IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 September 2023, accepted 11 September 2023, date of publication 13 September 2023,
date of current version 19 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3315235

RESEARCH ARTICLE

SAMStyler: Enhancing Visual Creativity With Neural Style Transfer and Segment Anything Model (SAM)

KONSTANTINOS PSYCHOGYIOS[1], HELEN C. LELIGOU[2], FILISIA MELISSARI[1], STAVROULA BOUROU[1], ZACHARIAS ANASTASAKIS[1], AND THEODORE ZAHARIADIS[1,3]

[1]Synelixis Solutions S.A., 34100 Chalkida, Greece
[2]Netcompany-Intrasoft S.A., 19002 Paiania, Greece
[3]Department of Agricultural Development, Agri-Food and Natural Resources Management, National and Kapodistrian University of Athens, GR15772 Athens, Greece

Corresponding author: konstantinos Psychogyios (psychogios@synelixis.com)

This work was supported in part by the European Union's Horizon Europe Research and Innovation Program through the DAFNE+ Project under Grant 101061548.

ABSTRACT Neural Style Transfer (NST) is a popular technique of computer vision where the content of an image is blended with the style of another, which results in a fused image with certain properties of both original images. This approach has practical applications in various domains and has garnered significant attention in both industry and academia. An interesting application of this technique is segmented style transfer where a segmentation algorithm is used to locate objects within an image and then the style transfer method is performed locally, producing images with different styles for different objects. This approach opens up possibilities for creating visually striking compositions by seamlessly blending various artistic styles onto specific objects within an image, allowing for a new level of creative expression. This paper proposes a novel method that combines Segment Anything Model (SAM), a state-of-the-art vision transformer-based image segmentation model developed by Facebook, with style transfer. Our approach includes performing localized style transfer in selected segmentation regions of an image using classical style transfer algorithms. To ensure smooth transitions between the stylized and non-stylized border we also develop our loss function with a border smoothing technique. Experimental results demonstrate the robustness and effectiveness of the proposed methodology, including the ability to infuse multiple artistic styles into different objects within an image. The contributions of this work include integrating SAM with style transfer, proposing a novel loss function, evaluating the segmented style transfer in multiple content regions, comparing with state-of-the-art approaches, and experimenting with multiple style images for diverse stylization. Our primary focus centers on creating a model that serves as a digital painter across a wide range of image genres and artistic styles.

INDEX TERMS Segment anything model, segment anything, segmentation, machine learning, style transfer.

I. INTRODUCTION

Machine learning (ML) is a powerful tool that allows computers to learn from data and make predictions [1], [2], [3]. When combined with computer vision (CV), ML enables computers to understand and interpret visual information [4], [5], [6]. CV specifically focuses on extracting meaningful information from images or videos, complementing ML's ability to learn and process data. Together, ML and CV form a synergistic partnership in the realm of visual understanding for computers. This powerful combination has led to numerous applications, such as object recognition [7], [8], motion tracking [9], [10], and medical imaging [11], [12], [13]. By continuously advancing these fields, we are able to enhance our interaction with the visual world and revolutionize various industries.

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

Neural Style Transfer is a technique in computer vision where the style of a particular image is used in combination with the content of another image to produce a blended third image. It is generally performed by neural networks where the representations of each layer are used to deconstruct and separate the components of an image [14], [15], [16], [17], [18]. More specifically, the feature maps of each layer are used to calculate the content and the style of an image which have been proven to be independent. This methodology has gained popularity because it has achieved generating artistic images that blend the content and style in visually striking ways [19], [20]. It has practical applications in art, design, and visual effects, and it has also become a popular topic in academic research and deep learning communities.

Another interesting application of the aforementioned method is its integration with semantic segmentation. The latter is the process of assigning semantic labels to different regions within an image, effectively segmenting it into meaningful parts. By leveraging this technique in conjunction with style transfer, we can achieve targeted and localized transformations. The combined approach incorporates first the object segmentation within an image and subsequently localized style transfer [21], [22]. For example, we could segment a picture of multiple cats and then perform style transfer only to specific, desired cats. This targeted style transfer enables us to selectively apply artistic styles to particular objects, creating visually striking and attention-grabbing images. By precisely delineating the regions of interest through semantic segmentation, we have finer control over the style transfer process, allowing for more personalized and tailored artistic effects. This integration paves the way for innovative applications in various domains, such as fashion, interior design, and advertising, where specific objects or regions can be stylized to enhance their visual appeal and convey desired aesthetics.

Concerning the segmentation part, Facebook released the Segment Anything Model (SAM) [23]. This is a Vision Transformer-based model trained on a very large dataset which is also created by Facebook named SA-1B. This dataset consists of 11M diverse, high-resolution images with 1.1B corresponding masks. These masks were automatically generated by the SAM mode. This model demonstrated groundbreaking results regarding the image segmentation task achieving state-of-the-art performance and generalizability. In more detail, SAM uses a vision transformer-based image encoder to extract image features and prompt encoders to incorporate user interactions, followed by a mask decoder to generate segmentation results and confidence scores based on the image embedding, prompt embedding, and output token [24].

In this work, we combine SAM model with style transfer to perform localized style transfer to selected segmentation regions of a content image. Using classical style transfer algorithms as a backbone we are able to perform accurate image synthesis where the style of a painting is transferred to segmentation areas of content images. We also develop and introduce a novel loss function which creates smooth shifts between stylized and non-stylized areas. This last part is crucial, since we noticed that even though SAM is very accurate in segmenting the objects in an image, the generated image contains a gap between the segmentation regions of different styles, which is not visual appealing. Our newly introduced approach leverages the mask generated by SAM, and subsequently employs dilation to generate an augmented segmentation mask, which serves to constrain the loss function.

An interesting application of this methodology could be robotic painting with AI [25], [26] which represents a fusion of cutting-edge technology and artistic expression. In this innovative approach, robots equipped with advanced artificial intelligence algorithms are capable of autonomously creating art. These AI-driven systems analyze data, including images and patterns, to generate unique and creative compositions. Our model has the potential to be utilized on a robot, enabling it to receive a content image and an artistic style for inspiration, along with specific instructions to perform localized style transfer within a designated region of the image.

We evaluate our approach for style transferring using one or more stylized images resulting in an image that may contain multiple artistic styles for different objects (e.g. Van Gogh for the sea and William Turner for the sky). Moreover, we compare our approach to another localized style transfer technique. Results indicate that our methodology is superior and robust yielding visually captivating outcomes. Our main contributions can be summarized as follows:

- We integrate SAM with the style transfer computer vision field.
- We propose a novel style transfer loss function with a smooth transition region.
- We evaluate our methods for the task of segmented style transfer in multiple regions of the content image.
- We compare with state-of-the-art approaches.
- We experiment with multiple style images to infuse multiple styles into the content image.
- We present a comprehensive methodology, accompanied by experiments blending landscapes and paintings.

The structure of the remaining sections in this paper is as follows: Section II provides an overview of the previous studies conducted on neural style transfer and its combination with segmentation techniques. In Section III, we delve into the methodology behind neural style transfer, the segment anything model, and our suggested loss function. Section IV presents the results of our experiments, which evaluate the performance of each individual component as well as the complete proposed approach. Lastly, in Section V, conclusions are drawn, and potential future directions for further research are outlined.

II. RELATED WORK

This section provides an overview of the related work regarding the field of computer vision known as neural style transfer. More specifically, we explore research works concerning classical style transfer techniques as well as their integration with segmentation models.

The technique of neural style transfer is first introduced by Gatys et al. [27]. The basic idea is that it is feasible to separate the style from the content of an image using the feature maps of a pre-trained neural network. Since then this has been an area within computer vision of active research [28], [29], [30]. Gupta et al. [31] introduced a novel loss function called the 'total variation loss', which encourages smoothness in the generated images and helps reduce artifacts and flickering. Additionally, they propose a new optimization algorithm that uses adaptive learning rates to control the stability and convergence of the style transfer process. Huang et al. [32] address the challenge of extending neural style transfer techniques to videos, where real-time performance is required. They propose an approach that enables the application of neural style transfer to video sequences in real-time, allowing for the artistic transformation of each frame based on a desired style. Deng et al. [33] proposed a transformer-based approach which takes into account the long-range dependencies of an image. Qualitative and quantitative experiments demonstrate the effectiveness of the proposed approach compared to classical CNN-based approaches that do not use attention. Yoo et al. [34] proposed a wavelet corrected transfer (WCT) based that allows features to preserve that structural information and statistical propoerties regarding the VGG feature space. Their model achieved remarkable results for high quality images (up to 1024 × 1024) with relatively small inference time ($\sim$ 4.5 seconds). Compared to the current state-of-the-art, their approach had a lesser GPU cost (Gigabytes). Huang et al. [35] proposed a novel adaptive instance normalization (AdaIN) layer that aligns the mean and variance of the content features with those of the style features. Results showed that their method achieves speed comparable to the fastest existing approach, without the restriction to a pre-defined set of style. Chandran et al. [36] introduced an extended version of AdaIN named Adaptive Convolutions (AdaConv). This method allows for the simultaneous transfer of both statistical and structural styles in real time. In addition, AdaConv is also applicable in style-based image generation which was demonstrated through experiments.

The integration of neural style transfer and semantic segmentation is also a topic that has already been explored by researchers [37], [38], [39]. Castillo et al. [22] successfully created a localized style transfer approach with a model that classified pixels as foreground or background. They used Mask R-CNN as a classification algorithm and Markov random fields (MRFs) to deal with the boundaries of the stylized segmentation area. While this approach yielded strong results in many scenarios, experimental findings indicated its limitations in segmenting camouflaged objects, such as those partially submerged underwater and partially visible outside the water. Matsuo et al. [40] proposed a complete system that can segment target object regions using a weakly supervised segmentation method and transfer a given texture style to only the segmented regions. While this method demonstrated strong performance on easily distinguishable objects, it exhibited limitations when applied to objects with rougher edges, as the style transfer technique tended to extend beyond the object boundaries. Virtusio et al. [41] proposed a framework that can selectively apply a given style onto an object using only 4 user-defined points. Their approach combines a style transfer module and an object segmentation module to synthesize the stylized image. Results indicate that their method can generate plausible results for multiple objects within an image as well as for multiple artistic styles. Hand et al. [42] proposed a methodology for local style transfer using masks. These masks were generated either with a histogram or based on the neighborhood of each pixel. Their algorithm produced visually pleasing results but was sensitive to the background of the image. Lin et al. [43] proposed an algorithm that extracts the semantic information of style image and content image automatically through a semantic segmentation network and uses the semantic information to guide the style transfer. Experiments on Celeba and Wikiart show that their method can automatically extract the semantic information of style image and content image. This method performs accurately and the segmentation is robust, however it is not generalizable since the segmentation model has been trained on a specific dataset and can't segment images from a different distribution. Kurzman et al. [21] proposed a class based style transfer technique that combines a segmentation model with a style transfer algorithm to perform real-time localized style transfer. Their approach is based on a pre-trained segmentation model named DBNet and a fast style transfer technique [44] achieving visually appealing results. Even though their approach can operate very fast, it is limited since the segmentation model has been trained on a specific dataset and they use also pre-trained style transfer models able to create images of only one style each.

A. RELATED WORK REVIEW FINDINGS

From the aforementioned related work review, we notice that most researchers use old and outdated segmentation methods which result in poor segmentation performance and thus suboptimal localized style transfer. Also many researchers use segmentation techniques that have been trained on a specific dataset and have constrained generalizability. In light of these limitations, our paper presents a novel contribution that addresses the shortcomings of existing approaches. We propose a segmentation model that seamlessly integrates with a style transfer algorithm while also incorporating a new loss function with a smoothing technique. Our segmentation model overcomes the limitations of weakly supervised methods by incorporating more robust and accurate segmentation techniques. Moreover, a considerable number of segmentation models are trained in advance on particular annotated datasets, yet their ability to generalize effectively to novel data is often limited.

Furthermore, we see that all methodologies simply combine segmentation and style transfer without ensuring a smooth blend between these two separate components. Instead, we introduce a novel border smoothing technique that effectively eliminates artifacts and jagged edges often associated with traditional style transfer methods. By carefully considering the border regions and introducing a novel loss function, we can seamlessly blend the stylized and original content, resulting in visually coherent and aesthetically pleasing images. Our approach thus advances the state-of-the-art in combining these two critical aspects of image processing.

III. METHODOLOGY

The proposed methodology is comprised of three distinct elements. To begin, Facebook's SAM model is employed to achieve precise image segmentation. Following this, the annotated image undergoes a transformation into an augmented segmentation mask, focusing on a particular object within the image. The final stage involves applying localized style transfer to this specific object through the utilization of a VGG convolutional neural network. This whole approach can be viewed on Figure 1. When we analyze this image, we observe that the content image is given to the segment anything model to created annotated segmentation regions. Subsequently, this annotate image is converted to a mask leveraging the user input which is the object of interest (in this case the sky). Following this, the mask along with the content and style image are given to the style transfer model which performs local style transfer only to the specified region.

A. NEURAL STYLE TRANSFER

Neural style transfer is an algorithmic technique that combines mathematics and deep learning to create captivating artistic transformations of images. The process involves optimizing an objective function using mathematical equations. Given a content image C and a style image S, the goal is to generate a new image X that captures the content of C while adopting the style of S. The algorithm leverages a pre-trained Convolutional Neural Network (CNN) to extract feature maps from C, S, and X at different layers. Let F_C, F_S, and F_X represent the respective feature maps. The content loss is computed as the mean squared error between F_C and F_X, given by:

$$L_{\text{content}} = \|F_C - F_X\|^2 \tag{1}$$

The style loss, which measures the differences in the Gram matrices of feature maps across layers, is calculated as:

$$L_{\text{style}} = \sum_l w_l \cdot \|G(S_l) - G(X_l)\|^2 \tag{2}$$

where S_l and X_l are the Gram matrices of F_S and F_X at layer l and w_l represents the weight for each layer. The total loss is a combination of the content and style losses, given by the equation:

$$L_{\text{total}} = \alpha \cdot L_{\text{content}} + \beta \cdot L_{\text{style}} \tag{3}$$

where α and β are hyper-parameters. By minimizing L_{total} using optimization algorithms like gradient descent, the neural style transfer algorithm iteratively adjusts the pixel values of X to converge towards a visually appealing fusion of content and style.

We apply this algorithm using increasing number of iterations to create images that have an ascending degree of fusion with the style image. Moreover, neural style transfer methods can sometimes introduce artifacts or distortions in the stylized images. To battle this, one can adjust the iterations parameter as well as the weights of the style, content loss to achieve the desired result.

Regarding the software implementation of this approach, we employed PyTorch with a pre-train VGG model and a simple optimization loop. The code within the loop is derived from the equations above.

B. SEGMENT ANYTHING MODEL (SAM)

The Segment Anything Model (SAM) is a model that performs segmentation tasks on images. It takes as input a set of points, each consisting of coordinates (x_i, y_i), object membership label l_i, and/or a bounding box defined by the top-left corner (x_{lt}, y_{lt}) and bottom-right corner (x_{rb}, y_{rb}). Additionally, an image I is provided as input.

SAM generates a mask, denoted as Γ, which represents the segmentation of the image according to the constraints specified by the input prompts. The mask Γ indicates the regions in the image that correspond to the foreground objects based on the provided points or bounding boxes.

During the online process, users can interact with SAM in real-time. After adding each point or bounding box, they receive updated segmentation results. If the resulting mask includes areas that are unexpected or should be considered as background, users can add more background points. On the other hand, if the mask lacks expected areas that should be considered as foreground, users can include additional foreground points.

The selection of content and style image masks occurs in pairs. Users specify which regions of the content image should adopt the style of which regions in the style image. This indicates that SAM supports style transfer functionality, where the style of one image can be applied to specific regions of another image.

Overall, the interaction with SAM is represented as a sequence of segmentation results $\Gamma = (\Gamma_{i,c}, \Gamma_{i,s})|\Gamma|i = 1$. Here, $\Gamma i, c$ represents the segmentation mask for the content image, and $\Gamma_{i,s}$ represents the segmentation mask for the style image at the i-th interaction step.

Regarding the software implementation of this step, we utilized the code available in the official SAM GitHub repository.[1] Integrating the model with Python is straightforward;

[1] https://github.com/facebookresearch/segment-anything

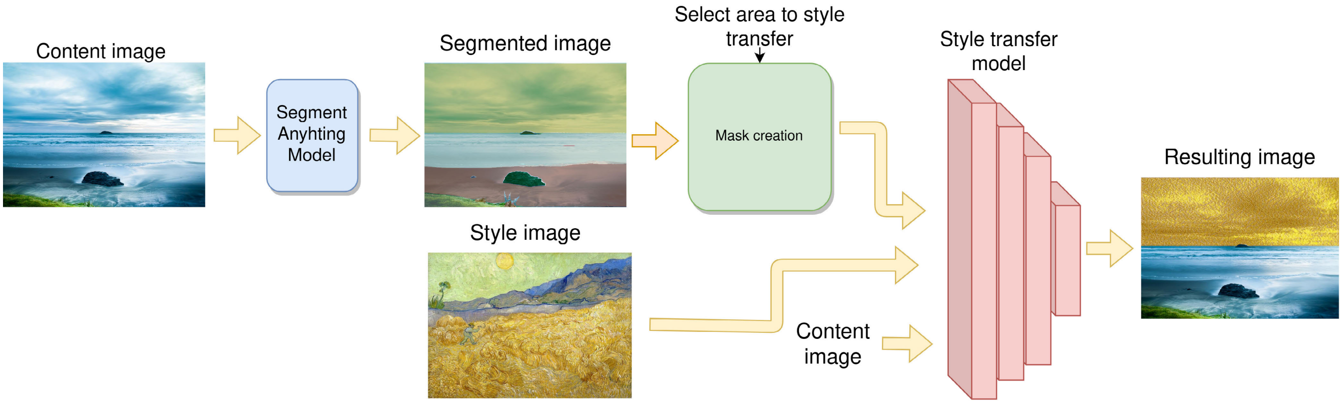


FIGURE 1. Complete architectural approach of SAMStyler.

you simply provide an image as input to the model, and it automatically generates all the masks.

C. LOCAL STYLE TRANSFER AND BORDER SMOOTHING

In this section, we describe how the style transfer algorithm is integrated with the SAM to perform local style transfer in a specific region of interest. After using SAM we end up with a segmentation mask (S) with pixels values of either 1 or 0, where 1 is the area corresponding to the object of interest and 0 elsewhere. An example of such a mask can be viewed on Figure 2. In this, we see an image of a house by the lake in the countryside. Using SAM, we extract the segmentation mask for the house (object). Based on this mask, we create a smooth transition region (M) using morphological operations on the segmentation mask. Specifically, dilation (D) expands the binary regions defined by S using a structuring element (B):

$$D(S) = S \oplus B. \tag{4}$$

This dilation can be configured using the *iterations* parameter which controls the number of times the dilation is applied. The higher this number, the thicker the smooth transition region is.

The smooth transition region is obtained by subtracting the original segmentation mask from the dilated map:

$$M = D(S) - S \tag{5}$$

We set the values of M to a number close to 1 (e.g., 0.9) and consequently add it to the original segmentation mask resulting in an augmented segmentation mask A:

$$A = M + S \tag{6}$$

An example of this is depicted in subfigure III-C. Here, we see that the original segmentation mask of the house has been dilated and the smooth transition area has a grey color (values of 0.9). On the other hand, the black area has values of 1 and the white values of 0.

The last step is to invert the augmented segmentation mask as follows:

$$I_A = 1 - A \tag{7}$$

Using this inverted augmented segmentation mask, we formulate a new loss function for the style transfer model, introducing an additional regularization component, as shown in Equation 8:

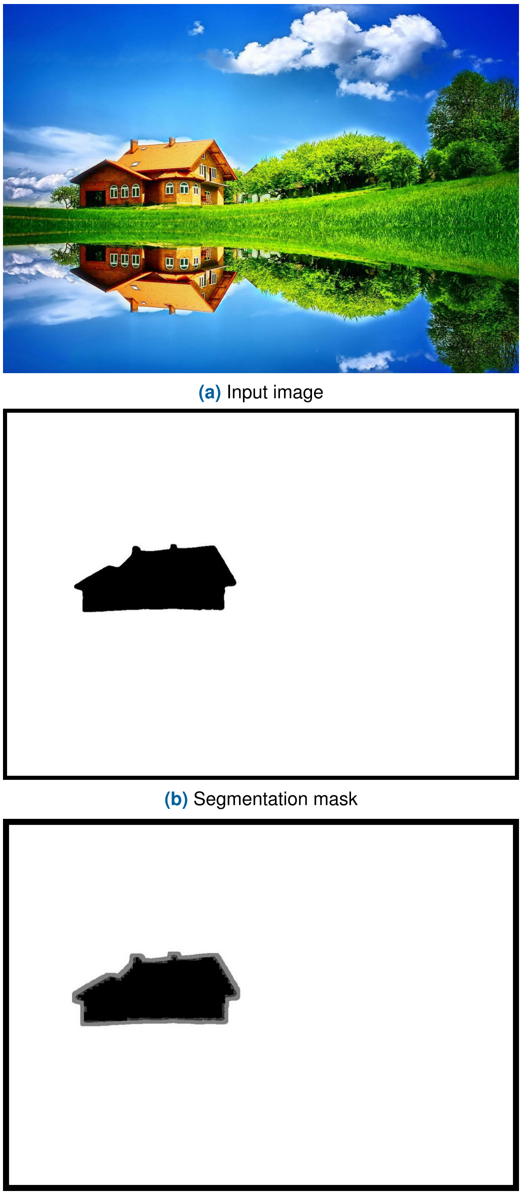
$$L_{\text{total}} = \alpha \cdot L_{\text{content}} + \beta \cdot L_{\text{style}} + c \cdot L_{\text{reg}} \tag{8}$$

where the L_{style} and $L_{content}$ remain the same. The last factor, namely L_{reg} is defined as:

$$L_{\text{reg}} = \|C \times I_A - G \times I_A\|^2 \tag{9}$$

where C is the original content image and G is the generated image by the style transfer process.To enforce regularization, we amplify this loss by a substantial factor, typically a large value like 10000, denoted as c. This loss is specifically computed for regions of the image that correspond to non-zero values in the inverted augmented segmentation mask. Also, one key difference here is that the staring image G is not random noise but rather is set to be the same as the content image C (in comparison to the original neural style transfer approach, section III-A).

The intuition behind this new loss function is that, since L_{reg} starts by being 0 (initially C and G are the same) and it has a very large scaling parameter, the optimizer will be forced to keep it close to 0 to minimize the total loss. This will subsequently compel the generated image to be the same as the content image outside of the augmented segmentation mask. Specifically, inside the object we expect the optimization to alter the pixel values to minimize the style loss and content loss (exactly like the original neural style transfer method) whereas outside the optimizer will keep the pixels intact. Regarding the in-between area (smooth transition region M, values set to $1-0.9 = 0.1$) the optimizer will alter the pixel values to minimize the style, content losses but with the mask M working as a regularization



(a) Input image

(b) Segmentation mask

(c) Augmented segmentation mask

FIGURE 2. Segmentation results and input.

(multiplication with 0.1) to limit the intensity of the style transfer process for this specific region.

In summary, the anticipation is that the model will perform intense style transfer within the segmentation region, slight style transfer within the smooth transition region, and no style transfer outside these designated areas.

The implementation of this loss in the software was achieved by utilizing the cv.dilate[2] function from openCV, operating on the output segmentation mask produced by SAM.

IV. RESULTS

Within this section we describe the outcomes of the proposed methodology, SAMStyler. As a starting point, we demonstrate results of the style transfer algorithm with varying degrees of intensity. Subsequently, we illustrate SAMStyler's results for both a content image and a painting as a stylized image, where we apply style transfer to a specific object within the image. In this configuration, we conduct experiments with varying numbers of dilation iterations to assess the impact of this modification on the model's performance. We also compare our approach with a state-of-the-art localized style transfer method to prove robustness. Lastly concerning this partial style transfer, we use multiple styling images to apply different artistic styles to different segmentation regions within the original content image.

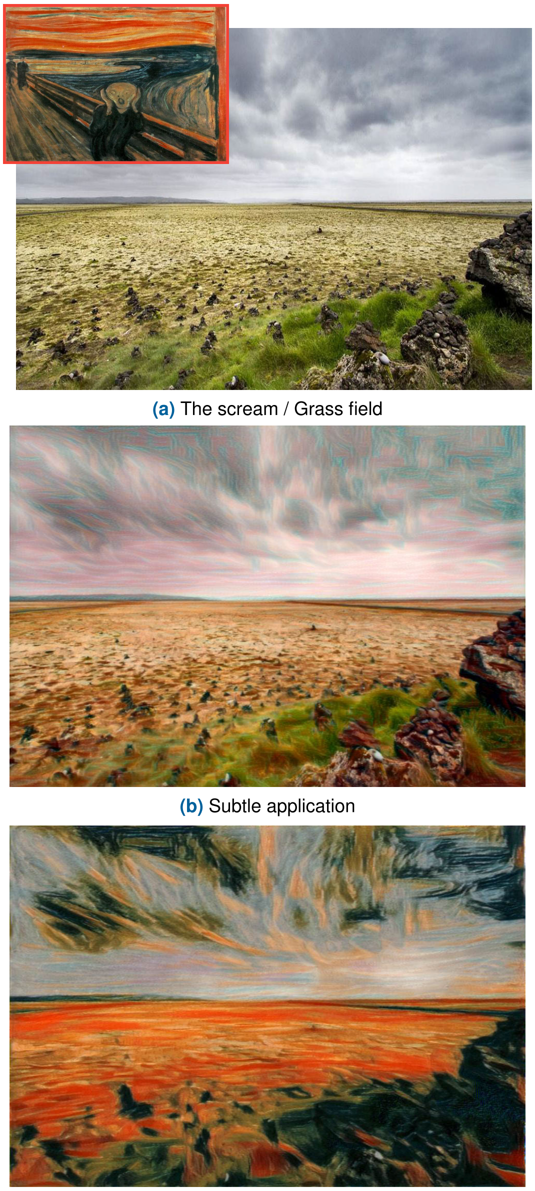
A. STYLE TRANSFER

In this subsection, we present the outcomes of style transfer with varying degrees of application. We demonstrate the effects of applying the style image to the content image in both subtle and pronounced manners showcasing the resulting transformations. The results of this process are exhibited in Figure 3. As mentioned before, the factor that determines the intensity of the style transfer operations are the epochs of the optimization process. To showcase diverse outcomes, we applied the identical process with both fewer and greater rounds. Here, we have a landscape content image depicting a grass field, accompanied by a style image, namely the ''The Scream'' painting of Edvard Munch. Using a VGG-based style transfer algorithm, we perform style transfer between these images in both a subtle and a strong manner. The middle picture demonstrates a slighter fusion of the painting and the right a more intense. The outcomes are indeed convincing, and the generated images exhibit Edvard Munch's characteristic traits of bold, dramatic brushwork and the compelling utilization of intense colors. Specifically, we notice that the brushwork characteristic of 'The Scream' has been assimilated into the image, resulting in an artistic expressionism style being adopted. These variations in style transfer intensity highlight the flexibility and artistic potential of our approach.

B. SAMSTYLER RESULTS FOR DIFFERENT DILATION ITERATIONS

In this subsection, we demonstrate SAMStyler results for a single styling image and different dilation iterations. The results of this process are demonstrated in subfigure 5. Unlike applying style transfer to the whole image, our approach involves performing style transfer in separate segmentation regions of the content image. To begin, the top left picture consists of the original content image which are two wolves in their natural habitat. Adjacent to it is the style image, ''Starry Night'', painted by Vincent Van Gogh (top-left). Based on this content image, we demonstrate SAM segmentation results on the top-right sub-figure. We observe that the model has correctly segmented the image in appropriate regions, namely each wolf separately, the forest background, and the logs at the bottom of the picture. It is also worth noting that these

[2] https://docs.opencv.org/3.4/db/df6/tutorial_erosion_dilatation.html



(a) The scream / Grass field

(b) Subtle application

(c) Strong application

FIGURE 3. Style transfer results.

segmentation regions can be overlapping (i.e. we can select a specific log as a separate object or all the logs combined).

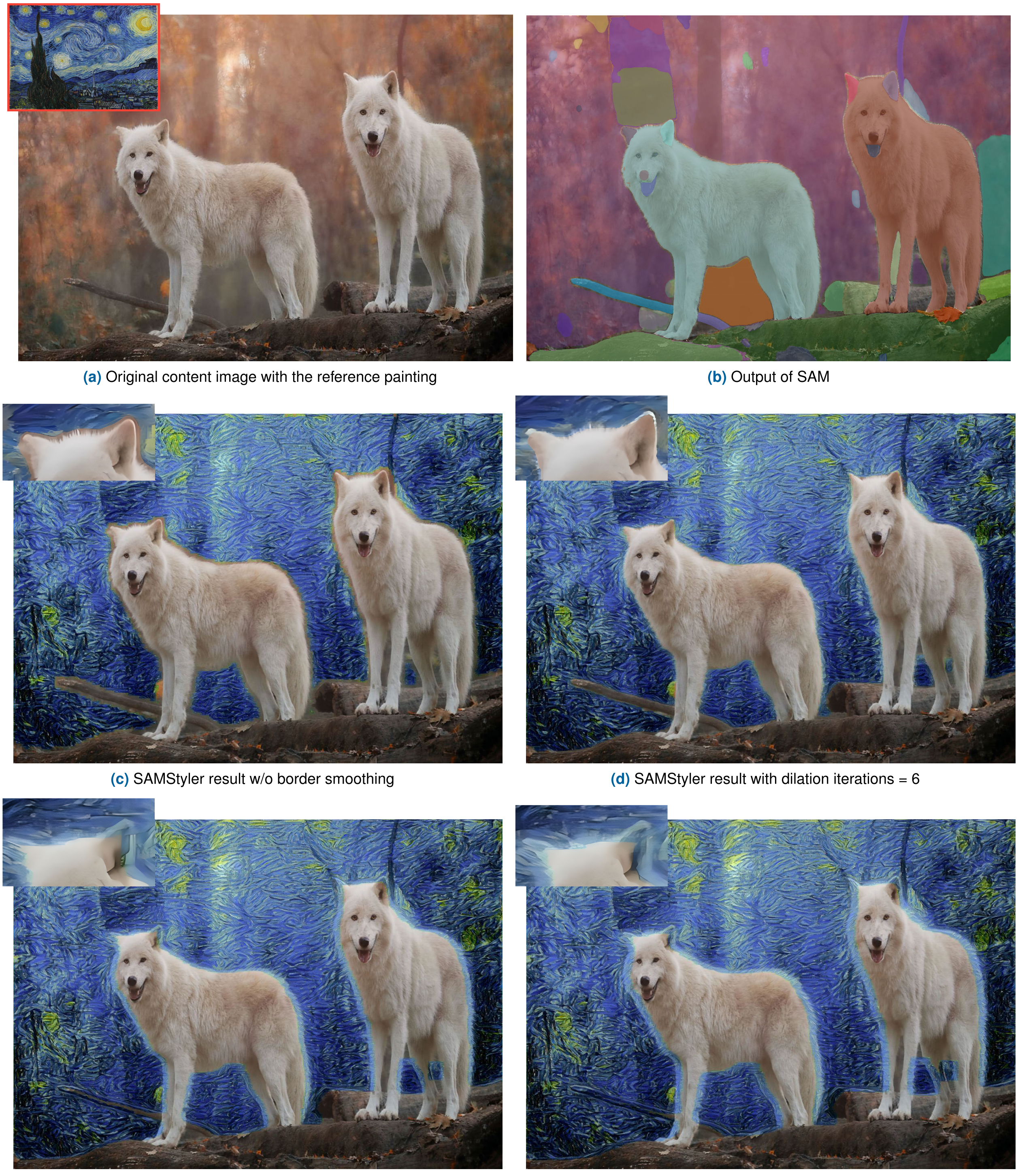
The rest of the sub-figures display the fusion results of SAMStyler. Specifically, Figure 4 displays a fusion image where the Starry Night painting has been applied to the forest background of the content picture. We notice that the application is successful and the painting has been applied to the background. However, if we zoom near the segmentation border (i.e. near the edge of the wolf's head and the background) we see that there is a small gap. We examine this phenomenon more closely and display it on the top-left of this fusion image. We observe that the segmentation is not precise enough and creates a contour around the wolf resulting in a rough transition. This experimental result confirms the needs of the augmented segmentation mask discussed in section III-C. On the other hand, subfigure IV-B displays the same fusion picture with the augmented segmentation mask instead of a simple segmentation mask. For this particular case, the smooth transition region has been created by dilating the segmentation mask using 6 iterations. The visual improvement of the resulting image and the blending of boundaries between stylized and non-stylized areas are clearly evident through the proposed smoothing technique. The outcome is visually attractive, exhibiting seamless transitions between the segmented areas. To highlight the effectiveness of our approach, we have included a small zoomed-in image in the top-right corner of this result. This small image can be directly compared to the one without the smoothing effect to emphasize visual difference.

The number of iterations serves as a delicate parameter and demands judicious selection to bolster image quality and establish a harmonious blending zone. In order to explore this facet and exhibit outcomes, we conduct a series of experiments with varying values for this parameter. Going into more depth, subfigure VI-B portrays the outcome of SAMStyler with 12 dilation iterations. Evidently, we observe a subtle infusion of the 'Starry night' style into the wolves. To clarify, we zoom in on the head of the left wolf to highlight the new boundary in this area, facilitating a direct comparison with other results. Furthermore, it's important to note the discernible impact of the regularization applied to the smooth segmentation region, as opposed to the segmentation mask. This is evident in the somewhat lesser degree to which the style has been applied there. Lastly, depicted in subfigure VI-B, the outcome for 18 dilation iterations is displayed. In this case, the style has been even more intricately integrated into the wolves, a fact readily apparent when zoomed in on the wolf's head. Consequently, tuning this parameter in alignment with the user's artistic vision is crucial for achieving the desired outcome.

C. SAMSTYLER COMPARISON WITH OTHER APPROACHES

Within this section, we undertake a comparison between our approach and the technique put forth by Kurzman et al. [21], specifically known as CBStyling. This method combines a pre-trained segmentation model (DBNet) with pre-trained style transfer models to perform local style transfer. The blending is done in a post-processing manner since the style transfer is initially performed in the whole image and later is restricted by a generated mask. It's essential to highlight that the segmentation model has been trained on a specific dataset, rendering it effective for images sharing the same distribution. Additionally, each individual style transfer model has undergone training to generate a distinct style and isn't conducive to generalizing across multiple styles. The implementation of this approach can be found on GitHub.[3] Our objective is to showcase the resilience of our model in contrast to a well-established method, shedding light

[3] https://github.com/IssamLaradji/CBStyling



(a) Original content image with the reference painting

(b) Output of SAM

(c) SAMStyler result w/o border smoothing

(d) SAMStyler result with dilation iterations = 6

(e) SAMStyler result with dilation iterations = 12

(f) SAMStyler result with dilation iterations = 18

FIGURE 4. **SAMStyler results for a single styling image.**

on the distinctive architectural strategies we employ when comparing with alternative algorithms.

The experiments conducted to compare CBStyling and SAMStyler is depicted in Figure 5. Every distinct experiment is presented in a separate row within the figure. Concerning the initial row, the left side showcases an urban image side by side with the painting utilized as the style reference. The primary focus revolves around the road, and we illustrate the outcomes achieved through the two approaches. To commence, it is notable that

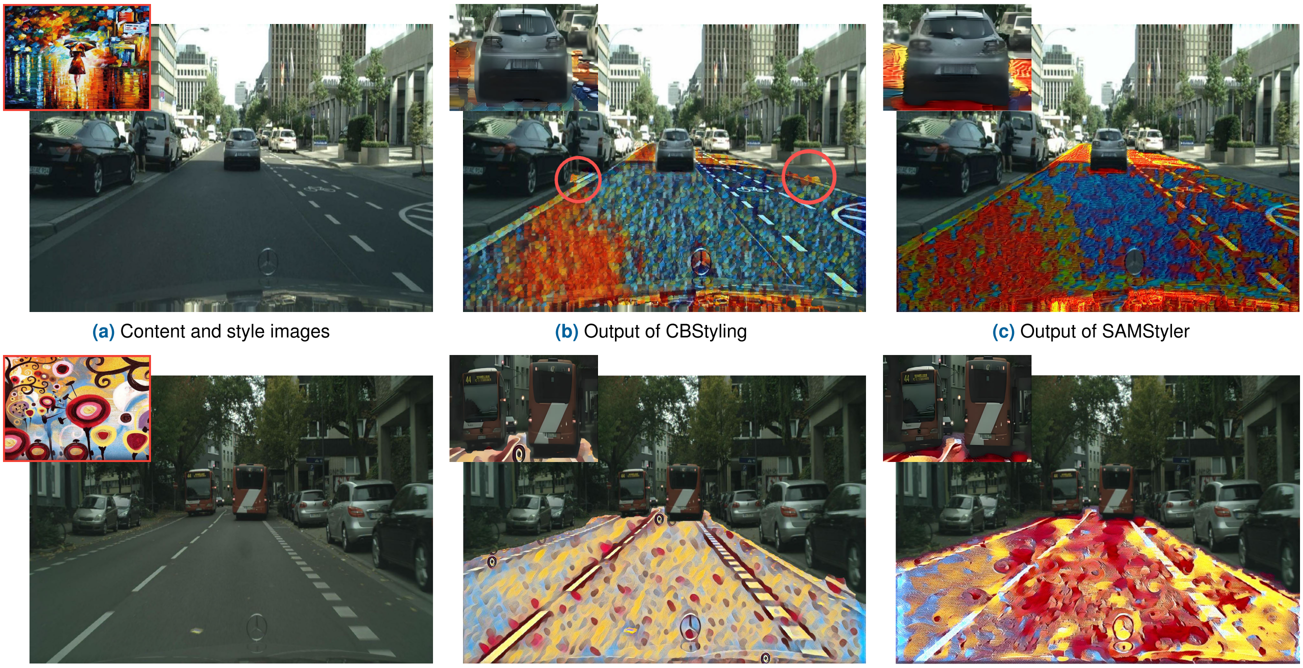

(a) Content and style images
(b) Output of CBStyling
(c) Output of SAMStyler
(d) Content and style images
(e) Output of CBStyling
(f) Output of SAMStyler

FIGURE 5. Comparison of CBStyling and SAMStyler.

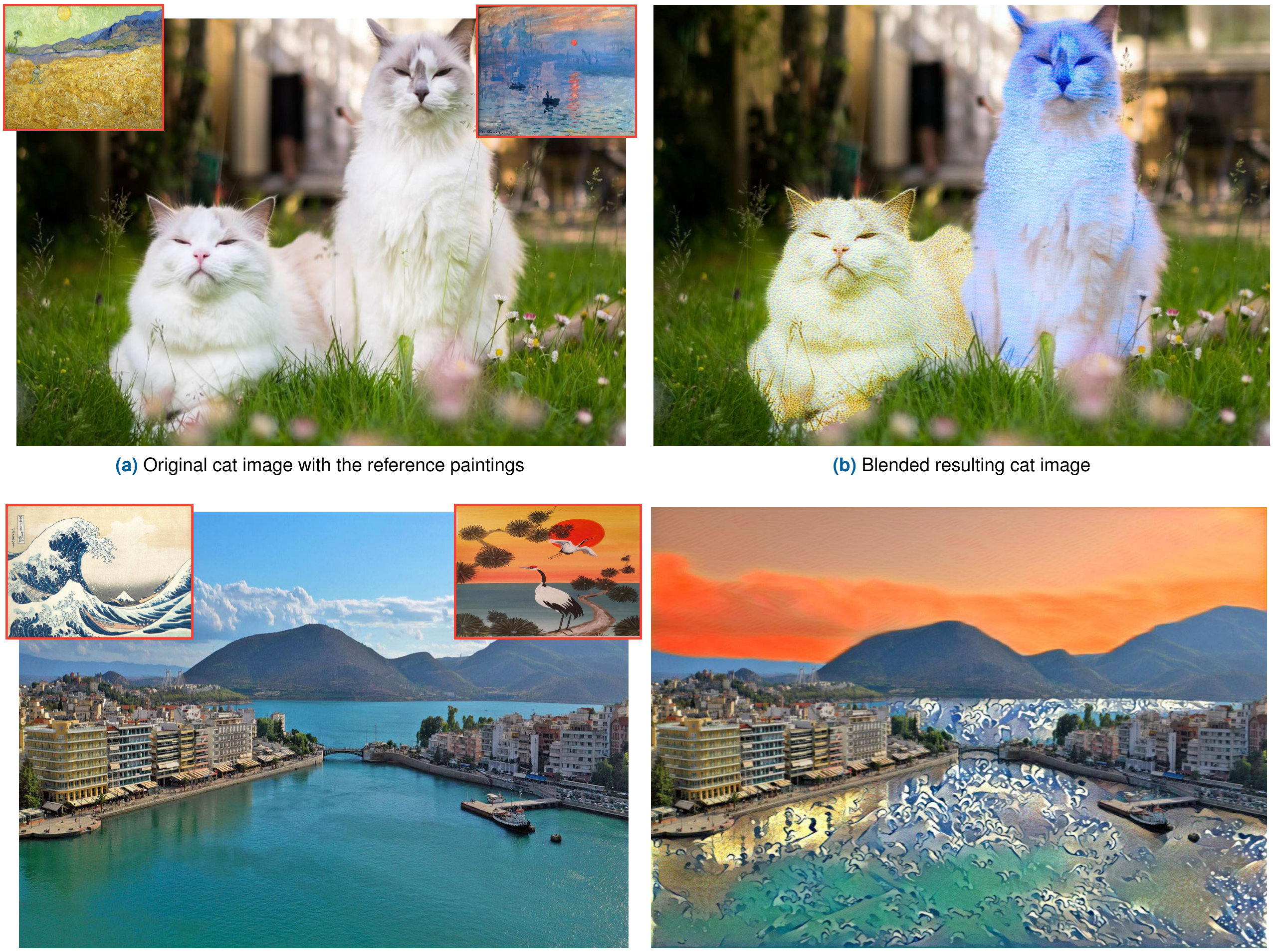
SAM's segmentation outperforms the alternative, effectively distinguishing the road from the surrounding elements. A closer examination of sub-figures VI-C, particularly the areas highlighted by red circles, reveals that DBNet mistakenly segments portions of the sidewalk as part of the road. Additionally, directing attention to the top-left corners of sub-figures VI-C and VI-C, we magnify the car to underscore the efficacy of the smooth transition region. The comparison highlights that CBStyling exhibits a perceptibly abrupt boundary between stylized and non-stylized regions, a challenge effectively addressed by SAMStyler. Analogous observations extend to the second experiment. Notably, in the zoomed-in image of the buses, the transition between the two zones is impeccably smooth in SAMStyler's output. This effect serves as a contour delineating the entire segmentation region.

Furthermore, it's worth noting that within the segmentation region, slight variations exist in the styling results—a consequence of employing diverse style transfer techniques.

D. SAMSTYLER FOR MULTIPLE STYLING IMAGES

In this subsection, we demonstrate SAMStyler results using multiple styling images where each styling image is used in a specific section of the segmented content image. The results encompass renowned paintings by Monet and Van Gogh, which can be observed in Figure 6. Regarding the first content picture, it features two cats in a grass yard. Also, we can see in the top left and right of the original picture the two paintings that were used for stylization. These renowned paintings are "Wheatfield With a Reaper" and "Impression, Sunrise". It is demonstrated that the two separate styles can be incorporated concurrently into the content image where each style is applied to a different object (cat). The result of SAMStyler is displayed on the right side of this figure. Upon examining the image, we observe that each cat has the stylistic features of the corresponding image without alternations in the background. The segmentation is performed smoothly and the outcome is aesthetically appealing. Moreover, it is worth noting that in this image the cats are in close proximity, however the distinction by the SAM model is highly accurate. Also, using our method of smooth transition the segmentation border between the two cats appears natural.

On the bottom-left of this figure we see a town as a content image and two reference paintings, namely 'The Great Wave off Kanagawa' and 'Japanese Sunset'. We apply the sunset painting to the sky of the content image and the wave artwork to the sea. The result of this choice is demonstrated in the bottom-right of the figure. It is evident that the style of the wave painting has been successfully transferred to the sea in the original town image, resulting in an visually satisfying outcome. Additionally, the application of the sunset style to the sky has transformed it from a daytime scene to resemble a genuine sunrise painting. It is also apparent, that the blending between the stylized and non stylized regions is smooth thanks to our introduced border smoothing technique. These examples showcase the versatility and effectiveness of SAMStyler in applying multiple styles to different regions of complex content images. This capability opens up a wide range of creative possibilities for artists and designers, allowing them to seamlessly integrate diverse



(a) Original cat image with the reference paintings

(b) Blended resulting cat image

(c) Original city image with the reference paintings

(d) Blended resulting city image

FIGURE 6. SAMStyler results for multiple styling images.

artistic styles into complex compositions while maintaining visual coherence.

V. CONCLUSION AND FUTURE WORK

Style transfer is a technique used to apply the artistic style of one image (referred to as the style image) to another (referred to as the content image). It aims to create a new one that combines the content of the content image with the visual style of the style image. This approach can be combined with segmentation to create locally stylized images that contain different styles for separate objects. In this work, we delve into this possibility further by integrating Segment Anything Model with the style transfer methodology. We also propose a loss function incorporating a border smoothing technique that creates a fluid transition between the areas that have been stylized and those that have not. Such a step ensures that the shift between the stylized and non-stylized regions is seamless and does not have a small gap. The combined architectural approach results in the creation of a model named SAMStyler. Results demonstrate that this combination is robust yielding visually attractive results where multiple styles have been fused into the content image. Through experiments, it is also proved that the proposed technique is performs better compared to state-of-the-art approaches, paving the way for enhanced and more versatile artistic stylization in digital image processing.

In the future, we aim to test this approach with more complex datasets that contain more objects that are not so easily distinguishable. As an example, hidden or camouflaged objects. To comprehensively assess performance under these conditions, we are also committed to employing relevant metrics tailored to the specific complexities of such datasets. We also aim to try different style transfer models that may produce more accurate results. Specifically, we would like to test transformer-based approaches using the above methodology. Additionally, we plan to evaluate the performance of our border smoothing technique in real-world scenarios, considering factors like varying lighting conditions and diverse backgrounds, to ensure its robustness and practical applicability. At last, we aim to conduct user studies to gather feedback and insights on the aesthetic quality and user satisfaction with the stylized outputs generated by our approach. This user-centric evaluation will enable us to further refine and improve our methodology, ensuring its effectiveness and usability in real-world creative applications.

REFERENCES

[1] K. Psychogyios, L. Ilias, C. Ntanos, and D. Askounis, "Missing value imputation methods for electronic health records," *IEEE Access*, vol. 11, pp. 21562–21574, 2023.

[2] K. Psychogyios, L. Ilias, and D. Askounis, "Comparison of missing data imputation methods using the Framingham heart study dataset," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 1–5.

[3] S. Chauhan, S. Manmohan, and A. Kumar, "Data science and data analytics: Artificial intelligence and machine learning integrated based approach," in *Data Science and Data Analytics: Opportunities and Challenges*. Boca Raton, FL, USA: Chapman & Hall, 2021.

[4] M. Dehghani, A. Gritsenko, A. Arnab, M. Minderer, and Y. Tay, "SCENIC: A JAX library for computer vision research and beyond," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21393–21398.

[5] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 1, pp. 105–117, Jan. 2023.

[6] G. K. Thiruvathukal and Y.-H. Lu, "Efficient computer vision for embedded systems," *Computer*, vol. 55, no. 4, pp. 15–19, Apr. 2022.

[7] Z. Huang, S. Yang, M. Zhou, Z. Li, Z. Gong, and Y. Chen, "Feature map distillation of thin nets for low-resolution object recognition," *IEEE Trans. Image Process.*, vol. 31, pp. 1364–1379, 2022.

[8] F. Ashiq, M. Asif, M. B. Ahmad, S. Zafar, K. Masood, T. Mahmood, M. T. Mahmood, and I. H. Lee, "CNN-based object recognition and tracking system to assist visually impaired people," *IEEE Access*, vol. 10, pp. 14819–14834, 2022.

[9] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu, "Physical inertial poser (PIP): Physics-aware real-time human motion tracking from sparse inertial sensors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13167–13178.

[10] U. G. Longo, S. D. Salvatore, M. Sassi, A. Carnevale, G. D. Luca, and V. Denaro, "Motion tracking algorithms based on wearable inertial sensor: A focus on shoulder," *Electronics*, vol. 11, no. 11, p. 1741, May 2022.

[11] S. E. Salcudean, H. Moradi, D. G. Black, and N. Navab, "Robot-assisted medical imaging: A review," *Proc. IEEE*, vol. 110, no. 7, pp. 951–967, Jul. 2022.

[12] B. Hu, S. Zhou, Z. Xiong, and F. Wu, "Cross-resolution distillation for efficient 3D medical image registration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 7269–7283, Oct. 2022.

[13] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 25–38, Jul. 2010.

[14] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[15] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6997–7005.

[16] X.-C. Liu, M.-M. Cheng, Y.-K. Lai, and P. L. Rosin, "Depth-aware neural style transfer," in *Non-Photorealistic Animation and Rendering*, H. Winnemoeller and L. Bartram, Eds. New York, NY, USA: Association for Computing Machinery, 2017.

[17] M. Garg, J. S. Ubhi, and A. K. Aggarwal, "Neural style transfer for image steganography and destylization with supervised image to image translation," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 6271–6288, Aug. 2022.

[18] Mallika, J. S. Ubhi, and A. K. Aggarwal, "Neural style transfer for image within images and conditional GANs for destylization," *J. Vis. Commun. Image Represent.*, vol. 85, May 2022, Art. no. 103483.

[19] N. Q. Tuyen, S. T. Nguyen, T. J. Choi, and V. Q. Dinh, "Deep correlation multimodal neural style transfer," *IEEE Access*, vol. 9, pp. 141329–141338, 2021.

[20] M.-M. Cheng, X.-C. Liu, J. Wang, S.-P. Lu, Y.-K. Lai, and P. L. Rosin, "Structure-preserving neural style transfer," *IEEE Trans. Image Process.*, vol. 29, pp. 909–920, 2020.

[21] L. Kurzman, D. Vazquez, and I. Laradji, "Class-based styling: Real-time localized style transfer with semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3189–3192.

[22] C. Castillo, S. De, X. Han, B. Singh, A. K. Yadav, and T. Goldstein, "Son of Zorn's lemma: Targeted style transfer using instance-aware semantic segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1348–1352.

[23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023, *arXiv:2304.02643*.

[24] J. Ma and B. Wang, "Segment anything in medical images," 2023, *arXiv:2304.12306*.

[25] A. I. Karimov, E. E. Kopets, V. G. Rybin, S. V. Leonov, A. I. Voroshilova, and D. N. Butusov, "Advanced tone rendition technique for a painting robot," *Robot. Auto. Syst.*, vol. 115, pp. 17–27, May 2019.

[26] E. S. Mikalonytė and M. Kneer, "Can artificial intelligence make art?: Folk intuitions as to whether AI-driven robots can be viewed as artists and produce art," *J. Hum.-Robot Interact.*, vol. 11, no. 4, pp. 1–19, Sep. 2022.

[27] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*.

[28] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 11, pp. 3365–3385, Nov. 2020.

[29] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite, and K. Kotecha, "Neural style transfer: A critical review," *IEEE Access*, vol. 9, pp. 131583–131613, 2021.

[30] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stereoscopic neural style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6654–6663.

[31] A. Gupta, J. Johnson, A. Alahi, and L. Fei-Fei, "Characterizing and improving stability in neural style transfer," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. Los Alamitos, CA, USA:. IEEE Computer Society, Oct. 2017, pp. 4087–4096.

[32] H. Huang, H. Wang, W. Luo, L. Ma, W. Jiang, X. Zhu, Z. Li, and W. Liu, "Real-time neural style transfer for videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7044–7052.

[33] Y. Deng, F. Tang, W. Dong, C. Ma, X. Pan, L. Wang, and C. Xu, "Stytr^2: Image style transfer with transformers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1–11.

[34] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1–10.

[35] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1510–1519.

[36] P. Chandran, G. Zoss, P. Gotardo, M. Gross, and D. Bradley, "Adaptive convolutions for structure-aware style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7972–7981.

[37] W. Ye, C. Liu, Y. Chen, Y. Liu, C. Liu, and H. Zhou, "Multi-style transfer and fusion of image's regions based on attention mechanism and instance segmentation," *Signal Process., Image Commun.*, vol. 110, Jan. 2023, Art. no. 116871.

[38] M. Reimann, M. Klingbeil, S. Pasewaldt, A. Semmo, M. Trapp, and J. Döllner, "Locally controllable neural style transfer on mobile devices," *Vis. Comput.*, vol. 35, pp. 1–17, Apr. 2019.

[39] X. Fang, "Neural style transfer with content feature segmentation," *Highlights Sci., Eng. Technol.*, vol. 34, pp. 53–59, Feb. 2023.

[40] S. Matsuo, W. Shimoda, and K. Yanai, "Partial style transfer using weakly supervised semantic segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 267–272.

[41] J. J. Virtusio, A. Talavera, D. S. Tan, K.-L. Hua, and A. Azcarraga, "Interactive style transfer: Towards styling user-specified object," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2018, pp. 1–4.

[42] A. Handa, P. Garg, and V. Khare, "Masked neural style transfer using convolutional neural networks," in *Proc. Int. Conf. Recent Innov. Electr., Electron. Commun. Eng. (ICRIEECE)*, Jul. 2018, pp. 2099–2104.

[43] Z. Lin, Z. Wang, H. Chen, X. Ma, C. Xie, W. Xing, L. Zhao, and W. Song, "Image style transfer algorithm based on semantic segmentation," *IEEE Access*, vol. 9, pp. 54518–54529, 2021.

[44] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV 2016*, B. Leibe, J. Matas, N. Sebe, M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 694–711.

KONSTANTINOS PSYCHOGYIOS received the M.Eng. degree in electrical and computer engineering with the National Technical University of Athens (NTUA), Greece. He is currently a Machine Learning Engineer with the industry, where he is implementing and researching solutions, such as IDS for federated systems and GAN for image generation. He has published and presented research papers at international conferences organized by reputable organizations, such as IEEE and Springer. His research interests include cybersecurity and bioinformatics, using machine learning techniques.

HELEN C. (NELLY) LELIGOU received the Dipl.-Ing. and Ph.D. degrees from the Department of Electrical and Computer Engineering, National Technical University of Athens, Greece. From 2007 to 2017, she was an Assistant Professor with the Technological Educational Institute of Sterea Ellada. She is currently an Associate Professor with the University of West Attica. She is also involved in blockchain technologies and their combination with artificial intelligence and federated learning techniques. She is/was a scientific coordinator of the LIFE-GENERA and H2020-ASSET Project. She has participated in more than 20 EU-funded ACTS, IST, ICT, and H2020 research projects in the above areas and also acts as an evaluator for national and EU-funded proposals. Her research interests include computer networks and information-and-communication-technologies, such as routing protocols and trust management in wireless sensor networks, control plane technologies in broadband networks, including HFC, PON, WDM metro, and core networks, embedded and network system design and development, and the IoT enabled solutions for different application sectors like energy efficiency/optimization in buildings and for affect detection in learning environments.

FILISIA MELISSARI received the degree in computer science and biomedical informatics from the University of Thessaly. She is currently a Software Engineer who is driven by a keen interest in emerging technologies and focuses her developments on blockchain technology and its applications. She has expertise in designing blockchain infrastructures, implementing smart contracts, and leveraging distributed ledger technology for various applications. She is committed to pushing the boundaries of knowledge in blockchain engineering and advancing its potential impact.

STAVROULA BOUROU received the M.Eng. degree in rural and surveying engineering from the National Technical University of Athens (NTUA), Greece, in 2015, and the M.Sc. degree in geodesy and geoinformation science with specialization in computer vision from Technical University Berlin, Germany, in 2019. Since 2019, she has been a Machine Learning Engineer. She has experience in researching different aspects of artificial intelligence (AI), including among others enhancement of IoT cybersecurity, GAN models in cybersecurity, federated learning, and privacy-preserving deep learning. In addition, she has hands-on experience in building complete AI workflows, from data collection and model creation until deployment to production. Her research interests include computer vision, AI methods for precision agriculture, and AI for satellite image analysis. She is a member of the European Commission sub-group on AI, connected products, and other new challenges in product safety.

ZACHARIAS ANASTASAKIS received the M.Eng. degree in electrical and computer engineering (ECE) from the National Technical University of Athens (NTUA), Greece, in 2023.

From 2022 to 2023, he was a Diploma Thesis Intern with the DeepLab, Athens, Greece. He is currently a Machine Learning Engineer with Synelixis Solutions S.A., Greece, focusing on cyber-security with machine learning and developing privacy-preserving techniques in federated learning systems. His research interests include deep learning, federated learning, computer vision, and natural language processing.

THEODORE ZAHARIADIS received the Dipl.-Ing. degree in computer engineering from the University of Patras, Greece, and the Ph.D. degree in electrical and computer engineering from the National Technical University of Athens, Greece. He is currently a Professor with the National and Kapodistrian University of Athens. Since 1997, he has been the Project Manager or the Technical Manager of many EU-funded projects. He has published more than 160 papers in magazines, journals, and conferences, and has more than 2500 citations (H-Index of 26). His research interests include energy efficiency, smart grids, network virtualization, and wireless sensor networks. He has been a member of the technical board in multiple scientific conferences and workshops and the lead guest editor in various magazines and journals.