

Portal to Other Dimensions: use of Computer Vision to create art work from day life images

Portal para Outras Dimensões: uso da Visão Computacional para criar obras de arte a partir de imagens do dia a dia

Luísa Ferreira^{1*}, Ana Osias¹, Lucas Vieira¹ and Michel Silva¹

Abstract: This work aims to combine classic and contemporary paintings with the external world by altering the interior of a door into another style in real-time photos and/or videos, creating a portal effect to provide an interactive experience with artworks. To achieve this, various techniques from Artificial Intelligence and Computer Vision were employed, primarily focusing on convolutional neural networks (CNNs) with supervised training. Models proposed for semantic segmentation problems were used for door detection and style transfer to stylize the input image. Finally, we conclude with an analysis of the results for each network individually and in combination, ensuring that the metrics and approaches were satisfactory. With our results, we were able to offer an accessible and interactive way to bring art into the daily lives of many people.

Keywords: Component — Formatting — Style — Styling

Resumo: Este trabalho tem como objetivo combinar pinturas clássicas e contemporâneas com o mundo externo, alterando o interior de uma porta para outro estilo em fotos e/ou vídeos em tempo real, criando um efeito de portal para proporcionar uma experiência interativa com obras de arte. Para alcançar isso, foram empregadas diversas técnicas de Inteligência Artificial e Visão Computacional, com foco principal em redes neurais convolucionais (CNNs) com treinamento supervisionado. Modelos propostos para problemas de segmentação semântica foram utilizados para detecção de portas e transferência de estilo para estilizar a imagem de entrada. Por fim, concluímos com uma análise dos resultados de cada rede individualmente e em combinação, garantindo que as métricas e abordagens fossem satisfatórias. Com nossos resultados, conseguimos oferecer uma maneira acessível e interativa de trazer arte para o cotidiano de muitas pessoas.

Palavras-Chave: Componente — Formatação — Estilo —Estilização

¹ Department of Informatics, Universidade Federal de Viçosa (UFV), Brazil

*Corresponding author: luisa.ferreira@ufv.br

DOI: <https://dx.doi.org/10.22456/2175-2745.143513> • Received: 24/10/2024 • Accepted: 02/12/2024

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

Throughout history, art has been a powerful form of expression, reflecting a wide range of themes from historical events to the most intimate feelings of its creators. This artistic production continues to be valued by diverse audiences and is an invaluable source of study for historians. Over the centuries, many artists have stood out, and one of the most influential is Vincent van Gogh, a 19th-century Dutch painter and the author of more than two thousand works. His creations are characterized by vibrant colors and emotionally charged representations, with one of his most famous pieces being "The Starry Night."

With the growing use of Machine Learning techniques, new avenues are being explored in art. For example, Chou *et al.* [1] applied machine learning techniques to extract head

pose angles and display types from 11,000 digital images of portraits, covering 500 years, from the Renaissance to the Modern Era, to uncover patterns and trends that reveal insights into artistic practices and their evolution over time. Another relevant study was conducted by Eriksson *et al.* [2], who proposed a deep learning approach to reconstruct images of lost works based on surviving information such as sketches, copies, and historical descriptions.

The increasing integration of technology with interior design has also become a significant topic. In particular, smart home systems are providing new possibilities for user interaction with their environments, fostering dynamic spaces that combine functionality and aesthetics [3]. Moreover, the principles of Wayfinding Design are fundamental to understanding how individuals navigate physical spaces organized

to facilitate navigation and enhance user experience [4].

Inspired by these advancements and principles, this work aims to explore how digital image processing techniques and neural networks can be applied to transform everyday objects into stylized artworks. Specifically, by converting the interior of a door into a creative portal, this project not only enhances the aesthetic of the space but also contributes to user orientation, making the door a more significant element in navigating and identifying different areas within an environment. In this way, the stylized door can serve as a visual reference point, guiding individuals intuitively and engagingly.

Finally, the role of art in education is widely recognized as essential for developing critical and creative thinking while promoting greater cultural understanding [5]. The accessibility of this technology, which requires only a computer and a camera, enables its implementation in schools and community centers with limited resources. By integrating these techniques into educational and public spaces, art can become a more tangible and interactive experience, fostering a deeper understanding and appreciation of different artistic movements. Moreover, the adaptability of these methods to various contexts and objects promises to uncover new ways of engaging with digital art, allowing users to explore and interact with the works of Vang Gogh and other celebrated artists in innovative and meaningful ways.

Our code is available at github.com/MaVILab-UFV/Portals-to-other-dimensions-WVC-2024.

2. Theoretical Foundation

In this work, we address the problem of door detection and segmentation in images using neural networks and computer vision techniques. To provide a solid theoretical foundation, we will explore the main architectures for object detection, semantic segmentation, and relevant datasets for the task.

2.1 Object Detection and Semantic Segmentation

Object detection architectures can be broadly categorized into two families: Region-based Convolutional Neural Networks (R-CNN) and the YOLO (You Only Look Once) family.

R-CNN [6] utilizes Selective Search to generate region proposals, processes each with a CNN to extract features, and then classifies and localizes objects. Although accurate, this method is computationally intensive. Fast R-CNN [7] improves efficiency by processing the entire image in a single CNN pass, but still relies on separate proposal generation. Faster R-CNN [8] accelerates the process by integrating a Region Proposal Network (RPN), eliminating the need for an external proposal generation step.

In contrast, YOLO [9] treats object detection as a single-stage regression problem. It divides the image into a grid and predicts bounding boxes and class probabilities directly from each cell, enabling real-time detection. While YOLO is efficient and accurate, it may struggle with very small or densely packed objects.

For semantic segmentation, YOLOv8 [10] extends its capabilities to include segmentation, in addition to detection, pose estimation, tracking, and classification. This makes it a versatile choice for tasks requiring both object detection and segmentation. Mask R-CNN [11] enhances Faster R-CNN by adding a branch for predicting object masks alongside bounding box predictions. Alternatively, models like U-net [12], though originally designed for medical image analysis, are effective for semantic segmentation tasks.

Given the need for both real-time detection and segmentation in our work, we prefer YOLOv8 for its comprehensive capabilities and efficiency.

2.2 Datasets

For computer vision models to operate effectively, it is essential to have a well-defined and comprehensive dataset that suits the problem at hand. Among the available datasets for door detection and segmentation, the following stand out:

2.2.1 DoorDetect Dataset and Door Detection Computer Vision Project

The DoorDetect Dataset [13] consists of 1,213 images annotated with object bounding boxes. The images exhibit significant diversity, often including complex scenes with multiple objects. This dataset covers various categories of doors, such as refrigerator doors and drawers, as well as handles. Although this variety makes DoorDetect valuable for general computer vision tasks, its broad scope may not be ideal for our application, which specifically focuses on detecting room passage doors.

The Door Detection Computer Vision Project [14] is another relevant dataset, containing 1,993 images. Similar to DoorDetect, this dataset also offers a wide range of door categories. However, like the previous one, the present diversity makes it less suitable for our application, which aims at detecting room doors.

Furthermore, both datasets are primarily focused on detection tasks. Considering that this work also addresses semantic segmentation, it is necessary to have a dataset that meets the specific requirements not covered by these two sets.

2.2.2 DeepDoors2 Dataset

The DeepDoors2 [15] Dataset stands out as a dataset specifically designed for the classification and segmentation of larger doors. The images were captured using a 3D Realsense camera in indoor environments, including universities, public places, and residences, offering diverse coverage of doors and their surroundings, with different textures and sizes. The dataset is divided into two sub-datasets: one focused on door detection and semantic segmentation, with 3,000 images, and another dedicated to door classification, also with 3,000 images. This structure makes DeepDoors2 the most suitable for the context of our work.

2.3 Transfer Style

Some style transfer networks proposed by Johnson *et al.* [16], Dumoulin *et al.* [17] and Ghiasi *et al.* [18] were studied for

the realization of this work. In addition to serving as a foundation, these works are also related to ours and will be explained in more detail in the related works section. The mentioned networks form a sequence of research that culminated in the work that contributed to our implementation. This implementation proposes a network¹ that transfers 16 fixed styles, which enhances the speed of stylization, enabling the desired applications.

3. Related Work

Just like us, other authors are also interested in the interdisciplinarity between art and computing. We highlight some works that perform style transfer tasks and others that create categorized datasets serving as a foundation for developing these studies.

3.1 Neural Networks for Style Transfer

Related work on the task of style transfer, primarily applied to the field of painting, has highlighted the convergence between art and computer vision. The following papers have proposed solutions for transferring the style of one image to another.

The approach presented by Johnson *et al.* [16] employs two networks. The first is an Image Transformation Network, which is responsible for transforming an original image into a stylized one using an encoder-decoder architecture. The original image is encoded to represent its context, and during decoding, the image is reconstructed with the defined style applied. This network streamlines the style application process in the subsequent network. The second is a Fixed Loss Network, which is a convolutional neural network (CNN). Different CNN layers capture different image features: layers

closer to the input detect less complex patterns, while layers closer to the output capture higher-level features. The fixed loss network uses this principle to extract and compare content and style separately. The network calculates the total loss by combining content loss, which compares high-level information between the original image and the output, and style loss, which compares stylistic features across the layers of the reference image and the output image.

Subsequent works, such as that proposed by Dumoulin *et al.* [17], extend this capability by manipulating normalization parameters, enabling the application of multiple artistic styles.

Finally, the network discussed by Ghiasi *et al.* [18] shares similar characteristics with the first described network but introduces a Style Prediction Network into its pipeline. This Style Prediction Network is responsible for making style application more flexible. It is trained on a variety of artistic images and textures, enabling it to predict the style of the desired image faster, even if the network was not explicitly trained for that style. Its output consists of a vector of parameters representing the style of the input style image, which is then passed to the style transformation network.

3.2 Artistic Dataset with Applications in Computer Vision

Van Zuijlen *et al.* [19] introduces the "Materials in Paintings" (MIP) dataset, consisting of 19,325 images of paintings, aimed at studying the pictorial representations of materials over time. This dataset includes 15 material categories for each painting. Khan *et al.* [20] proposed a large-scale dataset of digital paintings by 91 different artists, featuring applications such as artist categorization, style classification, and saliency detection. Additionally, Ju *et al.* [21] present a dataset with 50,000 high-quality images and over 123,000 instances

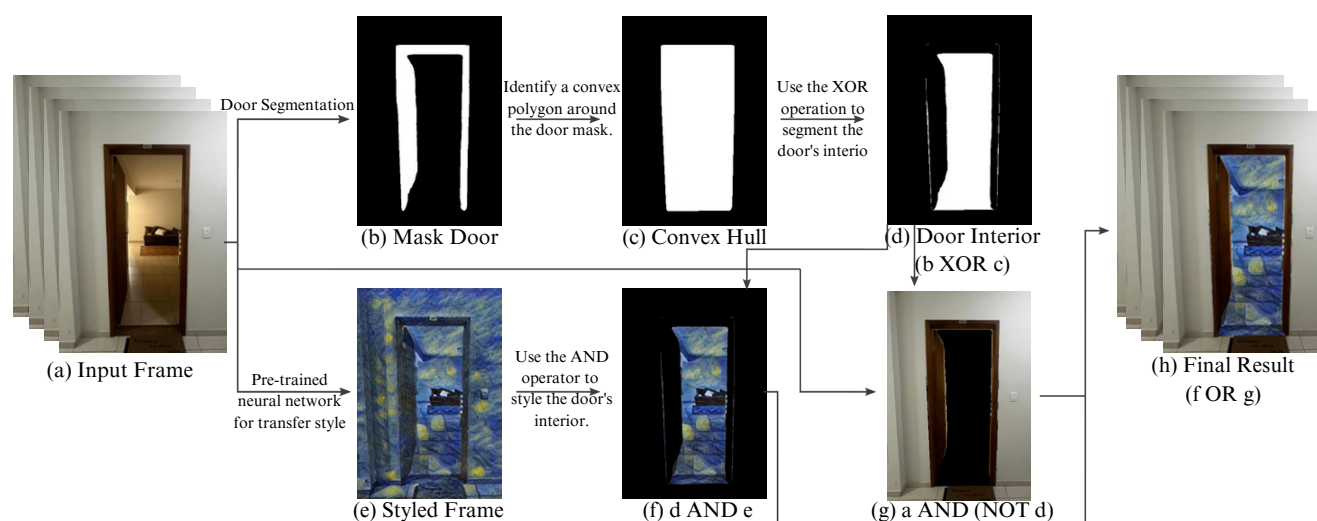


Figure 1. The proposed method for transforming the input frame into a result image with the styled interior of the door. First, (b) the door frame is segmented, and the convex hull algorithm is applied to obtain a convex polygon of the door's mask (c). The entire frame is also stylized (e). After applying digital image processing techniques, the final result is achieved, as shown in (f).



Figure 2. Examples of images with the ground truth from the DeepDoors2 dataset.

of people across 5 natural and 15 artificial scenarios. The images are annotated with bounding boxes, keypoints, self-contact points, and text information for humans in both 2D and 3D. This dataset is designed to be versatile and useful for various computer vision tasks, including human detection, pose estimation, image generation, and motion transfer.

4. Method

In this section, we outline our approach for using computer vision to transform everyday images into artwork by styling a door to appear as a portal. Our methodology involves three key steps, depicted in Fig. 1: First, we perform semantic segmentation on each video frame to identify the door. Second, we apply style transfer techniques to the original image or frame. Finally, we use digital image processing to ensure the style is applied exclusively to the door's interior.

Semantic Segmentation. Firstly, to train a model for object segmentation, it is crucial to assemble a dataset that thoroughly covers the specified problem domain. In our case, we utilized the DeepDoors2 dataset [15] (see in Fig. 2), which consists of 3,000 RGB images sized at 480x640 pixels. This dataset includes doors labeled as closed, open, and semi-open, divided into 2,400 images for training, 300 for validation, and 300 for testing purposes.

In addition to semantic segmentation, our objective extended to door detection and classification based on their states: open, closed, or semi-open. This distinction was necessary to avoid unnecessary processing on closed doors. To achieve this, we employed a pre-trained YOLOv8 model [10], which we further fine-tuned using the DeepDoors dataset. The result of the semantic segmentation is Fig. 1-b. We refer to this result as the "mask," a binary image where white pixels represent the door object, and black pixels represent areas not part of the door.

Transfer Style. For applying style transfer to an image, we chose two approaches: the first one enables arbitrary style transfer from any provided image ². The second is an adapta-

tion of Ghiasi *et al.* [18], which utilizes a pre-trained network with 16 specific styles ³ (the reader is referred to Section 3.1 for more details). Each approach offers distinct advantages and disadvantages, which will be discussed in Section 5. As a result of the style transfer, we have the original image completely stylized, as represented in Fig. 1-e.

Applying Style only to the Door Interior. In the previous sections, we discussed how we identified doors using semantic segmentation and applied style transfer to the entire image. Now, we need to combine this information to apply the style exclusively to the interior of the door. To achieve this, in addition to the techniques already described, we utilized digital image processing techniques, which we will describe throughout the flow of this methodology.

The input to our workflow is a video, which can be processed in real time. For each frame, we perform semantic segmentation to generate a mask that identifies the door, as shown in Fig. 1-b. Since our goal is to apply styling exclusively to the door's interior, and the semantic segmentation provides a mask of the entire door frame, we employed a convex hull algorithm. This algorithm refines the mask by converting it into a convex polygon that precisely outlines the door's interior, as shown in Fig. 1-c. To isolate the door's interior, we then apply a bitwise XOR operation between the original door mask (Fig. 1-b) and the convex hull (Fig. 1-c), resulting in a segmentation that accurately represents only the interior of the door, as illustrated in Fig. 1-d.

For each frame, we also apply style transfer to the original frame, as shown in Fig. 1-e. To style only the interior of the door, we combine the styled image with the mask of the door interior using a bitwise AND operation, resulting in an image where only the interior of the door is styled, as seen in Fig. 1-f. Then, we take the original frame and use a bitwise AND operation with the negative of the mask of the door interior to get the original frame without the door interior, as seen in Fig. 1-g. After that, we combine these two images

image-stylization-v1-256/2

³Architecture available at <https://github.com/ryanwongsa/Real-time-multi-style-transfer>

²Architecture available at <https://tfhub.dev/google/magenta/arbitrary->

with a bitwise OR operation to merge them, generating the final result, Fig. 1-h, where only the interior of the door is styled, giving the appearance of a portal.

5. Experiments and Results

In this section, we will analyze the results of our fine-tuning on YOLOv8, discuss the advantages and disadvantages of two pre-trained networks that perform style transfer on images (as mentioned in Section 4), and demonstrate how our workflow can selectively apply style to only the interior of a door.

As previously discussed, we used YOLO v8, we employed the "yolov8n-seg" architecture, a nano segmentation structure pre-trained with COCO (Common Objects in Context) weights, containing 3.4 million learned parameters. The images were resized to 640x640 pixels before inputting them into the model, and the output classes were reduced from 80 to 3 (closed door, open door, or semi-open door).

The training was conducted over 60 epochs, including hyperparameter optimizations of the architecture, such as an adaptive learning rate (Adam optimizer) set at 0.001429 and a batch size of 16. As a result, we achieved a mean Average Precision at 50% Intersection over Union (mAP50) of 0.96 on the dataset.

For the image stylization results mentioned in Section 4, we utilized two approaches. The first approach, demonstrated in Fig. 3, applies the style of any input image to another image. The key advantage of this method is its ability to use any artwork for stylization, aligning with the primary goal of this work. On a laptop using an NVIDIA RTX 3050 Ti, this process used 3GB of VRAM and took 2 seconds per frame to apply the style. This makes real-time infeasible on



Figure 3. Using the pre-trained network for arbitrary style transfer from input image

most portable computers; therefore, since this project was designed to be set up for showing computer vision applications in schools and museums with portable computers, we tried to minimize dependencies from high-end GPUs, as such advanced hardware is unlikely to be available in such spaces.

Aiming for real-time processing, we used another approach with a pre-trained network featuring 16 specific styles. Despite the limitation in the number of styles, this method still allowed us to create our portals with 16 different styles. Most importantly, it improved and enabled the real-time performance we desired, as it required only 1GB of VRAM. Fig. 4 shows the 16 different styles applied to an image with an open door. It is noteworthy that 1GB of VRAM is available in not high-end GPUs. We tested in a NVIDIA GTX 1060 and it still runs in real time, which means that the model can be executed in more affordable notebooks. With this model, we achieved our goal of minimizing reliance on computational power to make the project more suitable for real-time presentations in schools and museums.

As a final result, we successfully achieved the desired stylization effect, which operates in real-time on both videos and images. The stylization effectively tracks the door's movement, as demonstrated in Fig. 5, creating a convincing portal effect. This project has significant potential to demonstrate how computer vision can be applied in everyday activities,

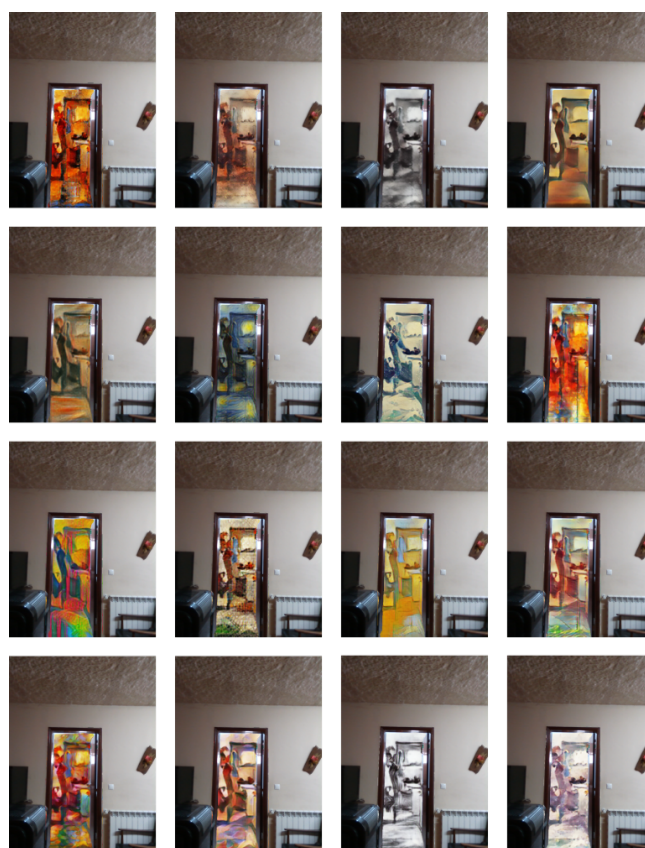


Figure 4. 16 specific styles from the pre-trained model applied to an example image of an open door.



Figure 5. Sequence of images showing the door in various positions during its opening, with the stylization tracking the movement.

bridging the gap between technology and art. Moreover, it highlights how machine learning and computer vision can be utilized with simple equipment, like a webcam and a portable desktop, to engage communities in interactive and enjoyable experiences. This approach could also be applied in educational settings to inspire students in both computer science and the arts.

5.1 Failure Cases

One of the failures we observed occurs when the segmentation of the door frame, for some reason, fails to include a part of it. This results in a polygon that does not completely encompass the interior area of the door, leading to incomplete stylization. As we can observe in Fig. 6-a, the segmentation has an interruption on the right side, and therefore, the interior of the door is not well captured, as observed in Fig. 6-b, resulting in incomplete stylization, as shown in Fig. 6-c.



Figure 6. Failure case when the segmentation fails.

Another potential failure occurs when the door opens toward the camera. If we consider the applied methodology, the problems encountered make sense. In our methodology, we identify the entire door and generate a polygon around it. This polygon is generated with all the content of the door, not just

the doorway itself or the frame. When the door opens away from the camera, this is not an issue, as the door itself does not alter the polygon's outline, and only the frame appears. However, in these other cases, there is a misformation of the polygon surrounding the door, which affects the final stylization. The resulting segmentation will form a closed convex polygon that encompasses not only the interior of the door but also the floor in front of the door and the wall above it. This can be observed in Fig. 7-b, where the convex polygon generated from the door mask also includes part of the floor and wall, resulting in undesired stylization, as shown in Fig. 7-c.

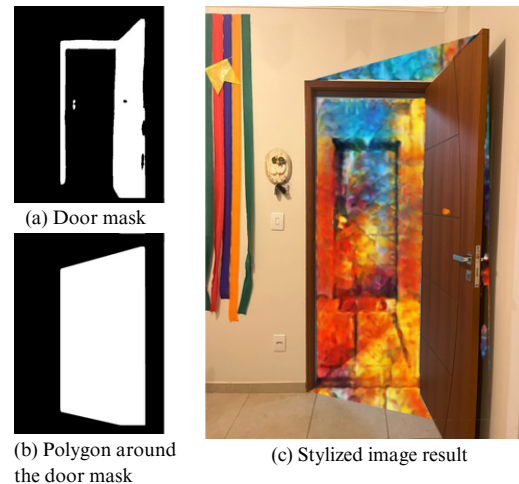


Figure 7. Failure case when the door opens outward.

6. Conclusion

The experiments conducted demonstrated that by adapting the style transfer methodology to use a pre-trained network with 16 specific styles, we achieved a significant reduction in VRAM consumption and execution time, making real-time application feasible. Additionally, the use of digital image processing techniques allowed for precise style application exclusively to the interior of the segmented doors, maintaining the original context of the image.

The results obtained with semantic segmentation using the YOLOv8 model were quite satisfactory, achieving an accuracy of 96.8% in classifying doors as closed, open, or ajar, and a mean Average Precision at 50% Intersection over Union (mAP50) of 95.5% in the semantic segmentation task. Therefore, the success of these three stages and the combination of different networks and techniques implemented proved to be effective both in terms of accuracy and computational efficiency.

This work not only demonstrates the potential of neural networks to transform everyday images into stylized artworks but also shows how we can make interaction with art more accessible, engaging, and dynamic. It is important to emphasize that the technology requires only a computer and a camera, allowing its implementation in schools, museums, and community centers with limited resources, thereby ex-

panding the reach of art. In educational settings, this tool is especially valuable, as physical and visual interaction can reinforce learning and increase student engagement, facilitating the understanding of different artistic movements. Furthermore, the integration of technology with interior design and the principles of Wayfinding Design is essential for creating spaces that enhance aesthetics and facilitate user orientation, promoting intuitive navigation and enriching the overall experience of individuals.

For future work, we suggest exploring a strategy to address the current limitations by focusing on mitigating the identified failure cases. One approach could be to use data from Deep Doors, which includes images of open and semi-open doors, to create a specialized dataset for segmenting only the door interiors, excluding the door frames. By employing a method designed to extract solely the door interior, it may be possible to avoid the annotation errors discussed in Section 5, shifting the focus from detecting door frames to detecting door interiors.

Additionally, expanding the range of available styles could provide more variety and customization, enhancing the user experience. Applying these techniques to different contexts and objects might also reveal new forms of interaction and offer novel insights into digital art appreciation.

Acknowledgment

The authors would like to thank CAPES, CNPq and FAPEMIG for supporting this project.

Author contributions

Luísa implemented the proposed method, developed the GitHub repository, and wrote parts of the paper. Ana conducted the literature review, found and prepared the dataset for YOLO fine-tuning, and assisted with the writing process. Lucas fine-tuned the YOLO model for detecting and segmenting doors and also contributed to the writing. Michel served as the advisor, offering critical feedback on the text and proposing essential ideas for implementing the main approach

References

- [1] CHOU, J.-P.; STORK, D. G. Computational tracking of head pose through 500 years of fine-art portraiture. *Electronic Imaging*, Society for Imaging Science and Technology, v. 35, p. 1–13, 2023. Disponível em: <https://doi.org/10.2352/ei.2023.35.13.cvaa-211>.
- [2] ERIKSSON, J. et al. Recovering lost artworks by deep neural networks: Motivations, methodology, and proof-of-concept simulations. *Electronic Imaging*, Society for Imaging Science and Technology, v. 35, p. 1–7, 2023. Disponível em: <https://doi.org/10.2352/ei.2023.35.13.cvaa-210>.
- [3] YANG, J. Design of architectural decoration based on smart home system. In: *2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. [s.n.], 2022. p. 899–902. Disponível em: <https://doi.org/10.1109/ICMTMA54903.2022.00183>.
- [4] ARTHUR, P.; PASSINI, R. *Wayfinding: people, signs, and architecture*. [s.n.], 1992. Disponível em: <https://doi.org/10.5860/choice.30-1301>.
- [5] GADSDEN, V. L. The arts and education: Knowledge generation, pedagogy, and the discourse of learning. *Review of research in education*, Sage Publications Sage CA: Los Angeles, CA, v. 32, n. 1, p. 29–61, 2008. Disponível em: <https://doi.org/10.3102/0091732x07309691>.
- [6] GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [s.n.], 2014. p. 580–587. Disponível em: <https://doi.org/10.1109/cvpr.2014.81>.
- [7] GIRSHICK, R. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [s.n.], 2015. p. 1440–1448. Disponível em: <https://doi.org/10.1109/iccv.2015.169>.
- [8] REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 6, p. 1137–1149, 2016. Disponível em: <http://arxiv.org/abs/1506.01497>.
- [9] 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*. [s.n.], 2016. p. 779–788. Disponível em: <https://arxiv.org/abs/1506.02640>.
- [10] JOCHER, G.; CHAURASIA, A.; QIU, J. *Ultralytics YOLOv8*. 2023. Disponível em: <https://github.com/ultralytics/ultralytics>.
- [11] HE, K. et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [s.n.], 2017. p. 2961–2969. Disponível em: <https://doi.org/10.1109/ICCV.2017.322>.
- [12] RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. 2015. p. 234–241. Disponível em: <https://arxiv.org/abs/1505.04597>.
- [13] ARDUENGO, M.; TORRAS, C.; SENTIS, L. Robust and adaptive door operation with a mobile robot. *Intelligent Service Robotics*, Springer Science and Business Media LLC, May 2021. ISSN 1861-2784. Disponível em: <http://dx.doi.org/10.1007/s11370-021-00366-7>.
- [14] ESPINOZA, N. Open Source Dataset, *Door detection Dataset*. Roboflow, 2022. <https://universe.roboflow.com/nathaly-espinoza/door-detection-zqt59>. Visited on 2024-08-19. Disponível em: <https://universe.roboflow.com/nathaly-espinoza/door-detection-zqt59>.

- [15] RAMÔA, J. G. et al. Real-time 2d–3d door detection and state classification on a low-power device. *SN Applied Sciences*, Springer, v. 3, n. 5, p. 590, 2021. Disponível em: <http://dx.doi.org/10.1007/s42452-021-04588-3>.
- [16] JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: SPRINGER. *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*. 2016. p. 694–711. Disponível em: <https://arxiv.org/abs/1603.08155>.
- [17] DUMOULIN, V.; SHLENS, J.; KUDLUR, M. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016. Disponível em: <http://arxiv.org/abs/1610.07629>.
- [18] GHIASI, G. et al. Exploring the structure of a real-time, arbitrary neural artistic stylization network. *arXiv preprint arXiv:1705.06830*, 2017. Disponível em: <http://arxiv.org/abs/1705.06830>.
- [19] ZUIJLEN, M. J. V. et al. Materials in paintings (mip): An interdisciplinary dataset for perception, art history, and computer vision. *Plos one*, Public Library of Science San Francisco, CA USA, v. 16, n. 8, p. e0255109, 2021. Disponível em: <https://doi.org/10.48550/arXiv.2012.02996>.
- [20] KHAN, F. S. et al. Painting-91: a large scale database for computational painting categorization. *Machine vision and applications*, Springer, v. 25, p. 1385–1397, 2014. Disponível em: <https://doi.org/10.1007/s00138-014-0621-6>.
- [21] JU, X. et al. Human-art: A versatile human-centric dataset bridging natural and artificial scenes. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [s.n.], 2023. p. 618–629. Disponível em: <https://doi.org/10.1109/CVPR52729.2023.00067>.