempo real demanda hardware com boa capacidade de processamento, o que pode limitar a aplicação em ambientes com recursos computacionais reduzidos (Zhang *et al.*, 2019).

Por isso, estratégias como a aplicação periódica do algoritmo em quadros selecionados (frame sampling) e o uso de técnicas de pré-processamento e pós-processamento tornam-se essenciais para equilibrar acurácia e desempenho.

METODOLOGIA

Este estudo foi conduzido por meio de uma revisão bibliográfica sistemática, com o objetivo de mapear e analisar o estado da arte dos telhados verdes, destacando as tendências e perspectivas futuras dessa

tecnologia. A pesquisa será baseada em artigos científicos, dissertações, teses e trabalhos acadêmicos, com foco naqueles publicados nos últimos cinco anos. As principais fontes de pesquisa incluiram bases de dados como ScienceDirect, Google Acadêmico, além de sites especializados em construção sustentável e telhados verdes.

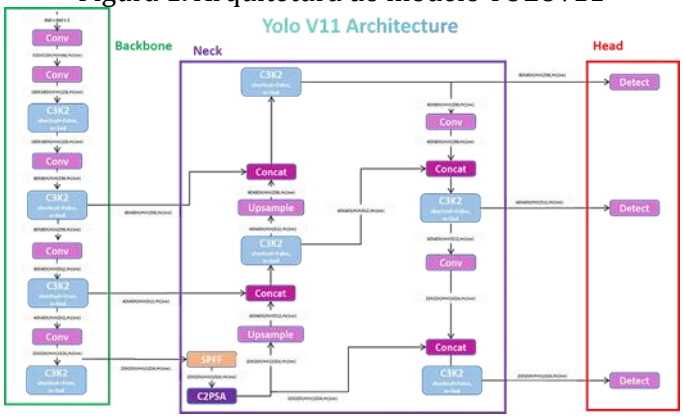
Foi utilizadas palavras-chave específicas para orientar a busca, tais como "telhado verde", "construção sustentável", "cobertura vegetada", "infraestrutura verde", "arquitetura sustentável", "mitigação urbana", "ecossistemas urbanos", e "soluções baseadas na natureza". As palavras-chave foram combinadas de diferentes maneiras para garantir uma abrangência adequada e a inclusão de estudos relevantes.

Os critérios de inclusão englobram publicações que discutam tanto aspectos técnicos quanto ambientais, além de abordagens inovadoras e tendências emergentes relacionadas aos telhados verdes. Trabalhos que ofereçam uma visão crítica sobre os desafios e as oportunidades desse sistema também serão priorizados.

Para a contagem automatizada de alunos em sala de aula, será utilizada uma câmera posicionada estrategicamente para captar toda a área ocupada pelos estudantes. Essa câmera estará conectada a um sistema computacional que fará o processamento das imagens em tempo real, empregando técnicas de visão computacional e aprendizado profundo.

O modelo central adotado para a detecção será o YOLOv11 (You Only Look Once, versão 11), uma das arquiteturas mais avançadas e eficientes para reconhecimento de objetos em vídeo. O YOLOv11 destaca-se pela capacidade de identificar pessoas mesmo em situações adversas, como oclusões parciais, variações de iluminação e movimentação, características comuns no ambiente de sala de aula. Sua arquitetura otimizada proporciona alta precisão e velocidade, aliadas a um consumo reduzido de recursos computacionais.

Figura 1. Arquitetura do modelo YOLOv11



Fonte: Adaptado de Rao (2023).

Arquitetura da Rede YOLOv11

A arquitetura da rede YOLOv11 é composta por três módulos principais: Backbone, Neck e Head, cada um com funções específicas no processo de detecção de objetos em imagens.

Backbone

O módulo Backbone, representado na cor verde, é responsável pela extração de características da imagem de entrada, cujo tamanho é de 640×640×3 (altura, largura e canais RGB). Essa etapa inicial é composta por diversas camadas convolucionais (Conv) e blocos C3K2, que representam uma variação dos blocos C3 com kernel 3×3 e duas repetições internas. Esses blocos são projetados para capturar padrões espaciais relevantes, mantendo a eficiência computacional da rede.

À medida que os dados percorrem as camadas do Backbone, ocorre uma redução progressiva nas dimensões espaciais da imagem e um aumento no número de canais (por exemplo: de 640×640×3 para 320×320×32, até atingir 20×20×512). Esse processo resulta na construção de representações progressivamente mais abstratas da imagem original.

Neck

O módulo Neck, destacado em roxo, tem como função refinar e combinar diferentes níveis de mapas de características extraídos pelo Backbone. O objetivo é aprimorar a capacidade da rede de detectar objetos em múltiplas escalas. Esse módulo incorpora os seguintes componentes:

- SPPF (Spatial Pyramid Pooling - Fast): responsável por realizar a fusão de informações espaciais em diferentes escalas;
- C2PSA: possivelmente uma variação de mecanismo de atenção, que visa enfatizar regiões mais relevantes da imagem;
- Upsample: camadas que aumentam a resolução espacial dos mapas de características, possibilitando a integração com informações de resoluções superiores;
- Concat: operação de concatenação de mapas de características oriundos de diferentes estágios da rede, promovendo o reaproveitamento de informações.

Head

O módulo Head, em vermelho, é encarregado de gerar as predições finais da rede, incluindo as caixas delimitadoras, as classes dos objetos detectados e os respectivos escores de confiança. Para isso, são utilizadas três camadas denominadas Detect, que atuam em diferentes resoluções: 80×80, 40×40 e 20×20. Essa abordagem multiescalar permite a detecção eficiente de objetos de diversos tamanhos, sendo as resoluções mais altas mais sensíveis a objetos pequenos, enquanto as menores são adequadas para objetos médios e grandes.

Resumo do Fluxo Operacional

A imagem de entrada é primeiramente processada pelo Backbone, onde são extraídas suas características por meio de convoluções e blocos especializados. As saídas intermediárias são então encaminhadas ao Neck, onde são refinadas, combinadas e redimensionadas. Por fim, o Head utiliza essas informações para realizar a detecção final de objetos em múltiplas escalas, fornecendo resultados precisos em diferentes níveis de resolução.

Além do YOLOv11, será utilizada a biblioteca OpenCV (Open Source Computer Vision Library), uma ferramenta de código aberto amplamente utilizada para o processamento de imagens e vídeos. O OpenCV será responsável por capturar os quadros de vídeo gerados pela câmera, realizar o pré-processamento dessas imagens (como redimensionamento e ajuste de brilho) e integrá-las ao modelo de detecção.

O sistema será desenvolvido utilizando a linguagem de programação Python, devido à sua robustez e ampla compatibilidade com bibliotecas de inteligência artificial e visão computacional. O fluxo de funcionamento será composto pelas seguintes etapas:

- Captura de vídeo: A câmera transmite os quadros em tempo real para o sistema computacional.
- Pré-processamento: As imagens são tratadas e preparadas para análise, garantindo qualidade e padronização.
- Detecção com YOLOv11: O modelo analisa cada quadro selecionado e identifica quantas pessoas estão presentes.
- Contagem e registro: O número de alunos detectado é armazenado e pode ser visualizado em tempo real ou exportado para relatórios.

Uso do COCO Dataset

Para treinar e validar o modelo YOLOv11, será utilizado o COCO Dataset (Common Objects in Context), um dos conjuntos de dados mais amplamente reconhecidos na área de detecção de objetos. O COCO Dataset contém mais de 330.000 imagens e 2,5 milhões de anotações de objetos distribuídos em 80 categorias, sendo a categoria “Pessoa” o foco principal deste estudo.

A divisão do dataset será feita da seguinte forma:

- Conjunto de treinamento: Utilizado para que o modelo aprenda a detectar pessoas nas imagens.
- Conjunto de validação: Empregado para avaliar o desempenho do modelo durante o treinamento e prevenir overfitting.
- Conjunto de teste: Usado para verificar a acurácia do modelo em cenários não vistos anteriormente.

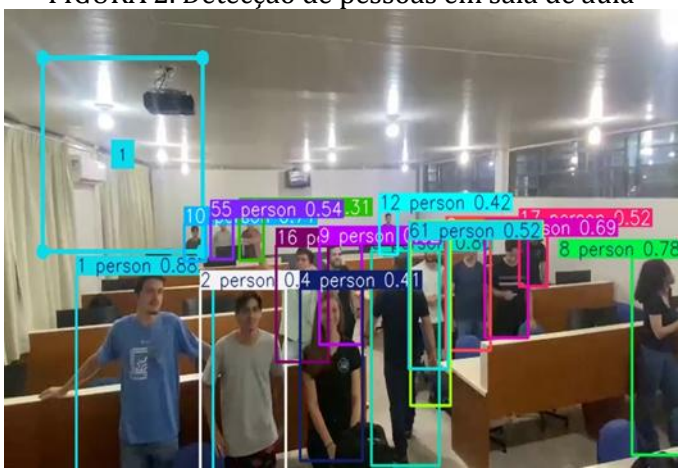
Esse método permite preservar a privacidade dos alunos, uma vez que o sistema não realiza reconhecimento facial ou identificação individual. O foco recai apenas na contagem total de pessoas no ambiente, garantindo a conformidade com princípios éticos e legais. Como resultado, a aplicação poderá ser utilizada para gerar relatórios de frequência automatizados, auxiliar professores e coordenadores pedagógicos, além de detectar situações anômalas, como superlotação ou evasão, com maior rapidez e precisão.

RESULTADOS E DISCUSSÃO

Os testes realizados com o modelo YOLOv11 demonstraram a capacidade do sistema em detectar e contar corretamente o número de pessoas em diferentes ambientes internos, incluindo salas de aula e corredores da instituição. As imagens analisadas foram capturadas em diferentes momentos e locais, evidenciando a eficácia do modelo mesmo diante de desafios como iluminação variada e oclusões parciais.

A Figura 2 mostra o ambiente de uma sala de aula com múltiplos alunos posicionados em pé, alguns parcialmente ocluídos por outros. Mesmo assim, o modelo foi capaz de identificar a maioria dos indivíduos com precisão satisfatória, apresentando valores de confiança superiores a 0.50 em grande parte das detecções.

FIGURA 2. Detecção de pessoas em sala de aula



Fonte da Figura: Elaborado pelo autor (2025).

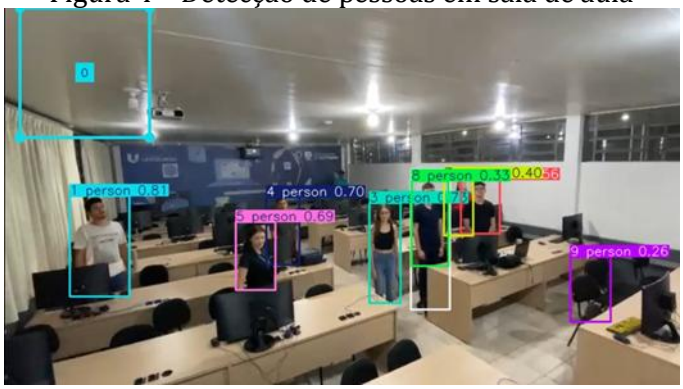
Na Figura 3, a detecção foi realizada em um corredor da instituição, com estudantes caminhando em direção à câmera. Apesar da movimentação, o modelo YOLOv11 manteve uma boa taxa de acerto, detectando corretamente todos os indivíduos com altos níveis de confiança.

A Figura 4 exibe outro cenário de sala de aula, desta vez com menos estudantes. Nesse caso, novamente é possível verificar a consistência do modelo, que detectou corretamente os indivíduos presentes, mesmo em diferentes posições dentro da sala e com iluminação artificial.

Figura 3 – Detecção de pessoas em corredor da instituição

Fonte: Elaborado pelo autor (2025).

Figura 4 – Detecção de pessoas em sala de aula



Fonte: Elaborado pelo autor (2025).

Esses resultados confirmam a viabilidade do uso do YOLOv11 para monitoramento e contagem de alunos em tempo real, mesmo em ambientes com variações de iluminação, ângulos de câmera e movimentação. Como limitação observada, destaca-se a sensibilidade do modelo a sobreposições extremas entre pessoas, o que pode afetar a precisão da contagem.

VISÕES FUTURAS

A evolução contínua dos modelos de detecção de objetos, especialmente a linha YOLO (You Only Look Once), abre um leque de possibilidades promissoras para aplicações em ambientes educacionais e outros domínios. As visões futuras apontam para um cenário onde a inteligência artificial se tornará ainda mais acessível, precisa e integrada ao cotidiano. Segundo Redmon *et al.* (2016), a proposta original do YOLO já visava conciliar precisão e velocidade, e os avanços atuais reforçam esse compromisso com aplicações em tempo real.

Uma das principais tendências é o refinamento da arquitetura YOLO por meio de técnicas modernas de aprendizado profundo, como o uso de modelos híbridos, aprendizado auto-supervisionado e redes neurais mais leves e eficientes. Isso permitirá que modelos como o YOLOv11 sejam executados com alto desempenho mesmo em dispositivos com menor capacidade computacional, como câmeras inteligentes, tablets e microcontroladores. De acordo com Liu *et al.* (2023), a miniaturização de modelos de detecção com alta acurácia é uma linha de pesquisa crescente para aplicações em dispositivos embarcados.

Além disso, espera-se que o aprendizado contínuo (continual learning) seja incorporado aos modelos, permitindo que sistemas de contagem de pessoas em salas de aula se adaptem automaticamente a novas configurações, padrões de iluminação, movimentações ou mudanças estruturais sem necessidade de re-treinamento completo. Isso tornaria a aplicação mais flexível e eficaz em longo prazo. Parisi *et al*. (2019) destacam que o aprendizado contínuo permite que modelos mantenham conhecimento prévio enquanto assimilam novas informações, algo essencial para ambientes dinâmicos.

Outro ponto de destaque é o avanço rumo a sistemas de visão computacional multimodal, capazes de integrar dados visuais com outras fontes, como áudio e texto. Esses sistemas poderiam, por exemplo, não apenas detectar a presença dos alunos, mas também analisar engajamento, expressões faciais e padrões de comportamento, sempre respeitando os limites éticos e legais de privacidade. Segundo Baltrušaitis *et al*. (2019), a multimodalidade é fundamental para uma interpretação mais rica e contextual dos dados no ambiente computacional.

Com a expansão da computação em borda (edge computing), vislumbra-se a possibilidade de processamento local e em tempo real, sem depender de servidores externos, o que aumenta a confiabilidade do sistema e reduz custos operacionais. Isso é particularmente relevante em instituições de ensino públicas ou em regiões com infraestrutura tecnológica limitada. Shi *et al*. (2016) apontam que o edge computing permite redução de latência e maior segurança de dados, sendo ideal para sistemas críticos.

Por fim, destaca-se a importância de desenvolver mecanismos de preservação da privacidade e transparência no uso de dados, como anonimização automática de rostos e registro claro dos propósitos da coleta. Esses recursos serão fundamentais para a aceitação social e institucional das tecnologias

baseadas em visão computacional. De acordo com Dufaux (2020), a privacidade em sistemas de visão é um dos principais desafios contemporâneos, e exige soluções técnicas e regulamentares simultâneas.

Dessa forma, as visões futuras indicam um cenário de maior integração entre inteligência artificial e educação, no qual modelos como o YOLO não apenas contarão alunos, mas apoiarão o planejamento escolar, a segurança e a personalização do ensino, sempre com foco na ética e na eficiência. Como observam Luckin *et al*. (2016), a IA tem potencial para transformar a educação, desde que usada de forma responsável e centrada no ser humano.

CONCLUSÕES

Os testes realizados com o modelo YOLOv11 demonstraram a capacidade robusta do sistema em detectar e contar com precisão o número de pessoas em ambientes internos variados, como salas de aula e corredores. As imagens analisadas, capturadas em momentos e locais distintos, evidenciaram que o modelo mantém um bom desempenho mesmo diante de desafios típicos, como variações na iluminação e oclusões parciais. Isso indica um avanço significativo na aplicação de tecnologias de visão computacional em ambientes dinâmicos e complexos, como o monitoramento em tempo real de alunos em uma instituição de ensino.

Os resultados mostraram que o YOLOv11 foi eficaz na detecção, com a maior parte das detecções apresentando níveis de confiança superiores a 0.50, o que reforça a precisão do modelo em condições variadas de iluminação e posicionamento das pessoas. A aplicação de detecção em corredores e salas de aula com diferentes níveis de movimentação também demonstrou a flexibilidade do sistema para lidar com cenários dinâmicos. No entanto, a limitação identificada, especialmente em situações de sobreposição extrema entre indivíduos, destacou uma área de melhoria para aumentar a precisão da contagem em cenários com alta densidade de pessoas.

Esses achados têm implicações práticas significativas para o desenvolvimento de sistemas de monitoramento automatizado em ambientes educacionais, oferecendo uma solução eficaz para contar e monitorar a presença de alunos em tempo real. Do ponto de vista teórico, os resultados reforçam a eficácia de modelos de detecção avançados, como o YOLOv11, para aplicações em ambientes internos dinâmicos e sujeitos a condições desafiadoras de visualização.

Apesar dos avanços, o estudo apresenta algumas limitações, como a sensibilidade a sobreposições extremas entre indivíduos, que pode comprometer a precisão da contagem em casos específicos. A aplicação do modelo em diferentes tipos de ambientes e condições de câmera pode proporcionar novos insights sobre sua escalabilidade e robustez. Para futuras pesquisas, recomenda-se a exploração de técnicas complementares, como o uso de redes neurais para o rastreamento de pessoas, a fim de reduzir as falhas em situações de oclusão e sobreposição. Além disso, o aprimoramento na configuração de câmeras e no ajuste de parâmetros do modelo pode aumentar ainda mais a precisão e a eficácia da contagem em ambientes mais desafiadores.

REFERÊNCIAS

AHMAD, M. W. *et al*. **Artificial intelligence techniques for analyzing and predicting student performance**. IEEE Access, v. 8, p. 170201–170215, 2020.

ALZUBAIDI, L. *et al*. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions**. Journal of Big Data**, v. 8, n. 1, 2021. Disponível em: https://doi.org/10.1186/s40537-021-00444-8.

BALTRUŠAITIS, T.; AULI, M.; MORENO, P. Multimodal machine learning: A survey and taxonomy. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 41, n. 2, p. 423–443, 2019.

BOCHKO, A.; KRASNOSHTAN, O. **Automated People Counting System Based on YOLO Object Detection and Tracking.** CEUR Workshop Proceedings, 2022.

CHENG, Y. *et al*. Vision-based classroom occupancy analysis using deep learning. **Sensors**, v. 21, n. 5, 2021. Disponível em: https://doi.org/10.3390/s21051684.

CHO, H.; CHEN, C.; WENG, J. Smart attendance monitoring system using computer vision. **Procedia Computer Science**, v. 141, p. 555–560, 2018.

COCO DATASET. **Coco** – Common Objects in Context. Disponível em: https://cocodataset.org/#home.

DUFAUX, F. **Video privacy protection: Trends and approaches.** In: CHEUNG, S. S. et al. (Org.).

Handbook of visual privacy. Cham: Springer, 2020. p. 15–39.

FANG, Y. *et al.* **Student engagement detection in e-learning environments using facial expression recognition.** Interactive Learning Environments, 2022. Disponível em: https://doi.org/10.1080/10494820.2022.2045202.

FARIA, R. L.; SANTOS, A. C.; OLIVEIRA, T. G. de. Sistema de controle de presença estudantil baseado em visão computacional e reconhecimento facial. **Revista Eletrônica de Iniciação Científica,** v. 12, n. 1, p. 112–121, 2019. Disponível em: https://revistas.unilasalle.edu.br/index.php/ric/article/view/5515.

FERREIRA, R. A. *et al*. Aplicações de visão computacional na educação: uma revisão sistemática. **Anais** do Workshop de Informática na Escola (WIE), v. 26, n. 1, p. 315–324, 2020.

GIRSHICK, R. *et al*. **Rich feature hierarchies for accurate object detection and semantic segmentation**. arXiv, 2015. Disponível em: https://arxiv.org/abs/1506.02640.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning. Cambridge**: MIT Press, 2016.

INGH, Dilbag; KUMAR, Yashpal; KALIA, Arvind; CHAUHAN, Neha. Computer vision and deep learning-based human detection and counting system for classroom monitoring. **Journal of Ambient Intelligence and Humanized Computing**, v. 12, p. 6841–6852, 2021. DOI: https://doi.org/10.1007/s12652-020-02587-2.

KANTIPUDI, M. V. *et al*. **Human detection and counting system using YOLO and Deep SORT**. Materials Today: Proceedings, 2022. Disponível em: https://doi.org/10.1016/j.matpr.2022.03.365.

KHAN, A. *et al*. Deep learning: Applications, challenges, and outlook in smart cities. **Sustainable Cities and Society,** v. 35, p. 612–624, 2017. Disponível em: https://doi.org/10.1016/j.scs.2017.09.018.

KUMAR, A.; VERMA, M.; SINGH, A. K.; RAI, H. M. Real-Time People Counting Using YOLO and DeepSORT in Classroom Environment. **Procedia Computer Science,** v. 171, p. 1571–1578, 2020. DOI: https://doi.org/10.1016/j.procs.2020.04.168.

KUMAR, R.; SHAH, M. Survey on Automated Attendance Monitoring Systems using Computer Vision Techniques. **International Journal of Computer Applications,** v. 179, n. 19, 2018.

LECUN, Yann *et al.* Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. DOI: https://doi.org/10.1038/nature14539.

LI, Y. *et al*. **CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes.** In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

LIN, Tsung-Yi *et al.* **Microsoft COCO: Common Objects in Context. In: European Conference on Computer Vision (ECCV).** Cham: Springer, 2014. p. 740–755. Disponível em: https://doi.org/10.1007/978-3-319-10602-1_48.

LIU, Y. *et al*. Lightweight object detection networks for edge devices: A survey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 34, n. 1, p. 144–163, 2023.

LUCKIN, R. *et al.* **Intelligence unleashed: An argument for AI in education**. London: Pearson Education, 2016.

MARTINS, R. M. *et al*. Inteligência Artificial Aplicada à Educação: Panorama e Tendências. **Revista Brasileira de Informática na Educação**, v. 28, n. 1, p. 25–42, 2020.

MATSUKAWA, T.; OKABE, T.; SUZUKI, E.; SUGIYAMA, M. **Hierarchical Gaussian descriptor for person re-identification**. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

MEDEIROS, J. P.; SILVA, M. F.; ALMEIDA, R. A. de. Sistema de contagem de pessoas com visão computacional em tempo real aplicado ao ambiente educacional. **Anais** do Simpósio Brasileiro de Informática na Educação (SBIE), v. 31, n. 1, p. 147–156, 2020. DOI: https://doi.org/10.5753/cbie.sbie.2020.147.

MDPI. YOLOv5: A Novel Real-Time Object Detection Algorithm. **Technologies**, v. 5, n. 4, 2021. Disponível em: https://www.mdpi.com/2504-4990/5/4/83.

NGUYEN, D. T. *et al.* Real-Time People Counting Using Deep Learning. **Journal of Communications and Networks,** v. 22, n. 4, p. 258–264, 2020.

PARISI, G. I. *et al*. Continual lifelong learning with neural networks: A review. Neural Networks, v. 113, p. 54–71, 2019.

PEREIRA, L. G. *et al*. Visão computacional na educação: Uma proposta de sistema inteligente para contagem de estudantes em sala de aula. **Revista Eletrônica de Iniciação Científica**, v. 17, n. 1, 2021.

RAO, Nikhil. **YOLOv11 Explained: Next-Level Object Detection With Enhanced Speed and Accuracy**. Medium, 2024. Disponível em: https://medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71.

REDMON, J. *et al*. **You Only Look Once: Unified, Real-Time Object Detection.** In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. p. 779–788. Disponível em: https://arxiv.org/abs/1506.02640.

REDMON, J.; FARHADI, A. YOLOv3**: An Incremental Improvement.** arXiv preprint, arXiv:1804.02767, 2018. Disponível em: https://arxiv.org/abs/1804.02767.

RUSSELL, Stuart J.; NORVIG, Peter; DAVIS, Ernest. **Artificial Intelligence: A Modern Approach**. 3. ed. Upper Saddle River: Pearson, 2015.

SALEH, M. A.; MOHAMED, A. A.; HELMY, Y. I. Real-time student attendance system using face recognition and tracking techniques**. Journal of King Saud University - Computer and Information Sciences,** 2021. Disponível em: https://doi.org/10.1016/j.jksuci.2021.01.012.

SHI, W. et al. Edge computing: Vision and challenges. **IEEE Internet of Things Journal**, v. 3, n. 5, p. 637–646, 2016.

TAN, H.; WEI, Z.; ZHANG, X. Vision-based monitoring system for classroom behavior analysis. **Computers & Education**, v. 164, 2021.

ULTRALYTICS. **Citation file for Ultralytics repository. GitHub,** 2024. Disponível em: https://github.com/ultralytics/ultralytics/blob/main/CITATION.cff.

ULTRALYTICS. **Ultralytics YOLOv11** – Citations and Acknowledgements. Ultralytics Docs, 2024. Disponível em: https://docs.ultralytics.com/pt/models/yolo11/#citations-and-acknowledgements.

UNIGUAÇU. Acervo digital **– Detalhes do documento.** UNIGUAÇU, 2025. Disponível em: https://academico.uniguacu.com.br/academico/biblioteca/acervo/detalhes/39574.

ZHANG, D. *et al*. Crowd counting with density map estimation: A review. **Neurocomputing**, v. 472, p. 336–353, 2022. Disponível em: https://doi.org/10.1016/j.neucom.2021.10.056.

ZHANG, Y. *et al.* **Single-Image Crowd Counting via Multi-Column Convolutional Neural Network**. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

ZHANG, Y.; ZHAO, Y.; ZHOU, J.; LI, X. Real-Time People Counting System Based on Deep Learning in Complex Scenes. **Sensors**, v. 21, n. 12, p. 4030, 2021. DOI: https://doi.org/10.3390/s21124030.

ZHENG, Y. *et al.* A YOLO-based deep learning model for student behavior detection in classrooms. **Sensors**, v. 21, n. 11, 2021. Disponível em: https://doi.org/10.3390/s21113927.

ZHAO, Z. Q.; ZHENG, P.; XU, S. T.; WU, X. Object detection with deep learning: A review. IEEE **Transactions on Neural Networks and Learning Systems**, v. 30, n. 11, p. 3212–3232, 2019. Disponível em: https://doi.org/10.1109/TNNLS.2018.2876865.