

Plataforma de Marcação de Imagens com Inteligência Artificial Integrada

**Arthur Siqueira da Cunha, 081210013@faculdade.cefsa.edu.br, Faculdade Engenheiro Salvador
Arena**

**Danilo Rodrigues Dantas, 081210027@faculdade.cefsa.edu.br, Faculdade Engenheiro Salvador
Arena**

Maik Soares Luiz, 081210040@faculdade.cefsa.edu.br, Faculdade Engenheiro Salvador Arena

Victor Inácio de Oliveira, 14724@cefsa.edu.br, Faculdade Engenheiro Salvador Arena

Resumo

Este artigo apresenta uma plataforma de marcação de imagem assistida por inteligência artificial (IA) para criação de conjuntos de dados personalizados. A proposta combina segmentação interativa com o modelo Segment Anything Model (SAM), permitindo sugestões automáticas de máscara que podem ser rapidamente ajustadas pelo usuário. Desenvolvido como aplicativo desktop com Tauri (React/JavaScript, Rust) e DuckDB, o sistema busca reduzir o tempo e o esforço da anotação, aumentar a consistência e diminuir riscos ocupacionais como LER/DORT e fadiga ocular. Espera-se que a solução beneficie equipes pequenas, trabalhadores da área e pesquisadores independentes.

Palavras-chave: Aprendizado de máquina; Visão Computacional; Marcação de Imagem; Anotação assistida.

Introdução

A inteligência artificial (IA) consolidou-se como uma das forças tecnológicas mais transformadoras do século XXI, possibilitando que máquinas executem tarefas antes restritas à cognição humana (Weng et al., 2024). Entre seus campos de aplicação, a visão computacional tem obtido destaque, permitindo a interpretação e análise automatizada de imagens para resolver problemas complexos em áreas como diagnóstico médico (Joseph et al., 2021), condução autônoma (Bojarski et al., 2016), segurança (Chen et al., 2025), varejo (Shin et al., 2022) e agricultura (Gohil et al., 2024). O avanço de modelos baseados em Redes Neurais Convolucionais (CNNs) e Transformers (Wang et al., 2024) impulsionou a qualidade e a diversidade dessas soluções, mas o desempenho desses modelos depende fortemente da disponibilidade de dados anotados com precisão.

O processo de anotação manual de imagens, embora essencial, apresenta desafios significativos: é demorado, propenso a inconsistências e pode impactar a saúde ocupacional, sendo associado a Lesões por Esforço Repetitivo (LER/DORT) (Hagberg et al., 1995) e fadiga ocular digital (Rosenfield, 2016). Estudos demonstram que a prevalência da fadiga ocular digital ultrapassa 75% em grupos específicos, como profissionais de tecnologia, o que reforça a importância dessas medidas preventivas (Sheppard, 2018). Essas condições são agravadas pela natureza repetitiva da tarefa e pelo tempo prolongado diante de telas, exigindo soluções que conciliem produtividade e ergonomia.

Ferramentas de anotação assistida por IA têm se mostrado promissoras na mitigação desses problemas, aumentando a eficiência e a consistência das marcações quando combinadas com supervisão humana (Dutta e Zisserman, 2019). Abordagens colaborativas, como o Fluid Annotation (Andriluka et al., 2018), demonstram que a interação homem-máquina pode acelerar o processo e reduzir a taxa de erros. Além disso, diretrizes ergonômicas estabelecidas por órgãos como o Ministério da Saúde (Brasil, 2023) e a Organização Mundial da Saúde (OMS, 2020) orientam práticas para minimizar os riscos à saúde.

Nesse cenário, a Plataforma de Marcação de Imagens com Inteligência Artificial Integrada foi concebida para unir recursos de segmentação interativa de ponta a uma interface intuitiva, incorporando recomendações ergonômicas e visando apoiar desde grandes projetos de anotação até iniciativas de equipes pequenas ou pesquisadores independentes.

Nesse cenário, a Plataforma de Marcação de Imagens com Inteligência Artificial Integrada foi concebida para unir recursos de segmentação interativa de ponta a uma interface intuitiva, incorporando recomendações ergonômicas e visando apoiar desde grandes projetos de anotação até iniciativas de equipes pequenas ou pesquisadores independentes.

Tabela 1 – Exemplos de sintomas relacionados a atividade	
LER/DORT	Fadiga Ocular
Redução de força muscular	Visão embaçada
Formigamento	Dores de cabeça
Limitação funcional	Cansaço ocular

Fonte: Própria (2025)

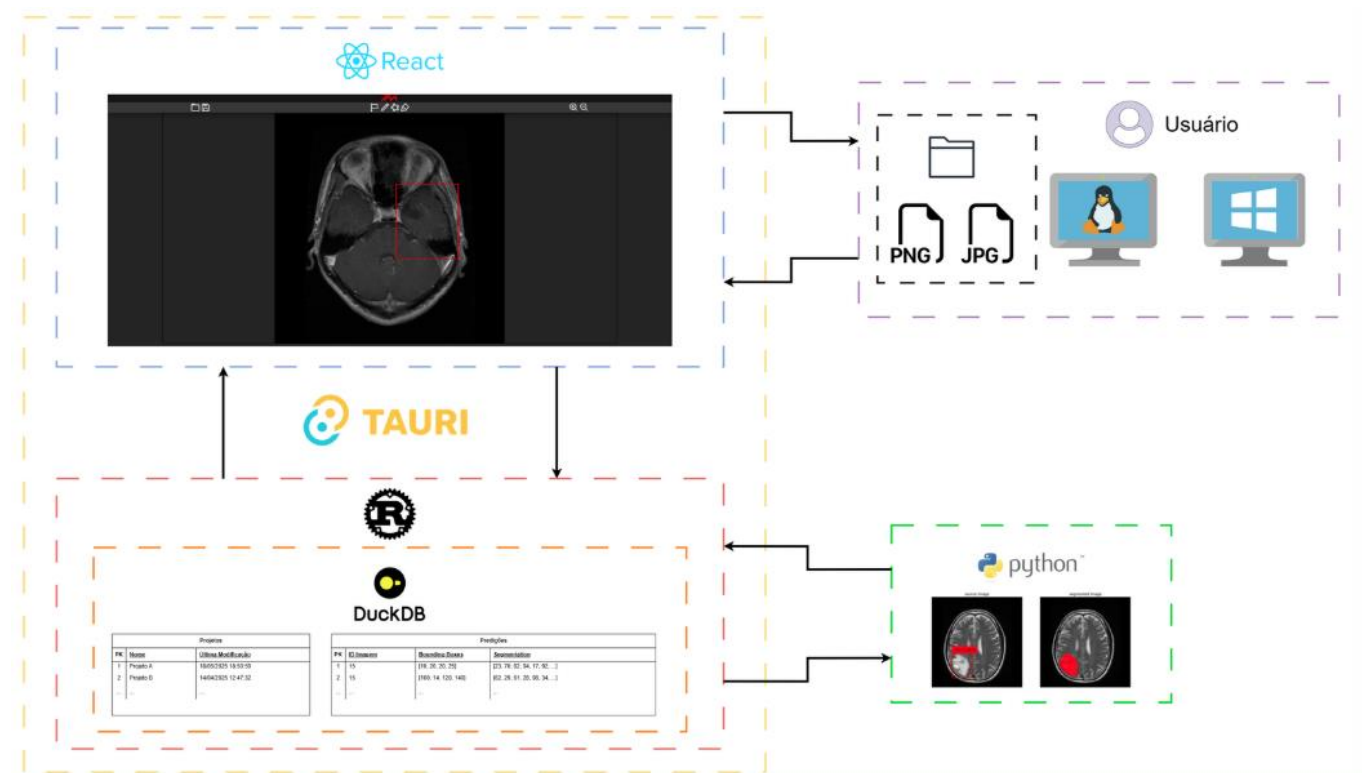
Metodologia

O desenvolvimento da plataforma adotou uma abordagem estruturada, envolvendo a definição da arquitetura, seleção de tecnologias, integração de módulos e elaboração de um fluxo de trabalho ergonômico. Inspirando-se na clareza e objetividade do projeto de estufa automatizada, buscou-se uma solução modular, escalável e de fácil manutenção.

A arquitetura da plataforma está organizada em quatro camadas principais, conforme ilustrado na Figura 1. A primeira delas é a Camada de Usuário, compatível com os sistemas operacionais Windows e Linux, cuja função é fornecer os arquivos de entrada à plataforma, especificamente imagens nos formatos

PNG e JPG. Em seguida, encontra-se a Camada de Interface (Frontend), implementada em React (JavaScript), que disponibiliza ferramentas de anotação e elementos de interação com o usuário, além de oferecer suporte a múltiplos formatos de exportação. A terceira camada é a Camada de Controle (Backend), desenvolvida em Rust, a qual atua como intermediária entre a interface e os módulos de processamento, garantindo elevado desempenho e segurança na execução local. Por fim, a Camada de Processamento de Inteligência Artificial, implementada em Python com o uso das bibliotecas PyTorch e SAM, é responsável por realizar a segmentação automática das imagens com base nos comandos fornecidos pelo usuário.

Figura 1 – Arquitetura do Projeto



Fonte: Própria (2025)

O armazenamento dos dados é gerenciado localmente pelo DuckDB, garantindo preservação e integridade sem depender de serviços externos. O fluxo de operação da plataforma segue etapas bem definidas: carregamento das imagens, geração inicial da segmentação pelo SAM, ajustes manuais refinados, validação dos resultados e exportação em formatos compatíveis com COCO e Pascal VOC.

Essa organização modular permite escalabilidade futura, facilidade na manutenção e integração de novos modelos de IA ou funcionalidades adicionais, mantendo a experiência do usuário intuitiva e eficiente.

Resultados preliminares

Em relação aos resultados preliminares da plataforma, foram considerados valores estimados para a operação assistida, conforme demonstrado na Figura 2. O total de objetos permanece em aproximadamente 57 milhões, enquanto o número de toques por objeto é reduzido de 35 + 1 no processo manual para cerca de 2 + 1 na abordagem com assistência. Essa redução implica uma diminuição significativa no número total de toques, de aproximadamente 2 bilhões para cerca de 172 milhões, representando uma melhoria estimada de 90% na eficiência da anotação.

Figura 2 – Resultados Estimados

LVIS (LARGE VOCABULARY INSTANCE SEGMENTATION)				
	Total de objetos	Toques por objeto	Toques totais	Diferença Percentual
Manual	~ 57 milhões	35 + 1	~ 2 bilhões	100%
Com assistência (estimado)	~ 57 milhões	2 + 1	~ 172 milhões	~ 10%

Fonte: Própria (2025)

Para validar e contextualizar esses resultados, eles foram comparados com o dataset LVIS (Large Vocabulary Instance Segmentation), reconhecido por sua diversidade de classes e complexidade em segmentação de instâncias. A comparação indica que a plataforma, mesmo em estágio preliminar, apresenta desempenho promissor em relação à anotação automatizada de grandes volumes de dados.

A Plataforma de Marcação de Imagens com Inteligência Artificial Integrada demonstrou potencial significativo para reduzir o esforço manual na anotação de imagens, aumentando a eficiência e a consistência dos dados gerados. Mesmo em estágio preliminar, os resultados estimados indicam uma melhoria expressiva no número de interações necessárias, o que pode contribuir para a diminuição de riscos ocupacionais e facilitar o trabalho de equipes pequenas ou pesquisadores independentes. Futuras etapas incluem a validação com datasets reais e a incorporação de novos modelos de IA para ampliar a robustez e a aplicabilidade da plataforma.

Referências

BOJARSKI, M.; TESTA, D. D.; DWORAKOWSKI, D.; et al. **End to End Learning for Self-Driving Cars**. arXiv, 25 abr. 2016. Disponível em: [arXiv:1604.07316](https://arxiv.org/abs/1604.07316).

CHEN, X.; LIU, Y.; LIU, Y.; et al. **A Survey of Security and Privacy Issues of Machine Unlearning**. AAAI, 2025. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/aaai.12209>.

GOHIL, H.; et al. **Computer Vision in Smart Agriculture and Precision Farming**. ScienceDirect, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2589721724000266>.

JOSÉPH, L.; et al. **What is New in Computer Vision and Artificial Intelligence in Medical Imaging**. PMC, 2024. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8245941/>.

SHIN, H.; et al. **AI-Driven Predictive Analytics in Retail: A Review of Emerging Trends and Customer Engagemee Strategies**. ResearchGate, 2023. Disponível em: <https://www.researchgate.net/>

WANG, Z.; et al. **Transformer-CNN Hybrid Network for Low-Light Image Enhancement**. ScienceDirect, 2024. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0097849324002243>.

WENG, J.; et al. **Artificial Intelligence and Social Media on Academic Performance**. ScienceDirect, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405844024055543>

HAGBERG, M.; et al. **Work-Related Musculoskeletal Disorders in the Upper Extremities: A Survey of the Prevalence and Risk Factors in the Swedish Workforce**. *Scandinavian Journal of Work, Environment & Health*, v. 21, n. 6, p. 429-437, 1995. Disponível em: <https://www.jstor.org/stable/40967800>.

ROSENFELD, M. **Computer Vision Syndrome (a.k.a. Digital Eye Strain)**. *Optometry in Practice*, v. 17, n. 1, p. 1-10, jan. 2016. Disponível em: https://www.college-optometrists.org/professional-development/college-journals/optometry-in-practice/volume-17%2C-issue-1/2016-02-computervisionsyndrome_a-k-a-digitaleyestr.

SHEPPARD, A. L.; WOLFFSOHN, J. S. **Digital Eye Strain: Prevalence, Measurement and Amelioration**. *BMJ Open Ophthalmology*, v. 3, n. 1, e000146, 16 abr. 2018. Disponível em: <https://bmjophth.bmj.com/content/bmjophth/3/1/e000146.full.pdf>.

DUTTA, A.; ZISSERMAN, A. **The VIA Annotation Software for Images, Audio and Video**. arXiv, abr. 2019. Disponível em: [arXiv:1904.10699](https://arxiv.org/abs/1904.10699).

ANDRILUKA, M.; UIJLINGS, J. R. R.; FERRARI, V. **Fluid Annotation: A Human-Machine Collaboration Interface for Full Image**

BRASIL. Ministério da Saúde. ***Diretrizes Ergonômicas para Prevenção de Lesões por Esforço Repetitivo (LER/DORT)***. Brasília, 2023. Disponível em: <https://www.gov.br/saude/pt-br>.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). ***Diretrizes para a Proteção da Saúde Ocular Digital***. Genebra, 2020. Disponível em: <https://www.who.int/pt>.