IMPACTO CIENTÍFICO
EDITORA

THE MARKER: ARTIFICIAL INTELLIGENCE-ASSISTED IMAGE MARKING TOOL

THE MARKER: FERRAMENTA DE MARCAÇÃO DE IMAGEM ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL

THE MARKER: HERRAMIENTA DE MARCADO DE IMÁGENES ASISTIDA POR INTELIGENCIA ARTIFICIAL

https://doi.org/10.56238/edimpacto2025.065-003

Arthur Siqueira da Cunha[1], Danilo Rodrigues Dantas[2], Maik Soares Luiz[3], Victor Inácio de Oliveira[4]

ABSTRACT

This undergraduate thesis aims to develop The Marker, an AI-assisted image-annotation tool designed to create customized datasets for computer-vision applications, grounding the study in concepts of artificial intelligence, machine learning, deep neural networks, and ergonomics while emphasizing the importance of image annotation in building effective computational models and the physical impacts associated with repetitive tasks such as RSI, WMSDs, and Computer Vision Syndrome. The applied methodology involved developing a modular application composed of a React graphical interface, processing modules in Rust, execution of the Segment Anything Model (SAM) via Python scripts, and secure storage with AES-GCM encryption; experimental tests were conducted to evaluate accuracy, interference time, the number of manual interactions required, and system performance across different image resolutions. The results indicate that the tool significantly reduces manual effort by automatically suggesting segmentation points, operates offline and on lower-powered machines, provides an improved ergonomic experience, and shows strong potential to accelerate the collaborative creation of visual datasets.

Keywords: Machine Learning. Computer Vision. Labeling.

RESUMO

O presente trabalho de conclusão de curso tem como principal objetivo desenvolver o The Marker, uma ferramenta de marcação de imagens assistida por inteligência artificial voltada à criação de conjuntos de dados personalizados para aplicações de visão computacional. O estudo fundamenta-se em conceitos de inteligência artificial, aprendizado de máquina, redes neurais profundas e ergonomia, destacando a importância da anotação de imagens na construção de modelos computacionais eficazes e os impactos físicos associados a atividades repetitivas, como LER, DORT e Síndrome de Visão computacional. A metodologia aplicada envolveu o desenvolvimento de uma aplicação modular composta por interface

[1] Undergraduated student in Computer Engineering. Faculdade Engenheiro Salvador Arena (FESA)
E-mail: arthursiqcunha@gmail.com

[2] Undergraduated student in Computer Engineering. Faculdade Engenheiro Salvador Arena (FESA)
E-mail: danilo_rodrigues90@hotmail.com

[3] Undergraduated student in Computer Engineering. Faculdade Engenheiro Salvador Arena (FESA)
E-mail: maik.masl@gmail.com

[4] Dr. in Control and Automation Engineering. Faculdade Engenheiro Salvador Arena (FESA).
Email: pro14724@cefsa.edu.br

gráfica em React, processamento em Rust, execução do modelo Segment Anything Model por meio de scripts em Python e armazenamento seguro com criptografia AES-GCM. Foram realizados testes experimentais para avaliar precisão, tempo de inferência, quantidade de interações manuais necessárias e desempenho do sistema em diferentes resoluções de imagem. Os resultados indicam que a ferramenta reduz significativamente o esforço manual ao sugerir pontos de segmentação automaticamente, funcionando em ambiente offline e em máquinas com menor poder de processamento, oferece experiência ergonômica aprimorada e demonstra potencial para acelerar a criação de bases de dados visuais de forma colaborativa.

Palavras-chave: Aprendizado de Máquina. Visão Computacional. Marcação de Imagem.

RESUMEN

El objetivo principal de este trabajo de fin de carrera es desarrollar The Marker, una herramienta de marcado de imágenes asistida por inteligencia artificial destinada a la creación de conjuntos de datos personalizados para aplicaciones de visión artificial. El estudio se basa en conceptos de inteligencia artificial, aprendizaje automático, redes neuronales profundas y ergonomía, destacando la importancia de la anotación de imágenes en la construcción de modelos computacionales eficaces y los impactos físicos asociados a actividades repetitivas, como LER, DORT y síndrome visual informático. La metodología aplicada implicó el desarrollo de una aplicación modular compuesta por una interfaz gráfica en React, procesamiento en Rust, ejecución del modelo Segment Anything Model mediante scripts en Python y almacenamiento seguro con cifrado AES-GCM. Se realizaron pruebas experimentales para evaluar la precisión, el tiempo de inferencia, la cantidad de interacciones manuales necesarias y el rendimiento del sistema en diferentes resoluciones de imagen. Los resultados indican que la herramienta reduce significativamente el esfuerzo manual al sugerir puntos de segmentación automáticamente, funciona en un entorno offline y en máquinas con menor potencia de procesamiento, ofrece una experiencia ergonómica mejorada y demuestra su potencial para acelerar la creación de bases de datos visuales de forma colaborativa.

Palabras clave: Aprendizaje Automático. Visión Artificial. Etiquetado de Imágenes.

1 INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most innovative and transformative technologies of the twenty-first century, and is defined as the ability of a machine to replicate the behavior generated by human intelligence (Wang *et al.*, 2024). Among the fields of artificial intelligence, computer vision (CV) stands out: the use of artificial intelligence to process and extract information from an image, as in the case of LeNet, who pioneered handwritten text recognition (Wang *et al*., 2024). The applications of computer vision by artificial intelligence, such as image classification, are diverse (Khalil *et al.*, 2023): Medical diagnoses as assistance in the detection of pneumonia (Kundu *et al.*, 2021), autonomous vehicle navigation systems (Bojarski *et al.,* 2016), automatic analysis of security services (Chen *et al*., 2025), product research in e-commerce (Shin *et al.,* 2022), and even plant disease detection in agriculture (Gohill *et al.,* 2024). The diversity of applications only highlights the potential of computer vision and its importance in the area of study (Wang *et al.,* 2024). There are a variety of models based on *Convolutional Neural Networks* (CNNs) and *Transformers* designed to address the various problems encountered in computer vision (Wang *et al.,* 2024).

A key ingredient for developing computer vision models is the dataset used in their training (Schuhmann *et al.,* 2022). An example is Microsoft *Common Objects in Context* (MS COCO), which has 328,000 images totaling 2,500,000 tagged objects (Lin *et al.*, 2014). This process occurs with the model receiving an image as input from the dataset, and resulting in an *output* with the information extracted from the image according to the model and the knowledge acquired from its training dataset (Khalil *et al.*, 2023). A key part in creating these datasets is the manual tagging of these images and objects (Papadopoulos *et al.*, 2017). Studies indicate that repetitive and prolonged tasks, such as manually marking images, are associated with a significant increase in the risk of developing Repetitive Strain Injuries (RSI) (Kovashka *et al*., 2016). The use of artificial intelligence tools in image annotation can improve the accuracy and efficiency of the tagging process, especially when combined with human supervision (Dutta and Zisserman, 2019). Implementing these technologies not only reduces the cost and time associated with image annotation but also minimizes worker fatigue, promoting a healthier and more productive work environment (Kovashka *et al*., 2016). Additionally, collaborative interfaces such as *Fluid Annotation* allow for more intuitive interaction between humans and machines, where the AI model provides suggestions that can be quickly adjusted by the human annotator, resulting in a faster and less error-prone annotation process (Andriluka *et al*., 2018).

This scenario reflects a growing movement in the world market, in which companies and institutions invest in solutions aimed at automated image tagging to accelerate the development of computer vision models. As was the case with the *National Geospatial-Intelligence Agency* (NGA) investment of US$ 700 million for the advancement of automated image tagging (Defense Scoop, 2024). Such as Meta, which announced initiatives to expand the collection and processing of visual data at scale (Time, 2024). There are tools already developed for this process, such as LabelImg, Label Studio, CVAT, VIA etc. However, we mapped their functionalities and analyzed that none simultaneously contemplates all the desired qualities among its functionalities.

Repetitive Strain Injuries (RSI) and Work-Related Musculoskeletal Disorders (WMSD) are conditions that affect the musculoskeletal system, resulting from work tasks that require repetitive movements, inadequate postures maintained for long periods, or physical overload (Sanar, 2024; Hagberg *et al*., 1995). Such injuries are often diagnosed in professionals who perform functions such as typing, assembly on production lines, or continuous use of manual devices (Armstrong *et al*., 1993; Hagberg *et al*., 1995). The most common symptoms include localized pain, tingling, reduced muscle strength, and functional limitation, especially in the upper limbs and cervical spine (Sanar, 2024; Armstrong *et al*., 1993).

Another side effect can be visual fatigue, also known as Computer Vision Syndrome (CVS) or digital eye strain which is defined as a set of visual and eye symptoms that arise after long periods of focusing on digital screens such as computers, *tablets,* and *smartphones* (Sheppard and Wolffsohn, 2018). The most common symptoms include blurry vision, dry eyes, headaches, eye strain, and general discomfort around the eye region (Rosenfield, 2016). The condition is recognized as a growing problem in the digital age, especially among workers who spend long hours in front of screens (Gowrisankaran and Sheedy, 2015). Activities with constant visual focus, such as data labeling, increase eye strain and can compromise accuracy and user comfort (Coles-Brennan *et al.,* 2019). Studies show that the prevalence of digital eye strain exceeds 75% in specific groups, such as technology professionals, which reinforces the importance of these preventive measures so that visual comfort and task accuracy are maintained (Sheppard, 2018).

Regulatory Standard No. 17 (NR-17) that deals with ergonomics in the work environment establishes minimum parameters for the adaptation of working conditions to the characteristics of workers, especially in activities that involve repetitive effort, such as typing, in order to preserve the health of the worker Among its recommendations, it is limited to 8,000 real touches per hour of work, aiming to prevent RSI and WMSD (Brazil, 2023). In the same sense, the World Health Organization (WHO) points out that prolonged use of screens without breaks can cause eye fatigue and musculoskeletal problems. Therefore, regular breaks of 10 to 15 minutes are recommended for every hour of continuous use of digital devices, as a way to prevent these effects (World Health Organization, 2020). Thus it takes a user approximately 4,000 hours to mark 1,000,000 objects.

Based on these conditions, The Marker is proposed, an image tagging tool assisted by artificial intelligence, with the objective of assisting the creation of personalized and encrypted database sets. The proposal seeks to reduce the number of actual touches per image tagging, directly impacting the total time needed to complete the activity, another desired effect is the incentive for smaller teams, such as independent researchers and users with less technological familiarity, to be able to develop their own datasets through a simple and accessible installation with an ergonomic interface, intuitive and collaborative.

2 THEORETICAL FRAMEWORK

2.1 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence (AI) and machine learning (*Machine Learning*) are fundamental fields of computer science, aimed at the development of systems capable of learning, generalizing, and making decisions based on data, these technologies have been widely explored and consolidated as pillars of digital transformation in various sectors, driving advances in areas such as health, security, education, transportation, industrial automation (Morais, 2020).

2.1.1 Fundamental concepts

Artificial intelligence (AI) can be defined as the field of computer science that seeks to simulate human cognitive processes through computational systems. This area is divided into two strands: the so-called strong AI, which seeks to fully reproduce cognition, and weak AI, focused on specific tasks (Russell; Norvig, 2010). While strong AI is still restricted to philosophical debates, it is weak AI that we find in practical applications such as virtual assistants and recommendation systems (Goodfellow; Bengio; Courville, 2016).

2.1.2 Machine Learning and Deep Learning

Machine learning is one of the core areas of AI, responsible for creating algorithms that learn from data. Mitchell (1997) defines it as the ability of a computer program to improve its performance in a given task as it acquires experience. This concept paved the way for advances in areas such as health and e-commerce (Lecun; Bengio; Hinton, 2015).

In recent years, deep learning has established itself as the main technique, thanks to the use of deep neural networks. Goodfellow, Bengio, and Courville (2016) explain that this learning model eliminates the need to manually create attributes for the data, enabling superior results in computer vision, machine translation, and speech recognition. This type of approach is fundamental for The Marker, as it makes it possible to implement automatic segmentation and suggest annotations to the user (Krizhevsky; Sutskever; Hinton, 2012).

2.2 COMPUTER VISION

2.2.1 Classical methods

Computer vision is the area of AI that seeks to enable systems to extract information from images and videos. For a long time, this task was performed by classical methods, such as SIFT and SURF, designed to identify points of interest in images with different scales and rotations (Szeliski, 2010). Although effective in controlled situations, these methods required manual parameterization and often failed in noisy environments (Everingham *et al*., 2010). Even with limitations, these algorithms created the basis for the adoption of more modern approaches, especially convolutional neural networks (Lecun; Bengio; Hinton, 2015).

2.2.2 Advances with neural networks

The big leap came in 2012, when the AlexNet network surpassed previous methods in the ImageNet challenge, becoming a game-changer for the area (Krizhevsky; Sutskever; Hinton, 2012). Since then, more advanced architectures, such as Faster R-CNN (Ren *et al*., 2015) and YOLO (*You Only Look Once*) (Redmon *et al*., 2016), have come to dominate the landscape, allowing real-time recognition. These advances have driven applications in safety, autonomous vehicles, and medical diagnostics (Szeliski, 2010). These networks represent the technological basis of assisted annotation algorithms, which enable automatic marking suggestions.

2.2.3 Intersection-over-Union

The *Intersection-over-Union* (IoU) metric evaluates the overlap of two areas, with the first area referring to the model's inference and the second area being the reference area. 100% overlap means that the areas are identical, and 0% overlap indicates that they are entirely different.

2.2.4 Segment Anything Model (SAM)

Meta AI introduced the *Segment Anything Model 2.1* (SAM), which introduced generalist segmentation: a single model capable of handling varied objects without the need for specific training (Kirillov *et al*., 2023). This feature reduces data preparation costs and expands the possibilities of use in different scenarios. SAM can be exploited to automatically suggest regions of interest in images, reducing manual effort and increasing the quality of annotations (Paszke *et al*., 2019).

2.3 ANNOTATION OF IMAGES IN ARTIFICIAL INTELLIGENCE

2.3.1 Importance of annotation

The process of annotating images is one of the most delicate steps in computer vision, as it defines the quality of the data used in model training. Inconsistent data results in unreliable algorithms, according to the logic "*garbage in, garbage out*" (Zhou, 2018). Sager, Janiesch, and Zschech (2021) point out that in many AI projects, annotation is more time-consuming than model development. This is even more evident in areas such as medicine, where only specialists can perform labeling (Rajpurkar *et al*., 2017).

2.3.2 Annotation Tools

Several tools have been created to facilitate annotation. LabelImg is widely used for marking *bounding boxes* (Tzelepis *et al*., 2021). LabelMe, developed at MIT, offers more flexible capabilities, allowing the creation of polygons for scenarios that require precision (Russell *et al*., 2008). The VGG Image Annotator (VIA) is lightweight and runs directly in the browser, with no need for installation (Dutta; Zisserman, 2019). Other solutions seek to add artificial intelligence to the process, such as Fluid Annotation (Andriluka; Uijlings; Ferrari, 2018) and Extreme Clicking (Papadopoulos *et al*., 2017). These initiatives point to a trend of collaboration between humans and machines in data annotation.

2.3.3 Dataset formats

Annotation formats also play a crucial role. Pascal VOC organizes information in XML (Everingham *et al*., 2010), while COCO uses JSON to enable richer descriptions, including segmentations and key points (Lin *et al.*, 2014). On the other hand, YOLO adopts TXT with normalized coordinates (Redmon *et al*., 2016). By supporting multiple formats, The Marker enables greater flexibility and compatibility with different AI workflows (Abadi *et al*., 2016).

2.4 ERGONOMICS AND OCCUPATIONAL HEALTH

Repetitive activities such as manual note-taking can cause negative health effects, including Repetitive Strain Injuries (RSIs) and Work-Related Musculoskeletal Disorders (WMSDs). Hagberg, Silverstein and Wells (1995) already highlighted these risks in typing activities and continuous use of devices. In Brazil, Regulatory Standard No. 17 establishes ergonomic parameters to prevent such conditions (Brasil, 2023). The World Health Organization (WHO, 2003) recommends regular breaks every hour of use of digital devices

Another recurring problem is digital eye strain, also called *Computer Vision Syndrome*. According to studies, more than 70% of IT workers report symptoms such as headaches and blurred vision (Rosenfield, 2016; Sheppard; Wolffsohn, 2018). The project proposal seeks to minimize these effects by reducing the number of clicks and manual interactions, contributing to a healthier work environment (Merrill; Alleman, 2012).

Based on the concepts presented about artificial intelligence, computer vision, image annotation and ergonomics, he defined the proposed architecture for The Marker. The integration of these fundamentals allowed us to design a tool capable of uniting computational efficiency and user well-being, reconciling technical aspects of image processing and segmentation with ergonomic and accessibility principles.

3 METHODOLOGY

The project architecture as shown in Figure 1 was defined to work in a modular way, with *front-end*, *back-end*, artificial intelligence and database modules. By making them connect in an agnostic way by not being connected as a monolithic or homogeneous piece of code, but by creating a scenario where they communicate through APIs so that each module can utilize the appropriate programming languages and *frameworks* for their specific functions.

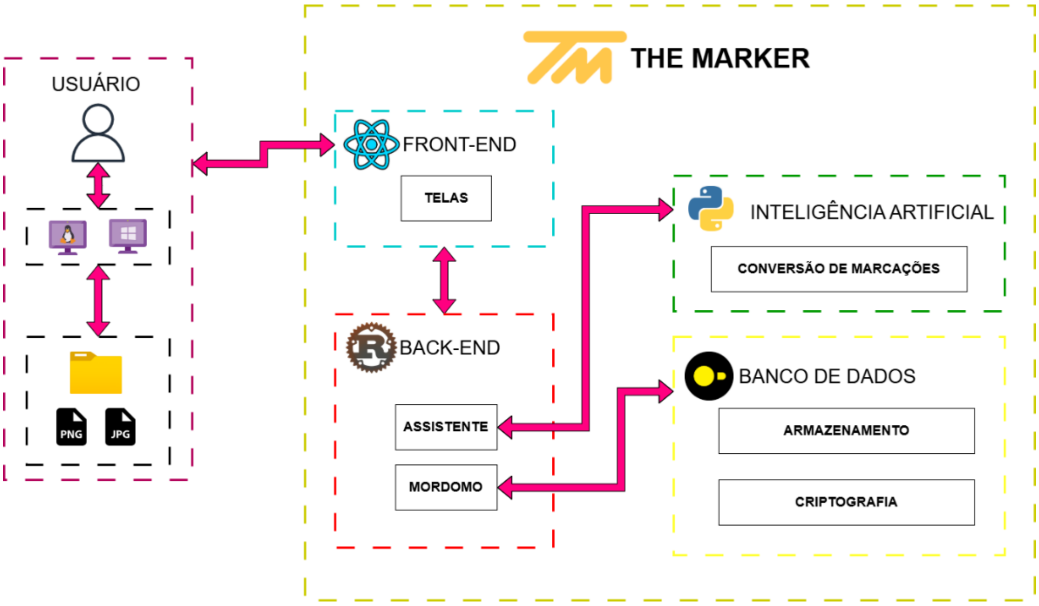
The *front-end* was developed with React for its features of creating modern and dynamic interfaces, also for its presence in Web development that offers a lot of study material and support for its use. In addition to the presence of the community that provides functional and graphical libraries that facilitate development. Used to create elements such as the infinite slide screen with the Konva library. The interface communicates directly with the processing layer through Tauri, which acts by interconnecting the *front-end* with the *back-end* layer, developed in Rust. In the *back-end layer*, there are the logical operations of the system in question of communication between modules. Managing and allowing them to communicate.

The database layer is structured using DuckDB, where data is stored in the structured tables and queried through SQL commands. Thus, each project generates a unique file that can be stored with or without encryption, promoting high performance and portability. Thus, future expansions can be implemented in isolation, without compromising the structure of other existing projects when they are created, updated or deleted.

Artificial intelligence exists as a layer of *Python scripts* that have been compiled into binaries and are queried by the main process through sub-processes of the operating system. The SAM model was chosen for its ability to propose masks of an area filled with objects without prior training. Capacity known as *zero-shot*.

Figure 1

Application architecture



Source: The authors (2025)

These components and modules are encompassed by the Tauri framework, which was chosen because it has a smaller size in the compiled file and better performance when compared to alternatives such as Flutter and Electron, as shown in Table 1. The criteria analyzed included: year of release, popularity among developers, size of files generated, and the number of dependencies required. During the configuration of Tauri, the use of React with JavaScript was defined.

Table 1

Comparison of frameworks sorted by year of publication

Framework	Year of publication	Stars on GitHub	Weekly downloads	Generated files	File size
Tauri	2022	92k	130k	1	2 MB
Flutter	2017	170k	?	100	50 MB
Electron	2013	117k	831k	20	47 MB

Source: The authors (2025)

For local data management, it was chosen to use DuckDB because it is an *open source* project, unlike traditional systems such as MySQL or PostgreSQL, it can be run locally and without the need for external servers, this decision contributed to the portability and independence of the system, allowing the user to manipulate the data individually, without additional facilities. The project was compiled as a cross-platform desktop application compatible with Windows and Linux-Debian, ensuring wide accessibility and practicality of use by end users. All the definitions of the tools used in each module are presented in Table 2.

Table 2

List of choices by function

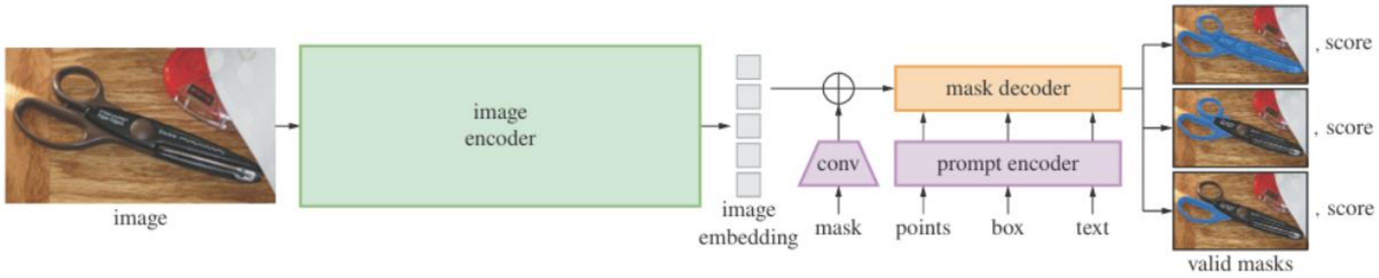
Function	Opted tool
Front-End	React
Back-End	Rust
Database	DuckDB
Operating system	Windows and Linux-debian

Source: The authors (2025)

The SAM model was developed by the team at Meta AI (formerly Facebook AI Research) and made available in 2023. The model was trained with more than 1 billion masks in 11 million images that uses a Pytorch-based architecture incorporating an *image encoder* (based on the Vision Transformer – ViT) and an *encoder prompt* as shown in Figure 2 that allows the model to receive instructions via points, bounding boxes or previous masks. This versatility allows SAM to perform segmentations in an interactive and real-time way, even without training for specific tasks, unlike conventional models that had limitations in scenarios not previously trained. The SAM proved to be versatile and robust, allowing greater autonomy in image marking. To manipulate the images and feed the model for inference, official model support through the Transformers library was used.

Figure 2

SAM Architecture



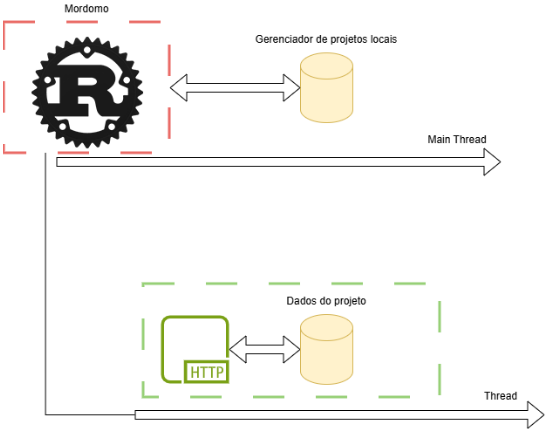
Source: Kirillov (2023)

During the development, some limitations and difficulties were identified that impacted the choice of the methodological process, the Electron *framework* was not used due to the high memory consumption and the final size of the generated packages and Flutter, however, presented a lower maturity in *desktop applications*, in addition to requiring additional packages that increased the complexity of the installation.

In order to have the option of synchronizing the data in a local network, the butler system represented in Figure 3 is used. This system exists as an asynchronous thread of the operating system that creates a server on an ephemeral network port, allowing the *back end* of the responsible machine and other machines on the local network to communicate via HTTP requests.

Figure 3

Butler system

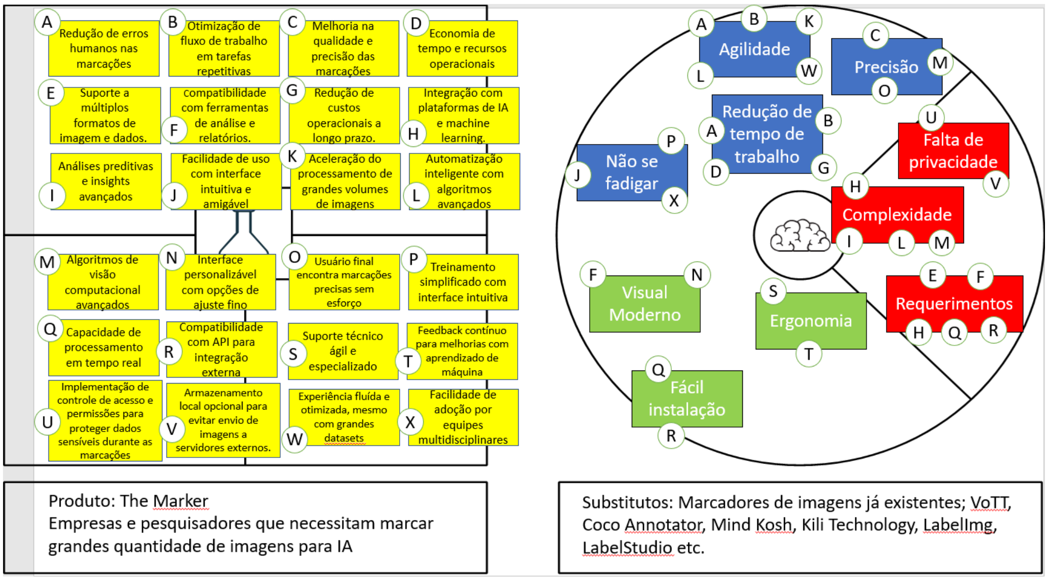


Source: The authors (2025)

The project began with the identification of the market opportunity aimed at creating more intuitive and efficient note-taking tools, the methodology was initially structured by defining the general architecture of the system, establishing the essential modules. From this architecture base, the value proposition of The Marker was elaborated, presented in figure 4, which highlights the importance of reducing operational effort through a modern, accessible and visually organized interface, combined with simplified interaction flows, the development of an experience that prioritized the reduction of repetitive movements was defined as a central guideline, visual clarity and ease of use, allowing researchers, students, and small teams to create personalized databases with less physical and cognitive load. This methodological guidance guided the design, usability, and implementation steps ensuring that The Marker was conceived from the ground up with a focus on efficiency, comfort, and affordability.

Figure 4

Value Proposition Analysis


A Redução de erros humanos nas marcações
B Otimização de fluxo de trabalho em tarefas repetitivas
C Melhoria na qualidade e precisão das marcações
D Economia de tempo e recursos operacionais
E Suporte a múltiplos formatos de imagem e dados.
F compatibilidade com ferramentas de análise e relatórios.
G Redução de custos operacionais a longo prazo.
H Integração com plataformas de IA e machine learning.
I Análises preditivas e insights avançados
J Facilidade de uso com interface intuitiva e amigável
K Aceleração do processamento de grandes volumes de imagens
L Automatização inteligente com algoritmos avançados
M Algoritmos de visão computacional avançados
N Interface personalizável com opções de ajuste fino
O Usuário final encontra marcações precisas sem esforço
P Treinamento simplificado com interface intuitiva
Q Capacidade de processamento em tempo real
R Compatibilidade com API para integração externa
S Suporte técnico ágil e especializado
T Feedback contínuo para melhorias com aprendizado de máquina
U Implementação de controle de acesso e permissões para proteger dados sensíveis durante as marcações
V Armazenamento local opcional para evitar envio de imagens a servidores externos.
W Experiência fluída e otimizada, mesmo com grandes datasets
X Facilidade de adoção por equipes multidisciplinares
A B K Agilidade L W
C Precisão M O
Redução de tempo de trabalho A B D G
U Falta de privacidade V
P Não se fadigar J X
H Complexidade I L M
E F Requerimentos H Q R
F N Visual Moderno
S Ergonomia T
Q Fácil instalação R

Produto: The Marker
Empresas e pesquisadores que necessitam marcar grandes quantidade de imagens para IA

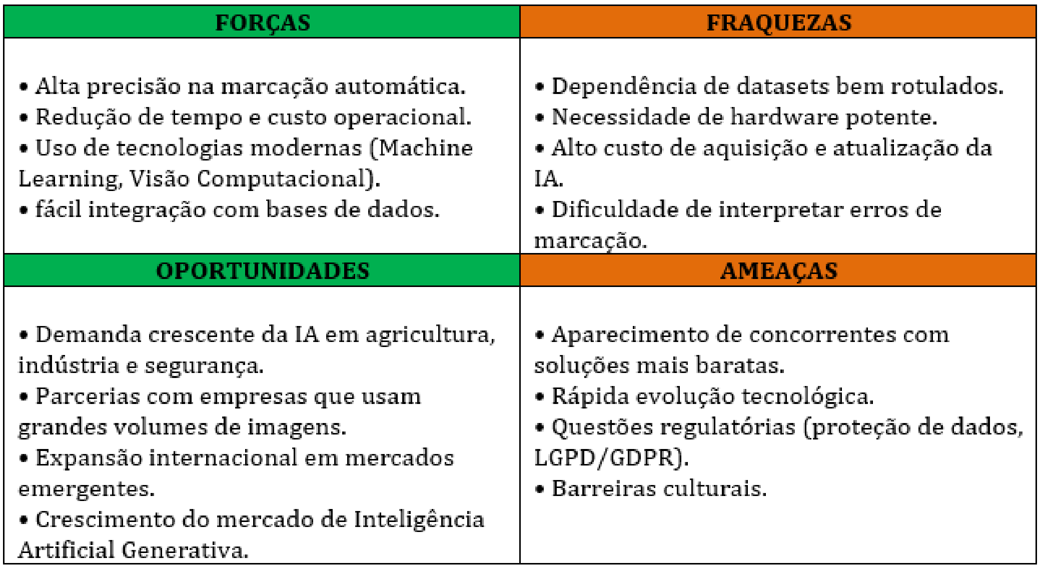
Substitutos: Marcadores de imagens já existentes; VoTT, Coco Annotator, Mind Kosh, Kili Technology, LabelImg, LabelStudio etc.

Source: The authors (2025)

Bringing with it the SWOT analysis (or SWOT) to analyze the strengths, opportunities, weaknesses and threats that the tool could face once available. As shown in Figure 5, it is possible to denote that the forces of The Marker contain high accuracy in marking, and an opportunity arises from the increased demand for the creation of visual data sets for image recognition processes.

Figure 5

SWOT Analysis



FORÇAS	FRAQUEZAS
• Alta precisão na marcação automática. • Redução de tempo e custo operacional. • Uso de tecnologias modernas (Machine Learning, Visão Computacional). • fácil integração com bases de dados.	• Dependência de datasets bem rotulados. • Necessidade de hardware potente. • Alto custo de aquisição e atualização da IA. • Dificuldade de interpretar erros de marcação.
OPORTUNIDADES	**AMEAÇAS**
• Demanda crescente da IA em agricultura, indústria e segurança. • Parcerias com empresas que usam grandes volumes de imagens. • Expansão internacional em mercados emergentes. • Crescimento do mercado de Inteligência Artificial Generativa.	• Aparecimento de concorrentes com soluções mais baratas. • Rápida evolução tecnológica. • Questões regulatórias (proteção de dados, LGPD/GDPR). • Barreiras culturais.

Source: The authors (2025)

Considering these aspects, a comparison table was created between The Marker's proposals and those already implemented by products implemented in the market. In Table 3 we can analyze the relationship of external projects in relation to their installation (the level of complexity to install or use the tool on a machine), the type of license (whether open for commercial use or not), whether it works *Offline* (or requires *an internet* connection), whether it has artificial intelligence assistance (or all appointments are manual) and whether it has synchronization between different machines (or if the projects can be accessed only on the main user's machines).

Table 3

Relationship between the market and needs

Tool	**Installation**	**License**	**Offline**	**Assistance**	**Synchronization**
LabelImg	Simple	Open	Yes	No	No
Label Studio	Simple	Closed	Yes	No	No
CVAT	Complex	Open	Yes	No	No
SuperAnnotate	Simple	Closed	No	Yes	Yes
Makesense.ai	Simple	Closed	No	Yes	Yes
VGG VIA	Complex	Open	Yes	No	No

Source: The authors (2025)

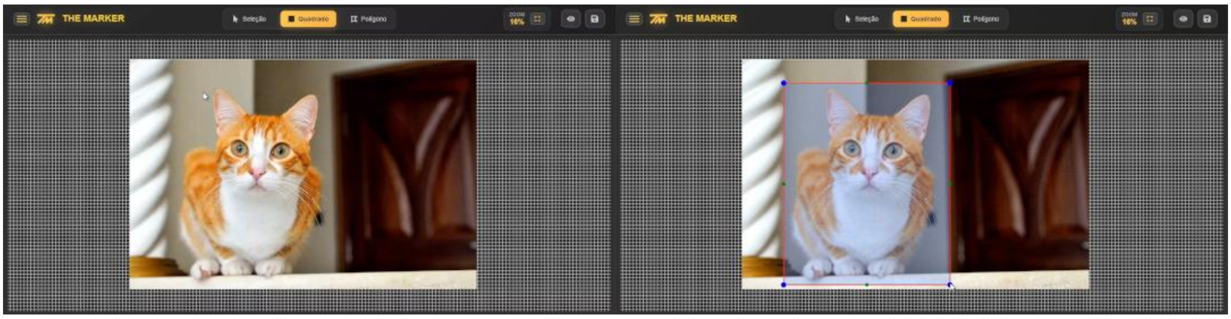
The results collected were made on a notebook machine with Intel I5 10400K processors, 12GB of RAM, 250GB of storage memory and running the Windows 11 operating system. The tests were done without an *available internet* connection .

4 RESULTS AND DISCUSSION

The first results of The Marker begin with the implementation of tagging tools for object detection and segmentation. Both options are available on the screen in the format of three buttons: "Selection", "Square" and "Polygon", being respective to the options of: "do not make markings", "make object detection markings" and "make object segmentation markings". The markings are made with blue and green dots and red lines for better contrast against the other options of the tool. As evidenced in Figure 6, the object detection process, in which the system identifies regions based on the markings made by the user.

Figure 6

Marking for object detection

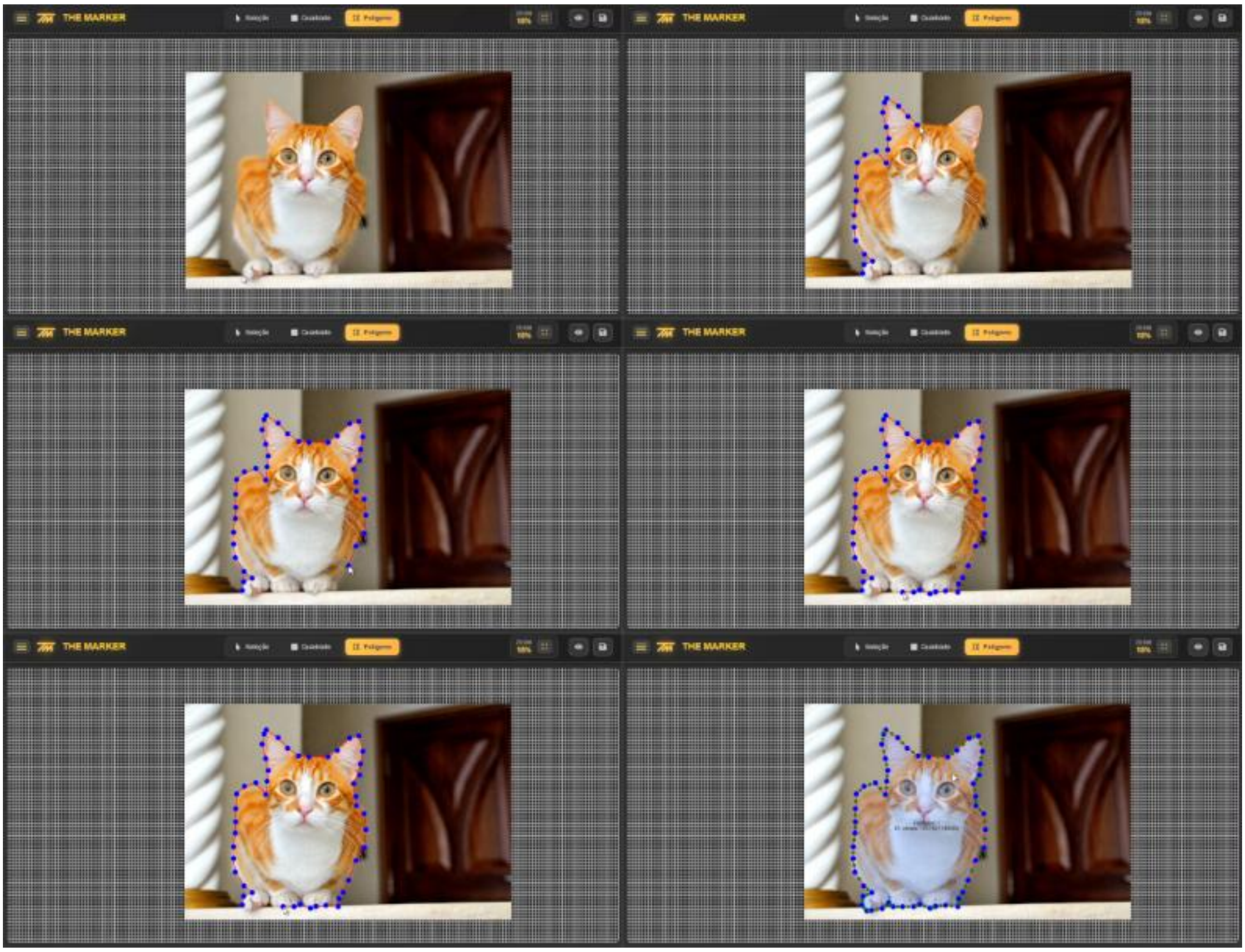


Source: The authors (2025)

Figure 7 presents the process of object segmentation, which aims to highlight more accurately the areas belonging to each element of the image. At this stage, it is essential to generate more detailed and consistent masks, fundamental in the creation of datasets, the system also displays the boundaries with contrasting colors to facilitate the distinction between different segmented regions, thus providing a more intuitive and organized visual experience.

Figure 7

Markup for object segmentation

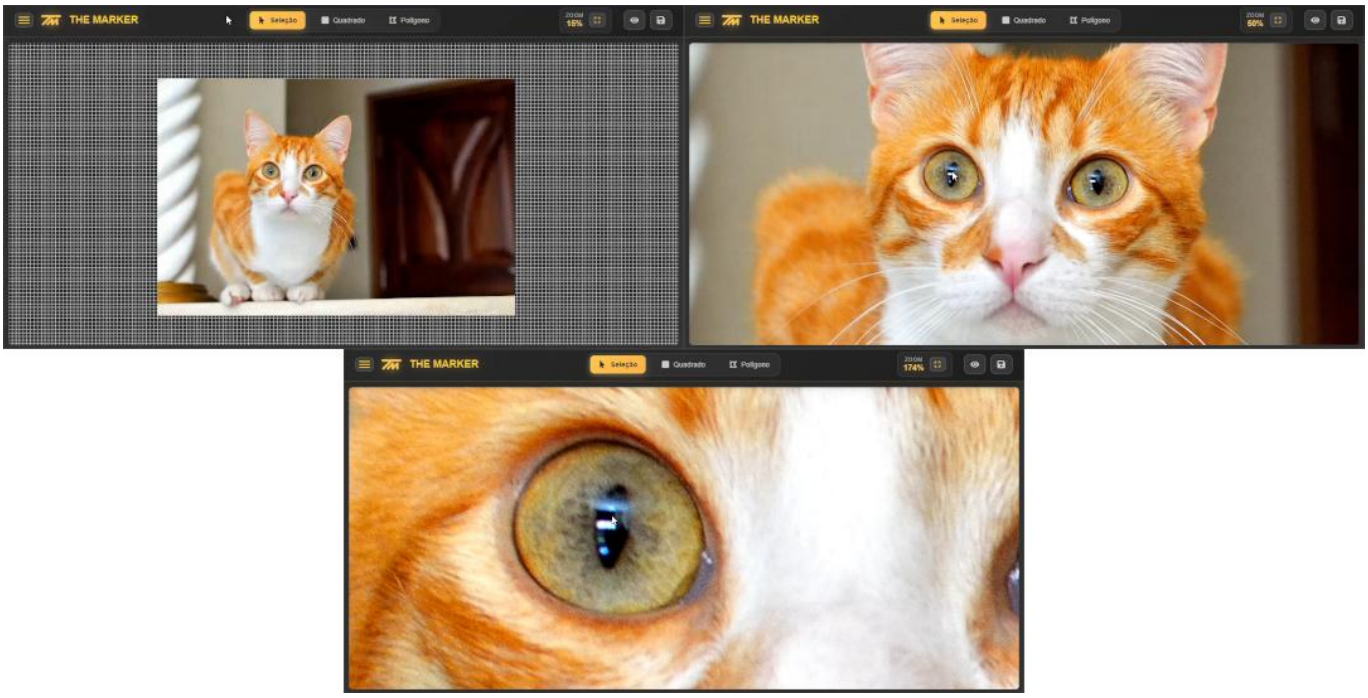


Source: The authors (2025)

In addition, the images are shown on an infinite screen as shown in Figure 8 that proposes vertical and horizontal freedom for the user to be able to adjust the position of the image on the screen as they prefer without losing the coordinate information of the markup. In addition, a magnification and zoom out functionality (popularly known as "*zoom-in*" and "*zoom-out") has been implemented*, which allows the user to focus on smaller details of the image to mark with better refinement.

Figure 8

Image magnification

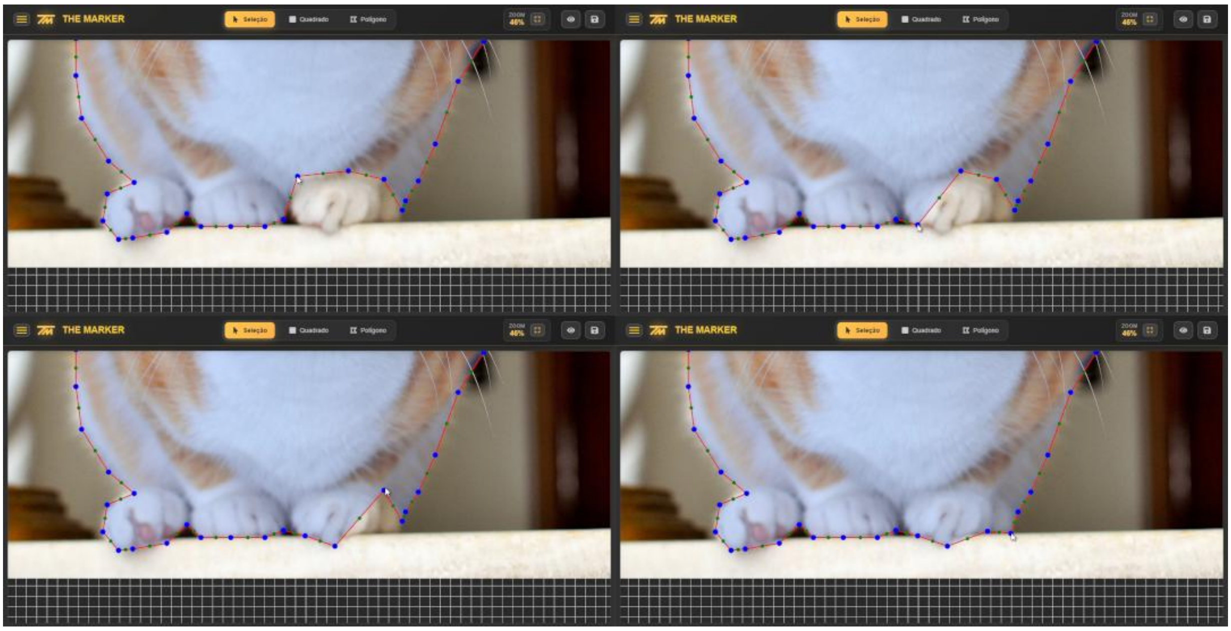


Source: The authors (2025)

In addition to the functionality of enlarging and reducing the image, the option to move marking points that have already been made has been implemented, allowing that even if mistakes have been made during the marking process, the points can be moved to the correct positions, as shown in Figure 9 where the cat's paw was purposely forgotten and then adjusted later.

Figure 9

Markup Edit

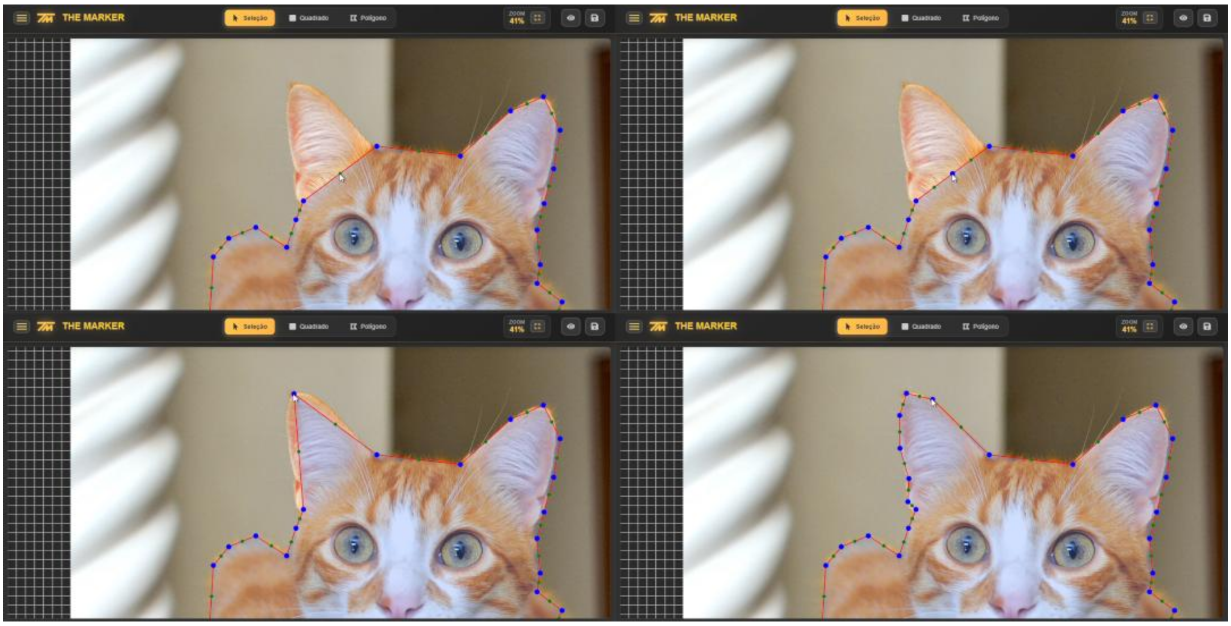


Source: The authors (2025)

In addition to editing the position of markings already made, there is the option for the user to create new intermediate points (represented by green dots) as shown in Figure 10. This way it is possible to increase the resolution of the polygon and be able to move its positions to fill areas that had been left out and that just moving the existing points would not fill.

Figure 10

Later addition of markup

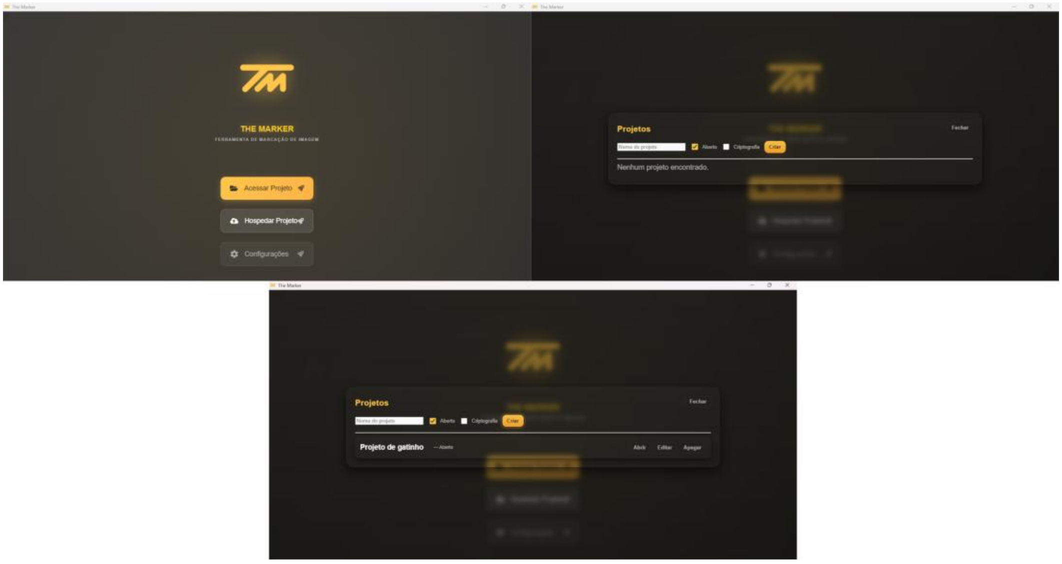


Source: The authors (2025)

The user also has the possibility to organize sets of working images, called "Projects", as shown in Figure 11, this functionality was developed to facilitate the management of different databases and tagging contexts within The Marker. On the main screen, we find the options "Host Project" that allows the user to create a new workgroup and make them available to access the network, we host it directly on their own machine, "Access Project" that makes it possible to connect to an existing workgroup hosted on other devices, promoting collaboration between different users and distributing the flow of appointments in an integrated way and "settings".

Figure 11

Later addition of markup



Nenhum projeto encontrado.

Projeto de gatinho — Aberto

Source: The authors (2025)

For the project, a database was created and with the option to encrypt the data storage with AES-GCM encryption using 256-bit keys. In this way, each project has a different key when stored and allowing that even if the equipment is shared among several people, only those who have the correct credentials can access it.

In Table 4 we can see the comparison of a test execution that compares a process repeated a hundred times of: create a database, create a table and fill this table with five items. It can be seen that even though the execution time in unencrypted mode is faster, the mode with encryption shows an increase of 4 milliseconds.

Table 4

Comparison between databases with and without encryption

Mode	Test duration (ms)
No encryption	6
With encryption	10

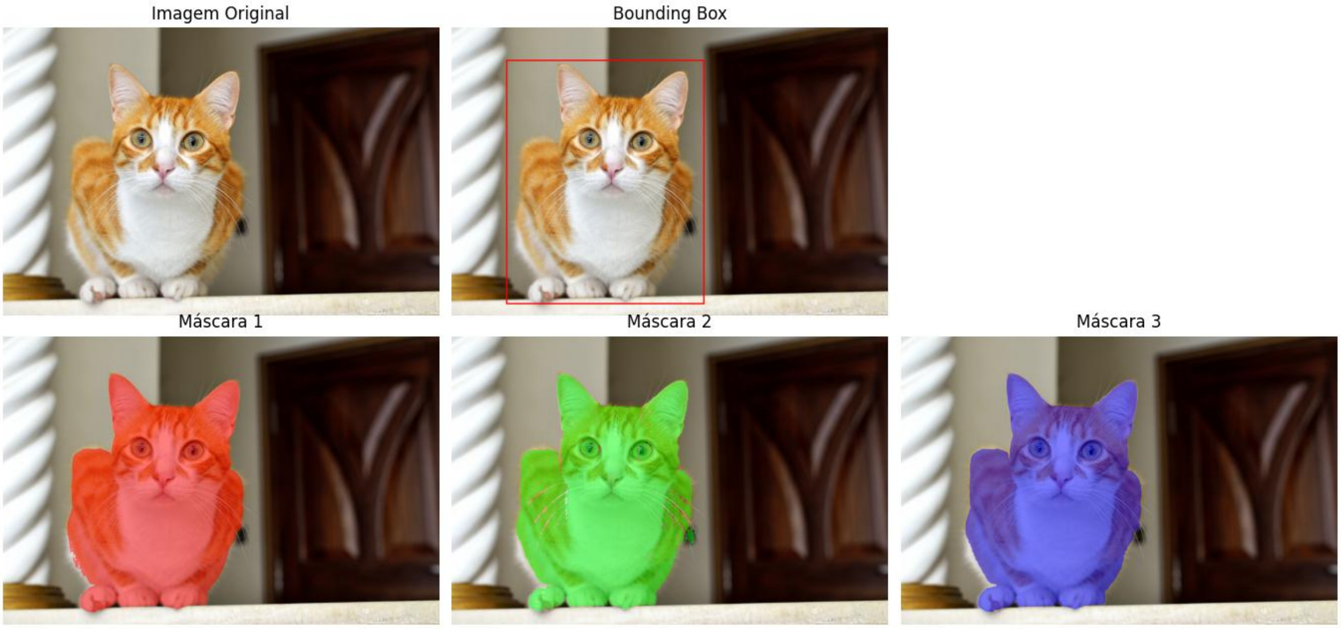
Source: The authors (2025)

This time increase can be considered negligible for The Marker scenario because the parsing time between database inserts is not expected to last less than one second, and considering the level of security that AES encryption provides with a 256-bit key.

With the implementation of SAM 2.1, the artificial intelligence model is able to analyze the region of a *manually tagged bounding box* and suggest three masks with the highest probability of being the desired object. As shown in Figure 12, where the *bounding box* demarcates the region where the cat is and the masks are proposed based on it.

Figure 12

Proposition of masks

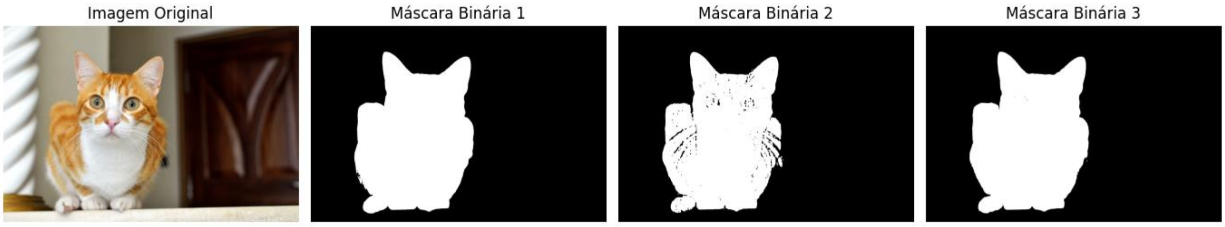


Source: The authors (2025)

Masks are image information with the same dimension as the original image, but they mark in black (value 0) where the mask does not exist and white (value 255) where the mask exists. It alone does not allow the project to identify the segmentation of the image for the user. This segregation of the mask and the background of the image can be seen in Figure 13.

Figure 13

Binary masks

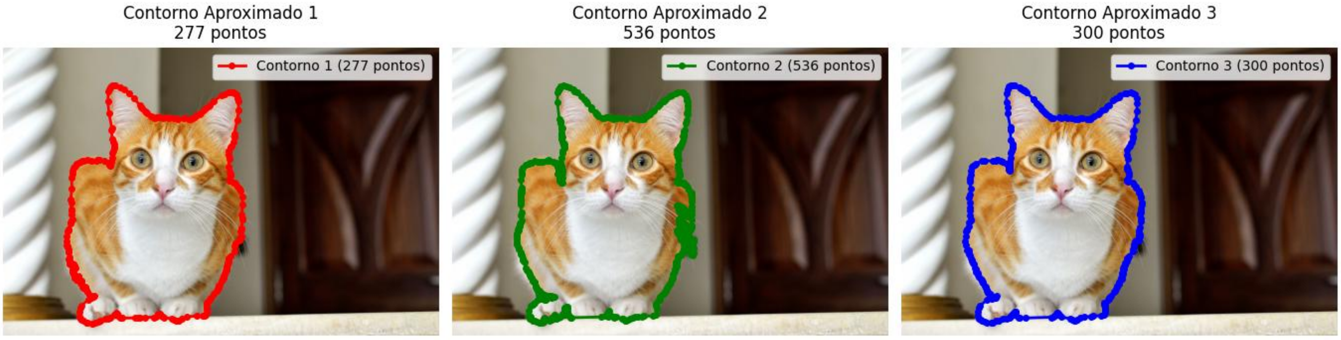


Source: The authors (2025)

To do this, the Python library *scikit-image* is used to apply image treatment to the proposed mask in order to convert the contours of the mask into marking points. As shown in Figure 14, this method is able to transform the bounding *box* into segmentation contours with hundreds of points that accurately track organic figures.

Figure 14

Converting masks to segmentation



Source: The authors (2025)

With the process defined, a study was set up that demonstrates the processing time of each step and the consumption of random access memory (RAM). As seen in Frame 5, the longest processing time took 8.58 seconds in CPU processing and consumed less than 1 GB of memory running locally. The entire process took 11.39 seconds given the sequential nature of the process.

Table 5

Consumption of the model in individual image

Mode	Duration(s)	RAM Used (MB)
Model Loading	2,10	9,92
Image loading	0,23	61,86
Image pre-processing	0,29	54,60
Model Inference	8,58	849,75
Post-processing of masks	0,19	48,27

Source: The authors (2025)

Using the LVIS dataset to do a massive analysis with a sample of a thousand images that mostly follow VGA (*Video Graphics Array*, resolution 640 *pixels* horizontally by 480 *pixels* vertically) quality, the results demonstrate that images from the *Segment Anything Model* It is divided into high-performance scenarios and low-performance scenarios. As in the one listed in Table 6. The lowest accuracies were identified in low-quality images, marking of small objects in relation to the image resolution (32 *pixels* or less) as in Figure 15, where we tried to mark the socket connected to the wall; or ambiguity in what may be the target of the marking, as shown in Figure 16, which is possible to be just the man with his newspaper or without his newspaper.

The accuracy is calculated using the IoU method and the level of variation in the model's accuracy in this test was due to the nature of LVIS having images in a range of resolutions and markings of objects in small and large sizes, even redundancy for markups (such as marking a large object, then marking small parts of it separately).

Table 6

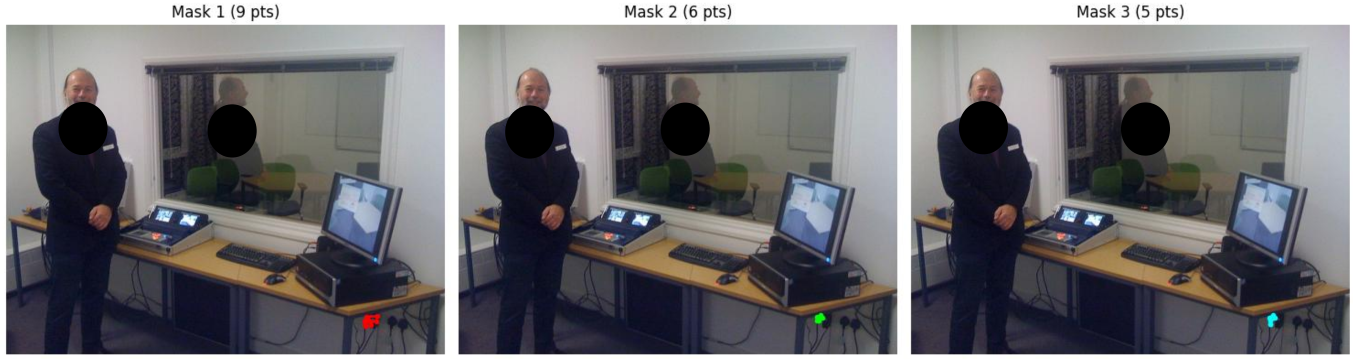
Massive results in accuracy

Category	Maximum precision	Minimum accuracy	Medium accuracy
Global	98%	22%	52%
10% Higher Accuracies	98%	87%	93%
10% Lower Accuracies	47%	22%	37%

Source: The authors (2025)

Figure 15

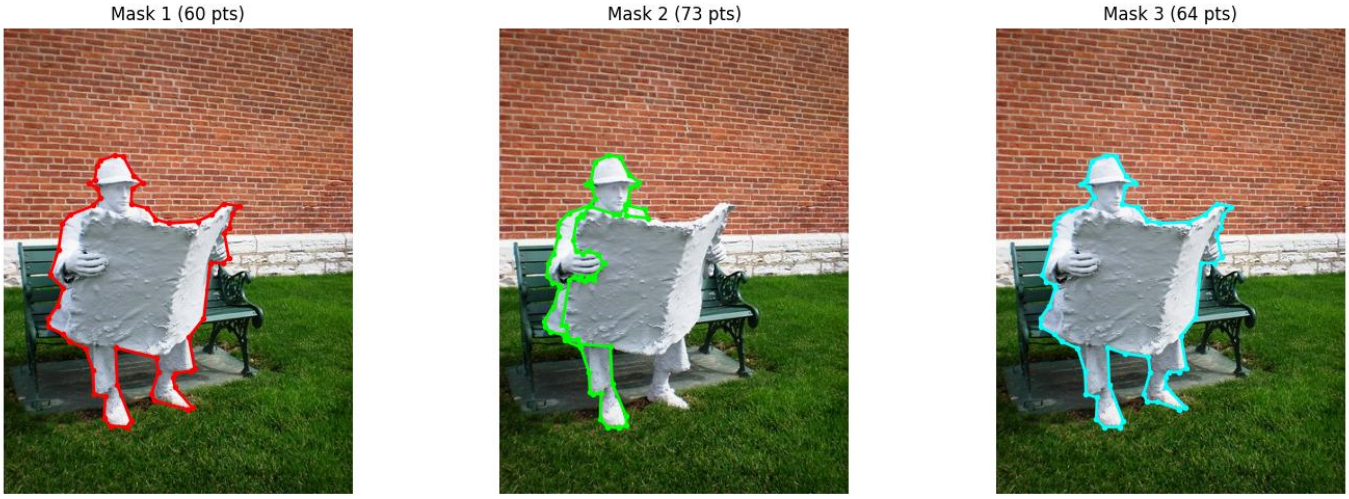
Marking small objects



Source: The authors (2025)

Figure 16

Ambiguity of marking



Source: The authors (2025)

There is also the difference that changes in resolution cause in processing time, where the amount of processing required in seconds increases in relation to resolution, where images of lower resolutions tend to consume less processing time in relation to images of higher quality as shown in Exhibit 7.

Table 7

Massive results at inference time

Category	Average Duration(s)
Global	8
10% Higher resolutions	13
10% Lower resolutions	6

Source: The authors (2025)

Regarding the number of points for the number of user touches, a reduction from the average of 35 touches per segmentation to only two touches per segmentation was expected. As shown in Table 8, the application of artificial intelligence allowed the creation of a greater number of points than estimated, generating the need for after-treatment to simplify visualization and user experience. Due to the inaccuracy of the model, a manual user curation is still required that adds 5 taps on average.

Table 8

Ratio of points by touches

Marking Category	Average touches per object	Average points per object
Detection	2	2
Manual annotation	35	35
Estimated targeting	2	35
Inference	2	127
Post-treatment inference	2	49
Inference with manual adjustment	7	49

Source: The authors (2025)

5 FINAL CONSIDERATIONS

The Marker project highlighted the possibility of a tool capable of integrating a simplified installation in an open source license that works *offline*, has artificial intelligence assistance and synchronization of projects between multiple users simultaneously in a scenario where the largest tools with similar functionalities are closed source and paid. The nature of utilizing web development tools allows for a modern look with long-term support of their technologies.

Even if The Marker project doesn't get results in individual categories like the simplified and lightweight installation of *LabelImg*, the open source community of Label Studio, the artificial intelligence assistance with models with dedicated training from Supervisely, or the synchronization of collaborative work from Diffgram, it was able to encompass and unite these categories into a single product along with a 256-bit AES-GCM encryption system for secure and private storage of information.

Among the limitations of the project, the absence of tests with teams that have the capacity to operate a *dataset* with ten thousand or more images to have extensive metrics of results in real use cases are highlighted. Also noteworthy is the limitation regarding the use of artificial intelligence that operates only with the *Segment Anything Model 2* in its lighter capacity (which requires less processing power).

In possible future developments it would be interesting to implement more artificial intelligence models and a test with a larger sample, however, it is assumed the possibility of adding the option of *fine-tuning* models within the tool to improve the accuracy of markings and automatic recognition of objects. Additionally: the elaboration of a model without a *front-end* that exists only as a manager of external access to the projects so that they can be hosted on servers with *IP* available on the *internet* instead of being limited to the local LAN network.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’16) (pp. 265–283). USENIX Association.

Andriluka, M., Uijlings, J. R. R., & Ferrari, V. (2018). Fluid annotation: A human-machine collaboration interface for full image annotation. arXiv. https://arxiv.org/abs/1806.07527

Armstrong, T. J., Buckle, P., Fine, L. J., Hagberg, M., Jonsson, B., Kilbom, Å., Kuorinka, I. A. A., Silverstein, B. A., Sjögaard, G., & Viikari-Juntura, E. (1993). A conceptual model for work-related neck and upper-limb musculoskeletal disorders. Scandinavian Journal of Work, Environment & Health, 19(2), 73–84.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., & Zieba, K. (2016). End to end learning for self-driving cars. arXiv. https://arxiv.org/abs/1604.07316

Brasil. Ministério da Saúde. (2023). Saúde do trabalhador: Notificações de LER/DORT no Brasil. https://www.gov.br/saude/pt-br/assuntos/saude-do-trabalhador

Brasil. Ministério do Trabalho e Emprego. (2023). Norma Regulamentadora nº 17: Ergonomia.

Chen, X., Zhang, K., Liu, Y., Li, X., Wang, Z., & Zhang, Y. (2022). Brain tumor classification based on neural architecture search. Scientific Reports, 12, Article 19206. https://doi.org/10.1038/s41598-022-22172-6

Chen, X., Li, Y., Zhang, Y., Liu, Z., & Wang, Z. (2025). UCVL: A benchmark for crime surveillance video analysis with large models. Neurocomputing, 600, 128–142.

Coles-Brennan, C., Sulley, A., & Young, G. (2019). Management of digital eye strain. Clinical and Experimental Optometry, 102(1), 18–29. https://doi.org/10.1111/cxo.12798

Defense Scoop. (2024, September 3). NGA awards $700M data labeling contract to advance computer vision models. DefenseScoop. https://defensescoop.com/2024/09/03/nga-700m-data-labeling-advance-computer-vision-models/

Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. arXiv. https://arxiv.org/abs/1904.10699

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) challenge. International Journal of Computer Vision, 88(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Gohill, H., Kaur, R., & Singh, A. (2024). A hybrid technique for plant disease identification and localisation in real-time. Computers and Electronics in Agriculture, 219, Article 108838. https://doi.org/10.1016/j.compag.2024.108838

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org

Gowrisankaran, S., & Sheedy, J. E. (2015). Computer vision syndrome: A review. Work, 52(2), 303–314. https://doi.org/10.3233/WOR-152162

Hagberg, M., Silverstein, B., Wells, R., Smith, M. J., Hendrick, H. W., Carayon, P., & Pérusse, M. (1995). Work related musculoskeletal disorders (WMSDs): A reference book for prevention. Taylor & Francis.

Khalil, K., Kimiafar, K., Zadeh, M. R., & others. (2023). Artificial intelligence literacy among healthcare professionals and students: A systematic review. Health Informatics Journal, 29(4), 1–15.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. arXiv. https://arxiv.org/abs/2304.02643

Kovashka, A., Russakovsky, O., Fei-Fei, L., & Grauman, K. (2016). Human-in-the-loop annotation. Foundations and Trends in Computer Graphics and Vision, 10(3), 187–278.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25).

Kundu, R., Das, R., Ghosh, S., & others. (2021). Pneumonia detection in chest X-ray images using an ensemble of convolutional neural networks. PLOS ONE, 16(9), Article e0256630. https://doi.org/10.1371/journal.pone.0256630

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In European Conference on Computer Vision (ECCV). https://arxiv.org/abs/1405.0312

Merrill, R. M., & Alleman, J. R. (2012). The relevance of ergonomic interventions for the prevention of musculoskeletal disorders. Journal of Occupational and Environmental Medicine, 54(4), 427–433.

Meta. (2013). React – A JavaScript library for building user interfaces. https://react.dev

Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

Morais, D. M. G., Silva, R. M., & others. (2020). O conceito de inteligência artificial usado no mercado de softwares, da educação tecnológica e na literatura científica. Educação Profissional e Tecnológica em Revista, 4(2), 98–109.

Organização Mundial da Saúde. (2003). Ergonomics in the workplace. World Health Organization.

Papadopoulos, D. P., Uijlings, J. R. R., Keller, F., & Ferrari, V. (2017). Extreme clicking for efficient object annotation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 4930–4939).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems (Vol. 32).

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv. https://arxiv.org/abs/1711.05225

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems (Vol. 28).

Rosenfield, M. (2016). Computer vision syndrome (a.k.a. digital eye strain). Optometry in Practice, 17(1), 1–10.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. International Journal of Computer Vision, 77(1–3), 157–173.

Russell, S., & Norvig, P. (2010). Artificial intelligence: A modern approach (3rd ed.). Pearson.

Sager, C., Janiesch, C., & Zschech, P. (2021). A survey of image labelling for computer vision applications. arXiv. https://arxiv.org/abs/2104.08885

Sanar. (n.d.). Lesões por esforço repetitivo (LER) e distúrbios osteomusculares relacionados ao trabalho (DORT): Conceitos e prevenção. https://www.sanar.com.br

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., … others. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. In Advances in Neural Information Processing Systems (Vol. 35, pp. 25278–25294).

Sheppard, A. L., & Wolffsohn, J. S. (2018). Digital eye strain: Prevalence, measurement and amelioration. BMJ Open Ophthalmology, 3(1), Article e000146. https://doi.org/10.1136/bmjophth-2018-000146

Shin, H., Lee, J., Kim, S., & others. (2022). Visual product search using deep learning. [Detalhes incompletos – completar se possível].

Szeleski, R. (2010). Computer vision: Algorithms and applications. Springer.

Tauri. (2022). Tauri documentation. https://tauri.app

Time. (2024, September 19). Meta scales up the AI data industry. Time. https://time.com/7294699/meta-scale-ai-data-industry/

Tzelepis, D., & others. (2021). Efficient bounding box annotation. Pattern Recognition Letters.

Wang, L., Zhao, X., Zhang, Y., Han, X., & Deveci, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57(4), Article 1–27.

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. National Science Review, 5(1), 44–53. Springer.