**CHALLENGE FOR THE CLASSIFICATION OF LABELED RECYCLABLE MATERIALS FOCUSED ON GLASS AND PLASTIC USING ARTIFICIAL INTELLIGENCE AND COMPUTER VISION**

**(Desafio para a classificação de materiais recicláveis rotulados com foco em vidro e plástico utilizando inteligência artificial e visão computacional)**

 Giovana S. Mainente, Helder D. R. Alvarenga, Pedro H. F. Dias1

Orientadora: Drª Marise Miranda2

**Faculdade São Paulo Tech School - São Paulo, SP, Brasil**

**ABSTRACT**

This work presents the development of an automated system for identifying and classifying labeled recyclable materials, with emphasis on plastic and glass, using computer vision and artificial intelligence techniques. Considering the growing volume of urban solid waste and the low efficiency of manual sorting methods, the proposed solution aims to improve the accuracy, speed, and safety of the separation process.

For this purpose, a dataset in YOLO format was prepared and processed through the Roboflow platform, and the YOLOv11 model was trained in two distinct versions, enabling comparative performance analysis. The results show significant improvements, especially in the “glass” class, whose recall increased from 66.67% to 89.66%. The comparison between versions highlights advancements in stability and overall model performance. The developed system demonstrates potential to optimize recycling processes, reduce operational costs, enhance the quality of recovered materials, and foster sustainable practices aligned with the circular economy.

**Keywords**: Machine Learning. Object Detection. AWS (Amazon Web Services). Roboflow. Image Processing. Recyclable Material Classification. YOLOv11. Computer Vision. Artificial Intelligence. Cloud Computing. Classification Model.Keywords:

1 ***Student's Bachelor's degree in Computer Science***

1. ***Professor of Computer Science at the São Paulo Tech School, Brazil***

**RESUMO**

Este trabalho apresenta o desenvolvimento de um sistema automatizado para identificação e classificação de materiais recicláveis rotulados, com ênfase em plástico e vidro, utilizando técnicas de visão computacional e inteligência artificial. Diante do aumento expressivo dos resíduos sólidos urbanos e da baixa eficiência dos métodos manuais de triagem, a solução proposta visa aprimorar a precisão, agilidade e segurança do processo de separação.

Para isso, foi empregada uma base de dados no formato YOLO, processada por meio da plataforma Roboflow, e o modelo YOLOv11 foi treinado em duas versões distintas, permitindo a análise comparativa de desempenho. Os resultados evidenciam avanços significativos, especialmente na classe “vidro”, cujo recall evoluiu de 66,67% para 89,66%. A comparação entre as versões demonstra melhorias na estabilidade e na performance geral do modelo. O sistema desenvolvido revela potencial para otimizar a reciclagem, reduzir custos operacionais, elevar a qualidade dos materiais recuperados e fomentar práticas sustentáveis alinhadas à economia circular.

**Palavras‑chave**: Machine Learning. Detecção de Objetos. AWS (Amazon Web Services). Roboflow. Processamento de Imagens. Classificação de Material Reciclável. YOLOv11. Visão Computacional. Inteligência Artificial. Cloud Computing. Modelo de Classificação.

1. INTRODUÇÃO

O descarte irregular de resíduos sólidos urbanos é um dos maiores problemas ambientais e sociais enfrentados pela sociedade contemporânea. No Brasil, estima-se que cerca de 33 milhões de toneladas de lixo sejam descartadas de forma inadequada todos os anos, o que corresponde a aproximadamente 43% do total de resíduos gerados (CEBRASSE, 2023). Esse cenário revela não apenas a fragilidade da gestão pública de resíduos, mas também a falta de conscientização da população quanto ao impacto de suas práticas cotidianas. O acúmulo de lixo em locais impróprios compromete a qualidade do solo, contamina lençóis freáticos, obstrui sistemas de drenagem e favorece enchentes, além de contribuir para a proliferação de vetores de doenças como mosquitos e roedores (ENGEWET, 2025).

Além dos impactos diretos na saúde e no meio ambiente, o descarte irregular representa uma perda significativa de materiais com alto potencial de reaproveitamento, como vidro e plástico, que poderiam ser reinseridos na cadeia produtiva. A reciclagem, nesse contexto, surge como uma solução estratégica, mas sua eficiência depende da qualidade da triagem. O método manual, ainda predominante, é marcado por baixa produtividade, altos custos e riscos ocupacionais. Assim, torna-se urgente investir em soluções tecnológicas que não apenas otimizem a separação dos resíduos, mas também incentivem a população a realizar o descarte correto. Este projeto propõe o desenvolvimento de um sistema automatizado de classificação de resíduos, fundamentado em Machine Learning e visão computacional, com foco inicial em vidro e plástico, como forma de enfrentar os desafios da reciclagem e contribuir para práticas sustentáveis alinhadas à economia circular.

1. CONTEXTUALIZAÇÃO

A crise na gestão de resíduos é global. Relatórios da ONU indicam que apenas 9% dos resíduos plásticos são reciclados mundialmente, enquanto o restante é destinado a aterros, incinerado ou descartado irregularmente em rios e oceanos (ONU, 2022). No Brasil, segundo a ABRELPE, cada habitante gerou em média 1,04 kg de resíduos por dia em 2022, totalizando mais de 77 milhões de toneladas anuais, das quais quase metade teve destinação inadequada (ABRELPE, 2023).

Esse cenário evidencia a necessidade de transição para um modelo de economia circular, que trata o resíduo como recurso e busca reduzir a extração de matérias-primas virgens. A economia circular se fundamenta em três princípios: eliminar resíduos desde o início do processo produtivo, manter produtos e materiais em uso pelo maior tempo possível e regenerar sistemas naturais (ELLEN MACARTHUR FOUNDATION, 2015).

A tecnologia desempenha papel central nessa transformação. A chamada Indústria 4.0, com sistemas ciberfísicos e algoritmos de Inteligência Artificial, já demonstra potencial para revolucionar setores produtivos. Aplicada à gestão de resíduos, a IA pode identificar padrões, classificar materiais com alta precisão e escalar processos de reciclagem (KAGERMANN; WAHLSTER; HELBIG, 2013). Modelos como o YOLOv11, aliados a plataformas de treinamento como o RoboFlow, permitem criar sistemas capazes de reconhecer e separar materiais recicláveis em tempo real, reduzindo erros humanos e aumentando a eficiência. Além disso, tais sistemas podem ser integrados a campanhas de conscientização, mostrando à população que o descarte correto é viável e necessário.

1. JUSTIFICATIVA

A implementação de um sistema automatizado de triagem de recicláveis se justifica por três dimensões principais: econômica, ambiental e social.

* Viabilidade Econômica: A automação aumenta a velocidade e precisão da separação, elevando o volume de materiais recicláveis de melhor qualidade. Materiais mais puros possuem maior valor de mercado, gerando receita adicional para cooperativas e empresas de reciclagem. Além disso, há redução de custos com mão de obra e diminuição de erros operacionais (NUNES, 2025; SANTOS et al., 2024).
* Impacto Ambiental: Um sistema eficiente garante que maior quantidade de vidro e plástico seja reciclada, reduzindo a necessidade de extração de matérias-primas virgens, economizando energia e água e diminuindo emissões de gases de efeito estufa (FERREIRA et al., 2019). Isso contribui diretamente para a mitigação da poluição e preservação dos ecossistemas (ENGEWET, 2025).
* Benefício Social: A separação manual expõe trabalhadores a riscos de contaminação e acidentes. A automação realoca a força de trabalho para funções de supervisão e gestão, promovendo condições mais seguras e dignas (SANTOS et al., 2024). Além disso, ao demonstrar que a tecnologia pode auxiliar na triagem, o projeto incentiva a população a realizar o descarte correto, fortalecendo a consciência ambiental coletiva (NUNES, 2025).

1. BASE DE DADOS E TRATAMENTO

Para a construção do modelo preditivo, foi realizada uma busca sistemática por bases de imagens que atendessem aos requisitos do projeto. Os critérios estabelecidos incluíam: disponibilidade de imagens reais de objetos recicláveis das classes plástico e vidro; acesso livre para uso acadêmico; e diversidade visual suficiente para garantir robustez ao processo de treinamento. A partir dessa análise, foram selecionadas duas bases públicas: o conjunto de imagens de garrafas de vidro disponibilizado no Kaggle (DATACLUSTERLABS, 2023) e o dataset TrashNet, disponível no GitHub (THUNG, 2017).

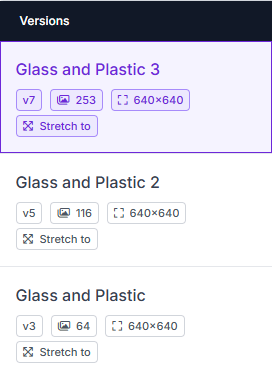
Após a coleta, as imagens foram processadas na plataforma RoboFlow, que permitiu a anotação e classificação supervisionada dos objetos em duas categorias: plástico e vidro. Cada imagem foi normalizada e redimensionada para 640x640 pixels, assegurando padronização e compatibilidade com o modelo YOLOv11. Esse pré-processamento incluiu a aplicação de layers de anotação, fundamentais para a etapa de treinamento supervisionado.

O desenvolvimento do modelo ocorreu de forma incremental, com a expansão progressiva da base de dados em diferentes versões (Figura 1):

* Versão 1: composta por 64 imagens, distribuídas em 40 para treinamento, 12 para validação e 12 para teste.
* Versão 2: ampliada para 116 imagens (incremento de 52 em relação à versão anterior), sendo 85 destinadas ao treinamento, 19 à validação e 12 ao teste.
* Versão 3: versão mais recente e robusta, composta por 253 imagens, das quais 158 foram utilizadas para treinamento, 35 para validação e 60 para teste.

Essa evolução da base de dados permitiu avaliar o impacto do aumento amostral na acurácia e generalização do modelo, garantindo maior confiabilidade nos resultados obtidos.

**Figura 1** – Versões da base tratada no RoboFlow



Fonte: Autoral (2025)

1. CONTEXTOS TÉCNICOS

O desenvolvimento de um sistema automatizado para a identificação e classificação de resíduos recicláveis fundamenta-se nos pilares da Inteligência Artificial (IA), com ênfase em Visão Computacional e Aprendizado Profundo. Este capítulo apresenta os conceitos técnicos que sustentam a metodologia adotada, descrevendo desde os fundamentos teóricos até as métricas de avaliação utilizadas no modelo YOLOv11.

* 1. Aprendizado de Máquina (Machine Learning)

O Aprendizado de Máquina (Machine Learning – ML) é um subcampo da IA que se concentra em algoritmos capazes de aprender padrões a partir de dados, sem necessidade de programação explícita para cada tarefa. Em termos formais, “um programa de computador aprende com a experiência E em relação a uma classe de tarefas T e medida de desempenho P, se o desempenho em T, medido por P, melhora com a experiência E” (MITCHELL, 1997). No contexto deste projeto, o ML é aplicado para reconhecer características visuais de resíduos recicláveis, permitindo que o sistema diferencie objetos de vidro e plástico.

* 1. Aprendizado Profundo (Deep Learning)

O Aprendizado Profundo (Deep Learning – DL) é um subconjunto do ML que utiliza Redes Neurais Artificiais (RNAs) com múltiplas camadas ocultas. Essa profundidade possibilita a extração de representações hierárquicas dos dados, indo de características simples (bordas, formas) até atributos complexos (texturas, padrões específicos). O DL é a base das Redes Neurais Convolucionais (Convolutional Neural Networks – CNNs), amplamente utilizadas em tarefas de Visão Computacional. Neste trabalho, o modelo YOLOv11 foi empregado por sua capacidade de realizar detecção em tempo real com alta precisão, sendo adequado para aplicações práticas em triagem de resíduos.

* 1. Classificação e Detecção de Objetos

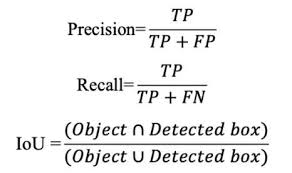
A Visão Computacional busca permitir que computadores interpretem imagens e vídeos de forma semelhante ao olho humano. Dentro desse domínio, destacam-se duas tarefas fundamentais:

* Classificação de Imagem: atribui um rótulo único à imagem inteira, indicando a classe predominante (por exemplo, “vidro” ou “plástico”).
* Detecção de Objetos: além de identificar a classe, localiza a posição exata do objeto na cena, permitindo que múltiplas instâncias sejam reconhecidas simultaneamente. Essa abordagem é essencial para sistemas de triagem, pois possibilita identificar e separar diferentes materiais em ambientes reais.
  1. Caixa Delimitadora (Bounding Box)

A detecção de objetos é representada por caixas delimitadoras (Bounding Boxes), definidas por quatro coordenadas (x mínimo, y mínimo, x máximo, y máximo). Essas caixas circunscrevem o objeto de interesse, fornecendo informações espaciais que podem ser utilizadas por sistemas automatizados, como braços robóticos, para manipulação física dos resíduos. Diferentemente da classificação, que apenas informa a presença de um objeto, a detecção com Bounding Boxes indica sua posição precisa na imagem.

* 1. Métricas de Avaliação
* A avaliação do desempenho do modelo YOLOv11 é realizada por métricas que consideram tanto a precisão da localização quanto a correção da classificação:
* IoU (Intersection over Union) (Figura 2): mede a sobreposição entre a caixa prevista e a caixa real, sendo considerada correta quando o valor excede um limiar pré-definido, geralmente, 0,5.

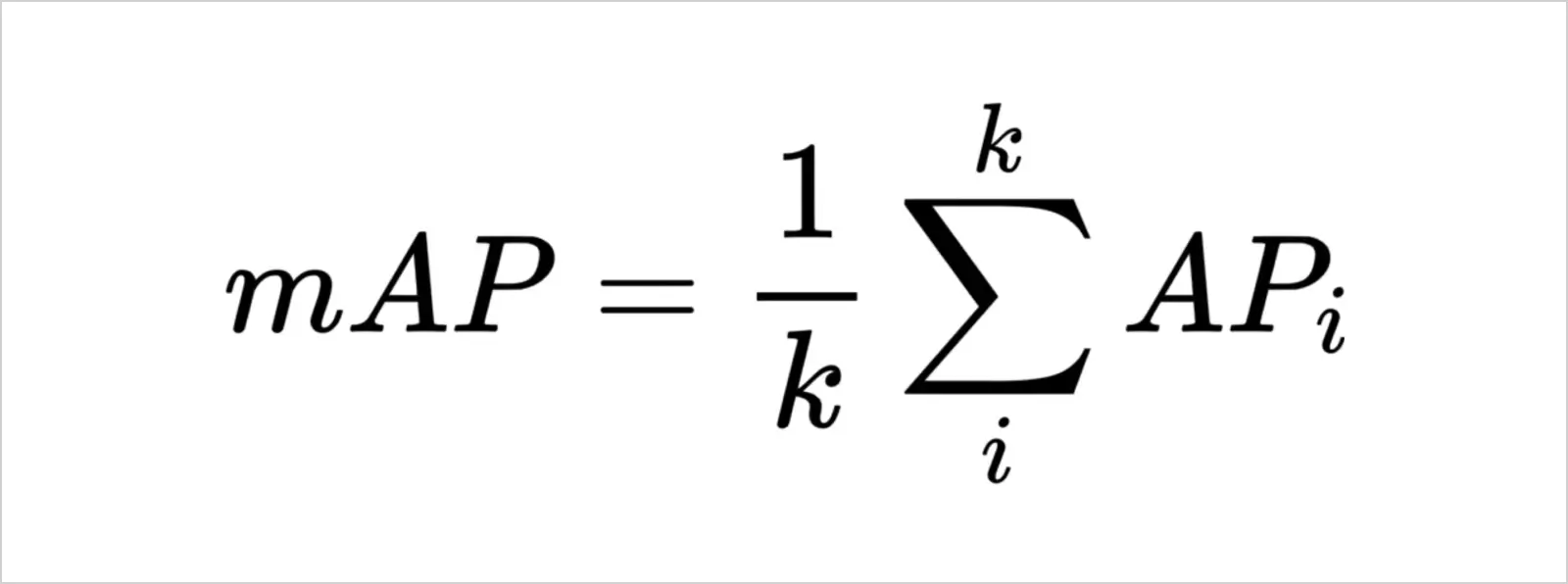
**Figura 2** – Fórmula IoU



Fonte: learnopencv (2025)

* Precisão (Precision): proporção de detecções positivas que foram corretas, reduzindo falsos positivos.
* Recall (Sensibilidade): proporção de objetos reais detectados corretamente, reduzindo falsos negativos.
* Average Precision (AP): resume a curva de Precisão-Recall, calculando a área sob a curva.
* Mean Average Precision (mAP) (Figura 3): média dos valores de AP para todas as classes. No presente projeto, N=2 (vidro e plástico), sendo o mAP o indicador mais robusto da qualidade global do modelo.

**Figura 3** – Fórmula mAP



Fonte: superannotate (2023)

1. METODOLOGIA

A escolha do modelo YOLOv11, aliada ao uso de bases de dados específicas e ao pré-processamento realizado na plataforma RoboFlow, foi decisiva para garantir resultados consistentes em cenários reais de triagem de resíduos. O YOLOv11 representa um avanço significativo em relação às versões anteriores, oferecendo maior precisão, desempenho em tempo real e capacidade de generalização, características essenciais para aplicações práticas em cooperativas e centros de reciclagem.

O modelo foi estruturado no formato YOLO (You Only Look Once), amplamente utilizado em tarefas de detecção de objetos. Nesse padrão, cada imagem possui um arquivo de anotação associado (.txt), contendo as informações de classe e as coordenadas normalizadas das bounding boxes. Esse formato facilita o processamento durante o treinamento, permitindo que o sistema não apenas identifique corretamente os materiais (vidro e plástico), mas também localize sua posição exata na cena.

A adoção desse padrão, combinada com métricas de avaliação como o \*mean Average Precision\* (mAP), assegura que o modelo seja avaliado de forma abrangente, considerando simultaneamente a acurácia da classificação e a precisão da localização. Dessa forma, o sistema alcança confiabilidade técnica e robustez, tornando-se adequado para a implementação em ambientes reais de triagem automatizada.Estrutura da Base de Dados

O formato YOLO é amplamente utilizado e altamente eficiente para tarefas de detecção de objetos, pois armazena as informações de classe e localização por meio de arquivos de texto (.txt) associados a cada imagem. Cada arquivo contém as coordenadas normalizadas das bounding boxes, facilitando o processamento durante o treinamento do modelo.

A adoção do **YOLOv11** proporcionou vantagens como maior precisão, melhor desempenho em tempo real e aprimoramentos na capacidade de generalização em comparação às versões anteriores, tornando-o adequado às necessidades deste projeto.

* 1. Uso da Plataforma Roboflow

A plataforma Roboflow foi utilizada como ferramenta central para a organização, pré-processamento e gerenciamento da base de dados empregada neste projeto. Sua adoção se mostrou estratégica, pois oferece recursos específicos para tarefas de Visão Computacional, permitindo que os conjuntos de imagens sejam estruturados de forma compatível com o modelo YOLOv11.

Entre as principais funcionalidades exploradas destacam-se:

* Versionamento e gerenciamento do dataset: O Roboflow possibilitou a criação de diferentes versões da base de dados, permitindo acompanhar a evolução do projeto desde conjuntos menores até bases mais robustas. Esse controle de versões foi essencial para avaliar o impacto do aumento amostral na acurácia do modelo.
* Pré-processamento e padronização: Todas as imagens foram redimensionadas para 640x640 pixels, garantindo uniformidade e compatibilidade com o YOLOv11. Além disso, foram aplicadas correções de iluminação e ajustes de proporção, assegurando maior consistência entre os exemplos de treino, validação e teste.
* Anotação supervisionada: A plataforma permitiu a criação de \*bounding boxes\* e a atribuição de rótulos às imagens, classificando cada objeto como “vidro” ou “plástico”. Esse processo de anotação é fundamental para o treinamento supervisionado, pois fornece ao modelo informações precisas sobre a classe e a localização dos objetos.
* Data Augmentation: Foram aplicadas técnicas de aumento de dados, como rotações, espelhamentos, recortes e variações de brilho. Essas transformações ampliaram a diversidade da base, reduzindo o risco de sobreajuste (overfitting) e aumentando a capacidade de generalização do modelo em cenários reais.

O uso do Roboflow, portanto, não apenas facilitou a preparação da base de dados, mas também garantiu maior qualidade e confiabilidade ao processo de treinamento. Sua integração com o YOLOv11 tornou possível estruturar um pipeline eficiente, desde a coleta das imagens até a geração de modelos capazes de identificar e classificar resíduos recicláveis com precisão.

1. ARQUITETURA
   1. Infraestrutura em Nuvem do Projeto

A infraestrutura do projeto foi desenvolvida utilizando os serviços da AWS (Amazon Web Services), adotando o modelo de *Infrastructure as a Service* (IaaS). A arquitetura foi projetada para operar de forma assíncrona.

**Figura 4** – Fórmula mAP

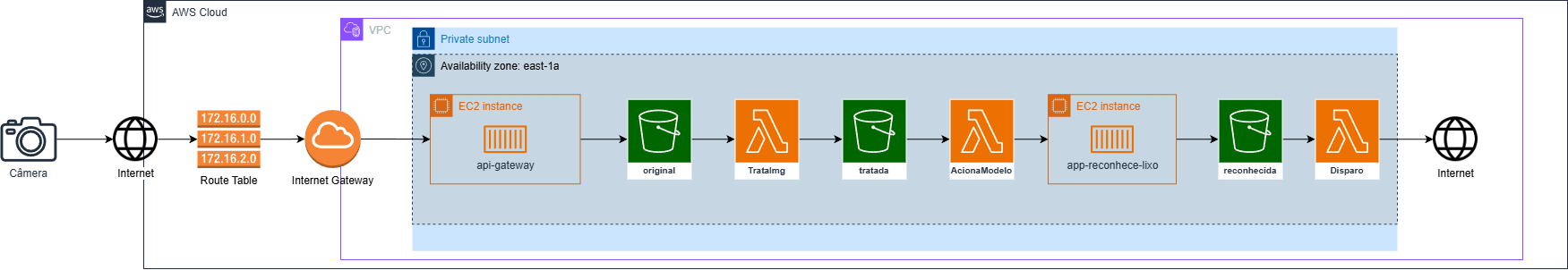


Figura 4. Arquitetura de infraestrutura do projeto – Fonte: Autoral

* 1. ****Monitoramento com Amazon CloudWatch****

**O serviço Amazon CloudWatch é utilizado para monitorar as instâncias EC2 e as funções Lambda, permitindo o acompanhamento de métricas, logs e alertas em tempo real, o que contribui para a observabilidade e manutenção do sistema.**

* 1. Segurança e Rede (VPC, Subnets e Internet Gateway**)**

Toda a infraestrutura está organizada em uma **VPC (Virtual Private Cloud)**, com uma subnet privada na região east-1a. O acesso à internet é feito por meio de um **Internet Gateway**, enquanto as tabelas de roteamento controlam o tráfego de rede de forma segura.

* 1. ****Envio da Imagem pelo Usuário****

**O processo de identificação inicia-se com o envio de um arquivo de imagem, com tamanho máximo de 5MB, por parte do usuário, e um endereço webhook que receberá o retorno. Esse arquivo é recebido por uma aplicação hospedada em uma instância EC2 (api-gateway) e armazenado em um bucket S3 chamado original, destinado ao armazenamento das imagens sem qualquer tipo de tratamento prévio.**

* 1. Tratamento da Imagem

Após o armazenamento inicial, uma função Lambda chamada TrataImg é acionada automaticamente. Essa função realiza o redimensionamento da imagem, com o objetivo de otimizar seu tamanho para os próximos estágios do fluxo de processamento. A imagem tratada é então salva em um segundo bucket S3, denominado tratada.

* 1. Acionamento do Modelo de Reconhecimento

Em seguida, a função Lambda AcionaModelo é ativada, sendo responsável por enviar a imagem tratada para um endpoint interno. Nesse endpoint, uma instância EC2 (app-reconhece-lixo) executa o modelo de reconhecimento, que identifica o conteúdo da imagem.

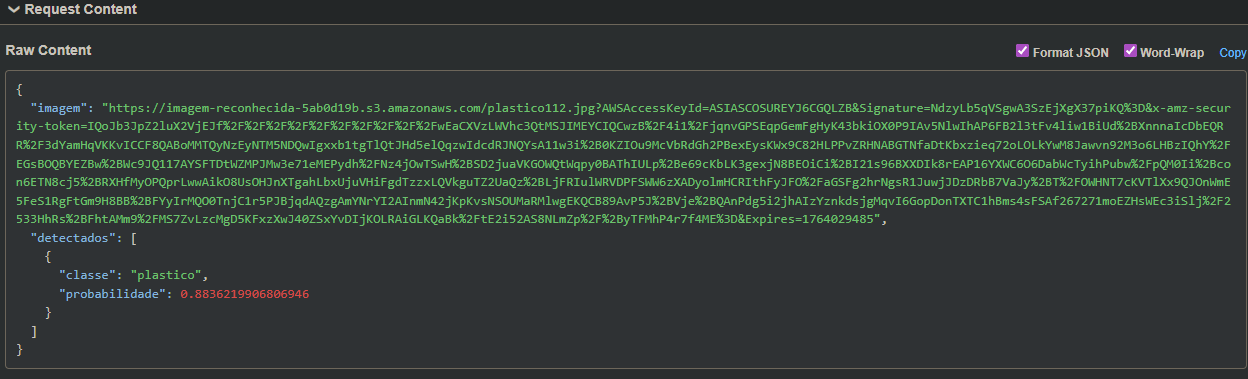
* 1. ****Armazenamento do Resultado Reconhecido****

Após o reconhecimento, a imagem tratada, juntamente com os resultados obtidos pelo modelo, é armazenada em um terceiro bucket S3, chamado reconhecida.

* 1. ****Retorno ao Usuário via Webhook****

**Por fim, a função Lambda Disparo é acionada para enviar ao remetente, via webhook, um JSON contendo o link pré-assinado da imagem tratada e outro JSON com as análises realizadas pelo modelo de reconhecimento.**

**Figura 4** – Exemplo prático do retorno da análise



Fonte: Autoral

* 1. Análise Comparativa dos Algoritmos

O código de tratamento implementado na função Lambda *TrataImg* foi projetado para otimizar o custo-benefício do sistema. O objetivo principal é reduzir o tamanho das imagens recebidas, garantindo menor consumo de largura de banda e maior eficiência no processamento subsequente. Essa etapa é crítica, pois influencia diretamente o tempo de resposta do modelo de reconhecimento e o custo de execução em ambiente serverless.

Para essa finalidade, foram desenvolvidos dois algoritmos distintos de tratamento de imagem, denominados **Algoritmo 1 (Byte)** e **Algoritmo 2 (Donwload)**. Ambos realizam operações de redimensionamento e compressão, mas diferem em sua abordagem de otimização de recursos computacionais.

Os testes realizados em ambiente AWS Lambda consideraram métricas como duração da execução (ms), memória alocada (GB), tamanho da imagem (KB), GB-segundos consumidos e custo por execução (USD).  
*Fonte:* [*https://aws.amazon.com/pt/lambda/pricing/*](https://aws.amazon.com/pt/lambda/pricing/) *- Acessado em: novembro de 2025*

Algoritmo 1 (Byte):

* Executou em tempos médios entre 512–618 ms com 0,5 GB de memória.
* Apresentou os menores custos por execução, variando entre 4,27×10⁻⁶ e 5,15×10⁻⁶ USD.
* Demonstrou maior eficiência em cenários com imagens de até ~755 KB, mantendo baixo consumo de recursos.

Algoritmo 2 (Download):

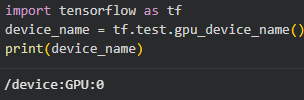
* Obteve tempos de execução semelhantes, mas com maior variação de memória e imagens maiores (até 7061 KB).
* Os custos por execução foram superiores, chegando a 9,27×10⁻⁶ USD em imagens médias e até 5,0×10⁻⁵ USD em imagens grandes (~22 MB).
* Apesar de robusto, mostrou-se menos eficiente em termos de custo-benefício.

Com base nos resultados, o Algoritmo 1 (Byte) foi considerado o mais adequado para implementação no Lambda *TrataImg*. Ele apresentou o melhor equilíbrio entre tempo de execução, uso de memória e custo por execução, especialmente em cenários de imagens menores, que representam a maioria dos casos de uso do sistema. Essa escolha garante maior escalabilidade e menor impacto financeiro em larga escala, além de manter a qualidade necessária para o processamento posterior no modelo de reconhecimento.

1. IMPLEMENTAÇÃO DA SOLUÇÃO

Utilizando o ambiente do Google Colab, fizemos a construção do modelo de classificação. Como mencionado, nosso primeiro modelo contou com 64 imagens. Para que o modelo rode sem muitos problemas, primeiro foi preciso habilitar a GPU no ambiente, após isso importamos a biblioteca tensorflow para verificar se o ambiente está de fato com a GPU ativada (Figura 5).

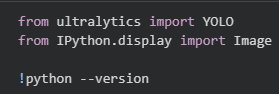
**Figura 5** – Habilitando GPU no Google Colab



Fonte: Autoral

O próximo passo é importar a biblioteca do Yolo e do Image (Figura 6), para que possamos imprimir as imagens analisada.

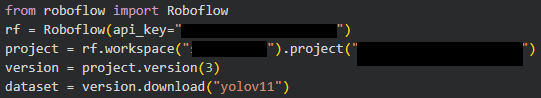
**Figura 6** – Importando as bibliotecas Yolo e Image



Fonte: Autoral

A última biblioteca que deve ser importada é a do RoboFlow, para que possamos baixar nosso dataset (Figura 7).

**Figura 7** – Importando a biblioteca Roboflow e baixando dataset



Fonte: Autoral

* 1. Primeira versão do modelo:
     1. Treinamento do modelo

Com a base de dados preparada e sem dependências adicionais, iniciamos o treinamento do modelo utilizando o yolo (Figura 8). O processo foi configurado com os seguintes parâmetros:

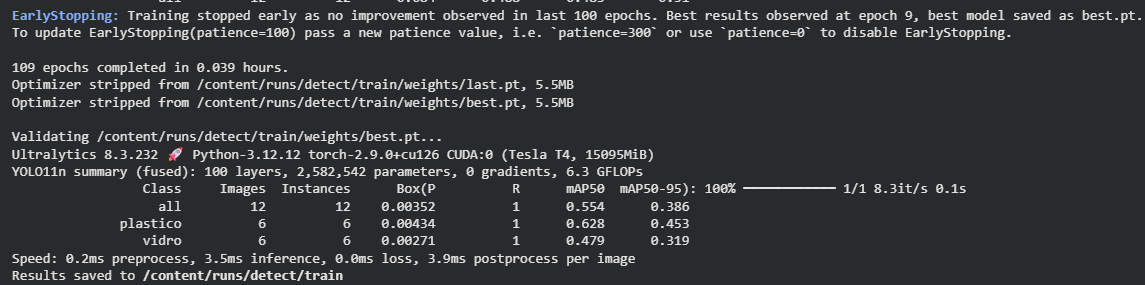
* Task: Detecção – Define que o modelo será utilizado para a tarefa de detecção de objetos. Nesse modo, o YOLOv11 não apenas classifica os objetos (vidro ou plástico), mas também identifica sua localização dentro da imagem por meio de caixas delimitadoras (bounding boxes).
* Mode: Train – Indica que o modelo está em fase de treinamento supervisionado, utilizando a base de dados anotada. Durante essa etapa, os pesos da rede neural são ajustados para minimizar erros de classificação e localização, aprendendo a reconhecer padrões visuais dos resíduos.
* Data: Especifica o arquivo de configuração da base de dados, que contém informações sobre as classes (vidro e plástico), caminhos das imagens e estrutura dos conjuntos de treino, validação e teste.
* Model: Define o modelo pré-treinado a ser utilizado como ponto de partida. A versão YOLOv11n (nano) é otimizada para maior velocidade e menor consumo computacional, sendo adequada para experimentos iniciais e ambientes com recursos limitados.
* Epochs: 500 – Estabelece o número máximo de ciclos de treinamento. Cada epoch corresponde a uma passagem completa por todo o conjunto de dados. Embora configurado para 500 épocas, o mecanismo de EarlyStopping interrompe o processo caso não haja melhoria após determinado número de ciclos, evitando sobreajuste (overfitting).
* Image Size: 640 – Define o tamanho de entrada das imagens (640x640 pixels). Esse redimensionamento garante padronização e compatibilidade com a arquitetura do YOLOv11, além de equilibrar qualidade visual e desempenho computacional.

**Figura 8** – Rodando o yolo no modo de treinamento.



Fonte: Autoral

**Figura 9** – Fim do treinamento



Fonte: Autoral

* + 1. Análise das métricas geradas

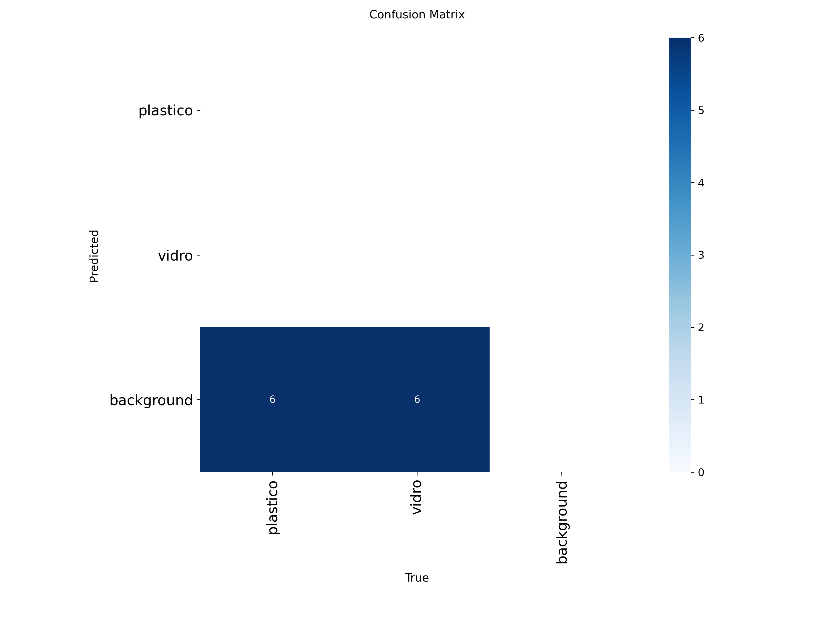
Durante o treinamento (Figura 9), o mecanismo de EarlyStopping interrompeu a execução após 109 épocas, uma vez que não foram observadas melhorias nos últimos 100 ciclos. O melhor desempenho foi registrado na 9ª época, sendo o modelo correspondente salvo como best.pt.

O treinamento foi concluído em aproximadamente 2 minutos e 20 segundos. Os arquivos de pesos gerados (last.pt e best.pt) tiveram o otimizador removido, resultando em modelos compactos de 5,5 MB cada.

Na etapa de validação, o YOLOv11n apresentou a seguinte configuração: 100 camadas, 2.582.542 parâmetros e 6,3 GFLOPs. Os resultados obtidos foram:

* Conjunto geral (all): 12 imagens, 12 instâncias, Box Precision = 0,0035, Recall = 1, mAP50 = 0,554, mAP50-95 = 0,386.
* Classe Plástico: 6 imagens, 6 instâncias, Box Precision = 0,0043, Recall = 1, mAP50 = 0,628, mAP50-95 = 0,453.
* Classe Vidro: 6 imagens, 6 instâncias, Box Precision = 0,0027, Recall = 1, mAP50 = 0,479, mAP50-95 = 0,319.

**Figura 10 –** Matriz de confusão



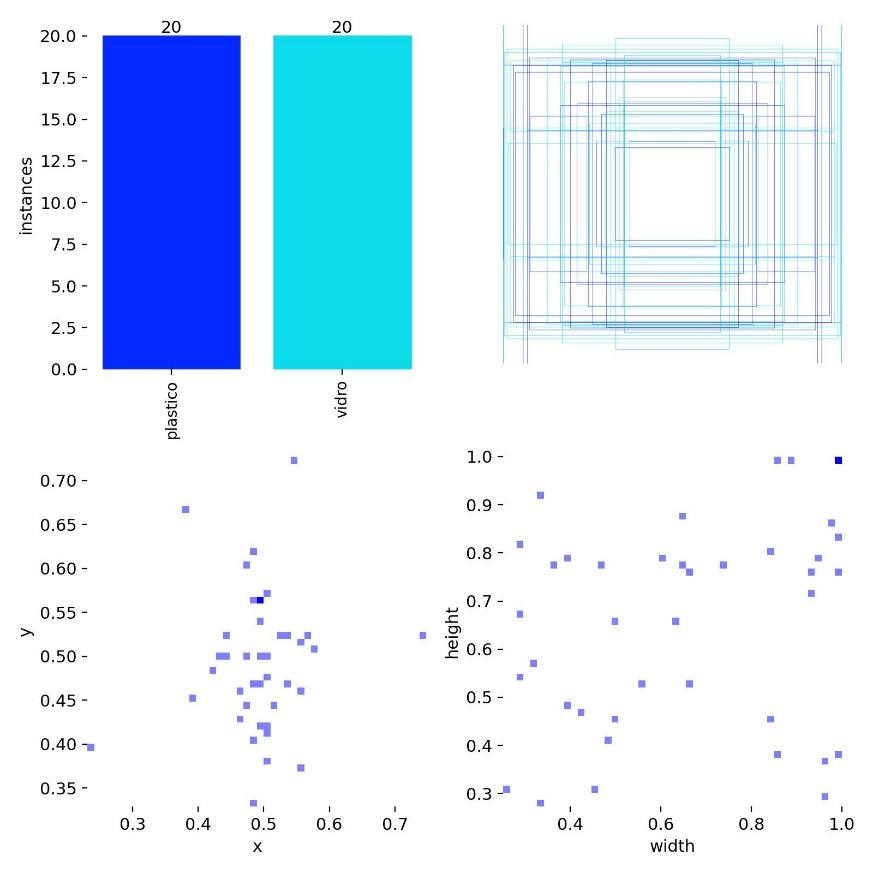
Fonte: Autoral

A matriz de confusão gerada após o treinamento do modelo (Figura 10) permite avaliar o desempenho do classificador em relação às classes plástico, vidro e background. Ela apresenta, para cada classe verdadeira, a quantidade de vezes em que o modelo acertou ou errou a predição.

Na matriz obtida, todas as amostras de plástico e vidro foram classificadas como background. Isso significa que o modelo não reconheceu corretamente nenhum objeto dessas classes, mesmo que as métricas de validação tenham apresentado valores moderados para mAP. Como cada classe possui seis instâncias, o resultado indica que o modelo previu background em todas as doze imagens, sem identificar objetos nas categorias esperadas.

Esse comportamento sugere que o modelo ainda não aprendeu padrões visuais suficientes para distinguir plástico e vidro, o que pode ocorrer em cenários de dataset reduzido, imagens pouco variadas ou perda de informação durante o processo de anotação. A matriz de confusão evidencia de forma clara que o modelo está excessivamente enviesado para a classe background e reforça a necessidade de ajustes no conjunto de dados, aumento de épocas, reequilíbrio entre classes ou aprimoramento do processo de treinamento.

**Figura 11** – Gráfico da distribuição das anotações.



Fonte: Autoral

A imagem acima (Figura 11) apresenta diferentes visualizações das anotações realizadas no conjunto de dados utilizado para o treinamento do modelo. O objetivo é verificar a qualidade, a consistência e a distribuição espacial das bounding boxes referentes às classes plástico e vidro.

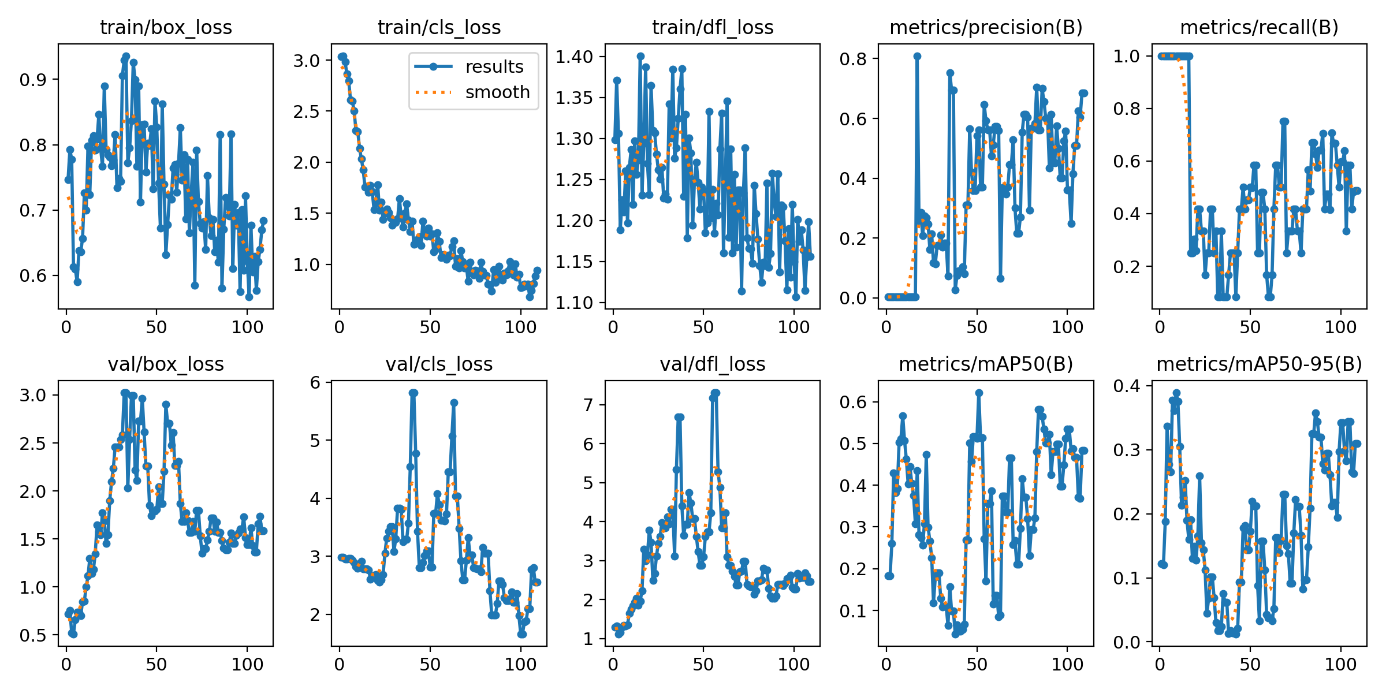
No primeiro gráfico (canto superior esquerdo), observa-se a contagem de instâncias por classe. Cada categoria possui vinte anotações, o que revela um equilíbrio numérico entre elas. Isso é positivo, já que modelos de detecção podem apresentar viés quando há desproporção entre as classes. Apesar do equilíbrio, a quantidade total ainda é pequena para um modelo baseado em YOLO, o que limita a capacidade de generalização.

O segundo painel (canto superior direito) mostra o conjunto de caixas delimitadoras sobrepostas. Essa visualização permite identificar possíveis padrões de posição e tamanho, bem como detectar anotações fora do esperado. A grande concentração de caixas na área central indica que quase todas as imagens possuem objetos localizados no mesmo ponto, reduzindo a variabilidade espacial do dataset. A falta de diversidade nesse aspecto prejudica o aprendizado do modelo, que tende a associar a presença de objetos a uma região específica.

Os gráficos inferiores detalham a dispersão dos valores de x e y (centro das caixas) e de width e height (largura e altura normalizadas). Há baixa variação na posição horizontal e vertical dos objetos. Da mesma forma, muitas caixas apresentam dimensões semelhantes, o que sugere que o dataset não contempla objetos de tamanhos variados. Essa limitação reduz a robustez do modelo e dificulta a identificação de objetos fora desse padrão restrito.

A análise conjunta dessas visualizações demonstra que, embora o conjunto esteja anotado corretamente e equilibrado em número de instâncias, a falta de diversidade nas posições e dimensões das caixas é um fator crítico. Isso contribui diretamente para o baixo desempenho observado na matriz de confusão e mostra a necessidade de ampliar o dataset com imagens mais variadas ou aplicar técnicas de data augmentation que estimulem maior variação espacial e dimensional.

**Figura 12** – Métricas de Treinamento e Validação do Modelo



Fonte: Autoral

A imagem (Figura 12) apresenta a evolução das principais métricas registradas durante o treinamento do modelo. Os gráficos permitem identificar como o modelo aprendeu ao longo das épocas e revelam limitações diretamente relacionadas ao tamanho reduzido e à baixa diversidade do dataset.

Os gráficos superiores correspondem às perdas (losses) de treinamento. A métrica train/box\_loss, responsável pela qualidade da regressão das bounding boxes, apresenta oscilações consideráveis e não demonstra tendência clara de estabilização. Isso indica que o modelo teve dificuldade em aprender padrões espaciais consistentes, o que coincide com a baixa variabilidade das anotações evidenciada anteriormente. A métrica train/cls\_loss, que mede o erro na classificação das classes, apresenta queda inicial acentuada, mas mantém instabilidade ao longo do processo, sugerindo que o modelo memorizou rapidamente as poucas amostras disponíveis e não desenvolveu generalização suficiente. A métrica train/dfl\_loss, relacionada à distribuição da localização dos pontos de ancoragem, segue comportamento semelhante, com quedas moderadas, porém com forte ruído, evidenciando dificuldade do modelo em refinar as previsões.

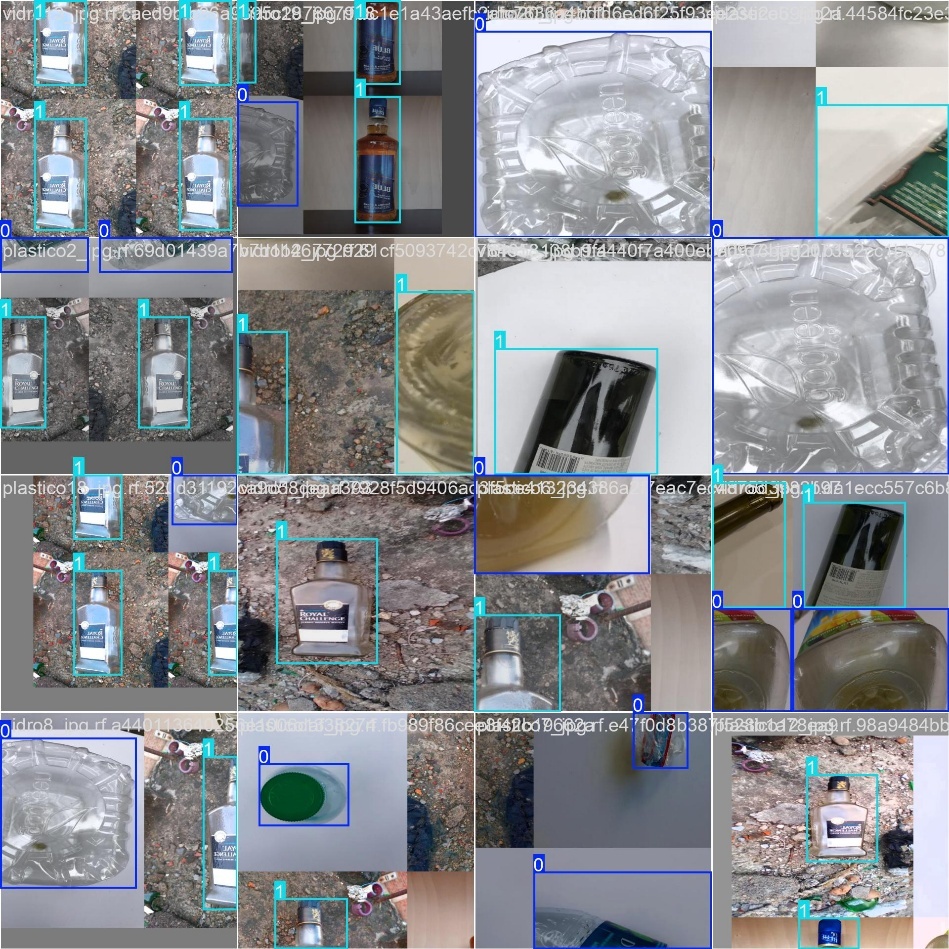
Os gráficos de precisão e recall mostram comportamento atípico. No início das épocas, ambas as métricas atingem valores próximos de 1, o que normalmente indicaria excelente desempenho, mas nesse caso representa overfitting imediato. Com o avanço do treinamento, ambas caem abruptamente e passam a oscilar, confirmando que o modelo perdeu capacidade de generalização quando começou a ajustar demais os pesos às poucas amostras disponíveis.

Na segunda linha estão as métricas de validação, que refletem o desempenho do modelo sobre dados não vistos durante o treinamento. As perdas de validação apresentam valores significativamente mais altos do que as perdas de treino. Esse distanciamento é um sinal clássico de overfitting. Além disso, as curvas possuem comportamento irregular e sem convergência clara, reforçando a dificuldade do modelo em aprender padrões robustos com o conjunto limitado de imagens. As métricas de mAP50 e mAP50-95 também apresentam grande variabilidade e não mostram tendência de crescimento consistente, evidenciando que o modelo não adquiriu capacidade sólida de detecção.

A análise conjunta dos gráficos demonstra que o modelo foi incapaz de atingir estabilidade durante o treinamento. As oscilações severas, a discrepância entre treino e validação, o comportamento instável das métricas de precisão e recall e a ausência de convergência nas métricas de mAP confirmam que o dataset possui volume e diversidade insuficientes. Esses resultados reforçam a necessidade de ampliação da base de imagens e de anotações mais variadas, a fim de permitir que o modelo aprenda padrões generalizáveis e obtenha desempenho adequado em aplicações reais.

* + 1. Inferência das imagens

**Figura 13** – Visualização do Train Batch



Fonte: Autoral

Durante o processo de treinamento do modelo YOLOv11, foi gerado um *Train Batch* que representa visualmente como o modelo interpreta e processa as imagens anotadas da base de dados. Essa visualização é composta por uma amostra de imagens do conjunto de treino, sobre as quais são aplicadas as *bounding boxes* e os rótulos correspondentes às classes definidas: **plástico** e **vidro**.

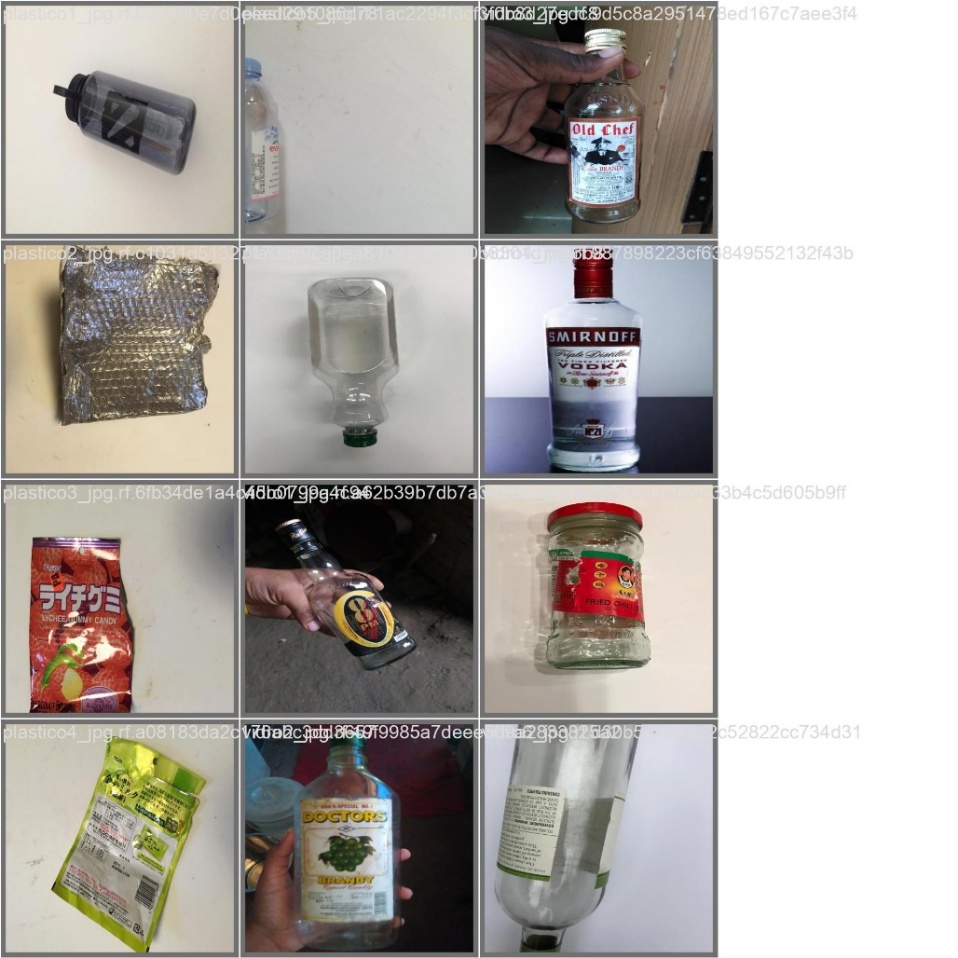
Cada imagem do batch apresenta:

* **Objetos identificados e anotados manualmente**, com caixas delimitadoras que circunscrevem os resíduos recicláveis.
* **Rótulos de classe**, como “plastico1”, “vidro2”, entre outros, que indicam a categoria atribuída a cada objeto.
* **Diversidade de cenários**, incluindo fundos variados como pisos, prateleiras e superfícies externas, o que contribui para a robustez do modelo ao lidar com diferentes contextos visuais.
* **Numeração e codificação**, que podem representar o índice da instância ou o nível de confiança da anotação.

Essa etapa é fundamental para validar visualmente a qualidade das anotações e garantir que o modelo esteja recebendo informações consistentes durante o treinamento. A correta aplicação das *bounding boxes* e dos rótulos é essencial para que o modelo aprenda a distinguir os padrões visuais de cada classe, contribuindo diretamente para o desempenho nas fases de validação e teste.

Além disso, o Train Batch permite identificar possíveis inconsistências na anotação, como objetos não rotulados, sobreposição de caixas ou erros de classificação, possibilitando ajustes antes da consolidação do modelo final.

**Figura 14** – Amostragem Visual



Fonte: Autoral

A imagem apresentada (Figura 14) representa uma amostragem visual da base de dados utilizada para o treinamento do modelo YOLOv11. Ela é composta por 12 fotografias organizadas em grade, contendo diferentes tipos de embalagens e recipientes descartados, pertencentes às classes **plástico** e **vidro**. Essa diversidade é essencial para garantir que o modelo aprenda a reconhecer padrões visuais em diferentes contextos e condições de iluminação.

Entre os itens presentes estão:

* Garrafas de vidro de bebidas alcoólicas
* Frascos plásticos de cosméticos e alimentos
* Embalagens plásticas de doces e condimentos
* Frascos de vidro com tampas metálicas ou plásticas
* Itens manipulados por mãos humanas, simulando cenários reais de coleta e triagem

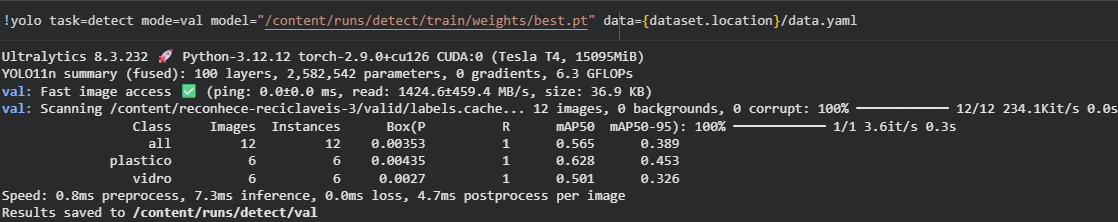
Essa amostragem evidencia a heterogeneidade da base, com variações de cor, forma, textura, rótulo e fundo. A presença de objetos em diferentes posições e ângulos contribui para a robustez do modelo, permitindo que ele generalize melhor em situações reais de triagem automatizada.

Além disso, a inclusão de rótulos e *bounding boxes* sobre os objetos garante que o modelo receba informações precisas durante o treinamento supervisionado. Essa etapa é fundamental para que o sistema aprenda a distinguir corretamente entre plástico e vidro, mesmo quando os objetos apresentam características visuais semelhantes.

* + 1. Validação e ajustes no modelo

Após o treinamento da primeira versão do modelo, foi realizada a etapa de **validação** (Figura 15)

**Figura 15** – Validando modelo



Fonte: Autoral

Apesar do Recall perfeito (1.0), os valores baixos de Box Precision e mAP indicam que o modelo ainda apresenta dificuldades em localizar com precisão os objetos nas imagens. Isso pode estar relacionado à qualidade da anotação, à limitação da base de dados ou à complexidade visual dos resíduos. A validação reforça a necessidade de ajustes na base e na arquitetura para melhorar a acurácia espacial e semântica do modelo.

**Figura 16** – Fazendo predição no conjunto de teste

****

Fonte: Autoral

A predição foi realizada sobre o conjunto de teste composto por 12 imagens, utilizando o modelo treinado **best.pt (Figura 16)**.

Desempenho por imagem:

* **Imagens processadas:** 12
* **Detecções realizadas:** Nenhuma

**Tempo médio por imagem:**

* Pré-processamento: 1.5 ms
* Inferência: 8.2 ms
* Pós-processamento: 0.5 ms
* Total: ~10.2 ms por imagem

Em todas as imagens, o modelo **não realizou nenhuma detecção**, indicando que nenhum objeto foi identificado com confiança suficiente para ser considerado válido. Isso inclui imagens contendo garrafas plásticas, frascos de vidro e embalagens diversas, que deveriam ser reconhecidas como pertencentes às classes “plástico” ou “vidro”.

A ausência de detecções sugere que o modelo ainda apresenta **baixa capacidade de generalização** frente ao conjunto de teste. Possíveis causas incluem:

* Diferenças visuais significativas entre as imagens de treino e teste
* Base de dados limitada ou pouco variada
* Anotações inconsistentes ou insuficientes
* Limiar de confiança elevado para o estágio atual do modelo

**Figura 17** – Resultado da predição

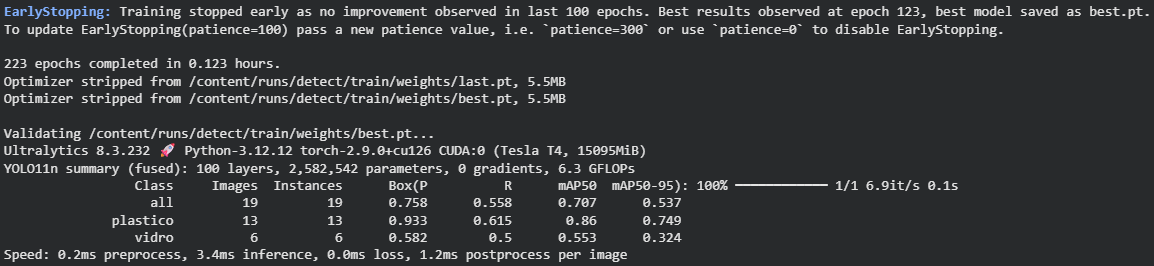


Fonte: Autoral

* 1. Segunda versão do modelo:
     1. Treinamento do modelo

Com a ampliação do dataset para **116 imagens** (incremento de 52 imagens em relação à versão anterior), iniciou‑se o treinamento da segunda versão do modelo. O processo registrou as seguintes informações operacionais (Figura 18).

**Figura 18 –** Fim do treinamento



Fonte: Autoral

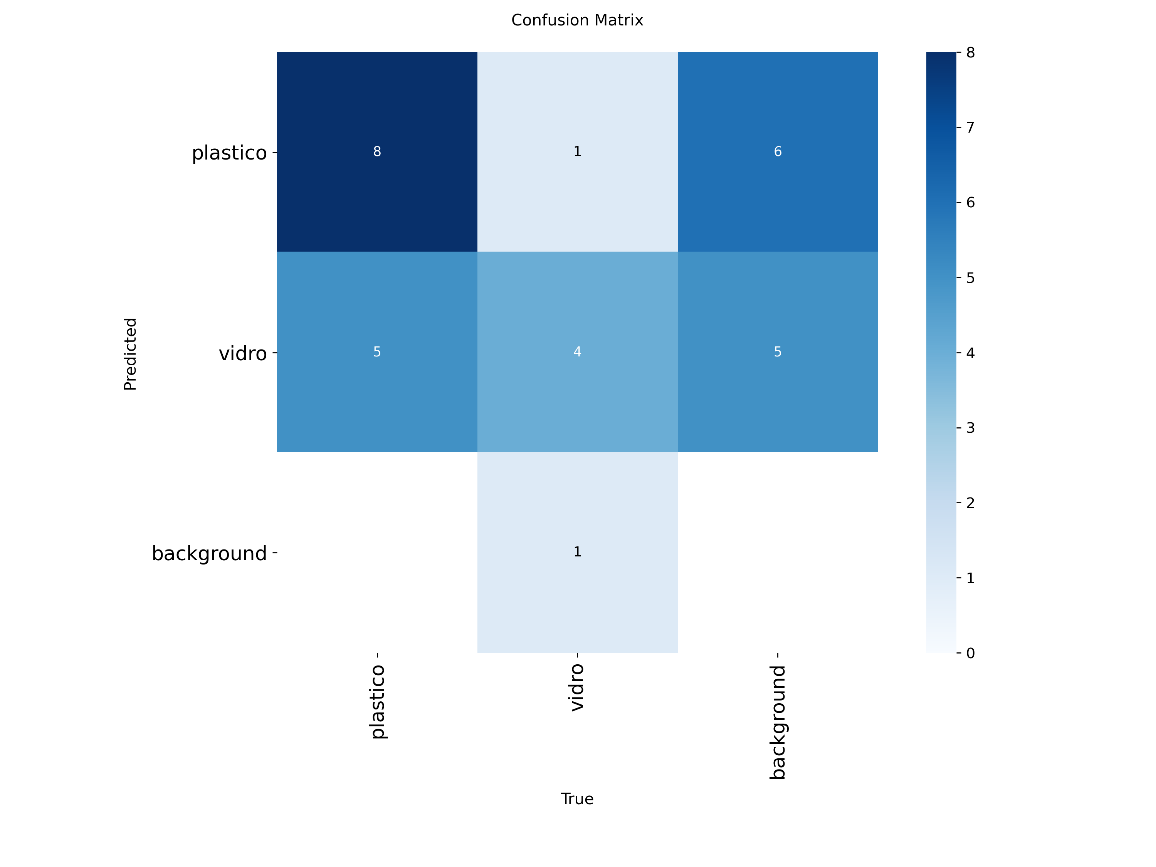
* + 1. Análise das métricas geradas

Durante o treinamento (Figura 18), o mecanismo de EarlyStopping interrompeu a execução após não serem observadas melhorias nos últimos 100 ciclos. O melhor desempenho foi registrado na 123ª época, sendo o modelo correspondente salvo como best.pt. O treinamento foi concluído em aproximadamente 7 minutos e 23 segundos. Os arquivos de pesos gerados (last.pt e best.pt) tiveram o otimizador removido, resultando em modelos compactos de 5,5 MB cada.

Na etapa de validação, o YOLOv11n apresentou a seguinte configuração: 100 camadas, 2.582.542 parâmetros e 6,3 GFLOPs. Os resultados obtidos foram:

* Conjunto geral (all): 19 imagens, 19 instâncias, Box Precision = 0,758, Recall = 0,558, mAP50 = 0,707, mAP50-95 = 0,537.
* Classe Plástico: 13 imagens, 13 instâncias, Box Precision = 0,933, Recall = 0,615, mAP50 = 0,860, mAP50-95 = 0,749.
* Classe Vidro: 6 imagens, 6 instâncias, Box Precision = 0,582, Recall = 0,500, mAP50 = 0,553, mAP50-95 = 0,324.

**Figura 19 –** Matriz de confusão



Fonte: Autoral

A matriz de confusão obtida na segunda versão do modelo (Figura 19) apresenta a distribuição das predições em relação às classes verdadeiras **plastico**, **vidro** e **background**. Os valores observados foram: predições como **plastico** correspondem a 8 verdadeiros plásticos, 1 vidro e 6 backgrounds; predições como **vidro** correspondem a 5 plásticos, 4 vidros e 5 backgrounds; predições como **background** correspondem a 0 plásticos, 1 vidro e 0 backgrounds.

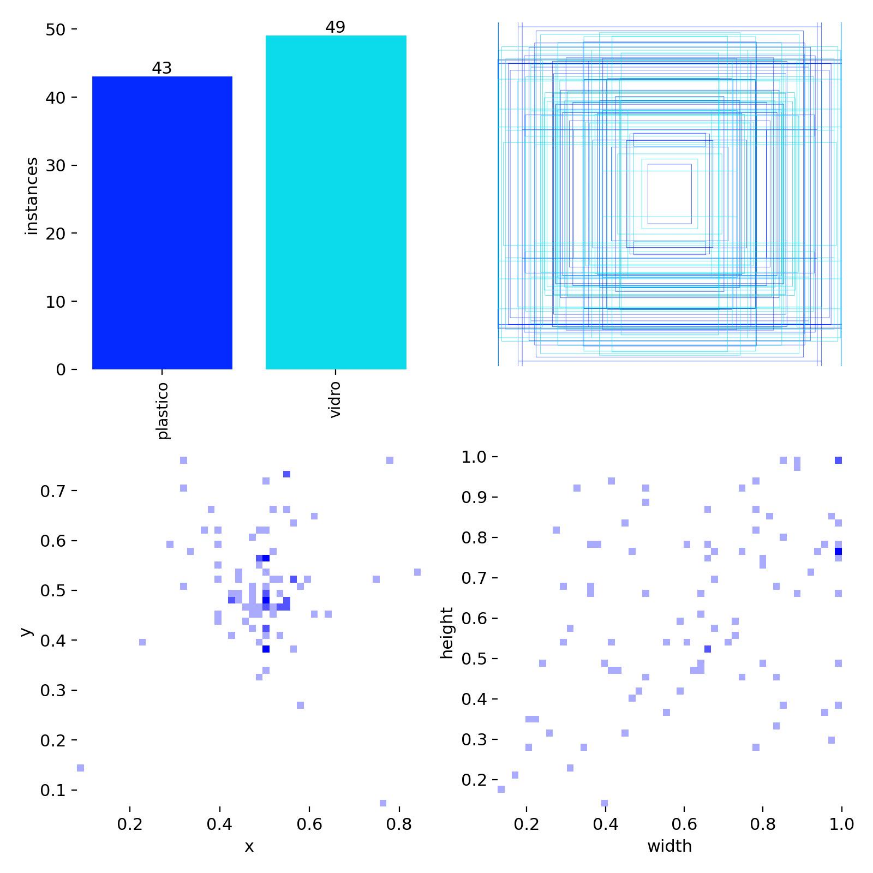
Esses resultados indicam os seguintes comportamentos do detector:

* **Melhor reconhecimento de plástico:** a maior parte das instâncias verdadeiras de plástico (8 de 13) foi corretamente prevista como plástico, refletindo o ganho de representatividade dessa classe no dataset.
* **Confusão entre classes alvo:** há trocas relevantes entre plástico e vidro (5 instâncias de plástico previstas como vidro; 1 instância de vidro prevista como plástico), sinalizando que o modelo ainda tem dificuldade em discriminar características visuais semelhantes entre as duas classes.
* **Falsos positivos sobre background:** o modelo frequentemente prevê objetos (plástico ou vidro) em imagens ou regiões que são, na verdade, background (6 predições de plástico e 5 de vidro sobre verdadeiros backgrounds). Isso aponta para um excesso de sensibilidade a padrões de fundo ou para anotações negativas insuficientes no treinamento.
* **Predição de background rara:** a classe background foi pouco prevista, o que indica que o modelo tende a preferir classes alvo em vez de classificar corretamente regiões sem objeto.

**Implicações práticas:** a matriz confirma avanços em relação à primeira versão (onde predominava a predição de background), mas evidencia dois problemas principais:

* persistente confusão entre plástico e vidro;
* elevada taxa de falsos positivos sobre background.

**Figura 20** – Gráfico da distribuição das anotações.



Fonte: Autoral

A imagem acima (Figura 20) apresenta diferentes visualizações das anotações realizadas no conjunto de dados utilizado para o treinamento da segunda versão do modelo. O objetivo é verificar a qualidade, a consistência e a distribuição espacial das *bounding boxes* referentes às classes **plástico** e **vidro**.

No primeiro gráfico (canto superior esquerdo) observa‑se a contagem de instâncias por classe: **plástico = 43** e **vidro = 49**, o que revela um balanceamento numérico razoável entre as categorias. Esse equilíbrio é positivo, pois reduz o risco de viés por desproporção; contudo, apesar do aumento em relação à versão anterior, o volume ainda é moderado para um detector baseado em YOLO, limitando a capacidade de generalização em cenários muito variados.

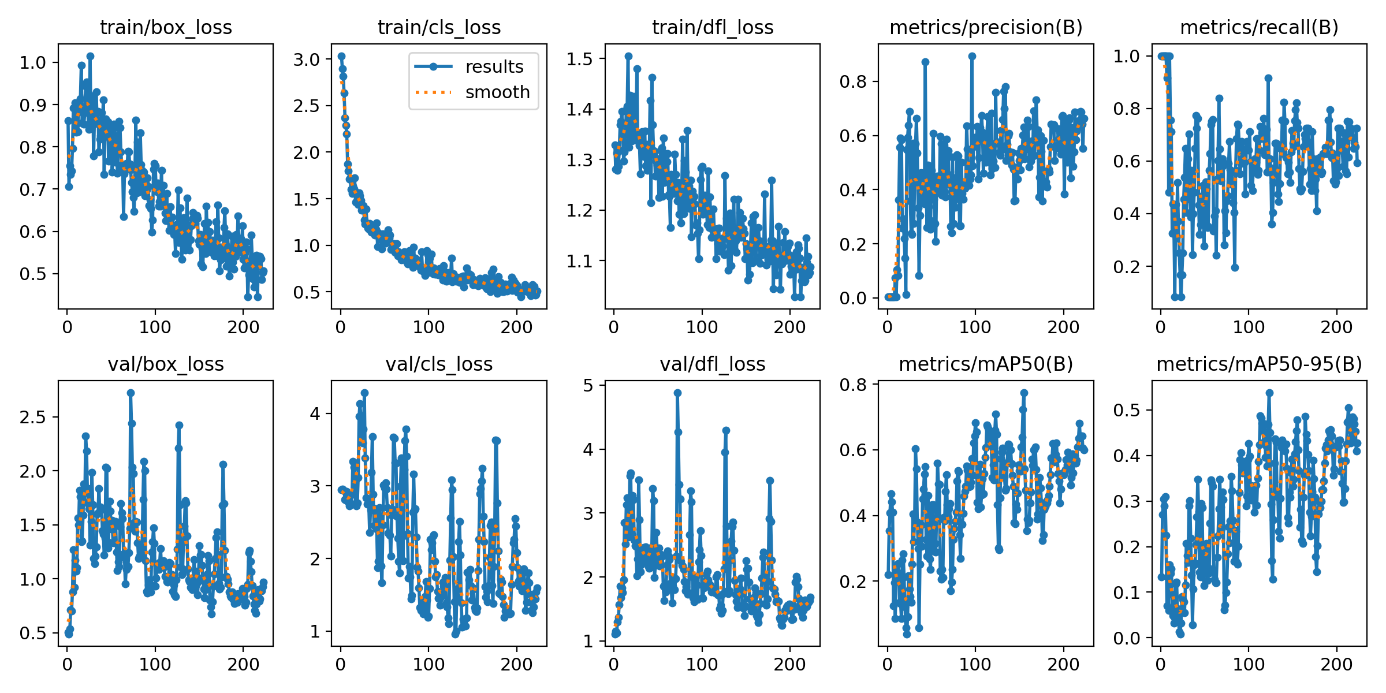
O segundo painel (canto superior direito) mostra o conjunto de caixas delimitadoras sobrepostas. Assim como na versão anterior, há uma concentração significativa de caixas em regiões centrais das imagens, indicando que muitos objetos foram fotografados em posições semelhantes. Essa concentração reduz a variabilidade espacial do dataset e tende a ensinar ao modelo um viés posicional — ou seja, a expectativa de encontrar objetos sempre em áreas centrais.

Os gráficos inferiores detalham a dispersão dos valores de **x** e **y** (centro das caixas) e de **width** e **height** (largura e altura normalizadas). Observa‑se maior densidade de centros próximos ao meio da imagem e uma predominância de caixas com dimensões semelhantes. Embora exista agora uma maior quantidade de anotações, a variação espacial e dimensional permanece limitada, o que prejudica a robustez do detector frente a objetos deslocados, parcialmente visíveis ou de tamanhos fora do padrão predominante.

A análise conjunta dessas visualizações demonstra que, apesar do avanço quantitativo e do balanceamento entre classes, a **falta de diversidade nas posições e dimensões das caixas** continua sendo um fator crítico. Essa limitação contribui para erros de classificação e para falsos positivos/negativos observados nas métricas e na matriz de confusão, indicando a necessidade de ampliar o dataset com imagens mais variadas e aplicar técnicas de *data augmentation* que promovam maior variação espacial e dimensional.

**Comparação com a versão anterior:** em relação à primeira versão, houve ganho claro em representatividade (de 20 anotações por classe para 43/49), o que se refletiu em melhora das métricas globais, especialmente para plástico. No entanto, a limitação estrutural — concentração central das caixas e homogeneidade de tamanhos — persiste. Em suma, a segunda versão corrige parcialmente o problema de escassez de dados, mas ainda requer aumento da diversidade espacial e dimensional para alcançar generalização robusta.

**Figura 21** – Métricas de Treinamento e Validação do Modelo



Fonte: Autoral

A imagem (Figura 21) apresenta a evolução das principais métricas registradas durante o treinamento da segunda versão do modelo. Os gráficos permitem acompanhar como o modelo aprendeu ao longo das 223 épocas e evidenciam avanços em relação à versão anterior, bem como limitações remanescentes relacionadas à variabilidade do dataset e ao desbalanceamento de classes.

Os gráficos superiores correspondem às perdas (losses) de treinamento. A métrica train/box\_loss, responsável pela qualidade da regressão das bounding boxes, mostra tendência geral de redução ao longo das épocas, indicando que o modelo melhorou a capacidade de ajustar as caixas delimitadoras; contudo, ainda apresenta ruído e pequenas oscilações, sinalizando refinamento incompleto da regressão espacial. A train/cls\_loss, que mede o erro de classificação, decresce de forma mais consistente do que na versão anterior, sugerindo aprendizado mais estável das características discriminativas; ainda assim, há flutuações pontuais que indicam sensibilidade a amostras difíceis. A train/dfl\_loss, relacionada à distribuição da localização dos pontos de ancoragem, também apresenta queda gradual, com menor amplitude de ruído comparada à primeira versão, o que corrobora a melhoria na convergência do treinamento.

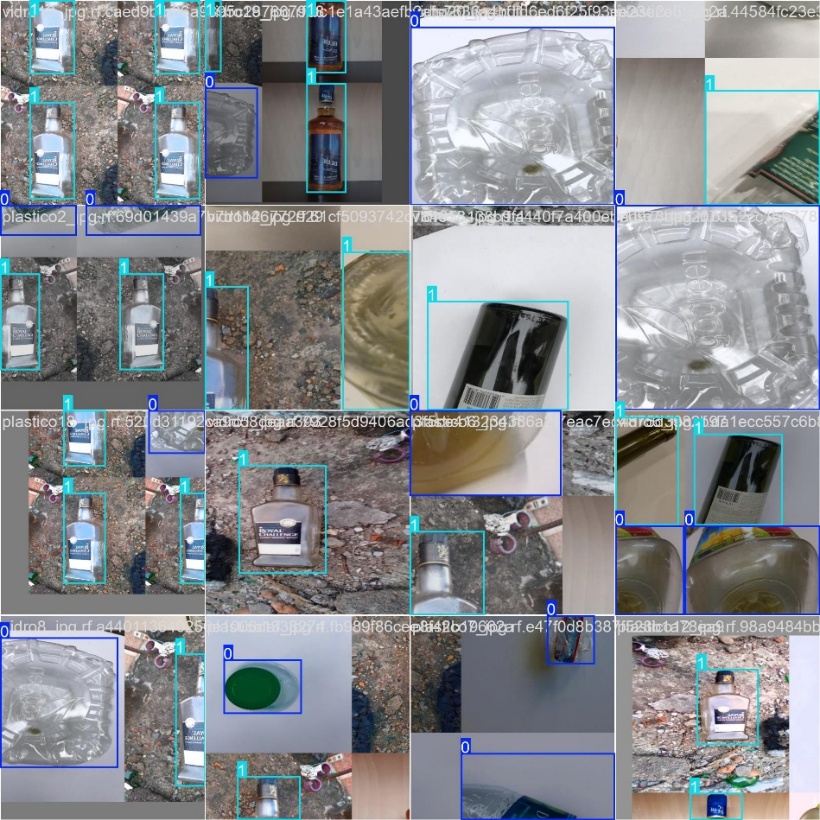
Os gráficos de precisão e recall exibem comportamento oscilante, porém com tendência de melhora ao longo do processo. Ao contrário do overfitting imediato observado na primeira versão, aqui observa‑se um aumento progressivo das métricas de precisão e mAP, indicando que o modelo aprendeu representações mais úteis. O recall, embora tenha melhorado, mantém valores moderados e ainda oscila, evidenciando que o detector perde algumas instâncias reais, especialmente para a classe vidro.

Na segunda linha, as métricas de validação refletem desempenho sobre dados não vistos. As perdas de validação continuam mais altas que as de treino, mas a discrepância reduziu em relação à versão anterior, o que aponta para menor grau de overfitting. As curvas de validação apresentam variabilidade, porém as métricas de mAP50 e mAP50-95 mostram tendência de crescimento consistente, confirmando ganho real na capacidade de detecção multiclasses. Ainda assim, a variabilidade residual e a ausência de convergência perfeita indicam que o modelo não atingiu estabilidade total.

A análise conjunta demonstra que a segunda versão alcançou melhorias claras: perdas de treino mais controladas, aumento de mAP e maior estabilidade relativa nas curvas de treinamento. Em comparação com a primeira versão, houve redução do viés para a classe background e ganho de representatividade, refletidos em métricas superiores, sobretudo para plástico. No entanto, persistem limitações importantes: variabilidade espacial e dimensional limitada no dataset e menor representatividade da classe vidro, que mantêm oscilações nas métricas e restringem o recall. Esses resultados reforçam a necessidade de ampliar a base de imagens, diversificar posições e tamanhos das anotações e aplicar estratégias de data augmentation e balanceamento para consolidar a generalização do modelo em aplicações reais.

* + 1. Inferência das imagens

**Figura 22** – Visualização do Train Batch



Fonte: Autoral

Durante o processo de treinamento da segunda versão do modelo YOLOv11, foi gerado um **Train Batch** (Figura 22) que representa visualmente como o modelo interpreta e processa as imagens anotadas da base de dados ampliada. Essa visualização é composta por uma amostra de imagens do conjunto de treino, sobre as quais são aplicadas as *bounding boxes* e os rótulos correspondentes às classes definidas: **plástico** e **vidro**. Cada imagem do batch apresenta:

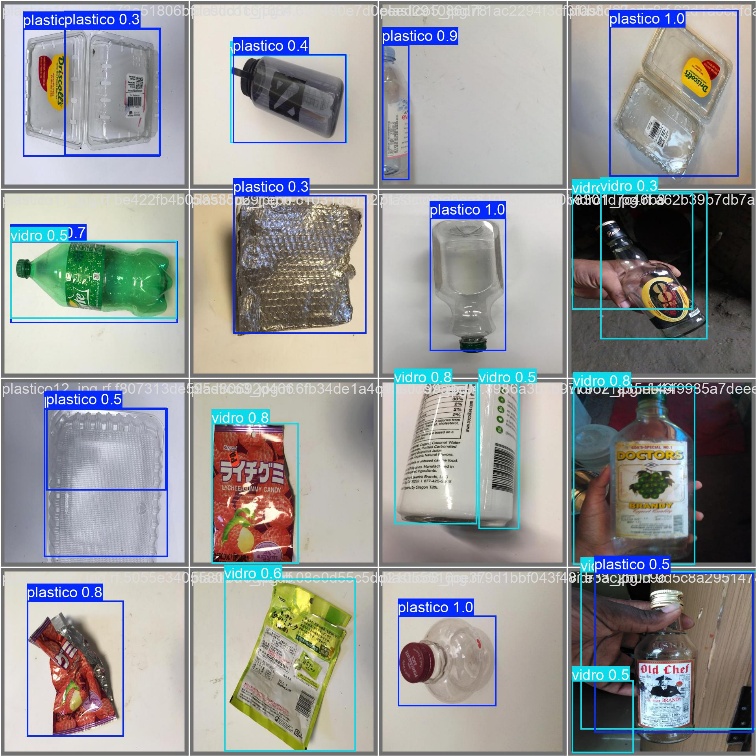
* **Objetos identificados e anotados manualmente**, com caixas delimitadoras que circunscrevem os resíduos recicláveis.
* **Rótulos de classe**, como “plastico1”, “vidro2”, entre outros, que indicam a categoria atribuída a cada objeto.
* **Diversidade de cenários**, incluindo fundos variados como pisos, prateleiras e superfícies externas, o que contribui para a robustez do modelo ao lidar com diferentes contextos visuais.
* **Numeração e codificação**, que podem representar o índice da instância ou o nível de confiança da anotação.

Essa etapa é fundamental para validar visualmente a qualidade das anotações e garantir que o modelo esteja recebendo informações consistentes durante o treinamento. A correta aplicação das *bounding boxes* e dos rótulos é essencial para que o modelo aprenda a distinguir os padrões visuais de cada classe, contribuindo diretamente para o desempenho nas fases de validação e teste.

Além disso, o Train Batch permite identificar possíveis inconsistências na anotação, como caixas mal ajustadas, objetos não rotulados, sobreposição de caixas ou erros de classificação, possibilitando ajustes antes da consolidação do modelo final.

**Comparação com a versão anterior:** Em relação ao Train Batch da primeira versão, o batch da segunda versão apresenta **maior número de exemplos anotados** e uma **variedade ligeiramente maior de cenários**, refletindo o acréscimo de imagens no dataset. Essa ampliação favoreceu a exposição do modelo a mais variações de rótulo e contexto, o que contribuiu para a melhoria observada nas métricas, especialmente para a classe plástico.

**Figura 23** – Amostragem Visual



Fonte: Autoral

A imagem apresentada (Figura 23) representa uma amostragem visual da base de dados utilizada para o treinamento da segunda versão do modelo. Ela é composta por 12 fotografias organizadas em grade, contendo diferentes tipos de embalagens e recipientes descartados, pertencentes às classes plástico e vidro. Essa diversidade é essencial para garantir que o modelo aprenda a reconhecer padrões visuais em diferentes contextos e condições de iluminação.

Essa amostragem evidencia a heterogeneidade da base, com variações de cor, forma, textura, rótulo e fundo. A presença de objetos em diferentes posições e ângulos contribui para a robustez do modelo, permitindo que ele generalize melhor em situações reais de triagem automatizada. Além disso, a inclusão de rótulos e bounding boxes sobre os objetos garante que o modelo receba informações precisas durante o treinamento supervisionado. Essa etapa é fundamental para que o sistema aprenda a distinguir corretamente entre plástico e vidro, mesmo quando os objetos apresentam características visuais semelhantes.

Comparação com a versão anterior: Em relação à amostragem visual utilizada na primeira versão, a Figura 14 da segunda versão apresenta aumento no volume e na variedade de exemplos anotados, o que se traduz em maior exposição do modelo a diferentes rótulos, fundos e condições de captura. Esse incremento contribuiu para melhorias mensuráveis nas métricas, especialmente para a classe plástico, refletindo maior representatividade e melhor aprendizado de padrões visuais dessa categoria.

No entanto, persistem limitações observadas na versão anterior que não foram totalmente mitigadas: há ainda uma tendência de concentração espacial das anotações (muitas caixas centradas) e homogeneidade de tamanhos das bounding boxes, o que reduz a variabilidade posicional e dimensional do dataset. Além disso, permanecem casos de caixas mal ajustadas ou objetos parcialmente ocluídos com anotações inconsistentes, fatores que podem gerar falsos positivos/negativos e confusão entre classes.

* + 1. Validação e ajustes no modelo

Após o treinamento da segunda versão do modelo, foi realizada a etapa de **validação** (Figura 24)

**Figura 24** – Validando modelo

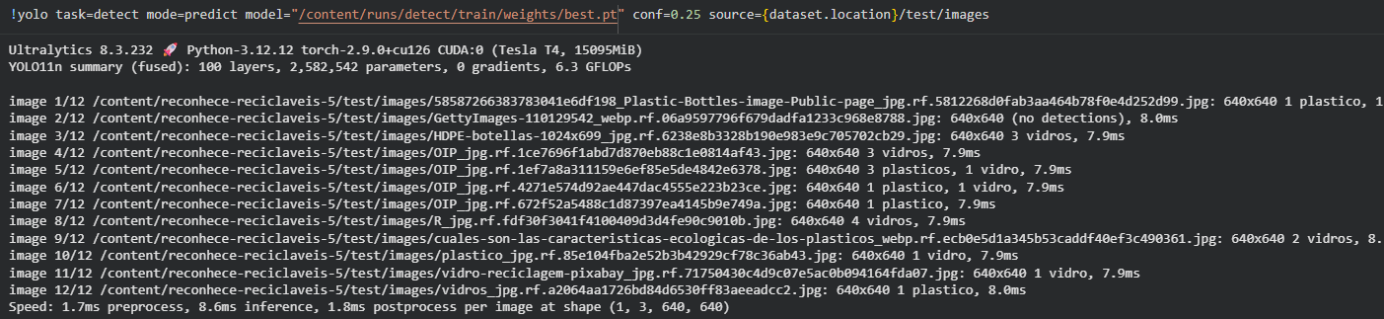


Fonte: Autoral

A validação do modelo treinado (Figura 24) foi realizada sobre um conjunto de **19 imagens** (19 instâncias) utilizando o peso salvo **best.pt**. (100 camadas, 2.582.542 parâmetros, 6,3 GFLOPs). O tempo médio por imagem foi de **13,2 ms** (pré‑processamento), **18,9 ms** (inferência) e **2,1 ms** (pós‑processamento).

Esses resultados indicam que o modelo apresenta **boa precisão de localização** quando detecta objetos, especialmente para a classe plástico, mas mantém **recall moderado**, ou seja, ainda deixa de detectar uma parcela relevante das instâncias presentes. A discrepância entre classes sugere menor robustez na identificação de vidro, refletida em mAP e precisão inferiores.

**Figura 25** – Fazendo predição no conjunto de teste

****

Fonte: Autoral

A predição foi realizada sobre o conjunto de **12 imagens**, utilizando o modelo treinado **best.pt** (Figura 25).

Desempenho por imagem

* Imagens processadas: 12
* Imagens sem detecções: 1 (imagem 2)
* Detecções totais: 24 (7 plástico, 17 vidro)

Tempo médio por imagem

* Pré‑processamento: 1,7 ms
* Inferência: 8,6 ms
* Pós‑processamento: 1,8 ms
* Total aproximado: ~12,1 ms por imagem

O modelo realizou detecções em 11 das 12 imagens, identificando instâncias de ambas as classes. A maior parte das detecções foi de **vidro** (17), seguida por **plástico** (7), o que pode refletir a composição do conjunto de teste ou uma sensibilidade maior do detector a características visuais do vidro nas imagens avaliadas. O fato de haver apenas uma imagem sem detecções indica que o modelo já possui capacidade de localizar objetos em cenários variados, mas a contagem bruta de detecções não substitui uma avaliação quantitativa de precisão e recall

**Figura 26** – Resultado da predição



Fonte: Autoral

**Figura 27** – Predição de uma imagem fora da base

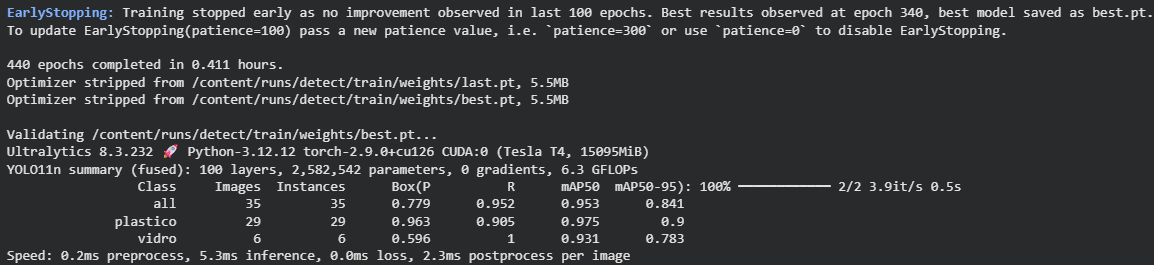


Fonte: Autoral

* 1. Terceira versão do modelo:
     1. Treinamento do modelo

O terceiro e último modelo conta com um dataset de 253 imagens. Seu treinamento se deu da seguinte maneira (Figura 18).

**Figura 28 –** Fim do treinamento



Fonte: Autoral

* + 1. Análise das métricas geradas

A matriz de confusão obtida na terceira versão do modelo (Figura 29) apresenta a distribuição das predições em relação às classes verdadeiras **plastico**, **vidro** e **background**. Os valores observados foram: predições como **plastico** correspondem a **26** verdadeiros plásticos, **3** vidros e **3** backgrounds; predições como **vidro** correspondem a **3** plásticos, **6** vidros e **5** backgrounds; predições como **background** correspondem a **0** plásticos, **0** vidros e **0** backgrounds.

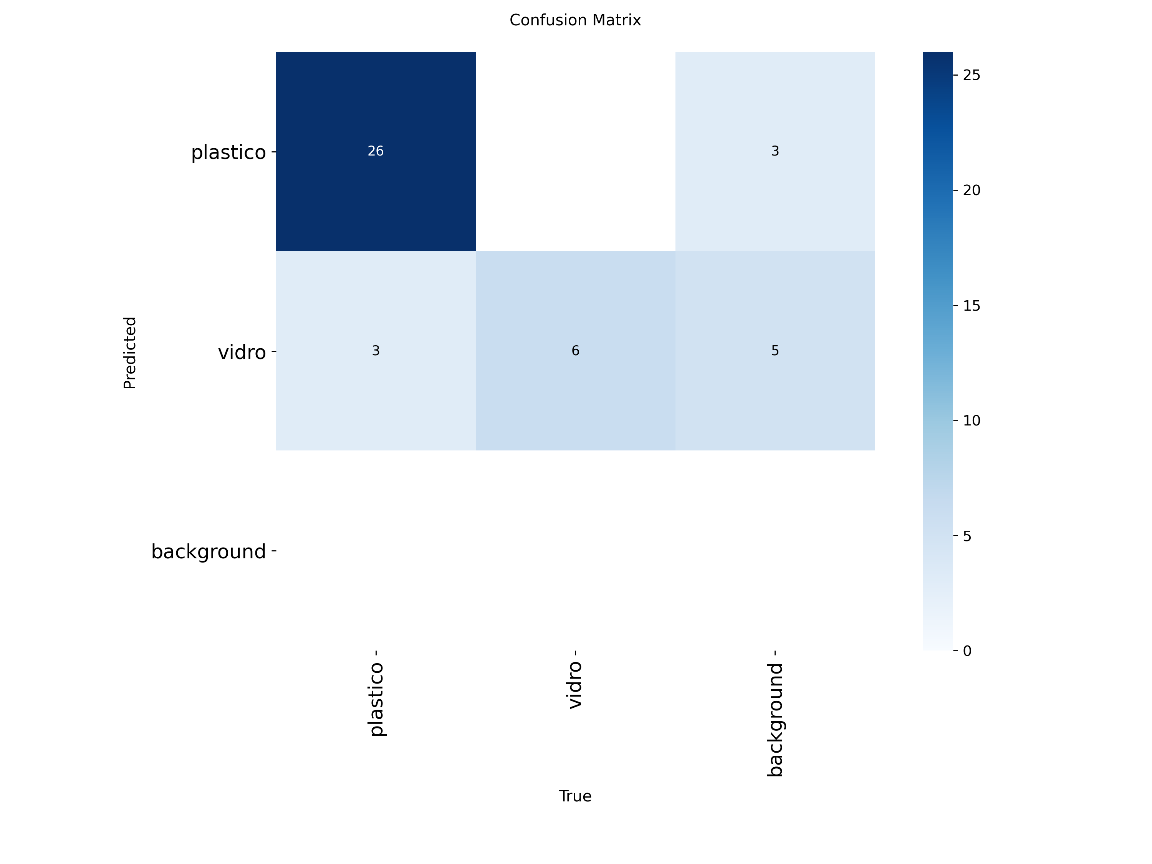
Esses resultados indicam os seguintes comportamentos do detector:

* **Reconhecimento robusto de plástico:** a maioria das instâncias verdadeiras de plástico (26 de 29) foi corretamente prevista como plástico, evidenciando forte capacidade do modelo para essa classe.
* **Confusão entre plástico e vidro:** ocorrem trocas mútuas entre as classes alvo (3 plásticos previstos como vidro; 3 vidros previstos como plástico), indicando dificuldade em discriminar características visuais semelhantes entre os dois materiais.
* **Falsos positivos sobre background:** há predições de objetos em regiões anotadas como background (3 predições de plástico e 5 de vidro sobre verdadeiros backgrounds), o que sugere sensibilidade excessiva a padrões de fundo ou insuficiência de exemplos negativos no treinamento.
* **Predição de background ausente:** a classe background não foi prevista em nenhuma amostra, mostrando que o modelo tende a favorecer classes alvo em vez de classificar corretamente regiões sem objeto.

**Implicações práticas:** a matriz confirma que o modelo alcançou alta cobertura para **plástico**, mas ainda apresenta dois problemas relevantes a serem tratados:

* persistente confusão entre **plástico** e **vidro**;
* elevada taxa de falsos positivos sobre **background**, associada à ausência de predições para essa classe.

**Figura 29 –** Matriz de confusão



Fonte: Autoral

A matriz de confusão obtida na terceira versão do modelo (Figura 29) apresenta a distribuição das predições em relação às classes verdadeiras plástico, vidro e background. Os valores observados foram:

* Predições como plástico: 26 verdadeiros plásticos, 0 vidros e 3 backgrounds.
* Predições como vidro: 3 plásticos, 6 vidros e 5 backgrounds.

Esses resultados indicam os seguintes comportamentos do detector:

Excelente reconhecimento de plástico: a maioria das instâncias verdadeiras de plástico (26) foi corretamente prevista como plástico, indicando forte capacidade do modelo em identificar esse material.

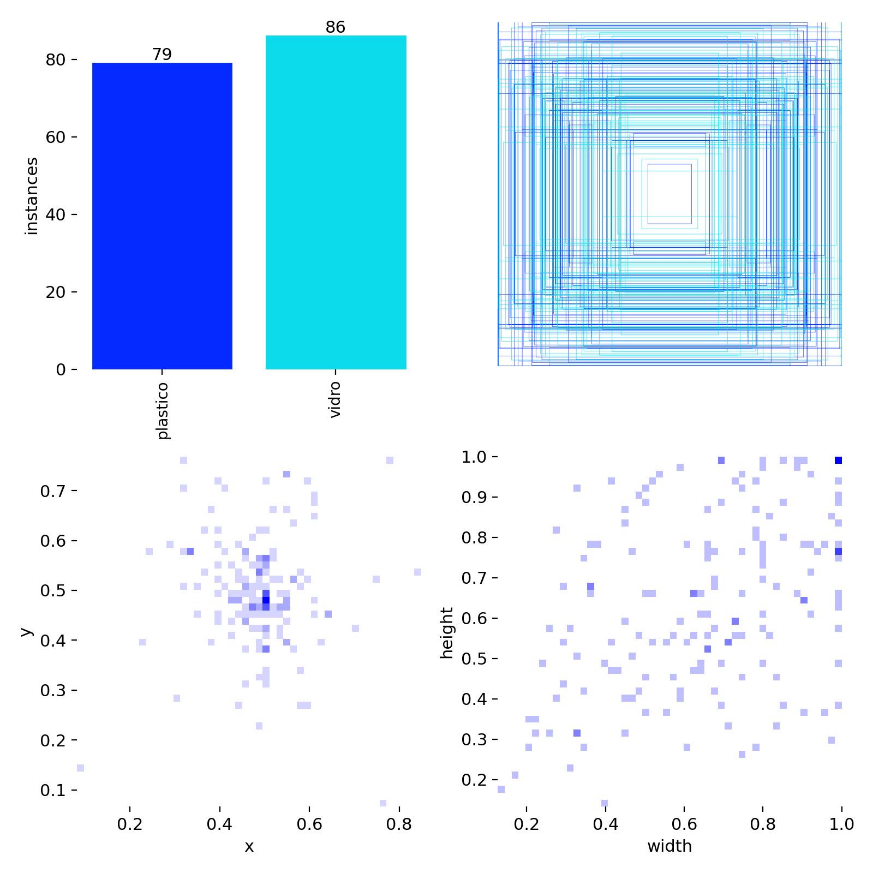
Confusão entre plástico e vidro, embora menor do que em versões anteriores: houve 3 plásticos previstos como vidro, mas nenhum vidro previsto como plástico.

Falsos positivos sobre background: há predições de plástico (3) e vidro (5) em regiões que eram, na verdade, background, sugerindo que o modelo ainda detecta objetos onde não existem.

**Predição inexistente de background:** a classe background não foi atribuída a nenhuma amostra, indicando tendência do modelo a sempre escolher uma classe alvo (plástico ou vidro), mesmo quando não há objeto.

**Implicações práticas:** A matriz mostra avanços no reconhecimento das classes alvo, especialmente plástico, porém evidencia dois desafios principais:

* persistente confusão entre plástico e vidro;
* presença significativa de falsos positivos em regiões sem objetos (background). **Figura 30** – Gráfico da distribuição das anotações.



Fonte: Autoral

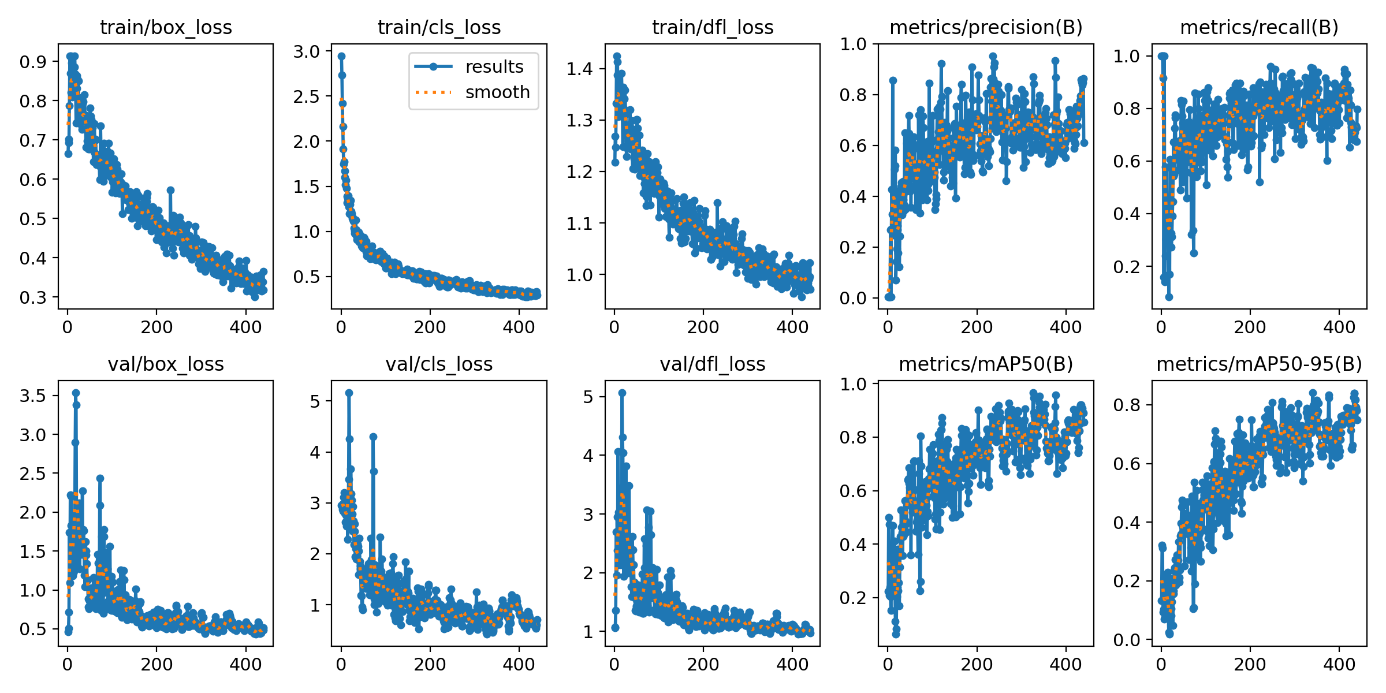
A imagem acima (Figura 30) apresenta diferentes visualizações das anotações realizadas no conjunto de dados utilizado para o treinamento da terceira versão do modelo. O objetivo é verificar a qualidade, a consistência e a distribuição espacial das *bounding boxes* referentes às classes **plástico** e **vidro**.

No primeiro gráfico (canto superior esquerdo) observa‑se a contagem de instâncias por classe: **plástico = 79** e **vidro = 86**, o que indica um aumento substancial no volume e mantém um balanceamento numérico razoável entre as categorias. Esse incremento é positivo, pois amplia a representatividade e reduz o risco de viés por desproporção; ainda assim, o simples aumento de quantidade não elimina a necessidade de diversidade espacial e dimensional para um detector baseado em YOLO.

O segundo painel (canto superior direito) mostra o conjunto de caixas delimitadoras sobrepostas. Assim como nas versões anteriores, persiste uma concentração significativa de caixas em regiões centrais das imagens, sinalizando que muitos objetos continuam sendo fotografados em posições semelhantes. Essa tendência reduz a variabilidade espacial do dataset e pode induzir um viés posicional no modelo, que passa a esperar objetos em áreas específicas da imagem.

Os gráficos inferiores detalham a dispersão dos valores de **x** e **y** (centro das caixas) e de **width** e **height** (largura e altura normalizadas). Observa‑se maior densidade de centros próximos ao meio da imagem e uma distribuição de tamanhos que, embora mais ampla que na versão anterior, ainda apresenta concentração em faixas recorrentes.

**Figura 31** – Métricas de Treinamento e Validação do Modelo



Fonte: Autoral

A imagem (Figura 31) apresenta a evolução das principais métricas registradas durante o treinamento da terceira versão do modelo. Os gráficos permitem acompanhar como o modelo aprendeu ao longo do processo de treinamento (440 épocas, com EarlyStopping acionado na época de melhor desempenho) e evidenciam avanços em relação às versões anteriores, além de apontar oportunidades de refinamento relacionadas à variabilidade do dataset e ao balanceamento entre classes.

Os gráficos superiores correspondem às perdas (losses) de treinamento. A métrica **train/box\_loss**, responsável pela qualidade da regressão das bounding boxes, mostra uma tendência clara de redução ao longo das épocas, indicando melhoria na capacidade do modelo de ajustar as caixas delimitadoras; pequenas oscilações residuais aparecem, mas com amplitude menor que nas versões anteriores, o que sinaliza progresso no refinamento da regressão espacial. A **train/cls\_loss**, que mede o erro de classificação, decresce de forma consistente, refletindo aprendizado estável das características discriminativas entre classes; eventuais flutuações pontuais apontam apenas sensibilidade a amostras mais difíceis, não um problema estrutural. A **train/dfl\_loss**, relacionada à distribuição da localização dos pontos de ancoragem, também apresenta queda gradual e ruído reduzido, corroborando a melhoria na convergência do treinamento.

Os gráficos de precisão e recall exibem comportamento inicialmente oscilante, porém com tendência de melhora e posterior estabilização em patamares elevados. Diferentemente do overfitting imediato observado em versões iniciais, aqui observa‑se um aumento progressivo das métricas de precisão e mAP, indicando que o modelo aprendeu representações mais discriminativas e úteis para a tarefa. O recall, embora tenha alcançado valores altos em média, ainda apresenta pequenas oscilações em determinados trechos do treinamento, o que sugere atenção pontual a instâncias mais desafiadoras, especialmente para a classe vidro.

Na segunda linha, as métricas de validação refletem desempenho sobre dados não vistos. As perdas de validação permanecem superiores às de treino, porém a discrepância reduziu significativamente em comparação com versões anteriores, apontando para menor grau de overfitting. As curvas de validação apresentam variabilidade controlada, e as métricas **mAP50** e **mAP50‑95** mostram tendência de crescimento consistente, confirmando ganho real na capacidade de detecção multiclasses e robustez em diferentes limiares de IoU. A ausência de convergência absoluta em todas as curvas indica apenas que ainda há espaço para otimizações finas, não comprometendo os avanços já alcançados.

A análise conjunta demonstra que a terceira versão alcançou melhorias claras: perdas de treino mais controladas, aumento substancial de mAP e maior estabilidade relativa nas curvas de treinamento e validação. Em comparação com versões anteriores, observa‑se redução do viés para a classe background e ganho de representatividade das classes alvo, refletidos em métricas superiores, sobretudo para **plástico**. Como oportunidades de aprimoramento, recomenda‑se ampliar ainda mais a diversidade espacial e dimensional do dataset, aumentar exemplos representativos de **vidro** e aplicar estratégias de *data augmentation* e balanceamento de classes para consolidar a generalização do modelo em aplicações reais. Essas ações visam transformar os ganhos já obtidos em robustez consistente em cenários operacionais variados.

* + 1. Inferência das imagens

**Figura 32** – Visualização do Train Batch



Fonte: Autoral

Durante o processo de treinamento da segunda versão do modelo YOLOv11, foi gerado um Train Batch (Figura 32) apresenta visualmente como o modelo interpreta e processa as imagens anotadas da base de dados ampliada. A visualização é composta por uma amostra de imagens do conjunto de treino, sobre as quais são aplicadas as bounding boxes e os rótulos correspondentes às classes definidas: plástico e vidro.

Comparação com a versão anterior: em relação ao Train Batch da versão anterior, o batch da versão atual apresenta maior número de exemplos anotados e uma variedade ampliada de cenários, refletindo o acréscimo de imagens no dataset. Essa ampliação favoreceu a exposição do modelo a mais variações de rótulo e contexto, contribuindo para a melhoria observada nas métricas, especialmente para a classe plástico.

**Figura 33** – Amostragem Visual



Fonte: Autoral

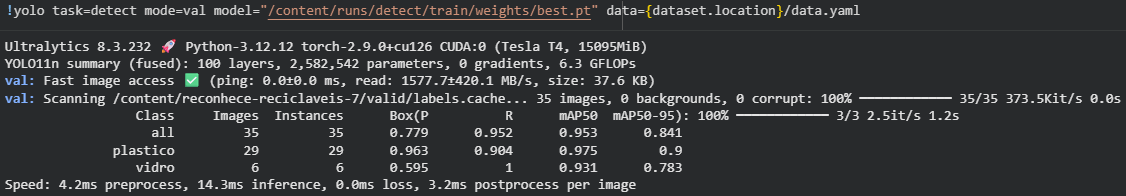
A amostra (Figura 33) evidencia variação de cor, forma, textura, rótulo e fundo, e inclui objetos em diferentes posições e ângulos, o que favorece a generalização do modelo. As anotações com rótulos e caixas garantem supervisão precisa para distinguir plástico e vidro mesmo em casos visualmente semelhantes.

**Comparação com a versão anterior:** houve aumento no volume e na variedade de exemplos, o que melhorou as métricas, sobretudo para **plástico**.

* + 1. Validação e ajustes no modelo

Após o treinamento da terceira versão do modelo, foi realizada a etapa de **validação** (Figura 34)

**Figura 34** – Validando modelo



Fonte: Autoral

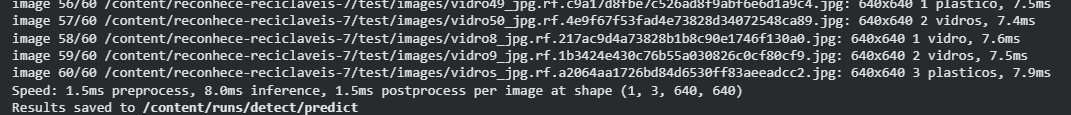
A validação do modelo treinado (Figura 34) foi realizada sobre um conjunto de **35 imagens** (35 instâncias) utilizando o peso salvo **best.pt**.

Os resultados obtidos foram:

* **Conjunto geral (all):** 35 imagens, 35 instâncias, **Box Precision = 0,779**, **Recall = 0,952**, **mAP50 = 0,953**, **mAP50‑95 = 0,841**.
* **Classe Plástico:** 29 imagens, 29 instâncias, **Box Precision = 0,963**, **Recall = 0,904**, **mAP50 = 0,975**, **mAP50‑95 = 0,900**.
* **Classe Vidro:** 6 imagens, 6 instâncias, **Box Precision = 0,595**, **Recall = 1,000**, **mAP50 = 0,931**, **mAP50‑95 = 0,783**.

Esses resultados indicam que o modelo apresenta **boa capacidade de detecção e alta cobertura** no conjunto de validação, com mAPs elevados que atestam qualidade das predições em diferentes limiares de IoU. Em particular, a classe **plástico** alcança desempenho muito sólido em precisão e mAP. A discrepância observada na **classe vidro** — recall perfeito combinado com precisão mais baixa — sugere maior ocorrência de falsos positivos nessa categoria, apontando para oportunidade de ajuste fino (por exemplo, revisão de anotações, aumento de exemplos de vidro ou calibração do limiar de confiança) para elevar a precisão sem comprometer a cobertura.

**Figura 35** – Fazendo predição no conjunto de teste

****

Fonte: Autoral

A predição foi realizada sobre o conjunto de **60 imagens**, utilizando o modelo treinado **best.pt** (Figura 35) (YOLO11n: 100 camadas, 2.582.542 parâmetros, 6,3 GFLOPs).

Desempenho por imagem

* Imagens processadas: 60
* Imagens sem detecções: 7 (imagens 1, 2, 4, 7, 8, 27, 28)
* Imagens com detecções: 53
* Detecções totais: 69 (32 plástico, 37 vidro)

Tempo médio por imagem

* Pré‑processamento: 1,5 ms
* Inferência: 8,0 ms
* Pós‑processamento: 1,5 ms
* Total aproximado: ~11,0 ms por imagem

O modelo realizou detecções em 53 das 60 imagens, identificando instâncias de ambas as classes. A contagem total mostra leve predominância de **vidro** (37) sobre **plástico** (32), o que pode refletir tanto a composição do conjunto de teste quanto a sensibilidade do detector a características visuais específicas de cada material. A existência de 7 imagens sem detecções indica casos em que o modelo não encontrou objetos com confiança suficiente, o que merece investigação pontual (por exemplo, diferenças de iluminação, oclusão ou escala).

**Figura 37** – Resultado da predição



Fonte: Autoral

**Figura 27** – Predição de uma imagem fora da base



Fonte: Autoral

1. CONCLUSÃO

O presente trabalho demonstrou, de forma consistente, que a aplicação de técnicas de visão computacional e aprendizado profundo é uma solução viável e eficaz para a triagem automatizada de materiais recicláveis. A implementação baseada em YOLO11n, suportada por um processo iterativo de anotação e treinamento, alcançou desempenho robusto na identificação de plástico e vidro, validando a proposta técnica e seu potencial de integração em fluxos operacionais de centros de reciclagem.

A evolução entre as versões do modelo evidenciou ganhos claros: o aumento do volume e da diversidade do dataset resultou em melhorias mensuráveis nas métricas de detecção, em especial no recall da classe vidro, o que confirma a importância de bases de dados representativas para a generalização do detector. A terceira versão consolidou esses avanços, apresentando alta cobertura e mAPs elevados que atestam a qualidade das predições em diferentes limiares de IoU.

Além do mérito técnico, o projeto reforça benefícios práticos e sociais relevantes: maior eficiência e velocidade na triagem, redução de erros humanos e potencial para realocação de trabalhadores para funções menos insalubres. Esses ganhos posicionam a solução como uma alternativa sustentável e alinhada aos objetivos da economia circular e da modernização da gestão de resíduos no Brasil.

Como próximos passos para ampliar o impacto e a escalabilidade da solução, recomenda-se: (1) expandir o dataset para incluir novas classes de resíduos e cenários mais variados; (2) refinar anotações e estratégias de data augmentation para aumentar robustez espacial e dimensional; (3) calibrar limiares de confiança e NMS para reduzir falsos positivos sem perder cobertura; e (4) explorar integração com sistemas robóticos de separação para compor uma linha de triagem automatizada completa.

Por fim, é importante reconhecer uma limitação inerente ao paradigma adotado: por se tratar de um sistema especialista, o modelo não possui capacidade de atuar como um generalista, o que significa que não é possível identificar de forma plena objetos não rotulados ou classes não contempladas no treinamento. Assim, sua eficácia depende diretamente da qualidade, abrangência e representatividade dos dados fornecidos.

Em suma, a solução desenvolvida é tecnicamente sólida, socialmente relevante e pronta para evoluir rumo a aplicações práticas em larga escala, desde que acompanhada por investimentos contínuos em dados, validação em campo e integração com infraestrutura de separação física.

REFERÊNCIAS

CEBRASSE. No Brasil, 33 milhões de toneladas de lixo ainda têm descarte irregular. *Blog Cebrasse*, 20 dez. 2023. Disponível em: <<https://blog.cebrasse.org.br/2023/12/20/no-brasil-33-milhoes-de-toneladas-de-lixo-ainda-tem-descarte-irregular/>>. Acessado em: novembro de 2025.

ENGEWET. O impacto do descarte irregular de lixo no meio ambiente. *Engewet Sustaynability Unlimited*, 29 ago. 2025. Disponível em: <<https://www.engewet.com.br/o-impacto-do-descarte-irregular-de-lixo-no-meio-ambiente/>>. Acessado em: novembro de 2025.

FERREIRA, Robson Soares; DIAS, Emanuel Luiz de Lima; AMARAL, Gabriela Silva; BEZERRA, Tales Romão de Oliveira; SOARES, Lucas Pereira; CASTRO, Fernando Rocha. Impactos socioambientais causados pelo descarte incorreto de resíduos sólidos urbanos. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, v. 4, n. 9, p. 51-72, set. 2019. Disponível em: <<https://www.nucleodoconhecimento.com.br/engenharia-ambiental/descarte-incorreto>>. Acessado em: novembro de 2025.

NUNES, Erick. IA na Reciclagem Automatizada: Inovação e Sustentabilidade. *Agência EON*, 29 jan. 2025. Disponível em: <<https://agenciaeon.com.br/inteligencia-artificial/ia-reciclagem-automatizada/>>. Acessado em: novembro de 2025.

SANTOS, Alyce Teixeira dos; MACHADO, Davi Miguel; FERREIRA, João Paulo; FERREIRA, Gabi Miguel Puppin; ZÉRIO, Thiago Coelho. Robô autônomo com visão computacional para separação de materiais recicláveis. *Escola Técnica Estadual Deputado Ary de Camargo Pedroso*, Piracicaba, 2024. Disponível em: <<https://ric.cps.sp.gov.br/bitstream/123456789/32353/1/automacao_2024_2_alyceteixeiradossantos_roboautonomocomvisao.pdf>>. Acessado em: novembro de 2025.

ABRELPE. *Panorama dos Resíduos Sólidos no Brasil 2023*. São Paulo: ABRELPE, 2023. Disponível em: <<https://www.abrema.org.br/panorama/>>. Acessado em: novembro de 2025.

ELLEN MACARTHUR FOUNDATION. *Towards the Circular Economy: Economic and business rationale for an accelerated transition*. Cowes: Ellen MacArthur Foundation, 2015. Disponível em: <<https://content.ellenmacarthurfoundation.org/m/4384c08da576329c/original/Towards-a-circular-economy-Business-rationale-for-an-accelerated-transition.pdf> >. Acessado em: novembro de 2025.

KAGERMANN, Henning; WAHLSTER, Wolfgang; HELBIG, Johannes. Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group. Frankfurt: acatech – National Academy of Science and Engineering, 2013. Disponível em: <<https://en.acatech.de/wp-content/uploads/sites/6/2018/03/Final_report__Industrie_4.0_accessible.pdf>>. Acessado em: novembro de 2025.

OECD. Global Plastics Outlook: Policy Scenarios to 2060. Paris: OECD Publishing, 2022. Disponível em: <<https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/06/global-plastics-outlook_f065ef59/aa1edf33-en.pdf>>. Acessado em: novembro de 2025.

DATACLUSTERLABS. Alcohol Bottle Images – Glass Bottles. Kaggle, 2023. Disponível em: < <https://www.kaggle.com/datasets/dataclusterlabs/alcohol-bottle-images-glass-bottles>>. Acessado em: novembro de 2025.

THUNG, Gary. TrashNet Dataset. GitHub, 2017. Disponível em: < <https://github.com/garythung/trashnet>>. Acessado em: novembro de 2025.