VectorPD: Artistic Portrait Drawing with Vector Strokes

Yiqi Liang, Ying Liu, Dandan Long, Ruihui Li[*]

College of Computer Science and Electronic Engineering, Hunan University, China
{yiqiliang, liu_ying, ddlong, liruihui}@hnu.edu.cn

Abstract

In this paper, we present a method, VectorPD, for converting a given human face image into a vector portrait sketch. VectorPD supports different levels of abstraction by simply controlling the number of strokes. Since vector graphics are composed of different shape primitives, it is challenging for rendering complex faces to accurately express facial details and structure. To address this, VectorPD employs a novel two-round optimization mechanism. We first initialize the strokes with facial keypoints, and generate a basic portrait sketch by a CLIP-based Semantic Loss. Then we complete the face structure through VGG-based Structure Loss, and propose a novel Crop-based Shadow Loss to enrich the shadow details of the sketch, achieving a visually pleasing portrait sketch. Quantitative and qualitative evaluations both demonstrate that the portrait sketches generated by VectorPD can produce better visual effects than existing state-of-the-art methods, maintaining as much fidelity as possible at different levels of abstraction.

Introduction

Visual abstraction is regarded as an abstract, minimal expression that can convey the salient visual features of an image more effectively and efficiently (Biederman and Ju 1988; Hertzmann 2020). It is a valuable visual tool for expressing ideas, concepts and actions (Fan, Yamins, and Turk-Browne 2018; Gryaditskaya et al. 2019; Xu et al. 2022). In the paper, we focus on a simple but important domain: portrait sketching.

Today, many works related to portrait sketching often reflect the main features of portrait in an abstract style by using rendering, such as watercolor (Rosin and Lai 2018), geometric abstract (Tian and Ha 2022), oil painting (Winnemöller, Kyprianidis, and Olsen 2012), point painting (Li and Mould 2011), etc. While these works delivered captivating portrait results, they were confined to generating raster images, limiting their pragmatic applicability in the field of computer graphics. In recent years, Scalable Vector Graphics (SVG) (Ma et al. 2022) has received a lot of attention due to its advantages of compact file size and resolution independence. More notably, vector images supply dispersed
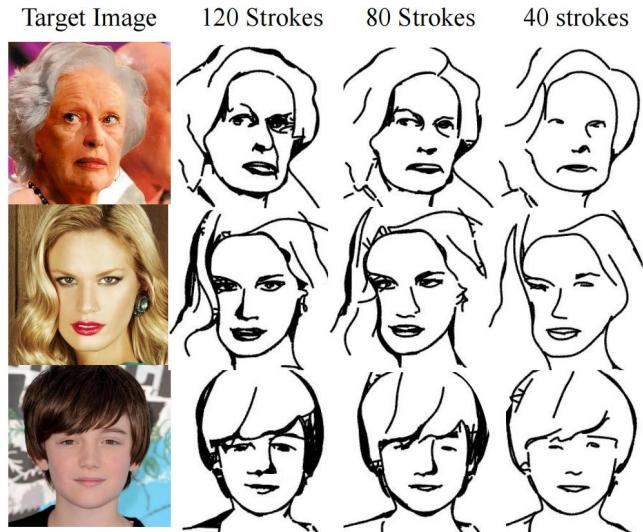
---

[*]Corresponding author.

Figure 1: Examples of our method. VectorPD is able to generates vector portrait sketches at different levels of abstraction L, containing facial features and structure as much as possible. Left to right: The total number of strokes S is set to 120, 80, 40 and the number of strokes in each optimization round is evenly divided.

visual data, empowering artists or users to undertake subsequent edits and abstract processing of portrait sketches (see Figure 1, Figure 2) with greater ease.

To further expand the field of portrait sketching, we propose a question: *Can computer rendering convert photos of human faces from raster to vector depictions?*

CLIPasso (Vinker et al. 2022) proposed an optimization-based photo-sketch generation technique without requiring an explicit sketch dataset. But it is a general study of broad visual abstraction. Compared with other types of images, generating portrait sketch is more difficult and uniquely challenging. Because the intricate nature of the human face, encompassing subtle features like eyes, nose, and mouth, etc. requires more finesse to capture and express these features. In addition, people have sensitivity to facial structures (Shao, Weng, and He 2017). Capturing the proportions and symmetry of the face accurately also require more pre-
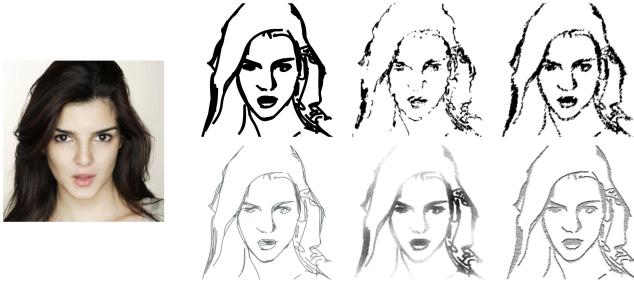
Figure 2: Editing the brush style on SVGs. Our method generates portrait sketches in vector form, which can be easily used by designers for further editing.

cise capabilities.

In this paper, we propose VectorPD, a novel method for generating vector-level portrait sketches. We define two optimization round of portrait sketching. In the first round, we utilize the face extractor to obtain the facial key points of the input human face image, and use the CLIP image encoder to guide these key points to generate basic semantic concepts of the human face. The second round is based on the sketch results from the previous round, strengthening the contour of the portrait sketch according to the key points provided by the contour extractor, and deepening some key lines, such as accessories, hair, jawline, etc., to highlight the structure and characteristics of human face.

Our portrait sketches are defined using a set of black Bézier curves as strokes placed on a white background. To improve fidelity, we uniformly set key point positions based on the point set from extractors. And the degree of abstraction of our portrait sketch can also be determined by the number and distribution of strokes used for face N_f and contour N_c.

Given a target human face image, we use a differentiable rasterizer (Li et al. 2020) to directly optimize the parameters (control points positions) of strokes with respect to a CLIP-based Semantic Loss. We combine the final and intermediate activations of a pretrained CLIP model to generate a base portrait sketch containing basic semantic information. In order to improve the integrity and artistry of the portrait sketch, we used VGG-based Loss in the second round of optimization to improve the structure of the sketch and proposed a novel distance loss, Crop-based Shadow Loss, to make the shadow elements more orderly incorporating into portrait sketches.

The resulting sketches illustrate the effectiveness of VectorPD. We compare the results of our approach with existing vector sketching methods. We additionally evaluate our results quantitatively and demonstrate that our sketch results can be successfully preserve the semantics and structure of the input human face at different levels of abstraction.

To summarize, the main contributions of our paper include:

- We are the first to combine vector graphics and portrait sketches. sketching, converting a human face image into a sketch. VectorPD allows to obtain vector-level portrait

sketching in a variety of styles, without requiring any datasets or training.

- We introduced a novel stroke optimization mechanism that can control the style of sketches through the number of strokes in different rounds and ensure the expressiveness and completeness of portrait sketches at different levels of abstraction.

- To improve the fidelity and visibility of portrait sketches, we propose a novel distance loss function (Crop-based Shadow Loss) to guide the orderly and gentle merging of shadow elements in portrait sketches.

Related Work

Face-Sketch Synthesis

Early methods learned explicit models to synthesize facial sketches (Portenier et al. 2018; Li et al. 2019; DeCarlo et al. 2003). After (Gatys, Ecker, and Bethge 2016) opened up the field of neural style transfer, it is intuitive to regard face sketch synthesis as a style transfer task (Johnson, Alahi, and Fei-Fei 2016; Ulyanov et al. 2016).

(Li and Wand 2016) proposed replacing the Gram matrix with a Markov Random Field (MRF) regularizer to model face style. (Wang et al. 2013) use semantic segmentation of input human face images to learn style models (stroke length, position, color, etc.). (Huang et al. 2022) use a similar portrait rendering model to generate portrait images and use them as local guidance. (Meng, Zhao, and Zhu 2010) considered the more unusual stylization of paper-cut. They matched facial feature to a dictionary of facial templates, achieving global consistency by a hierarchical graph model. These methods are usually effective, but there are constraints on the sketching effects they can produce. In addition, they yield results at the pixel level, restricting their applicability in certain situations.

Vector-based Generation

Traditional vector-based generation methods that typically require vector-based datasets. However, based on the work of (Li et al. 2020), an increasing number of works (Reddy et al. 2021; Shen and Chen 2021) are starting to use differentiable renderers to bypass this limitation and optimize to match images at evaluation time. Rather than directly learning image generation networks, various works manipulate or synthesize vector content by using raster-based loss functions (Mo et al. 2021; Bessmeltsev and Solomon 2019; Frans and Cheng 2018; Huang, Heng, and Zhou 2019; Li et al. 2020; Simo-Serra, Iizuka, and Ishikawa 2018; Wang, Ren, and Zemel 2021; Xu et al. 2019). We use the method of (Li et al. 2020) to synthesis portrait sketches, as it can handle a wide range of curves and strokes, including Bézier curves.

CLIPasso

The most similar to our work is CLIPasso (Vinker et al. 2022), which is designed for object sketching at multiple abstraction levels. They define a sketch as a set of Bézier curves and optimize the stroke parameters and generate sketches through a CLIP-based (Radford et al. 2021) similarity loss between input images. And multi-level abstract

sketches can be achieved by reducing the number of strokes used to compose the image.

Compared with CLIPasso, VectorPD focuses on faces, which requires to adopt a unique processing approach to better meet the needs of generating portrait sketches. In this aspect, we focus on the facial details and the structures of the human face, making our method more capable of capturing facial expressions, proportions and subtleties. Additionally, while (Vinker et al. 2022) examined only a single form of abstraction, we take a more nuanced approach by dividing the abstraction into two key aspects: faces and contours. This mechanism not only gives us precise control over the accuracy of the face but also lets us adjust the completeness of the sketch, offering more flexibility and personalized choices.
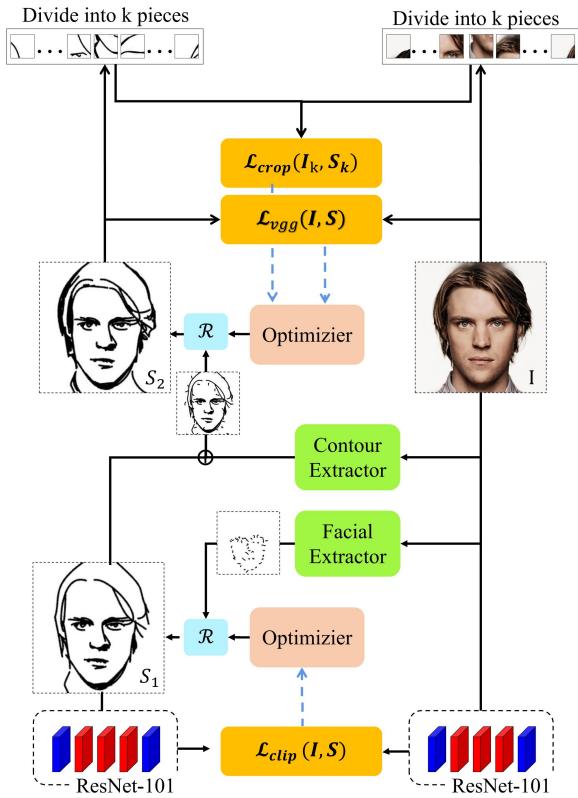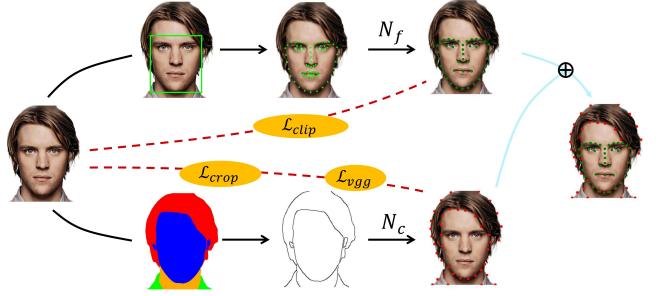

Figure 4: More details on VectorPD. The top path represents the processing of the face extractor based on the user-defined number of facial strokes N_f. The path below shows the processing process of the contour extractor based on the user-defined number of contour strokes N_c. We employ different loss functions to guide the iterative optimization of strokes located at these key points.

Method

Overview

Given a target human face image $\mathcal{I}$, our goal is to generate the corresponding vector portrait sketch $\mathcal{S}$. We define the stroke s of the portrait sketch as a two-dimensional Bézier curve with four control points $s_i = \{p_i^j\}_{j=1}^4$. For simplicity, we only optimize the position of control points and choose to keep the number of control points, color, degree, width and opacity of each stroke fixed. However, it is possible to use these parameters to achieve variations in style (see Figure 2).

An overview of our method can be seen in Figure 3. In order to preserve facial feature details, we first obtain user-defined facial key points through the face extractor as the initial stroke positions for the first round of optimization. After obtaining the basic portrait sketch, we add the contour key points extracted from the contour extractor on the sketch from the previous round and perform the optimization iteration again. In each round of optimization, the position information of these key points is converted into vector parameters and provided to the differentiable rasterizer $\mathcal{R}$. For the sake of face fidelity and sketch integrity, we define different loss functions in each round of optimization to guide the optimization of strokes. Note that we are not optimizing the position of the initial stroke, but rather the position of the control points on the stroke to change the shape.

Figure 3: An illustration of our framework. Given a target image $\mathcal{I}$, our method initiates by employing a facial extractor to derive essential facial keypoints within the facial region. These keypoints serve as the initial positions for strokes in the initial optimization round. The iterative optimization process is then guided by a CLIP-based Semantic Loss. Following the generation of the initial sketch results $\mathcal{S}_1$, we overlay the keypoints onto the facial contour obtained from contour extractor, initiating a second optimization round. The optimization continues iteratively, incorporating VGG-based Structure Loss and Crop-based Shadow Loss guidance until convergence is achieved, resulting in the final portrait sketch $\mathcal{S}_2$.

Strokes Initialization

For the fidelity of the portrait and the completeness of the sketch, VectorPD uses a face extractor and a contour extractor to obtain the initial position of the strokes.

Face Extraction. The main challenge in portrait sketching is preserving the details of facial features. To address this, we use a facial extractor to obtain the accurate locations of facial keypoints. These key points can identify and locate different parts of the human face, such as eyes, nose, mouth, eyebrows, etc., and can be used to describe the shape, structure, and expression of the face.

Specifically, we use a convolutional neural network (CNN) to detect the facial region of the human face in the
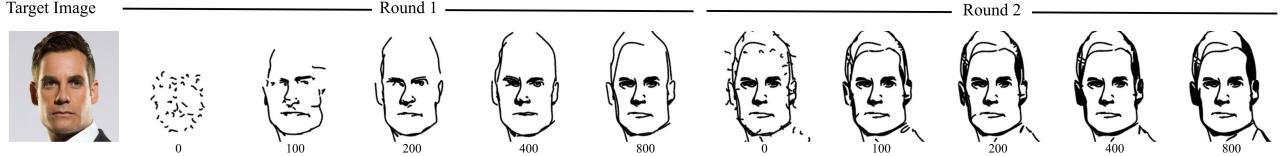
Figure 5: The optimization iteration process reflects two-round optimization mechanism of VectorPD.

Figure 6: Stroke positions in face extractor extracted by FPS algorithm as the number of strokes increases.

Figure 7: Stroke positions in contour extractor extracted by FPS algorithm as the number of strokes increases.

Figure 8: The shadow loss relies on a masked image of the input image. It can be seen that compared to L2, the shadow elements based on Crop-based Shadow Loss will be more orderly. k here takes 4 and 16. I can be found that when k takes 16, the shadow effect is the most effective and orderly.

target image and extract facial landmarks from it (Kazemi and Sullivan 2014). For a specific number of strokes, denoted as N_f, we specify that the placement of strokes must align with the positions of facial landmarks (depicted as green dots in Figure 4). This initialization method not only allows us to preserve facial features as much as possible at different levels of abstraction, but also ensures that the overall sketch is sufficiently expressive and artistic.

Contour Extraction. In a portrait sketch, the contour of the human provides visual completeness and symmetry. In the absence of contours, portrait sketches tend to lose their fidelity to the input face image. To enhance the integrity of the face sketch, we choose to mask each input human face image using the instance segmentation model MaskGAN (Lee et al. 2020). This instance segmentation model can effectively produce segmentation masks corresponding to facial attributes, including all facial components and accessories, such as skin, nose, eyes, eyebrows, ears, lips, hair, hat, glasses, neck etc.

After obtaining the segmentation mask, we use the Canny edge detector (Ding and Goshtasby 2001) on the masked image to extract the mask edges and extract edge key points from them. Likewise, for a specific number of strokes N_c, the initial position of the strokes in contour extraction also should be aligned with the positions of these edge key points set (depicted as red dots in Figure 4).

Stroke Selection. Selecting strokes involves choosing key points from facial landmarks and contour key points. In enhancing semantic description, CLIPasso (Vinker et al. 2022) positioned initial strokes based on salient areas. Nevertheless, this approach doesn't work well for portrait sketching. Because integrity (symmetry and proportion) are crucial in portrait sketches. Placing strokes solely based on salient areas might compromise the accurate capture of facial sym-
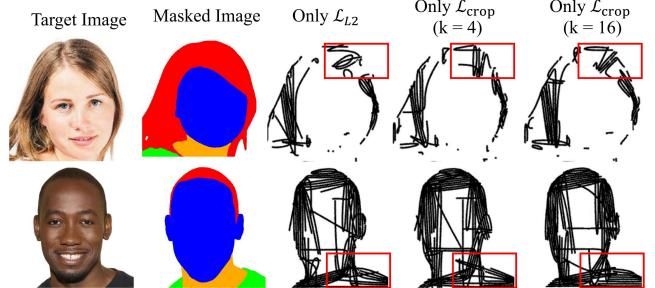
metry and proportions, neglecting the equal significance of diverse facial features. Therefore, VectorPD uses the Farthest Point Sampling algorithm (FPS) to choose a suitable set of points. This ensures that the selected points are evenly spaced and represent the entire set well (see Figure 6 and Figure 7).

Stroke Optimization

VectorPD has a two-round optimization mechanism. The first round is used to optimize facial key points, and the second round is used to optimize contour key points. All parameters in the second round will be superimposed with the basic vector result of the previous round (see Figure 5).

After obtaining the initial stroke position of each round in extractors, VectorPD will convert the position information into a stroke parameter set $\{s_i\}_{i=1}^n$ in vector form. In each round of optimization, we use off-the-shelf gradient-based solving techniques to compute loss gradients with respect to the stroke parameters. We repeat optimization process until convergence, when the optimization error does not change significantly and the portrait sketch results stabilize (taking typically 800 iterations).

Loss Function

As each optimization round serves a distinct purpose, we employ different loss functions for gradient calculation in each round. The first round is steered by CLIP-based Semantic Loss, focusing on generating foundational sketches that retain essential facial information. The second round is directed by VGG-based Structure Loss and Crop-based Shadow Loss, with the goal of enhancing the structure of sketch, adding visual effects and enriching artistic expression.

CLIP-based Semantic Loss. As our portrait sketch consists solely of vector strokes, pixel-level metrics are insufficient for gauging the distance between the target face image and the sketch result. To address this, we refer to the CLIP loss in CLIPasso (Vinker et al. 2022) and use layers 2, 3, and 4 of the CLIP model (ResNet-101) to guide the result generation. Consequently, the embedding distance between the image $\mathcal{I}$ and the sketch $\mathcal{S}$ is determined by:

$$\begin{aligned}\mathcal{L}_{\text{clip}} = dist(CLIP(\mathcal{I}), CLIP(\mathcal{S})) \\ + L_2(CLIP_{l_{2,3,4}}(\mathcal{I}), CLIP_{l_{2,3,4}}(\mathcal{S}))\end{aligned} \quad (1)$$

VGG-based Structure Loss. To further complete the portrait sketch, we compute the Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018) between the sketch features and the mask image obtained from the instance segmentation model based on VGG16 (Simonyan and Zisserman 2014). This helps preserve the structure of the sketch and makes the resulting sketch more reflective of the shape and contours of the original image. I_{mask} is the masked image obtained by processing input human face image I.

$$\mathcal{L}_{vgg} = LPIPS(VGG(I_{mask}), VGG(S)) \quad (2)$$

Crop-based Shadow Loss. Referring to Clipasso (Vinker et al. 2022), we observed that the L2 loss function primarily promotes pixel filling. To enhance the clarity of portrait sketches while retaining image shadows, we experimented with using L2 loss to fill the sketch shadows. As illustrated in Figure 8, using only L2 loss results in distorted and disorderly filling, impacting the expressiveness of portrait sketches. To achieve a softer shadow fill, we introduced cropped blocks for calculating distance loss:

$$\mathcal{L}_{crop} = \|\mathcal{I}_i - \mathcal{S}_i\|_2^2 + \frac{\sum_i^k \|\mathcal{I}_i - \mathcal{S}_i\|_2^2}{k} \quad (3)$$

where k represents the number of cropped blocks. The impact of the number of blocks is shown in Figure 8. Since our images are all 224x224, in order to ensure that the side length of the cropped block is an integer, the value of k is constrained. See the supplementary material for more details.

The ultimate goal of optimization is defined as:

$$\mathcal{L}_{sum} = \mathcal{L}_{clip} + \mathcal{L}_{vgg} + \mathcal{L}_{crop} \quad (4)$$

Experiments

Qualitative Evaluation

VectorPD is different from the usual sketch generation techniques. We create vector sketches and specifically emphasize generating portrait sketches. The complexity of human face demands a distinctive method in the sketch generation process. In Figure 1, we demonstrate our ability to generate portrait sketches at different levels of abstraction. The fewer strokes, the greater the need to focus on facial features and structure, capturing the abstract essence of the human face. In addition to stroke adjustment, changing the stroke style (see Figure 2) can also achieve different sketching results.
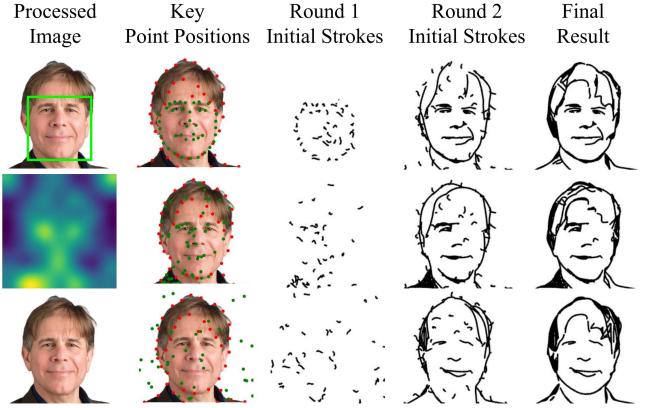

Figure 9: Different stroke selection methods in the first round of optimization. Top to bottom: FPS, salient area and random.

Figure 10: Trade-off between facial strokes N_f and contour strokes N_c. As shown in the sketch, N_f increases and the face of the portrait sketch becomes more detailed. As N_c increases, the contour of the portrait sketch becomes more complete.

In Figure 9 we show the results using different stroke initialization methods in the first stroke initialization. It is easy to recognize that the initial position based on facial key points can better preserve the facial features after stroke optimization than salient areas and random selection. And the second optimization round performs well in preserving the silhouette of the character. Note that we did not change the contour keypoint positions in Figure 9. More results will be presented in the supplementary material.

In Figure 10, we illustrate how the portrait sketch transforms when adjusting the number of facial and contour strokes while keeping the total strokes constant. Shifting to the right enhances the portrayal of contours, resembling a masked image. Shifting to the left emphasizes conveying the facial semantic information of the input human face image.

Comparison with Existing Methods

In Figure 11, we compare two state-of-the-art vector-level sketch generation methods. Virtual Sketching (Mo et al. 2021) introduces a general framework for generating line drawings from a variety of images, but its limitations is that it is difficult to generalize well to complex photos, such as faces, where its results can be seen that a nose is connected to the eyebrows, with a noticeable lack of detail. Clipasso (Vinker et al. 2022), is a photo sketch synthesis method. We present three sketches generated by Clipasso and our approach, depicting three representative levels of
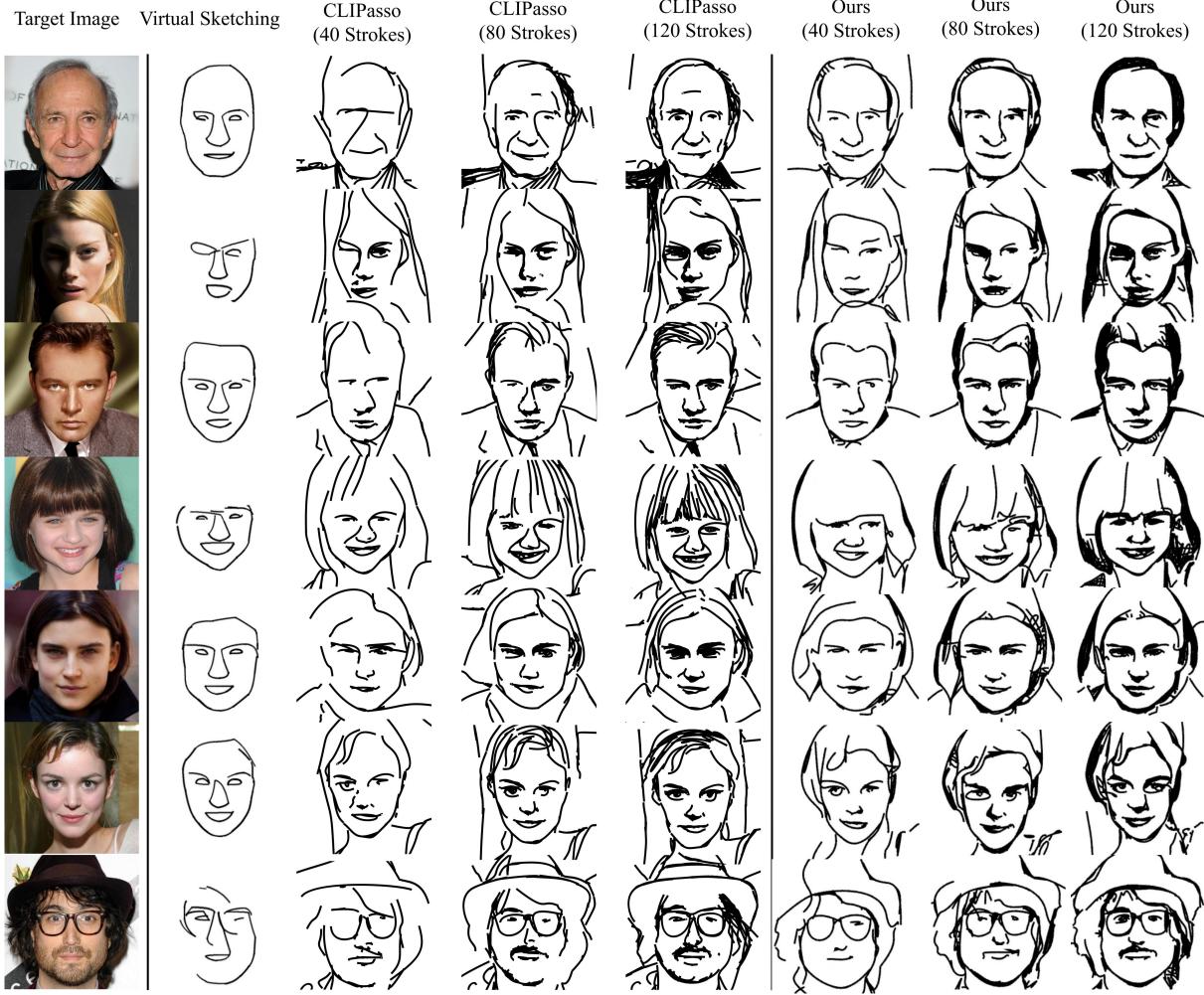
Figure 11: Portrait sketching results and comparisons. From left to right, the images showcase the results from Virtual Sketching, CLIPasso and VectorPD. CLIPasso and our method generate three distinct sketches, depicting three levels of abstraction.

	Virtual Sketching	CLIPasso			Ours		
		40	80	120	40	80	120
Recognition rate	96.6%	97.9%	99.3%	99.7%	99.7%	99.9%	99.9%

Table 1: User study results – recognition rates. The higher the recognition rate, the sketch generated by this work can better retain the features of the input human face image.

abstraction for more comparison. When Clipasso uses only a few strokes, the details of human face might not show up. But with more strokes, a more complete portrait sketch will be generated. In comparison, our sketch results seems more visually cleaner and tidier. In addition, we note that our method is also better able to ignore the interfering background of faces than Clipasso. However, our disadvantage is the long running time, which takes 6 minutes to generate a clear portrait sketch on a commodity GPU.

In Figure 12, we compare our method with (Berger et al. 2013). Their code is not open source, and the paper only displays examples at a certain level of abstraction. For a fair comparison, we use these examples as a reference and select an appropriate number of strokes with an even distribution to generate a portrait sketch. We also apply similar pencil styles to match the artist's touch. The comparison highlights that, at various abstraction levels, our approach consistently maintains a closer resemblance to the input human face image compared to (Berger et al. 2013) and Clipasso.

Quantitative Evaluation

In this section, we provide a quantitative evaluation of the ability of our method to generate portrait sketches that preserve facial features and structure.
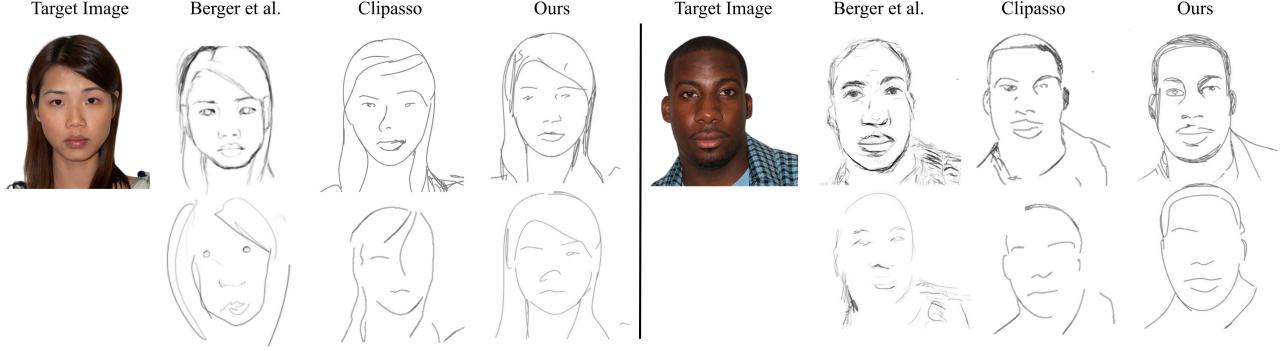
Figure 12: Comparison of portrait sketches at different levels of abstraction. For the woman on the left, the top result is generated with 40 strokes and the bottom result is generated with 20 strokes. For the man on the right, the top result is generated with 60 strokes and the bottom result is generated with 20 strokes.

| | 40 Strokes | | | 80 Strokes | | 120 Strokes | |
	Virtual Sketching	CLIPasso	Ours	CLIPasso	Ours	CLIPasso	Ours
Reference rate	2.1%	26.7%	71.2%	35.6%	64.4%	40.3%	59.7%

Table 2: User study results – preference rates. Higher preference rate means users more appreciate the portrait sketch results from the work. The results show that most people prefer the portrait sketch effect generated by our method compared to other methods.

| | Virtual Sketching | CLIPasso | | | Ours | | |
		40	80	120	40	80	120
SSIM ↑	0.309	0.292	0.290	0.290	0.305	0.306	0.307

Table 3: The average scores of SSIM. The higher the score, the higher the structure similarity between the portrait sketch generated by the method and the input human face image.

User Study. We conducted a user study with 50 participants and set up different research experiments to compare the recognizability and visual sense of Virtual Sketching (Mo et al. 2021), CLIPasso (Vinker et al. 2022) and our VectorPD in portrait sketching quantitative results on.

Table 1 gives the face sketch recognition rate results under different methods and different number of strokes. Specifically, we randomly selected 30 human face images in the CelebA-HQ dataset (Karras et al. 2017). We used these human face images as input of these works to generate a total of 120 portrait sketches. Each portrait will be provided to the user for recognition together with the portrait sketches generated from 5 similar face images, such as age, gender and other basic attributes are the same. Users were required to choose the one they think best matches the sketch.

Table 2 shows the preference results of users. We randomly choose another 30 human face images in the CelebA-HQ dataset. We used Clipasso and our method to generate portrait sketches at three different abstraction levels of 40, 80 and 120 strokes, and divided the comparison group into three groups according to the number of strokes. For the sake of fairness, we put Virtual Sketching into the sketch comparison group with 20 strokes. Users need to choose the sketch result they like most among the sketches in each group.

Structural Similarity Index. To measure the structural integrity of the portrait sketch, we calculate the Structural Similarity Index (SSIM) (Wang et al. 2004) score between each input human face image and the corresponding sketch result. Table 3 shows the average result scores. Virtual Sketching (Mo et al. 2021) exhibits superior structure as it follows a linear drawing approach with stringent structural demands, making it an ideal baseline. In comparison, our method surpasses CLIPasso (Vinker et al. 2022) in retaining the structure of human face images. At various abstraction levels, our method is proved that our sketch results could maintain a stable structure, evident in minimal SSIM score fluctuations closely aligned with the baseline.

Discussions

Limitations and Future Works. VectorPD presents vectorized results from huamn face images, which can be used for further style editing or other applications. However, there are still some issues we can discuss. First, multi-round optimization is not as efficient as single-round optimization. Exploring the combination of optimized generalization capabilities with efficient inference in deep models will be a future focus. Second, optimization is a black box. Exploring the principles of the optimization process will be of great

help in the work of vector graphics generation. Third, using other different shape primitives to generate portrait sketches will also be an interesting topic. We leave these to future work.

Potential Negative Impact. Converting human face images into vector sketches may pose potentially negative consequences. Unauthorized conversion and copying of vectorized portrait sketches may result in privacy breach. Secondary creation based on portrait sketches may also lead to unethical use. To mitigate these issues, achieving transparency in the use of portrait sketching and adhering to industry guidelines and regulations will help to aid ethical practices. Overall, maintaining a balance between technological advancement and ethical considerations is critical to preventing misuse and ensuring responsible use of face image conversion technology.

Conclusions

VectorPD is the first work to combine vector graphics and portrait sketch generation. We propose a two-round optimization mechanism to optimize the shape of vector strokes, first for obtaining the basic semantic information of the face, and second for superimposing the contour of the face. We employ CLIP-based Semantic Loss, VGG-based Structure Loss, and Crop-based Shadow Loss across different optimization round, which ensure that our portrait sketches maintain the basic visual features and structures of a human face at challenging abstraction levels as much as possible.

References

Berger, I.; Shamir, A.; Mahler, M.; Carter, E.; and Hodgins, J. 2013. Style and abstraction in portrait sketching. *ACM Transactions on Graphics (TOG)*, 32(4): 1–12.

Bessmeltsev, M.; and Solomon, J. 2019. Vectorization of line drawings via polyvector fields. *ACM Transactions on Graphics (TOG)*, 38(1): 1–12.

Biederman, I.; and Ju, G. 1988. Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, 20(1): 38–64.

DeCarlo, D.; Finkelstein, A.; Rusinkiewicz, S.; and Santella, A. 2003. Suggestive contours for conveying shape. In *ACM SIGGRAPH 2003 Papers*, 848–855.

Ding, L.; and Goshtasby, A. 2001. On the Canny edge detector. *Pattern recognition*, 34(3): 721–725.

Fan, J. E.; Yamins, D. L.; and Turk-Browne, N. B. 2018. Common object representations for visual production and recognition. *Cognitive science*, 42(8): 2670–2698.

Frans, K.; and Cheng, C.-Y. 2018. Unsupervised image to sequence translation with canvas-drawer networks. *arXiv preprint arXiv:1809.08340*.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2414–2423.

Gryaditskaya, Y.; Sypesteyn, M.; Hoftijzer, J. W.; Pont, S. C.; Durand, F.; and Bousseau, A. 2019. OpenSketch: a richly-annotated dataset of product design sketches. *ACM Trans. Graph.*, 38(6): 232–1.

Hertzmann, A. 2020. Why do line drawings work? a realism hypothesis. *Perception*, 49(4): 439–451.

Huang, Z.; Heng, W.; and Zhou, S. 2019. Learning to paint with model-based deep reinforcement learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8709–8718.

Huang, Z.; Peng, Y.; Hibino, T.; Zhao, C.; Xie, H.; Fukusato, T.; and Miyata, K. 2022. dualface: Two-stage drawing guidance for freehand portrait sketching. *Computational Visual Media*, 8: 63–77.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, 694–711. Springer.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

Kazemi, V.; and Sullivan, J. 2014. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1867–1874.

Lee, C.-H.; Liu, Z.; Wu, L.; and Luo, P. 2020. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5549–5558.

Li, C.; and Wand, M. 2016. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2479–2486.

Li, H.; and Mould, D. 2011. Structure-preserving stippling by priority-based error diffusion. In *Proceedings of Graphics Interface 2011*, 127–134.

Li, M.; Lin, Z.; Mech, R.; Yumer, E.; and Ramanan, D. 2019. Photo-sketching: Inferring contour drawings from images. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1403–1412. IEEE.

Li, T.-M.; Lukáč, M.; Gharbi, M.; and Ragan-Kelley, J. 2020. Differentiable vector graphics rasterization for editing and learning. *ACM Transactions on Graphics (TOG)*, 39(6): 1–15.

Ma, X.; Zhou, Y.; Xu, X.; Sun, B.; Filev, V.; Orlov, N.; Fu, Y.; and Shi, H. 2022. Towards layer-wise image vectorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16314–16323.

Meng, M.; Zhao, M.; and Zhu, S.-C. 2010. Artistic papercut of human portraits. In *Proceedings of the 18th ACM international conference on Multimedia*, 931–934.

Mo, H.; Simo-Serra, E.; Gao, C.; Zou, C.; and Wang, R. 2021. General virtual sketching framework for vector line art. *ACM Transactions on Graphics (TOG)*, 40(4): 1–14.

Portenier, T.; Hu, Q.; Szabo, A.; Bigdeli, S. A.; Favaro, P.; and Zwicker, M. 2018. Faceshop: Deep sketch-based face image editing. *arXiv preprint arXiv:1804.08972*.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Reddy, P.; Gharbi, M.; Lukac, M.; and Mitra, N. J. 2021. Im2vec: Synthesizing vector graphics without vector supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7342–7351.

Rosin, P. L.; and Lai, Y.-K. 2018. Watercolour rendering of portraits. In *Image and Video Technology: PSIVT 2017 International Workshops, Wuhan, China, November 20-24, 2017, Revised Selected Papers 8*, 268–282. Springer.

Shao, H.; Weng, X.; and He, S. 2017. Functional organization of the face-sensitive areas in human occipital-temporal cortex. *Neuroimage*, 157: 129–143.

Shen, I.-C.; and Chen, B.-Y. 2021. Clipgen: A deep generative model for clipart vectorization and synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 28(12): 4211–4224.

Simo-Serra, E.; Iizuka, S.; and Ishikawa, H. 2018. Mastering sketching: adversarial augmentation for structured prediction. *ACM Transactions on Graphics (TOG)*, 37(1): 1–13.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tian, Y.; and Ha, D. 2022. Modern evolution strategies for creativity: Fitting concrete images and abstract concepts. In *Artificial Intelligence in Music, Sound, Art and Design: 11th International Conference, EvoMUSART 2022, Held as Part of EvoStar 2022, Madrid, Spain, April 20–22, 2022, Proceedings*, 275–291. Springer.

Ulyanov, D.; Lebedev, V.; Vedaldi, A.; and Lempitsky, V. 2016. Texture networks: Feed-forward synthesis of textures and stylized images. *arXiv preprint arXiv:1603.03417*.

Vinker, Y.; Pajouheshgar, E.; Bo, J. Y.; Bachmann, R. C.; Bermano, A. H.; Cohen-Or, D.; Zamir, A.; and Shamir, A. 2022. Clipasso: Semantically-aware object sketching. *ACM Transactions on Graphics (TOG)*, 41(4): 1–11.

Wang, A.; Ren, M.; and Zemel, R. 2021. Sketchembednet: Learning novel concepts by imitating drawings. In *International Conference on Machine Learning*, 10870–10881. PMLR.

Wang, T.; Collomosse, J. P.; Hunter, A.; and Greig, D. 2013. Learnable stroke models for example-based portrait painting. In *British Machine Vision Conference (BMVC)*.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Winnemöller, H.; Kyprianidis, J. E.; and Olsen, S. C. 2012. XDoG: An eXtended difference-of-Gaussians compendium including advanced image stylization. *Computers & Graphics*, 36(6): 740–753.

Xu, P.; Hospedales, T. M.; Yin, Q.; Song, Y.-Z.; Xiang, T.; and Wang, L. 2022. Deep learning for free-hand sketch: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(1): 285–312.

Xu, X.; Xie, M.; Miao, P.; Qu, W.; Xiao, W.; Zhang, H.; Liu, X.; and Wong, T.-T. 2019. Perceptual-aware sketch simplification based on integrated VGG layers. *IEEE transactions on visualization and computer graphics*, 27(1): 178–189.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.