



UNIVERSIDADE FEDERAL DO PARÁ  
CAMPUS UNIVERSITÁRIO DE TUCURUÍ  
FACULDADE DE ENGENHARIA DE COMPUTAÇÃO

MARCUS VINICIUS CARVALHO FERREIRA

**ARTE ACESSÍVEL: UMA ANÁLISE COMPARATIVA DE ALGORITMOS DE  
DETECÇÃO DE OBRAS DE ARTE**

TUCURUÍ  
2025

MARCUS VINICIUS CARVALHO FERREIRA

**ARTE ACESSÍVEL: UMA ANÁLISE COMPARATIVA DE ALGORITMOS DE  
DETECÇÃO DE OBRAS DE ARTE**

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Engenheiro de Computação.

Orientador(a): Prof. Dr. Iago Lins de Medeiros

MARCUS VINICIUS CARVALHO FERREIRA

**ARTE ACESSÍVEL: UMA ANÁLISE COMPARATIVA DE ALGORITMOS DE  
DETECÇÃO DE OBRAS DE ARTE**

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Engenheiro de Computação.

Data da aprovação: 17/07/2025

Conceito: EXCELENTE

**BANCA EXAMINADORA**

Prof. Dr. Iago Lins de Medeiros

Orientador

FECOMP/CAMTUC/UFPA

Prof. Dr. Hugo Leonardo Melo dos Santos

Avaliador Externo

DSCI/Campus VX/UEPA

Prof. Dr. Lucas de Lima Bastos

Avaliador Externo

FEC/IGE/UNIFESSPA

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD  
Sistema de Bibliotecas da Universidade Federal do Pará  
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

---

C331a Carvalho Ferreira, Marcus Vinicius.  
ARTE ACESSÍVEL : UMA ANÁLISE COMPARATIVA DE  
ALGORITMOS DE DETECÇÃO DE OBRAS DE ARTE / Marcus  
Vinicius Carvalho Ferreira. — 2025.  
xxvii, 27 f. : il. color.

Orientador(a): Prof. Dr. Iago Lins de Medeiros  
Trabalho de Conclusão (Graduação) - Universidade Federal do  
Pará, Campus Universitário de Tucuruí, Faculdade de Engenharia  
da Computação, Tucuruí, 2025.

1. Visão Computacional. 2. Acessibilidade. 3. Detecção de  
Objetos. I. Título.

CDD 006.37

---

*“Quando a tecnologia conecta as pessoas à arte, ela não apenas informa, mas também transforma. A acessibilidade é o vínculo entre criação e inclusão.”*

## **AGRADECIMENTOS**

Agradeço, acima de tudo, a Deus, por me sustentar nos dias difíceis e me conceder discernimento para seguir em frente.

À minha mãe, meu maior exemplo de força, amor e dedicação. Agradeço por todo o incentivo, cada gesto de carinho e cada sacrifício feito ao longo do caminho. Seu apoio incondicional e sua presença constante foram fundamentais para que eu chegasse até aqui. Esta conquista também é sua.

Ao meu orientador, pela escuta atenta, pela paciência diante das minhas dúvidas e pelo compromisso em orientar com seriedade e respeito. Sua orientação foi essencial para transformar ideias soltas em um trabalho com propósito.

Aos amigos que, com companheirismo e bom humor, ajudaram a tornar essa caminhada mais leve. A troca de experiências, o suporte nos momentos desafiadores e as conversas fora da rotina acadêmica tiveram um valor enorme.

Aos professores e à instituição de ensino, pelos ensinamentos e contribuições que marcaram minha formação.

E a todos que, de alguma forma, fizeram parte dessa trajetória, deixo minha sincera gratidão.

## RESUMO

A visão computacional tem se mostrado uma ferramenta promissora no desenvolvimento de tecnologias assistivas, especialmente para promover acessibilidade a pessoas com deficiência visual. Este trabalho propõe a análise comparativa de algoritmos de detecção de objetos com foco na identificação de obras de arte pintadas, utilizando redes neurais convolucionais. Foram avaliados os modelos YOLOv8, YOLOv11, RetinaNet e *Faster R-CNN* quanto à sua eficácia e viabilidade em dispositivos móveis. O dataset utilizado foi criado a partir da coleta automatizada de imagens da internet, envolvendo obras consagradas como *Mona Lisa*, *Meisje met de Parel* e *The Starry Night*. As imagens foram rotuladas manualmente com o auxílio da ferramenta LabelImg. Os modelos foram treinados e avaliados com base em métricas como precisão média (mAP), *recall*, tempo de inferência e consumo de recursos computacionais. Os resultados demonstraram que os modelos da família YOLO, especialmente o YOLOv8, obtiveram o melhor equilíbrio entre desempenho e eficiência, atingindo mAP de 0,992 e *recall* de 0,987. Além disso, apresentaram menor tempo de inferência, o que os torna particularmente adequados para aplicações em tempo real e em dispositivos com recursos limitados. A pesquisa contribui para o avanço de soluções acessíveis no campo da arte e inclusão digital, demonstrando o potencial da inteligência artificial como aliada da acessibilidade cultural.

**Palavras-Chave:** Visão Computacional; Acessibilidade; Detecção de Objetos.

## ABSTRACT

*Computer vision has proven to be a promising tool in the development of assistive technologies, particularly in promoting accessibility for people with visual impairments. This work presents a comparative analysis of object detection algorithms focused on the recognition of painted artworks, using convolutional neural networks. The evaluated models include YOLOv8, YOLOv11, RetinaNet, and Faster R-CNN, with emphasis on their effectiveness and feasibility for deployment on mobile devices. The dataset was created through the automated collection of images from the internet, featuring renowned artworks such as Mona Lisa, Meisje met de Parel, and The Starry Night. All images were manually annotated using the LabelImg tool. The models were trained and assessed based on metrics such as mean Average Precision (mAP), recall, inference time, and computational resource usage. The results showed that the YOLO-based models, especially YOLOv8, achieved the best balance between performance and efficiency, reaching a mAP of 0.992 and recall of 0.987. Additionally, they demonstrated shorter inference times, making them particularly suitable for real-time applications and devices with limited hardware capabilities. This research contributes to the advancement of accessible solutions in the fields of art and digital inclusion, highlighting the potential of artificial intelligence as an ally in promoting cultural accessibility.*

**Keywords:** *Computer Vision; Accessibility; Object Detection.*

## LISTA DE FIGURAS

Figura 1 – Etapas do Desenvolvimento Arte Acessível .....	pág. 16
Figura 2 – Obras de Arte que Compõem o Dataset.....	pág. 16
Figura 3 – Aumento de Dados do Dataset. ....	pág. 17
Figura 4 – Distribuição das Imagens por Classe.....	pág. 18
Figura 5 – Comparação Entre Modelos/Versões. ....	pág. 20
Figura 6 – Comparação Entre Perda de Validação e Treino.....	pág. 21
Figura 7 – Comparação de mAP50.....	pág. 23
Figura 8 – Comparação de Recall.....	pág. 24
Figura 9 – Comparação de Tempo de Inferência.....	pág. 24

## **LISTA DE TABELAS**

Tabela 1 – Resultados do Refinamento (fine tuning) dos Modelos.....pág. 21

Tabela 2 – Recursos Computacionais Utilizados por Algoritmos. ....pág. 22

## SUMÁRIO

1. INTRODUÇÃO.....	pág. 12
1.1 OBJETIVO GERAL.....	pág.13
1.2 OBJETIVO ESPECÍFICO.....	pág.13
1.3 ORGANIZAÇÃO DO TEXTO.....	pág.13
2. TRABALHOS RELACIONADOS.....	pág.15
3. ARTE ACESSÍVEL: DETECTOR DE OBRAS DE ARTE.....	pág.16
3.1 CRIAÇÃO DO DATASET.....	pág.16
3.2. ALGORITMOS AVALIADOS DE REDES NEURAIS CONVOLUCIONAIS.....	pág.18
4. ANÁLISE E DISCUSSÃO DE RESULTADOS.....	pág.19
4.1. MÉTRICAS AVALIADAS E AMBIENTE DE SIMULAÇÃO.....	pág.19
4.2. AVALIAÇÕES INTRA-ALGORITMOS.....	pág.20
4.3. AVALIAÇÕES INTERALGORITMOS.....	pág.20
5. CONCLUSÃO.....	pág.25
6. REFERÊNCIAS.....	pág.26
APÊNDICE A – CÓDIGO DO PROJETO.....	pág.27

## INTRODUÇÃO

Com o avanço contínuo das tecnologias digitais, surgem ferramentas inovadoras capazes de amenizar desafios que acompanham a sociedade ao longo do tempo, como as questões relacionadas à acessibilidade e à inclusão de pessoas com algum tipo de deficiência visual. Conforme dados do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2022, cerca de 18,6 milhões de brasileiros com 2 anos ou mais apresentam algum tipo de deficiência, correspondendo a 8,9% da população. Entre essas deficiências, a dificuldade visual, mesmo com o uso de óculos ou lentes de contato, foi a segunda mais relatada, afetando 3,1% dos entrevistados (IBGE, 2023).

Quando nos referimos a pessoas com deficiência visual, é necessário compreender que se trata de um grupo heterogêneo, cuja condição pode variar significativamente, pois a perda da visão é uma experiência subjetiva, manifestando-se de formas distintas entre os indivíduos (MARTINS, 2008). No que tange ao campo das artes e da visualidade, é fundamental reconhecer que vivemos em um mundo predominantemente visual. Assim, a apreciação das obras artísticas ocorre majoritariamente pela observação visual, já que os olhos desempenham um papel central na compreensão do fenômeno artístico (PATEL et al., 2025).

Dessa forma, o ensino da arte para todos deve ser considerado de extrema importância, visto que a arte exerce uma função ativa na produção do conhecimento e no desenvolvimento pessoal dos indivíduos. Negar o acesso ao ensino artístico a grupos específicos, especialmente pessoas com deficiência, significa desperdiçar o potencial intrínseco desses sujeitos. Portanto, a responsabilidade de reduzir a exclusão desse público deve ser assumida pelos diversos núcleos sociais em que eles estão inseridos, como escolas, famílias, o Estado e a sociedade em geral (DE MILANO; HONORATO, 2010).

Entre as tecnologias que contribuem para superar as limitações impostas pela deficiência visual, destaca-se a visão computacional. Este campo da ciência permite que computadores capturem e interpretem o ambiente ao seu redor por meio de imagens obtidas por câmeras, scanners ou outros dispositivos capazes de fornecer dados visuais (DE MILANO; HONORATO, 2010).

Sistemas com capacidade de reconhecimento de objetos a partir de imagens podem ser aplicados em múltiplos contextos, incluindo veículos autônomos, controle de dispositivos por gestos, monitoramento de espaços públicos, entre outros. Para garantir a eficácia dessas aplicações, é fundamental realizar estudos que conciliem o desempenho dos algoritmos com o custo computacional envolvido (BERNARDO, 2019). Este trabalho tem como objetivo

integrar ferramentas de visão computacional e algoritmos de inteligência artificial para identificar obras de arte pintadas, oferecendo suporte a pessoas com deficiência visual e promovendo maior inclusão e acessibilidade em ambientes culturais, como museus e exposições.

### **1.1. Objetivo Geral**

O objetivo geral deste trabalho é desenvolver e analisar algoritmos de visão computacional baseados em redes neurais convolucionais para a detecção precisa de obras de arte pintadas, visando a criação de soluções acessíveis que promovam a inclusão de pessoas com deficiência visual em ambientes culturais, como museus e exposições.

### **1.2. Objetivo Específico**

- Coletar e organizar um dataset representativo contendo imagens de obras de arte relevantes, com suas respectivas anotações para treinamento dos modelos.
- Implementar e treinar modelos de detecção de objetos utilizando arquiteturas como YOLOv8, YOLOv11, RetinaNet e Faster R-CNN.
- Avaliar o desempenho dos modelos através de métricas quantitativas, como precisão média (mAP), recall, tempo de inferência e consumo de recursos computacionais.
- Comparar os resultados obtidos entre os diferentes modelos, identificando o mais adequado para aplicação em dispositivos móveis.
- Propor um protótipo funcional capaz de identificar obras de arte em tempo real, contribuindo para a acessibilidade e inclusão de pessoas com deficiência visual.

### **1.3. Organização do Texto**

Este trabalho é organizado em cinco capítulos. O Capítulo 2 expõe os trabalhos correlacionados, compilando estudos e iniciativas que investigam a aplicação da visão computacional voltada à acessibilidade, particularmente em situações que envolvem a identificação de objetos e obras de arte. O Capítulo 3, a seguir, aborda o desenvolvimento do projeto proposto, detalhando a criação do conjunto de dados, a seleção das obras, as ferramentas empregadas na coleta e anotação das imagens, além dos algoritmos de redes neurais convolucionais selecionados para a análise comparativa.

O Capítulo 4 é voltado para a análise e discussão dos resultados dos experimentos conduzidos com os modelos de detecção, onde são avaliadas métricas como precisão, recall, tempo de inferência e uso de recursos computacionais, além de comparar o desempenho dos

diversos algoritmos. Por último, o Capítulo 5 traz as considerações finais, ressaltando os principais achados da pesquisa e sugerindo caminhos para futuras investigações, com ênfase na implementação prática dos modelos em soluções acessíveis destinadas a espaços culturais e museus.

## 2. TRABALHOS RELACIONADOS

Com a popularização e simplificação crescente de ferramentas que implementam recursos de visão computacional, diversos projetos têm buscado ampliar a acessibilidade por meio da detecção de objetos visuais. Por exemplo, o estudo realizado por Olivatto (2021) emprega uma rede neural convolucional utilizando o modelo YOLOv4 para identificar rampas de acessibilidade em áreas públicas, alcançando uma precisão média de 75%. Entretanto, o modelo enfrentou dificuldades na detecção, atribuídas à falta de padronização das rampas, sobretudo em cidades com menor população (OLIVATTO, 2021).

No trabalho de de Carvalho e do Nascimento (2025), os autores analisam as transformações culturais e tecnológicas na era digital, focando na importância das tecnologias assistivas para a inclusão digital e o empoderamento de pessoas com deficiência em ambientes museológicos. Esse estudo é relevante para o desenvolvimento de novas soluções assistivas, pois oferece perspectivas inovadoras sobre a integração dessas tecnologias no contexto cultural contemporâneo (DE CARVALHO; DO NASCIMENTO, 2025).

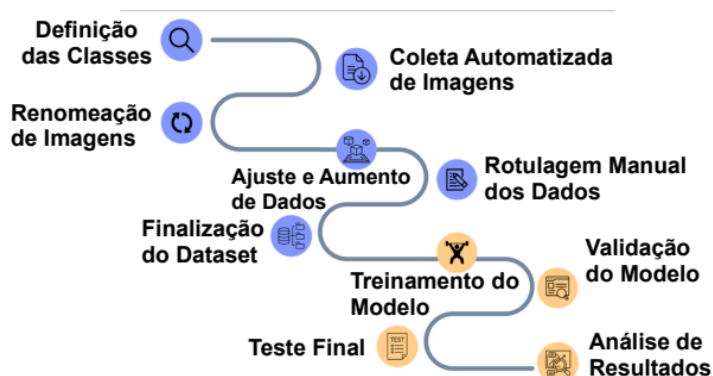
Khan et al. (2014) desenvolveram um conjunto de dados de alta resolução denominado Paintings-100, contendo imagens de pinturas para a classificação de artistas e estilos artísticos. A avaliação preliminar desse dataset, utilizando redes neurais convolucionais (CNNs), apresentou resultados promissores na tarefa de classificação, com precisão de até 38% por meio da combinação de decisões de modelos que analisam tanto a imagem completa quanto fragmentos aleatórios. Esse trabalho destaca a importância de bases de dados diversificadas e de qualidade para o avanço das técnicas de classificação no domínio artístico (KHAN et al., 2014).

Os trabalhos mencionados evidenciam a aplicabilidade da visão computacional em diferentes áreas, desde a acessibilidade urbana até a criação de bases para classificação de obras de arte. No entanto, observa-se uma lacuna na literatura no que diz respeito ao uso dessas tecnologias para aprimorar a experiência e promover a inclusão de pessoas com deficiência visual no contexto artístico. Dessa forma, o presente estudo propõe a comparação de diversos algoritmos para a detecção de objetos artísticos visuais, com o intuito de identificar a solução mais adequada para futuras implementações voltadas à acessibilidade.

### 3. ARTE ACESSÍVEL: DETECTOR DE OBRAS DE ARTE

A Figura 1 apresenta as etapas da proposta desenvolvida, composta por dez fases que abrangem todo o processo: desde a coleta de dados, aprimoramento, treinamento dos modelos, realização de testes e, por fim, a análise dos resultados obtidos.

Figura 1 – Etapas do Desenvolvimento Arte Acessível



Fonte: autoria própria, 2025.

#### 3.1.Criação do Dataset

A elaboração do dataset iniciou-se com a definição das obras de arte que o comporiam. Foram selecionadas obras consagradas que representam distintos aspectos culturais da humanidade. Outro critério relevante foi a popularidade dos museus onde essas obras estão expostas, destacando-se o Museu do Louvre, responsável pela maior parte das imagens utilizadas, que recebeu cerca de 8,7 milhões de visitantes no ano de 2024 (NOGUEIRA, 2025). A lista com as obras selecionadas está representada na Figura 2 e inclui: 1. Les Noces de Cana, 2. Liberty Leading The People, 3. Meisje Met de Parel, 4. Mona Lisa, 5. Virgin Of The Rocks, 6.The Creation Of Adam, 7. The Great Wave off Kanagawa, 8. The Night Watch e 9. The Starry Night.

Figura 2 – Obras Arte que Compõem o *Dataset*.

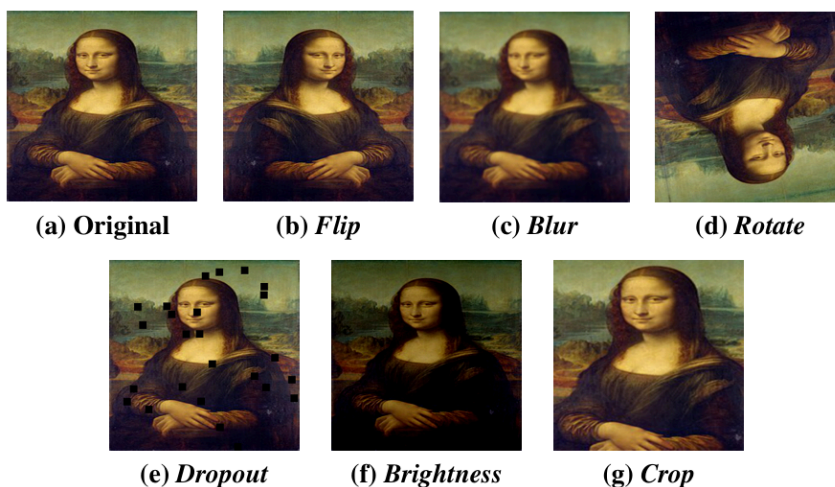


Fonte: autoria própria, 2025.

Para a coleta das imagens dessas obras, foi desenvolvido um algoritmo em Python que realiza automaticamente buscas na internet com base no nome da obra. Esse processo foi facilitado pelo uso da biblioteca Bing Image Downloader. O código desenvolvido encontra-se disponível para consulta [Link anonimizado para fins de revisão duplo-cega.](#). A etapa de rotulação das imagens foi conduzida com o auxílio da ferramenta LabelImg, um software de código aberto amplamente utilizado para anotação manual de imagens (TZUTALIN; CONTRIBUTORS, 2024).

Durante o processo de montagem do dataset, constatou-se uma escassez de imagens em algumas classes. Para solucionar essa limitação, foi empregada a biblioteca Albumentations, também por meio de um script em Python. Essa biblioteca destaca-se por sua capacidade de gerar novas amostras a partir das imagens existentes, aplicando transformações como espelhamento (flip), desfoque (blur), rotação (rotate), inserção de ruído (dropout), ajuste de brilho (brightness) e recorte (crop) (ALBUMENTATIONS TEAM, 2024), conforme ilustrado na Figura 3.

Figura 3 – Aumento de Dados do *Dataset*.

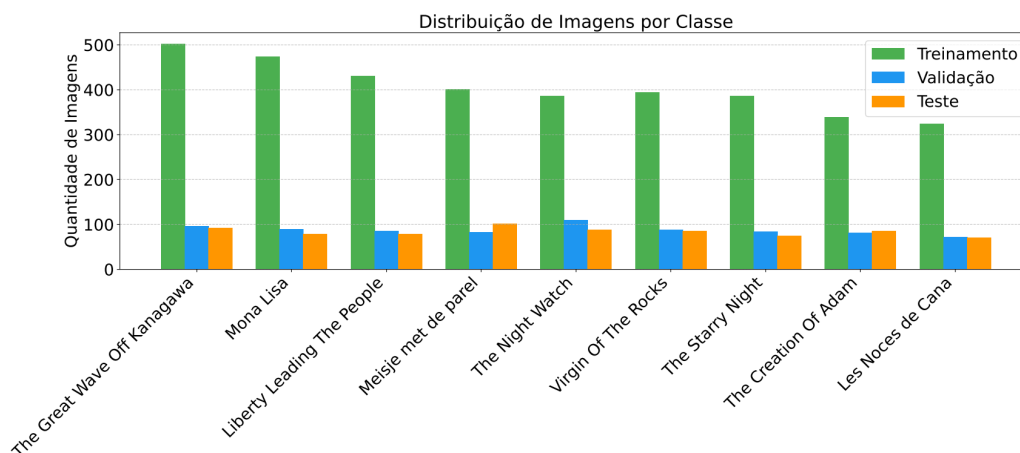


Fonte: autoria própria, 2025.

Após a expansão do conjunto de dados, uma nova rotulação manual foi necessária, visto que as imagens aumentadas exigem marcação individual. O dataset final resultante consistiu em 3.529 imagens para o treinamento, 770 para validação e 750 para teste.

A Figura 4 apresenta a distribuição das imagens por classe. Os dados foram divididos proporcionalmente: 70% para o treinamento, 15% para a validação e 15% para os testes. Essa divisão equilibrada entre as classes garantiu representatividade e robustez na avaliação dos modelos.

Figura 4 – Distribuição das Imagens por Classe.



Fonte: autoria própria, 2025.

### 3.2. Algoritmos Avaliados de Redes Neurais Convolucionais

Com o dataset estruturado, foi possível realizar a avaliação comparativa entre os algoritmos mais adequados para a tarefa de detecção de obras de arte. O processo envolveu o treinamento de diferentes modelos, validação cruzada e testes finais, seguidos da análise dos resultados obtidos. Neste trabalho, foram utilizados três algoritmos amplamente reconhecidos: YOLO, RetinaNet e Faster R-CNN, todos baseados em Redes Neurais Convolucionais (CNN).

O YOLO (You Only Look Once) é um modelo de detecção de objetos que se diferencia por tratar a tarefa como um problema de regressão unificada. Em vez de etapas sequenciais como nas abordagens R-CNN, o YOLO prediz diretamente, a partir dos pixels da imagem, as coordenadas das caixas delimitadoras (bounding boxes) e as probabilidades das classes associadas, o que proporciona rapidez e eficiência no processo de inferência (REDMON et al., 2016).

O RetinaNet busca equilibrar desempenho e custo computacional. Sua arquitetura é composta por três elementos principais: um backbone baseado em Feature Pyramid Network (FPN), que fornece uma representação multiescalar eficiente; uma sub-rede de classificação de objetos; e uma sub-rede de regressão para prever as caixas delimitadoras (LIN et al., 2017).

Já o Faster R-CNN foi proposto como uma evolução do R-CNN, visando superar limitações de velocidade e precisão. Ele opera em duas etapas principais: a geração de propostas de regiões de interesse (Região Proposal Network – RPN) e a classificação dessas regiões com suas respectivas bounding boxes. Trata-se de uma abordagem robusta e precisa, amplamente utilizada em contextos que exigem alta acurácia (BOESCH, 2024).

#### 4. ANÁLISE E DISCUSSÃO DE RESULTADOS

Esta seção apresenta os resultados dos experimentos realizados com diferentes modelos de detecção de objetos, com foco no desempenho do YOLO e sua comparação com outros algoritmos de redes neurais, como RetinaNet e Faster R-CNN. A primeira análise objetiva determinar quais versões dos algoritmos são mais adequadas para a tarefa de detecção de pinturas, considerando as variantes mais completas e lentas, e as mais enxutas e rápidas. A segunda parte da análise compara diretamente os modelos para determinar qual apresenta melhor desempenho geral.

##### 4.1. Métricas Avaliadas e Ambiente de Simulação

Para compreensão dos resultados, são destacadas as principais métricas utilizadas: recall, precisão, mAP (mean Average Precision), tempo médio de treinamento/inferência e perda (loss). Métricas como tempo e consumo de recursos apresentam melhor desempenho com valores mais baixos, enquanto recall, precisão e mAP são desejáveis em valores mais altos.

O *recall* mede a proporção de objetos reais corretamente identificados, enquanto a precisão avalia a proporção de predições positivas corretas. A partir dessas, constrói-se a curva precisão-recall, cuja área sob a curva corresponde à AP de cada classe; o mAP é a média das APs de todas as classes.

$$Recall = \frac{TP}{TP+FN} \quad (1.1)$$

$$Precisão = \frac{TP}{TP+FP} \quad (1.2)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N T_i \quad (1.3)$$

O tempo médio de treinamento/inferência representa o desempenho temporal por imagem. A função de perda mede o erro entre as saídas previstas e os rótulos reais, sendo essencial para avaliar a convergência do modelo.

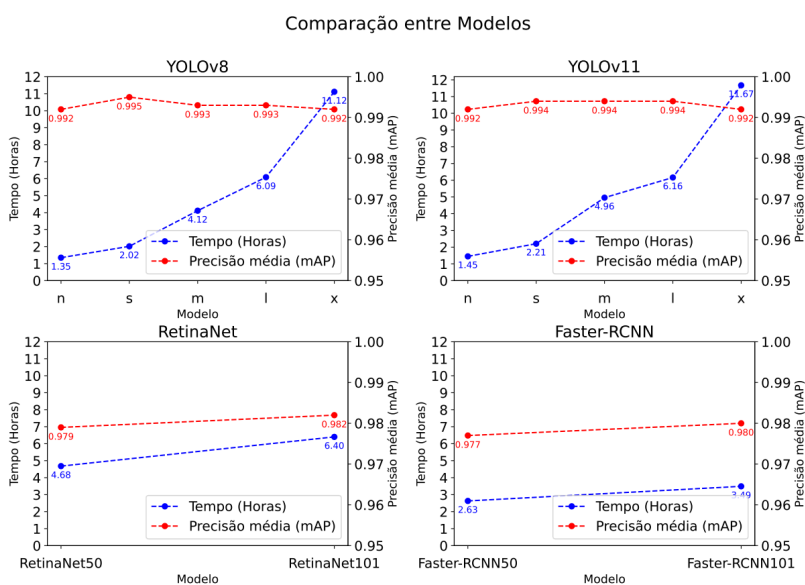
$$T_{médio} = \sum_{i=1}^N T_i \quad (1.4)$$

Os experimentos foram executados na plataforma *Kaggle*, que fornece um ambiente com CPU Intel Xeon 2GHz, 31 GB de RAM e GPU Tesla P100 com 16 GB de VRAM. O treinamento utilizou 100 épocas, 12 *workers* e batches de 16 imagens.

#### 4.2. Avaliações Intra-Algoritmos

Foram analisadas versões distintas dos algoritmos YOLOv8 e YOLOv11, RetinaNet e Faster R-CNN, com foco no custo-benefício entre desempenho e eficiência computacional. Para o YOLOv8, o modelo YOLOv8n apresentou o melhor custo-benefício, alcançando alta precisão média com baixo tempo de treinamento. Da mesma forma, o YOLOv11n foi o destaque entre as versões do YOLOv11, sendo ambos selecionados para as comparações interalgoritmos.

Figura 5 – Comparação Entre Modelos/Versões.



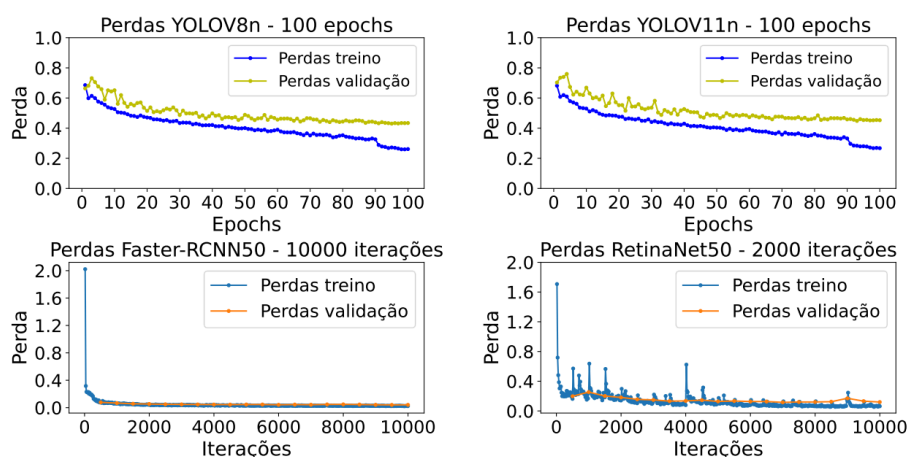
Fonte: autoria própria, 2025.

O RetinaNet, avaliado nas versões com backbones ResNet-50 e ResNet-101, teve melhor desempenho com a ResNet-50 (RetinaNet50), por apresentar menor custo computacional com boa capacidade de detecção. O *Faster R-CNN*, também testado com backbones ResNet-50 e ResNet-101, teve melhor desempenho com ResNet-50, sendo essa a versão escolhida para as comparações.

#### 4.3. Avaliações Interalgoritmos

Para avaliar a eficiência dos modelos treinados, é fundamental analisar o comportamento deles ao longo do processo de treinamento, especialmente com base em métricas como a perda de treinamento e a perda de validação. A partir da análise da Figura 6, observa-se que os modelos YOLOv8 e YOLOv11 demonstram uma estabilização nos valores de perda a partir da epoch 40, indicando que, nesse ponto, já captaram as principais características do conjunto de dados. Por outro lado, os modelos RetinaNet e Faster R-CNN só apresentam essa mesma tendência de estabilização após aproximadamente 2000 iterações. Esse comportamento sugere que tanto o número de epochs quanto o número de iterações pode ser otimizado, reduzindo o tempo total de treinamento sem prejuízo ao desempenho final, o que representa uma vantagem em termos de economia de recursos computacionais.

Figura 6 – Comparação Entre Perda de Validação e Treino.



Fonte: autoria própria, 2025.

A Tabela 1 reforça essa diferença ao demonstrar que, para os modelos YOLO, treinamentos com 40 épocas foram suficientes para alcançar resultados satisfatórios, sem prejuízo no desempenho, ao passo que para RetinaNet e Faster R-CNN, o melhor custo-benefício foi alcançado com cerca de 2000 iterações. Extensões desses treinamentos como 100 épocas para YOLO ou 10.000 iterações para os demais não demonstraram ganhos proporcionais, indicando possível saturação do aprendizado.

Tabela 1 – Resultados do Refinamento (*fine tuning*) dos Modelos.

Modelo	Epochs/Iterações	Tempo de Treinamento	MAP:50
YOLOv8	100	1:21:04	0.992
YOLOv11	100	1:26:56	0.995

RetinaNet50	1000	4:40:43	0.979
<i>Faster R-CNN50</i>	1000	2:37:48	0.977
YOLOv8	40	0:30:08	0.992
YOLOv11	40	0:34:48	0.992
RetinaNet50	2000	1:15:36	0.936
<i>Faster R-CNN50</i>	2000	0:42:14	0.941

Fonte: autoria própria, 2025.

Em relação ao uso de recursos computacionais, os modelos YOLOv8 e YOLOv11 se destacaram como os mais eficientes. Mesmo operando com menor consumo de memória e tempo de processamento, mantiveram elevados níveis de desempenho, evidenciados por um mAP50:95 de 0.992, valor significativamente superior ao obtido por RetinaNet (0.936) e Faster R-CNN (0.941).

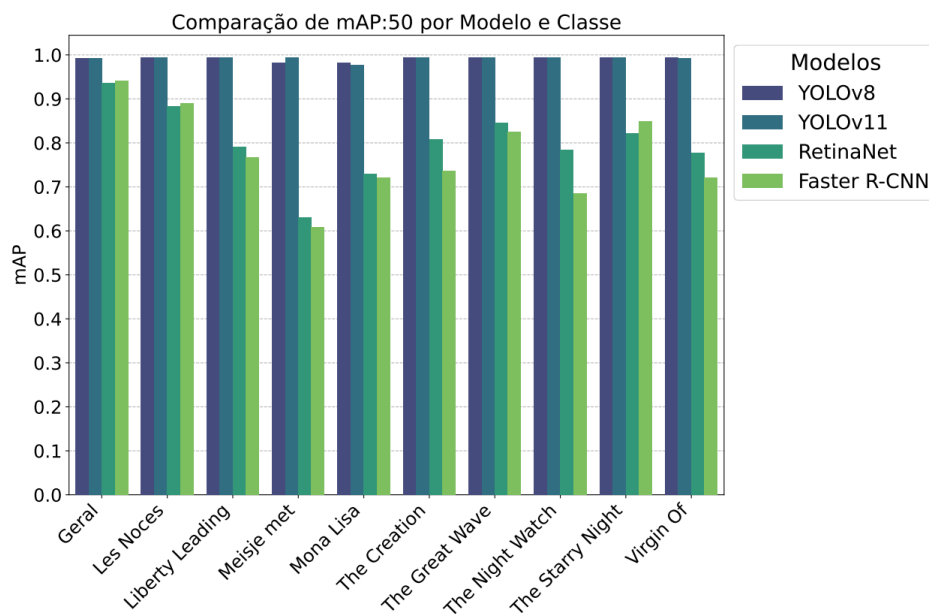
Tabela 2 – Recursos Computacionais Utilizados por Algoritmos.

Modelo	Batch	Workers	Disco (GiB)	RAM(GiB)	Memó. GPU(GiB)	mAP50:95
YOLOv8	16	12	2.4	5.1	2.8	0.911
YOLOv11	16	12	2.4	4.9	3.1	0.918
RetinaNet50	8	4	2.6	2.1	8.8	0.786
Faster R-CNN50	8	4	2.6	2.1	8.4	0.757

Fonte: autoria própria, 2025.

Uma das principais métricas utilizadas na avaliação de modelos de detecção de objetos é a *Mean Average Precision* (mAP), por permitir medir o quão precisas são as predições em relação às anotações reais do conjunto de teste. A Figura 7 apresenta a comparação dos valores de mAP@50 entre os modelos treinados, na qual se observa que os modelos YOLOv8 e YOLOv11 obtiveram os melhores resultados, ambos atingindo a marca de 0,992. Já os modelos RetinaNet e *Faster R-CNN* alcançaram valores de 0,936 e 0,941, respectivamente, o que representa uma leve inferioridade frente aos modelos da família YOLO.

Figura 7 – Comparação de mAP50.



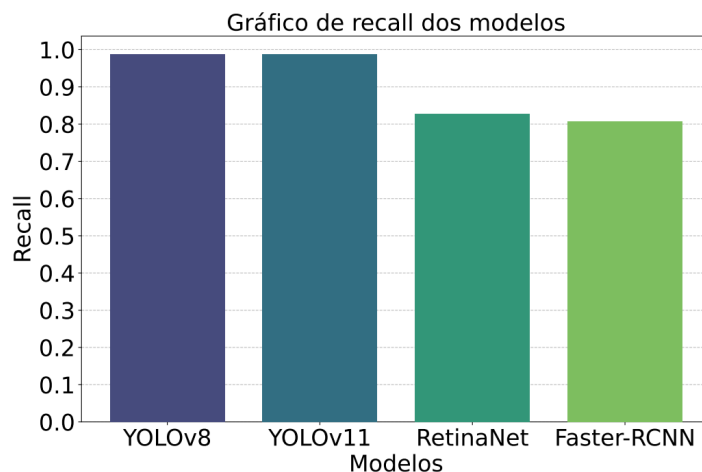
Fonte: autoria própria, 2025.

No entanto, ao se analisar o desempenho por classe, observa-se que essa diferença se acentua em algumas categorias específicas, como nas obras *Meisje met de Parel* e *Mona Lisa*. Nessas classes, os modelos YOLO mantiveram valores de mAP superiores a 0,9, enquanto o RetinaNet e o *Faster R-CNN* apresentaram quedas mais expressivas, com valores abaixo de 0,8. Esses resultados reforçam a maior robustez dos modelos YOLO na detecção de obras com características visuais sutis e complexas.

A Figura 8 apresenta os valores de *recall* obtidos pelos modelos avaliados. Observa-se que os modelos YOLOv8 e YOLOv11 se destacaram com os maiores índices, ambos alcançando 0,987, o que indica alta capacidade na identificação correta dos objetos presentes nas imagens. Em contrapartida, os modelos RetinaNet e *Faster R-CNN* obtiveram desempenhos mais modestos, com valores de *recall* de 0,827 e 0,808, respectivamente.

Essa diferença sugere que esses dois últimos modelos são mais propensos a produzir falsos negativos, ou seja, deixar de detectar objetos que realmente estão nas imagens. Esse comportamento impacta diretamente a confiabilidade da detecção, especialmente em contextos que exigem alta taxa de acerto na identificação.

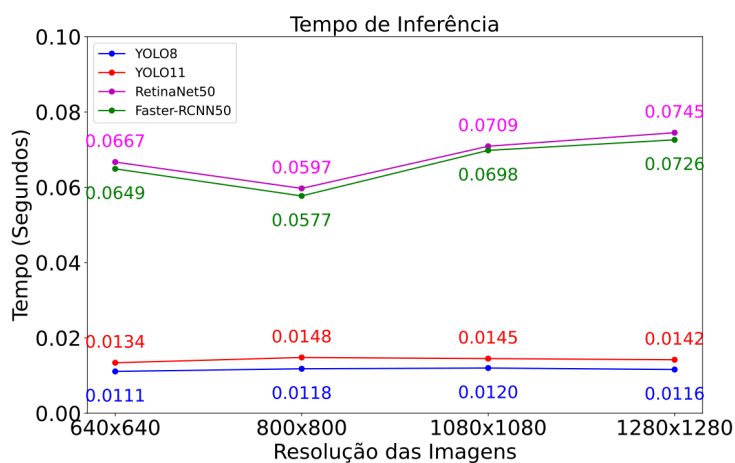
Figura 8 – Comparação de Recall.



Fonte: autoria própria, 2025.

A análise de desempenho também incluiu o tempo de inferência em diferentes resoluções de entrada ( $640 \times 640$ ,  $800 \times 800$ ,  $1080 \times 1080$  e  $1280 \times 1280$ ). Os modelos YOLO, em especial o YOLOv8n, demonstraram tempos de inferência consideravelmente menores em todas as resoluções, o que os torna particularmente indicados para aplicações em tempo real. Em contraste, RetinaNet e Faster R-CNN apresentaram maiores latências, especialmente em resoluções mais altas. Contudo, ambos demonstraram desempenho mais competitivo em  $800 \times 800$ , possivelmente por essa resolução estar mais próxima de sua configuração nativa de treinamento.

Figura 9 – Comparação de Tempo de Inferência.



Fonte: autoria própria, 2025.

## 5. CONCLUSÃO

Este trabalho apresentou uma análise comparativa entre distintos algoritmos de detecção de objetos, com foco em seu potencial de aplicação na promoção da acessibilidade para pessoas com deficiência visual. Os modelos avaliados (YOLOv8, YOLOv11, RetinaNet e *Faster R-CNN*) foram submetidos a uma série de experimentos envolvendo métricas fundamentais como a precisão média (*mAP*), *recall*, tempo de inferência e exigência de recursos computacionais durante o processo de treinamento.

Os resultados obtidos evidenciaram o desempenho superior dos modelos da família YOLO, especialmente o YOLOv8 e YOLOv11, que se destacaram por alcançar elevados índices de *mAP* e *recall*, mantendo ao mesmo tempo um equilíbrio eficiente entre desempenho e consumo de recursos. Essa combinação de precisão e leveza computacional torna esses modelos altamente adequados para aplicações que exigem processamento em tempo real, como é o caso de sistemas voltados à acessibilidade em ambientes culturais e museológicos.

Neste contexto, a utilização de modelos YOLO em uma aplicação voltada à identificação de obras de arte em formato de pintura mostra-se promissora, permitindo a mediação entre o público com deficiência visual e o acervo artístico, contribuindo para democratizar o acesso à arte e à cultura.

Como desdobramento futuro, propõe-se a integração desses modelos em plataformas móveis acessíveis, como aplicativos para *smartphones*, com o objetivo de proporcionar uma interação mais autônoma e inclusiva em espaços expositivos. Além disso, vislumbra-se a possibilidade de expandir e especializar os modelos para abarcar coleções completas de museus ou instituições culturais específicas, permitindo uma experiência de visita assistida ainda mais rica, personalizada e sensível ao contexto histórico e artístico das obras.

## 6. REFERÊNCIAS

- BERNARDO, F. T. *Desenvolvimento de uma interface homem-máquina utilizando visão computacional para acessibilidade digital*. 2019.
- BOESCH, G. *The fundamental guide to Faster R-CNN*. 2024. Disponível em: <https://roboflow.com>. Acesso em: 21 mar. 2025.
- CARVALHO, L. E. S. de; NASCIMENTO, H. A. D. do. Humanidades digitais: performatividades na cultura digital. In: FERNANDES, A. (org.). *Os inventores das obras-primas científicas: uma viagem através dos tempos e das tecnologias*. Goiânia: Publicação CIAR, UFG, 2025. cap. 5, p. 37–49.
- MILANO, D. de; HONORATO, L. B. *Visão computacional*. Campinas: Faculdade de Tecnologia, Universidade Estadual de Campinas, 2010.
- IBGE. Pessoas com deficiência têm menor acesso à educação, ao trabalho e à renda. 2023. Disponível em: <https://www.ibge.gov.br>. Acesso em: 16 mar. 2025.
- KHAN, F. et al. Painting-91: A large scale database for computational painting categorization. *Machine Vision and Applications*, v. 25, p. 1385–1397, 2014.
- LIN, T.-Y. et al. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017. p. 2980–2988.
- MARTINS, P. I. S. R. *A inclusão pela arte: museus e públicos com deficiência visual*. 2008. Dissertação (Mestrado) — Universidade de Lisboa, Portugal.
- NOGUEIRA, M. E. Três museus brasileiros estão entre os 100 mais visitados em 2024, mas nenhum deles é o MASP. *Seu Dinheiro*, 2025. Disponível em: <https://www.seudinheiro.com>. Acesso em: 28 jun. 2025.
- OLIVATTO, T. F. *Identificação automática de rampas de acessibilidade apoiada por visão computacional a partir de imagens panorâmicas street-level*. 2021.
- PATEL, P. et al. Enhancing accessibility through machine learning: A review on visual and hearing impairment technologies. *IEEE Access*, v. 13, p. 33286–33307, 2025.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. p. 779–788.
- TEAM, A. *Albumentations documentation*. 2024. Disponível em: <https://albumentations.ai>. Acesso em: 24 nov. 2024.
- TZUTALIN; CONTRIBUTORS. *LabelImg: A graphical image annotation tool*. 2024. Disponível em: <https://github.com/tzutalin/labelImg>. Acesso em: 11 nov. 2024.

## APÊNDICE A – CÓDIGO DO PROJETO

- <https://github.com/MarcusCarvalho21/ArteAcessivel.git>