

Aplicação do algoritmo YOLO para eficiência energética: detecção de pessoas em salas de aula no gerenciamento de ares-condicionados

Application of the YOLO algorithm for energy efficiency: people detection in classrooms for air conditioner management

Aplicación del algoritmo YOLO para la eficiencia energética: detección de personas en aulas para la gestión de aires acondicionados

Maria Luiza de Lucena e Melo Pierini¹
Eduardo André Mossin²

Resumo: O crescente consumo de energia elétrica no Brasil, impulsionado, dentre outros pelo uso intensivo de aparelhos de ar-condicionado, reforça a necessidade de soluções inteligentes para otimizar a climatização em ambientes de alta ocupação, como salas de aula. Nesse contexto, previamente, foi desenvolvido no IFSP, *Campus Sertãozinho*, o sistema *AirCon*, que funciona como controle remoto dos aparelhos de ar-condicionado, monitorando temperatura e umidade. Este trabalho propõe integrar ao *AirCon* uma funcionalidade de acionamento e desligamento automático, fundamentada na detecção de pessoas. Para isso, utilizaram-se técnicas de visão computacional e de inteligência artificial. Com imagens capturadas pelo microcontrolador ESP32-CAM, foi construído um *dataset* personalizado com variações de luminosidade e ocupação, processado por redes neurais convolucionais baseadas no algoritmo YOLO. O modelo YOLO11 foi treinado em diferentes versões, aplicando *transfer learning* e treinamento com dados próprios. Os resultados destacaram a versão *extra-large* pré-treinada como mais robusta, possibilitando acionamento inteligente e promovendo eficiência energética.

Palavras-chave: Detecção de pessoas. ESP32-CAM. Redes Neurais. YOLO. Sustentabilidade energética.

Abstract: The growing electricity consumption in Brazil, driven by the intensive use of air conditioners, reinforces the need for intelligent solutions to optimize climate control in high-occupancy environments, such as classrooms. In this context, the *AirCon* system was previously developed at IFSP, Sertãozinho campus, functioning as a remote controller for air conditioners, while monitoring temperature and humidity. This work proposes to integrate into *AirCon* an automatic on/off functionality, based on people detection. For this purpose, computer vision and artificial intelligence techniques were employed. Using images captured by the ESP32-CAM microcontroller, a customized dataset was constructed with variations in lighting and occupancy, processed by convolutional neural networks based on the YOLO algorithm. The YOLO11 model was trained in different versions, applying both transfer learning and training with proprietary data. The results highlighted the pre-trained extra-large version as the most robust, enabling intelligent activation and promoting energy efficiency.

Keywords: People Detection. ESP32-CAM. Neural Networks. YOLO. Energy Sustainability.

¹Graduanda em Engenharia Elétrica. Instituto Federal de São Paulo. ORCID: <https://orcid.org/0009-0004-0408-5256>. E-mail: maria.pierini@aluno.ifsp.edu.br

² Doutor em Engenharia Elétrica. Instituto Federal de São Paulo. ORCID: <https://orcid.org/0000-0001-5144-518X>. E-mail: emossin@ifsp.edu.br

Resumen: El aumento del consumo de energía eléctrica en Brasil, impulsado por el uso intensivo de aires acondicionados, resalta la necesidad de soluciones inteligentes que optimicen la climatización en ambientes de alta ocupación, como las aulas. En este contexto, se propuso la integración al sistema AirCon, desarrollado en el IFSP – *Campus Sertãozinho*, de una función de encendido y apagado automático basada en detección de personas. Para ello, se emplearon técnicas de visión por computadora e inteligencia artificial. A partir de imágenes capturadas por el microcontrolador ESP32-CAM, se construyó un conjunto de datos personalizado con variaciones de iluminación y ocupación, procesado mediante redes neuronales convolucionales basadas en el algoritmo YOLO. El modelo YOLO11 fue entrenado en distintas versiones, aplicando *transfer learning* y datos propios. Los resultados indicaron que la versión extra-large preentrenada mostró mayor robustez, posibilitando un accionamiento eficiente y contribuyendo al ahorro energético.

Palabras-clave: Detección de Personas. ESP32-CAM. Redes Neuronales. YOLO. Sostenibilidad Energética.

Submetido 10/09/2025

Aceito 08/01/2026

Publicado 10/03/2026

Considerações iniciais

O consumo de energia elétrica no Brasil tem apresentado crescimento expressivo nos últimos anos, fortemente associado ao uso de sistemas de climatização em decorrência do aumento das temperaturas médias em diferentes regiões do país. No primeiro trimestre de 2024, o Brasil registrou um consumo médio de 72.416 MWm, refletindo um crescimento de 5% em relação ao ano anterior (Electra Energy, 2024). Já no primeiro quadrimestre de 2025, o consumo médio subiu para 73.567 MWm, representando um aumento de 1,3% em relação a 2024, segundo dados da Câmara de Comercialização de Energia Elétrica – CCEE (Canal Energia, 2025). Esse cenário evidencia um desafio significativo, pois o aumento no consumo energético compromete não apenas a sustentabilidade financeira, mas também o equilíbrio ambiental.

Diante disso, torna-se essencial o desenvolvimento de soluções capazes de promover maior eficiência energética em ambientes climatizados, sobretudo em espaços de alta rotatividade, como instituições de ensino, onde aparelhos de ar-condicionado frequentemente permanecem ligados mesmo em salas desocupadas. O desperdício resultante eleva custos operacionais e intensifica impactos ambientais, reforçando a necessidade de mecanismos de gestão mais sustentáveis. Nesse cenário, o presente projeto está em consonância com a Política Nacional de Conservação e Uso Racional de Energia (Lei nº 9.991/2000), que estabelece a obrigatoriedade de concessionárias, permissionárias e autorizadas do setor elétrico a investir em projetos de pesquisa, desenvolvimento e eficiência energética, promovendo a modernização do uso da energia no país. Além disso, alinha-se, também, ao Programa Nacional de Conservação de Energia Elétrica (Procel), iniciativa governamental voltada a fomentar o uso racional de energia por meio de ações técnicas, educativas e normativas.

Concomitantemente, os avanços recentes em visão computacional e inteligência artificial (IA) têm impulsionado o desenvolvimento de sistemas automatizados. A detecção de pessoas destaca-se como uma das aplicações centrais dessa evolução da automação, possibilitando soluções robustas e precisas baseadas em tecnologias inteligentes. Entretanto, ainda persistem desafios técnicos na implementação de ferramentas de rastreamento e detecção de objetos em tempo real, como a variabilidade nas formas dos objetos, oclusões parciais ou totais e variações nas condições de iluminação (Kumar; Valarmathi, 2023).

Estudos recentes têm demonstrado o potencial da visão computacional associada a redes neurais, especialmente modelos da família YOLO, para aplicações em detecção de pessoas em

ambientes fechados. Pesquisas como as de Alruwaili *et al.* (2024), em um estudo voltado à aplicação em sistemas ubíquos para pessoas com necessidades especiais, utilizaram YOLOv3 e YOLOv5 para detectar e rastrear essas pessoas, alcançando precisão de 91,5% e 88,5%, respectivamente, o que evidencia a aplicabilidade desses modelos em cenários específicos e desafiadores. Já em outro trabalho, comparações de diferentes arquiteturas YOLO, demonstram que a versão YOLOv8 obteve melhor desempenho (90,7%) em relação ao YOLOv5 (88,5%) e YOLOv7 (90,6%), confirmando a importância da evolução arquitetural para o aprimoramento da detecção de pessoas (Alruwaili *et al.*, 2023).

No campo da eficiência energética, pesquisas também têm destacado a relevância do uso de sistemas inteligentes. Em um estudo de caso sobre o consumo de energia em um edifício educacional (Aziz *et al.*, 2012), identificou-se que a climatização representa parcela significativa da demanda elétrica, ressaltando a necessidade de conscientização e adoção de estratégias para racionalizar o uso de ar-condicionado. No trabalho de Refaat, Abu-Rub e Mohamed (2017), foi proposto um sistema de climatização inteligente e adaptativo, capaz de operar em tempo real para reduzir tanto o consumo energético quanto os custos diários, considerando aspectos como preferências do usuário, conforto térmico e saúde.

Com base nessas evidências, e considerando a associação entre os avanços em visão computacional, inteligência artificial e estratégias de gerenciamento energético, destaca-se a necessidade de construir sistemas mais eficientes, integrados e capazes de atuar de forma autônoma. Nessa perspectiva, anteriormente ao desenvolvimento da pesquisa apresentada neste artigo, foi desenvolvido, no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), *Campus Sertãozinho*, o sistema denominado *AirCon*, uma solução de baixo custo voltada para o controle remoto de aparelhos de ar-condicionado, que integra *hardware* e *software* para otimizar a gestão do consumo energético.

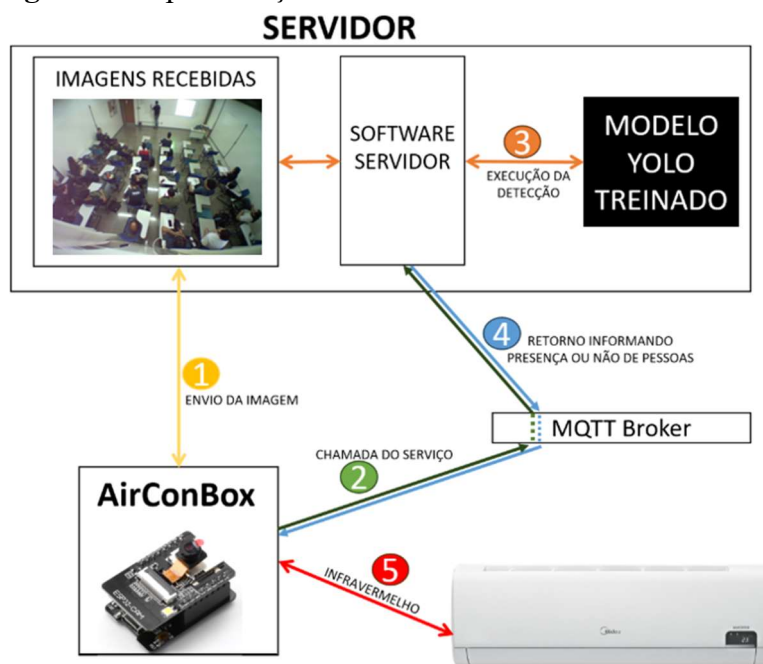
O sistema é composto pelo módulo *AirConBox*, baseado em um microcontrolador ESP32, equipado com sensores de temperatura e um emissor de infravermelho. Esse módulo é responsável por enviar comandos diretamente aos aparelhos, simulando o controle remoto tradicional. A comunicação entre os dispositivos ocorre via protocolo *Message Queuing Telemetry Transport* (MQTT), permitindo que o usuário, por meio de uma interface *web*, acompanhe em tempo real as condições de temperatura, umidade, histórico de funcionamento e controle dos aparelhos climatizadores.

Apesar de funcional, a versão original do *AirCon* não contempla a capacidade de identificar pessoas nos ambientes e, conseqüentemente, não realiza o desligamento automático dos equipamentos quando não há ocupação, um fator que limita sua contribuição para a redução de desperdício energético.

Diante disso, o presente trabalho propõe a evolução e implementação de uma nova funcionalidade ao *AirCon*, incorporando técnicas de visão computacional e inteligência artificial, assim, possibilitando a detecção automática de pessoas. Para isso, foram utilizadas imagens capturadas pelo microcontrolador ESP32-CAM, transmitidas a um servidor na nuvem e processadas por redes neurais convolucionais baseadas algoritmo YOLO (*You Only Look Once*). Essa abordagem permitiu que o sistema tomasse decisões autônomas inteligentes sobre o acionamento ou desligamento dos aparelhos de ar-condicionado, em razão da presença ou ausência de indivíduos no ambiente, elevando significativamente sua eficácia no uso racional da energia.

O funcionamento completo do sistema está ilustrado na Figura 1. Inicialmente, o módulo *AirConBox* (ESP32-CAM) captura e envia imagens ao servidor (1). A solicitação de análise da imagem é encaminhada via protocolo MQTT, comunicação entre dispositivos (2). Em seguida, o servidor processa a imagem utilizando o modelo YOLO desenvolvido neste projeto (3) e retorna o resultado da análise sobre a detecção ou ausência de pessoas no ambiente de volta ao servidor MQTT broker (4). Por fim, o *AirConBox* decide sobre o acionamento ou desligamento do ar-condicionado, enviando o comando por infravermelho (5).

Figura 1: Representação do sistema *AirCon*



Fonte: Autores (2025).

Cabe destacar que o foco deste artigo está na detecção de pessoas na sala de aula, ou seja, etapas 1, 2 e 3 da Figura 1. Para isso, foi desenvolvido um *dataset* personalizado do ambiente de sala de aula. Para sua construção, um módulo ESP32-CAM foi posicionado na parede, em altura próxima ao aparelho de ar-condicionado, a fim de capturar imagens representativas do espaço. Com isso, esse *dataset* serviu como base para o treinamento de diferentes versões do modelo YOLO11. Foram adotadas duas estratégias de treinamento, sendo elas a aplicação de *transfer learning*, utilizando pesos previamente treinados do conjunto de dados da base COCO, e o treinamento utilizando a configuração base do YOLO11 *extra-large*, apenas com dados próprios. Dessa forma, foi possível comparar o desempenho e a confiabilidade de cada abordagem na tarefa de detecção da classe denominada *person*.

Assim, a integração da nova funcionalidade de detecção de pessoas ao sistema *AirCon*, por meio de visão computacional, permitiu o controle automático dos aparelhos de climatização de acordo com a presença ou ausência de indivíduos no ambiente. Em suma, espera-se que essa solução contribua para a redução do consumo de energia elétrica em salas de aula e, consequentemente, em instituições de ensino, promovendo ganhos econômicos e benefícios ambientais, alinhados às demandas atuais de eficiência energética e sustentabilidade.

Metodologia

Este trabalho enquadra-se como uma pesquisa de natureza aplicada, já que seu foco está em resolver um problema prático, ou seja, reduzir o desperdício de energia elétrica causada pelo uso inadequado de aparelhos de ar-condicionado em salas de aula. Do ponto de vista dos objetivos, a pesquisa é explicativa, pois busca entender como o uso de diferentes versões do modelo YOLO afeta a precisão na detecção de pessoas e, por consequência, a eficiência do gerenciamento automático do climatizador. Para alcançar tais resultados, foram adotados procedimentos experimentais, por meio da realização de testes de treinamento e validação de redes neurais, e também de campo, já que o conjunto de imagens foi construído a partir da coleta de fotos capturadas em salas de aula reais. A análise foi conduzida sob uma perspectiva quantitativa, apoiada em indicadores objetivos (*precision*, *recall*, etc), que permitiram comparar o desempenho dos modelos.

Nesse projeto, foram utilizadas as imagens capturadas pelo microcontrolador ESP32-CAM, que integra conectividade *Wi-Fi* e *Bluetooth* e está equipado com a câmera modelo DC-5M21-5640-8 O3-V1, que possui lente *fish-eye* com campo de visão de aproximadamente 160° na diagonal. O dispositivo combina processamento embarcado, baixo consumo de energia e suporte à comunicação sem fio, sendo amplamente empregado em projeto de Internet das Coisas (IoT), sistemas de monitoramento e aplicações de visão computacional em tempo real, como neste trabalho. Além disso, a característica de uma câmera de 160° permite ampliar significativamente a área monitorada, garantindo a captura de um panorama abrangente da sala de aula. Desse modo, possibilita registrar múltiplos pontos de ocupação em um único enquadramento, contribuindo para a construção de um *dataset* personalizado, compondo um conjunto de dados representativo.

Para capturar as imagens, um código embarcado, desenvolvido em linguagem de programação C, foi carregado no dispositivo utilizando a plataforma de desenvolvimento Arduino (Arduino.cc, 2023), via conexão cabo micro USB. As imagens foram enviadas diretamente para o *Google Drive*, por meio de um *script* desenvolvido no *Google APIs (Google Apps Script*, 2022), estabelecendo uma interface de comunicação entre o sistema embarcado (ESP32-CAM) e os serviços em nuvem do *Google*, garantindo segurança na transmissão dos dados.

Em seguida, foi criado e instalado um suporte para fixação do ESP32-CAM na parede ao fundo da sala de aula, em altura adequada para capturar todo o ambiente. Vale ressaltar que, nesse posicionamento, as imagens registravam as pessoas de costas. As capturas das imagens foram realizadas em intervalos de 3 minutos entre cada uma. Inicialmente, as imagens foram capturadas com o *flash* embutido do microcontrolador ativado, no entanto, visando obter um cenário mais natural, com variações reais de luminosidade e minimizar a predominância de tonalidades arroxeadas, desativou-se o *flash* do microcontrolador.

A coleta das fotos da sala de aula ocorreu em um período de três semanas, em intervalos distribuídos entre 8h e 17h. A coleta não ocorreu de forma contínua todos os dias, mas foi planejada de modo a contemplar variações de cenário, iluminação e quantidade de pessoas presentes, conforme apresentado na Tabela 1.

Tabela 1- Cenários sala de aula.

ILUMINAÇÃO – LUZES ³	PROJETOR	CORTINAS	DETECÇÃO DE PESSOAS ⁴
Frente e fundo acesas	Desligado	Fechadas	Sim
Somente frente acesa	Desligado	Fechadas	Sim
Somente fundo aceso	Desligado	Fechadas	Sim
Frente e fundo acesas	Ligado	Fechadas	Sim
Somente frente acesa	Ligado	Fechadas	Sim
Somente fundo aceso	Ligado	Fechadas	Sim
Frente e fundo acesas	Desligado	Parcialmente abertas	Sim
Apagadas	Desligado	Fechadas	Não
Apagadas	Ligado	Fechadas	Não
Apagadas	Ligado	Parcialmente abertas	Não

Fonte: Autores (2025).

Durante essa etapa de desenvolvimento, foram observados desafios significativos na detecção de pessoas em determinadas condições de cenário, como em ambientes de baixa luminosidade, que resultaram em imagens excessivamente escuras ou com coloração distorcida (arroxeadas), bem como em situações de superexposição causadas por feixes intensos de luz.

³ Conjunto de lâmpadas posicionadas na frente e no fundo da sala, podendo ser acionadas de forma independente (somente frente ou somente fundo) ou simultaneamente (frente e fundo).

⁴ “Sim” indica que há condições favoráveis para o modelo realizar a detecção de pessoas. Enquanto, “Não” significa que a detecção não é possível devido à falta de iluminação adequada nas imagens.

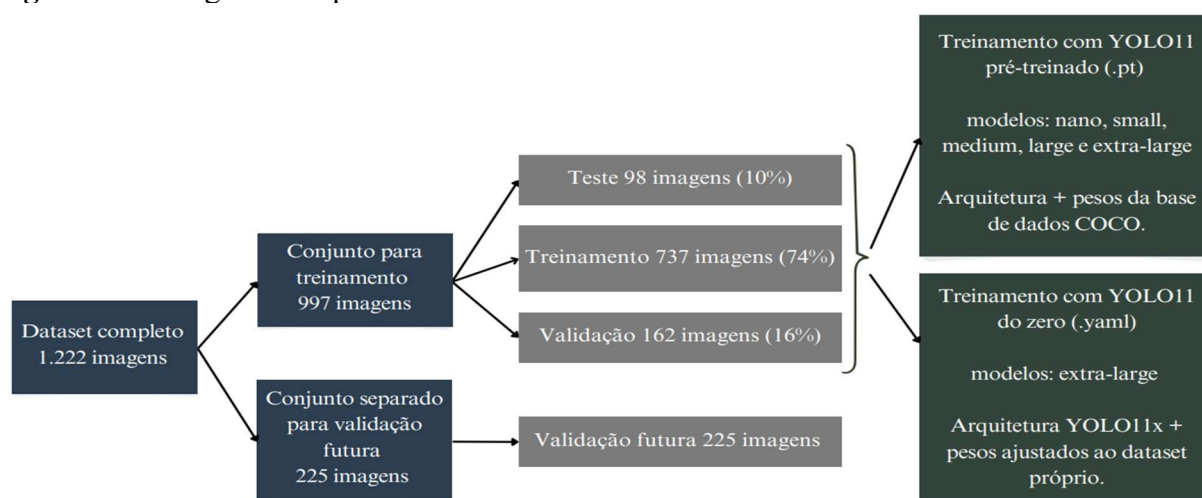
Nesses casos, a visibilidade dos indivíduos nas imagens foi severamente comprometida, impossibilitando sua identificação por meio do modelo de detecção. Em função disso, tais imagens foram excluídas do *dataset*. Possíveis soluções para contornar essas limitações incluem a utilização de sensores de luminosidade e câmeras com visão noturna.

Em seguida, todas as imagens foram anotadas por meio da criação de caixas delimitadoras (*bounding boxes*) e rotuladas (*labels*) associando-as à classe “*person*”, utilizando o *software Roboflow*. No total, foram 1.222 imagens, sendo 997 destinadas ao processo de treinamento do modelo e as demais 225 reservadas para uma etapa posterior de validação com dados desconhecidos. Vale ressaltar que, para uma validação mais robusta, essas imagens foram capturadas em dias diferentes daqueles utilizados para o treinamento.

As 997 imagens foram divididas aleatoriamente e de forma automática pelo *software Roboflow* em três conjuntos essenciais para o processo adequado de treinamento do modelo, da seguinte forma: 98 imagens destinadas ao teste (10%), 737 ao treinamento (74%) e 162 à validação (16%). Cada conjunto de dados contém imagens no formato PNG, acompanhadas pelos respectivos arquivos de anotação no formato TXT.

Em sequência, iniciou-se a fase de treinamento dos modelos com a versão 11 do modelo YOLO e suas variações técnicas de tamanho, a qual utiliza a biblioteca YOLO, desenvolvida em Python pela *Ultralytics* (Ultralytics, 2025). Todo o esquemático dessa etapa está ilustrado na Figura 2.

Figura 2: Fluxograma do processo de treinamento do modelo.



Fonte: Autores (2025).

Para o desenvolvimento do projeto e treinamento dos modelos, foi feito o uso de um computador com capacidade computacional adequada para tarefas de aprendizado profundo. A máquina utilizada possui uma placa de vídeo *NVIDIA GeForce RTX 3060*, equipada com 12 GB de memória dedicada, o que possibilitou acelerar o processo de treinamento por meio do uso de GPU. O desenvolvimento e a execução dos scripts ocorreram no ambiente *PyCharm*, uma IDE (*Integrated Development Environment*) voltada para programação em Python, no qual os códigos incluíam parâmetros e métricas para monitorar o desempenho do modelo.

Concluída a etapa de anotação e rotulação, o treinamento foi conduzido em duas abordagens distintas para possibilitar uma análise comparativa do modelo. Na primeira, adotou-se a técnica de *transfer learning* a partir do modelo YOLO11 pré-treinado (YOLO11.pt). O treinamento contemplou todas as versões do YOLO11 (*nano, small, medium, large e extra-large*), cuja nomenclatura reflete o tamanho e a complexidade da arquitetura, em que modelos menores oferecem maior velocidade e menor precisão, enquanto os maiores proporcionam maior acurácia ao custo de maior consumo computacional.

Durante o treinamento dos modelos YOLO11, foram adotados, em sua maioria, os hiperparâmetros padrão (*default*) definidos pela biblioteca *Ultralytics*, os quais controlam aspectos fundamentais do processo de aprendizado. O tamanho do lote (*batch size*) foi fixado em 16 imagens, valor padrão da plataforma, representando a quantidade de amostras utilizadas em cada etapa de atualização dos pesos do modelo durante o processo de treinamento. As imagens de entrada foram redimensionadas para 640x640 pixels, resolução amplamente adotada em tarefas de detecção de objetos por equilibrar precisão e custo computacional.

A taxa de aprendizado inicial (*learning rate*) foi definida como 0,01, valor *default* da biblioteca, responsável por determinar a magnitude das atualizações aplicadas aos pesos da rede neural durante o processo de otimização. O treinamento utilizou o otimizador selecionado automaticamente pela plataforma (*optimizer = auto*), que adota estratégias adaptativas para garantir estabilidade e convergência do aprendizado.

O número total de épocas foi definido em 1000, permitindo múltiplas passagens completas pelo conjunto de dados. Para evitar *overfitting*, empregou-se o mecanismo de *early stopping*, controlado pelo parâmetro *patience* igual a 100 épocas, que interrompe automaticamente o treinamento quando não há melhora significativa nas métricas de validação.

A segunda abordagem consistiu no treinamento do modelo sem pesos pré-treinados utilizando exclusivamente o conjunto próprio de imagens, ou seja, apenas o *dataset* personalizado com fotografias do ambiente da sala de aula, tornando o treinamento mais específico ao contexto estudado. Nesse momento, foi escolhido o modelo YOLO11 *extra-large* configurado a partir do arquivo *.yaml*, que descreve a arquitetura da rede (número de camadas, tipos de blocos convolucionais, número de parâmetros, entre outras). Esse arquivo não está vinculado a nenhum *dataset* específico, como o COCO, sendo sua função unicamente estrutural, definindo o modelo a ser treinado sem impor restrições quanto à base de dados utilizada. Assim, o modelo gerado a partir deste processo foi denominado YOLO11x_custom.pt, treinado exclusivamente com dados próprios do projeto.

Posteriormente ao treinamento, realizou-se uma análise detalhada dos resultados obtidos avaliados por meio de métricas amplamente reconhecidas em tarefas de detecção de objetos. A métrica *precision* quantifica a proporção de predições positivas corretas em relação ao total de predições positivas realizadas, sendo expressa por:

$$Precision = True\ Positive / (True\ Positive + False\ Positive) \quad (1)$$

essa equação reflete a confiabilidade das detecções efetuadas pelo modelo, enquanto a métrica *recall* mede a capacidade do modelo em identificar todos os objetos reais presentes nas imagens, sendo sensível a falsos negativos, e é definida por:

$$Recall = True\ Positive / (True\ Positive + False\ Negative) \quad (2).$$

O mAP-50 corresponde à média da precisão, considerando um limiar de sobreposição (IoU- Intersecção sobre União) de 50%, indicando o quão bem o modelo localiza os objetos com uma sobreposição mínima aceitável. Por sua vez, o mAP50-95 representa uma métrica mais abrangente e rigorosa, calculada sobre múltiplos limiares de IoU, de 50% a 95%, medindo a localização espacial da detecção dentre de níveis de tolerância.

Além das métricas quantitativas tradicionais, a avaliação do desempenho dos modelos YOLO11 foi complementada pela análise da matriz de confusão e de gráficos de desempenho gerados automaticamente pela biblioteca *Ultralytics* durante a etapa de validação. A matriz de confusão permite avaliar o comportamento do modelo ao relacionar predições corretas e incorretas, evidenciando os valores de verdadeiros positivos (TP), falsos positivos (FP), falsos negativos (FN) e verdadeiros negativos (TN). Essa análise possibilita a identificação de padrões de erro, como detecções indevidas ou falhas na identificação de pessoas.

No contexto deste trabalho, os verdadeiros positivos correspondem às situações em que o modelo identificou corretamente a presença de pessoas nas imagens analisadas. Os falsos positivos referem-se às detecções incorretas, nas quais o modelo indicou a presença de pessoas em imagens onde não havia indivíduos presentes. Os falsos negativos representam os casos em que o modelo não conseguiu detectar pessoas que estavam presentes na sala de aula efetivamente. Por sua vez, os verdadeiros negativos correspondem às situações em que o modelo corretamente não realizou detecções em imagens nas quais não havia pessoas.

Paralelamente, foram analisados gráficos de desempenho gerados na etapa de validação, incluindo as curvas *Precision Curve*, *Recall Curve* e *Precision–Recall Curve*. Esses gráficos permitem observar o comportamento do modelo em diferentes limiares de confiança, evidenciando o equilíbrio entre falsos positivos e falsos negativos. Adicionalmente, a curva *F1* foi considerada para avaliar o compromisso entre precisão e *recall*, auxiliando na identificação de limiares que maximizam o desempenho global do modelo.

Para concluir e assegurar uma análise comparativa realista e confiável do desempenho dos modelos treinados, foi conduzida uma etapa adicional de validação. Nessa fase, utilizou-se o conjunto previamente separado para futura validação, composto por 225 imagens anotadas, conforme ilustrado no fluxograma da Figura 2. Essas imagens foram processadas por um código específico que gerou métricas quantitativas essenciais para a escolha do melhor modelo. Essa etapa também reflete o funcionamento prático do sistema, no qual as imagens capturadas são armazenadas na nuvem e posteriormente analisadas pela rede neural treinada neste projeto.

Por fim, a integração completa com o sistema *AirCon* permitiu que, com base nas detecções realizadas, fossem tomadas decisões autônomas e inteligentes sobre o acionamento ou desligamento dos equipamentos de climatização, conforme a presença ou ausência de pessoas no ambiente da sala de aula.

Análise dos dados e resultados

Diante das análises realizadas, foi possível comparar o desempenho das diferentes versões do modelo YOLO11, com o objetivo de identificar a arquitetura mais adequada para a solução proposta. Dessa maneira, verificou-se que todas as variantes apresentaram desempenho satisfatórios, com variações associadas principalmente à complexidade das imagens de entrada e ao número de parâmetros de cada modelo.

Na primeira abordagem, baseada em *transfer learning*, as versões *nano*, *small*, *medium*, *large* e *extra-large*, inicialmente pré-treinadas na base COCO e posteriormente refinadas com um *dataset* personalizado de salas de aula, obtiveram métricas semelhantes. Nessas condições, arquiteturas mais compactas foram capazes de capturar os padrões essenciais sem exigir redes profundas, como apresentado na Tabela 2, que sintetiza os resultados obtidos na validação com 225 imagens previamente separadas.

Tabela 2- Comparação das métricas na etapa de validação dos modelos treinados.

MODELO YOLO11	TIPO	mAP50 (train)	mAP50 (val)	mAP50-95	Precision	Recall
<i>Nano_pretrained</i>	Pré-treinado	0,965	0,940	0,484	0,923	0,919
<i>Small_pretrained</i>	Pré-treinado	0,973	0,941	0,504	0,932	0,907
<i>Medium_pretrained</i>	Pré-treinado	0,974	0,941	0,503	0,925	0,904
<i>Large_pretrained</i>	Pré-treinado	0,974	0,954	0,527	0,945	0,926
<i>Extra_large_pretrained</i>	Pré-treinado	0,974	0,945	0,510	0,938	0,897
<i>Extra_large_custom</i>	Sem pesos pré-treinados	0,970	0,939	0,478	0,901	0,902

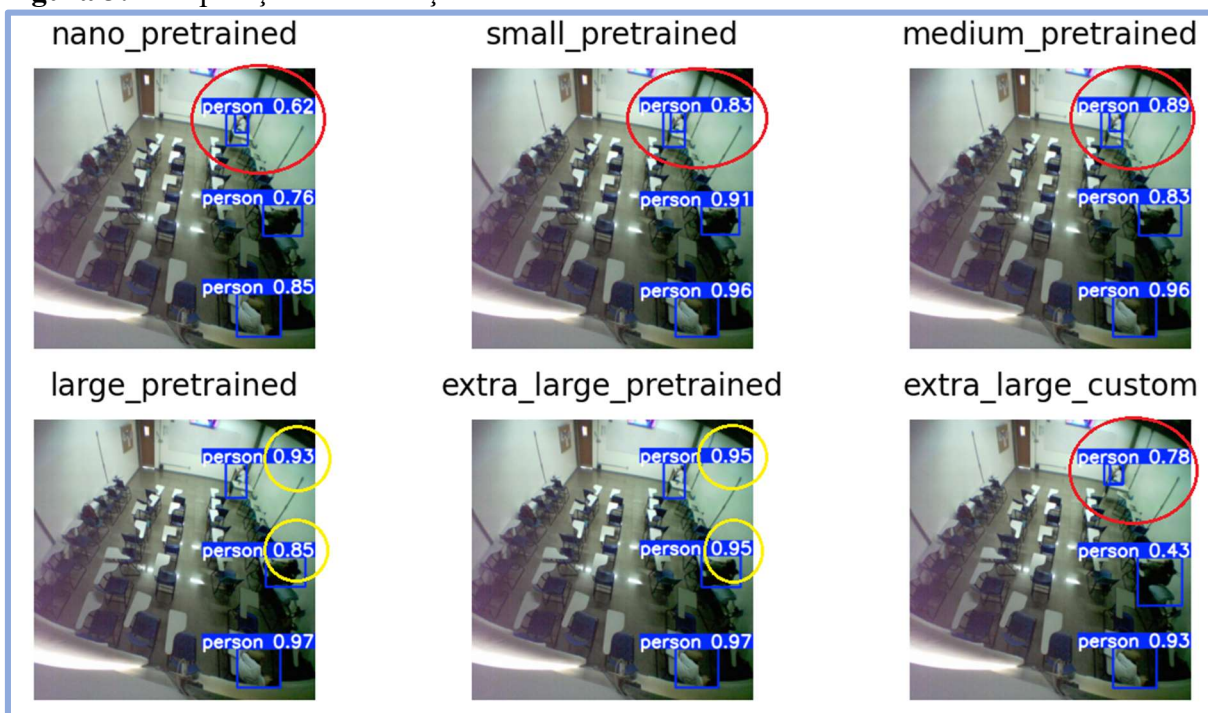
Fonte: Autores (2025).

Embora o modelo YOLO11 *extra-large* possua maior robustez estrutural, seus resultados não foram substancialmente superiores aos das versões menores. Em muitos casos, os desempenhos apresentam valores muito próximos, o que pode ser explicado pelo fato de que modelos menos complexos generalizam com maior facilidade em cenários de baixa variabilidade, enquanto arquiteturas com grande número de parâmetros apresentam maior suscetibilidade ao *overfitting*.

Por outro lado, análises qualitativas visuais demonstraram que o modelo *extra-large* apresentou vantagens em imagens mais complexas, sobretudo em situações que exigiam maior precisão. A Figura 3 exemplifica esse comportamento, na qual estão destacadas, em círculos vermelhos, as sobreposições das caixas delimitadoras geradas por modelos menores. Já os círculos amarelos indicam comparações no nível de confiança dos modelos, uma vez que,

enquanto o modelo YOLO11 *large* pré-treinado obteve valores de 93%, 85% e 97% para três pessoas detectadas na imagem, o YOLO11 *extra-large* alcançou 95%, 95% e 97%, demonstrando maior confiabilidade e justificando sua seleção como o modelo referência entre os pré-treinados.

Figura 3: Comparação das detecções de todos os modelos treinados.



Fonte: Autores (2025).

Com base nas métricas quantitativas, matrizes de confusão, gráficos de desempenho e a avaliação visual com imagens inéditas, concluiu-se que o modelo de *transfer learning* que apresentou os melhores resultados foi o YOLO11 *extra-large* (designado YOLO11x_pretrained.pt). Nesse panorama, o modelo apresentou métricas de desempenho com altos níveis de confiança, destacando-se, por exemplo, a métrica *precision*, que avalia a proporção de detecções corretas em relação ao total de previsões realizadas, a qual atingiu 96% no treinamento, além de apresentar maior consistência na localização das *bounding boxes*. Adicionalmente, o modelo destacou-se no nível de confiança atribuído à classificação dos objetos como pertencentes à classe “*person*”. Enquanto outros modelos treinados apresentaram limiares em torno de 93% acertos nas previsões, o YOLO11 *extra-large* alcançou um índice de aproximadamente 98% de confiança na identificação correta de pessoas, reforçando sua

superioridade em termos de segurança na predição. As detecções obtidas pelo modelo são apresentadas na Figura 4, na qual é possível observar que o nível de confiança para a classe “person” permanece acima de 90% na maioria dos casos.

Figura 4: Detecções feitas pelo modelo YOLO11x_pretrained.pt



Fonte: Autores (2025).

Durante o desenvolvimento, foi observado um aumento na métrica *val/dfl_loss*, responsável por medir o erro na predição das coordenadas das *bounding boxes* na validação. Como se trata de uma métrica de erro, valores mais altos poderiam indicar perda de precisão na localização e um possível indício de *overfitting*. Esse comportamento, entretanto, foi associado à elevada similaridade entre os conjuntos de treinamento e validação, que compartilham características como cenário, ângulos de captura e condições de iluminação semelhantes. Para confirmar se isso poderia impactar o desempenho em situações reais, foi então realizada a etapa adicional de validação com imagens capturadas pelo ESP32-CAM em outro momento. Os resultados obtidos, apresentados na Tabela 2, confirmaram que o modelo manteve bom desempenho, evidenciando a capacidade de generalização adequada ao cenário proposto.

Além do *transfer learning*, foram analisados os resultados da segunda abordagem do treinamento do modelo em configuração base, ou seja, sem utilizar o aprendizado prévio oriundo do conjunto de dados COCO, empregando exclusivamente imagens de um *dataset* personalizado. Nessa abordagem, foi escolhida e utilizada a arquitetura do modelo YOLO11 *extra-large*; o desempenho do modelo YOLO11x_custom.pt foi também considerado satisfatório, destacando-se métricas elevadas, como 97% de *precision* na detecção de pessoas e 97% de precisão na localização das *bounding boxes*, indicando alta confiabilidade tanto na identificação quanto na delimitação dos objetos de interesse. Essas métricas foram extraídas durante a etapa de treinamento.

Por fim, para a definição do modelo mais adequado ao projeto, foram selecionadas para comparação crítica as duas variações do modelo *extra-large*, tanto o YOLO11x_pretrained, que se destacou entre os modelos baseados em *transfer learning*, quanto o YOLO11x_custom, treinado integralmente a partir de sua configuração base. A análise comparativa evidenciou que o YOLO11x_pretrained apresentou desempenho superior nas principais métricas avaliadas, como *mAP50*, *mAP50-95* e *precision*, confirmando sua elevada acurácia na detecção e maior confiabilidade na localização das *bounding boxes*. Embora o modelo treinado sem pesos prévios (YOLO11x_custom) tenha registrado uma métrica *recall* ligeiramente superior, essa vantagem mostrou-se pouco relevante diante da consistência global do modelo pré-treinado.

Cabe destacar um ponto relevante referente às condições desafiadoras de captura das imagens. Apesar de a lente de 160° da câmera ter proporcionado uma ampla cobertura do ambiente, certas regiões, como os cantos inferiores, registraram cortes parciais de indivíduos. Essa característica é compreendida como uma dificuldade inerente ao cenário monitorado. Consequentemente, o modelo foi exposto a exemplos em que apenas partes do corpo estavam visíveis, o que aumentou a complexidade do processo de detecção. Ainda assim, as métricas obtidas evidenciam que ambos os modelos foram capazes de generalizar para novas imagens, confirmando a robustez do treinamento mesmo diante de condições de captura menos favoráveis.

Em suma, os resultados confirmam que o modelo YOLO11x_pretrained.pt representa a alternativa mais adequada para o projeto, destacando-se como a solução mais robusta para integração ao sistema *AirCon* e garantindo a detecção confiável da presença de pessoas em ambientes de sala de aula.

No que se refere ao desempenho dos modelos, embora o YOLO11x_*pretrained* tenha se destacado como a melhor alternativa para o projeto, o YOLO11x_*custom* também apresentou métricas próximas, demonstrando capacidade de generalização para novas imagens e desempenho consistente. Essa proximidade sugere que o modelo customizado, treinado a partir da configuração base e exclusivamente com o conjunto de dados desenvolvido para este estudo, pode constituir uma alternativa estratégica em cenários nos quais haja restrições ao uso de modelos pré-treinados ou indisponibilidade de bases de dados públicas. Assim, o YOLO11x_*custom* configura-se como uma segunda opção viável, a depender das exigências e particularidades do contexto de aplicação.

Considerações finais

Em síntese, os resultados obtidos confirmam a viabilidade da integração entre a visão computacional e o sistema *AirCon* para o gerenciamento automatizado e inteligente da climatização em salas de aula, a partir da detecção de pessoas. Dessa forma, o desenvolvimento do trabalho permitiu a criação de um *dataset* abrangente e devidamente anotado, que serviu de base para o treinamento e a avaliação de diferentes versões do modelo YOLO11.

Um aspecto relevante é que a diferença de desempenho entre o modelo YOLO11 *extra-large* e versões reduzidas, como o YOLO11 nano, mostrou-se pouco expressiva nos testes realizados. Esse resultado é bastante pertinente, pois demonstra que, em contextos nos quais existem limitações de *hardware*, seja em termos de memória, processamento ou consumo energético, modelos mais compactos podem ser empregados sem perdas significativas na qualidade da detecção. Nesses casos, a escolha por arquiteturas menores revela-se vantajosa, proporcionando respostas mais rápidas e menor demanda computacional, ao mesmo tempo em que mantém níveis de acurácia adequados às exigências práticas. Desse modo, a decisão pelo modelo a ser utilizado deve considerar não apenas o desempenho absoluto, mas também o equilíbrio entre eficiência operacional e as restrições do ambiente de aplicação.

A exploração tanto do *transfer learning* quanto do treinamento a partir da configuração base do modelo sem pesos pré-treinados possibilitou comparar estratégias distintas, evidenciando que o modelo YOLO11x_*pretrained* apresentou o melhor desempenho dentre os modelos, resultando em uma maior robustez da rede capaz de generalizar adequadamente frente a novas imagens e atender à proposta do projeto. A partir disso, a integração da nova

funcionalidade ao *AirCon* potencializou sua aplicação prática, garantindo o acionamento e desligamento automatizado e inteligente do ar-condicionado conforme a ocupação do ambiente.

Além de aspectos técnicos, a adoção de um sistema automatizado de detecção de presença para controle de climatização, como o proposto neste trabalho, apresenta potencial para gerar benefícios significativos no âmbito ambiental. Em termos energéticos, a desativação automática dos aparelhos de ar-condicionado em ambientes desocupados pode reduzir substancialmente o consumo de eletricidade, contribuindo para a diminuição da demanda na rede elétrica e, conseqüentemente, para a redução de custos operacionais nas instituições de ensino.

Portanto, a solução desenvolvida e apresentada nesse trabalho mostra-se promissora para mitigar desperdícios energéticos e impulsionar o avanço de sistemas autônomos baseados em visão computacional. A integração do modelo de rede neural YOLO11 com sistemas embarcados, como o ESP32-CAM, atendeu aos objetivos propostos, demonstrando eficiência, confiabilidade e aplicabilidade prática, alinhando-se, assim, às demandas contemporâneas de eficiência e sustentabilidade.

Agradecimentos

Agradecimento ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que está financiando e apoiando o projeto. Os autores agradecem também ao Instituto Federal de São Paulo, *Campus Sertãozinho*, pela infraestrutura disponibilizada e incentivo à pesquisa.

Referências

ALRUWAILI, Madallah; SIDDIQI, Muhammad Hameed; ATTA, Muhammad Nouman; ARIF, Mohammad. Deep learning and ubiquitous systems for disabled people detection using YOLO models. **Computers in Human Behavior**, Londres: Elsevier Ltd, v. 154, 2024.

ALRUWAILI, M.; et al. Deep learning-based YOLO models for the detection of people with disabilities. **IEEE Access**, v. 12, p. 2543-2566, 2024.

ARDUINO.CC. **Arduino Software**. Disponível em: <https://www.arduino.cc/en/software/>. Acesso em: 25 set. 2024.

AZIZ, M. B. A.; ZAIN, Z. M.; BAKI, S. R. M. S.; HADI, R. A. Air-conditioning energy consumption of an education building and its building energy index: A case study in engineering complex, UiTM Shah Alam, Selangor. **IEEE Control and System Graduate Research Colloquium**, Shah Alam, Malaysia, 2012. p. 175-180.

BRASIL. MINISTÉRIO DE MINAS E ENERGIA. **Procel – Programa Nacional de Conservação de Energia Elétrica**. Brasília, 2019. Disponível em: <https://www.gov.br/mme/pt-br/assuntos/secretarias/sntep/procel>. Acesso em: 15 dez. 2024.

BRASIL. Lei nº 9.991, de 24 de julho de 2000. **Política Nacional de Conservação e Uso Racional de Energia**. Brasília, 2000. Disponível em: <https://www.gov.br/mme/pt-br/acesso-a-informacao/legislacao/leis/lei-n-9-991-2000.pdf/view>. Acesso em: 02 jul. 2025.

CANAL ENERGIA. **CCEE aponta aumento de 1,3% no consumo de energia do Brasil nos primeiros meses de 2025**. 2025. Disponível em: <https://www.canalenergia.com.br/noticias/53296173/ccee-aponta-aumento-de-13-no-consumo-de-energia-do-brasil-nos-primeiros-meses-de-2025>. Acesso em: 23 ago. 2025.

ELECTRA ENERGY. **Consumo de energia no Brasil subiu 5% no 1º trimestre de 2024, aponta CCEE**. 2024. Disponível em: <https://electraenergy.com.br/consumo-de-energia-no-brasil-subiu-5-no-1o-trimestre-de-2024-aponta-ccee/>. Acesso em: 28 jan. 2025.

GOOGLE APP SCRIPT, 2022. Disponível em: <https://developers.google.com/apps-script>. Acesso em: 12 nov. 2024.

KUMAR, Y. M. Jaswanth; VALARMATHI, P. YOLO based real-time human detection using deep learning. In: **National Conference on Communication Systems (NCOCS)**, 4., 2022, Chennai, India. Journal of Physics: Conference Series, IOP Publishing, v. 2466, p. 012034, 2023.

REFAAT, S. S.; ABU-RUB, H.; MOHAMED, A. Smart and energy efficient air-conditioning system. In: **19th European Conference on Power Electronics and Applications (EPE'17 ECCE Europe)**, Warsaw, Poland, 2017. p. P.1-P.6.

ULTRALYTICS. **YOLO – State-of-the-Art Computer Vision Models**. Disponível em: <https://ultralytics.com/yolo>. Acesso em: 16 jul. 2025.