Brushes in Motion: Vector Guided Strokes for Computational Painting

Jeripothula Prudviraj **Vikram Jamwal**
Computational Creativity Group
TCS Research
INDIA
{prudviraj.jeripothula, vikram.jamwal}@tcs.com

Abstract

Understanding and computationally modeling the creative process behind artistic creation remains a fundamental challenge in Creative AI. While existing neural painting methods focus on replicating final visual outcomes, they largely ignore the sequential, hierarchical decision-making that characterizes human artistic workflow - the dynamic motion of brushes across canvas that brings art to life. We introduce a novel approach to computationally decompose and reconstruct the painting process itself, revealing how artworks emerge through systematic region-aware brushstroke sequences that mirror a more natural artistic practice. Our method leverages image vectorization to extract semantic painting regions and develops algorithms to estimate brushstroke parameters and sequencing strategies that progress from global compositional structure to localized detail refinement. This enables generation of stroke-by-stroke painting animations that expose the underlying creative process in motion, with applications ranging from immersive museum experiences to collaborative AI-assisted art creation platforms.

1 Introduction

The creative process underlying artistic expression has long captivated researchers in computational creativity. Unlike static museum displays, the true essence of artistic creation lies in the dynamic, evolving process through which ideas transform into visual form - brushes in motion, each stroke building upon the last. Recent advances in AI-assisted artistic tools are revolutionizing how art is created and experienced, yet current approaches primarily focus on visual transformation rather than revealing the underlying creative process and the temporal dance of brush movements that generate artistic expression.

Human artists approach creation through systematic, hierarchical processes that reflect deep understanding of visual composition and aesthetic principles. They set their brushes in motion strategically, progressing from establishing global compositional elements before refining localized regions with increasingly detailed brushwork. This semantic-region-based approach reflects strategic creative decision-making that organizes complex visual information into coherent, expressive form through the coordinated movement of brush across canvas.

Existing stroke-based rendering (SBR) approaches treat painting primarily as optimization problems, learning stroke parameters to minimize pixel-wise reconstruction loss Hu et al. [2024], de Guevara et al. [2024], Singh et al. [2022], Hu et al. [2023]. While visually compelling, they fundamentally misunderstand artistic creation by focusing on final visual fidelity rather than the creative process and the natural motion patterns that guide artistic expression. The strokes produced often appear spatially incoherent, jumping between regions without the systematic progression that characterizes human artistic practice - brushes moving without purpose or natural flow.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

We propose a fundamentally different approach that prioritizes understanding and modeling artistic *process* over optimizing visual output, capturing the essence of brushes in motion during artistic creation. Rather than learning stroke parameters through pixel-wise loss minimization, we investigate whether these parameters can be derived by analyzing the underlying semantic and geometric structure of painting regions themselves. Our key insight is that meaningful artistic stroke sequences - the natural motion of brushes during creation - emerge from the hierarchical organization of visual content.

Our approach leverages Scalable Vector Graphics (SVG) representation as a bridge between structured geometric understanding and expressive artistic creation. By decomposing input artworks into vectorized regions through a two-stage process - global layer-by-layer vectorization for compositional structure and segment-based patch vectorization for detailed refinement - we extract stroke parameters that reflect natural artistic workflow and the organic motion of brushes during painting.

The contributions of our work are: (1) We investigate a novel two-stage image vectorization framework that captures both coarse compositional structure and fine-grained artistic detail, enabling hierarchical stroke sequence generation that better mirrors the natural motion of brushes during human artistic practice. (2) We propose algorithms for estimating stroke parameters directly from SVG-vectorized regions, emphasizing semantic-region-based painting through region-aware stroke estimation and sequencing that captures the flow and movement of artistic creation. (3) We point to potential applications across diverse creative domains, establishing foundations for AI systems that engage meaningfully with artistic process and the dynamic nature of brushes in motion during creative expression.

2 Computational Modeling of Artistic Process

Understanding how human artists construct paintings provides the foundation for our computational approach. Artists typically work in a systematic, hierarchical manner: first establishing the overall composition and major structural elements, then progressively refining specific regions with increasingly detailed brushwork. This creative workflow reflects not random mark-making, but strategic decision-making that organizes visual information into coherent artistic expression.

Our method computationally models this artistic process through a two-stage framework that mirrors human painting workflow. Given an input artwork image X, we compute a sequence of brushstroke parameters $\{B_p^t\}_{t=1}^T$ that enables progressive reconstruction of the painting process. The **Global stage** analyzes the entire composition to establish broad structural elements, while the **Segment stage** focuses on individual semantic regions for detailed refinement - directly paralleling how artists move from compositional planning to localized detail work.

2.1 Region-Aware Image Decomposition

Human artists inherently organize their work around meaningful visual regions -backgrounds, objects, facial features - rather than arbitrary pixel patches. To capture this semantic organization computationally, we employ image vectorization that decomposes artworks into structured geometric regions that correspond to natural painting units.

Global Compositional Analysis: At the global stage, we apply layer-by-layer vectorization to capture the overall compositional hierarchy. This process decomposes the image X into a stack of visual layers $\{H_0, H_1, ..., H_k\}$ where top layers represent the broadest compositional elements and lower layers capture progressively finer structural details:

$$V_G^{(0)}{}_j = \text{Vect}_{\text{Layer}}(X), \quad j = 1, 2, ..., J \tag{1}$$

This hierarchical organization mirrors how artists establish overall composition before adding detail, enabling our system to generate stroke sequences that build paintings systematically from abstract structure to concrete form.

Semantic Region Refinement: The segment stage focuses on individual semantic regions, reflecting how artists work region-by-region to develop local details. We first segment the input into meaningful

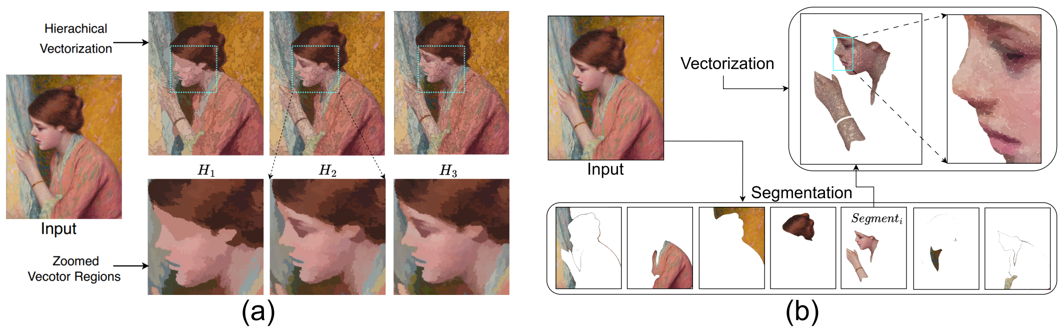


Figure 1: **Hierarchical artistic decomposition through vectorization.** Our method strives to mirror human artistic workflow better by analyzing images at multiple levels of detail. **(a) Global stage:** Layer-by-layer vectorization captures the compositional hierarchy (H_0 to H_3) that artists use to establish overall structure, progressing from broadest regions to medium-scale elements. **(b) Segment stage:** Patch-based vectorization within semantic segments enables detailed region-specific analysis, respecting the natural boundaries that guide artistic decision-making. This two-stage approach computationally models how artists systematically organize visual information from abstract composition to concrete detail.

semantic units, then apply patch-based vectorization within each segment:

$$S_i = \text{Segmentation}(X), \quad i = 1, 2, ..., I \tag{2}$$

$$V_S^{(i)}{}_j = \text{Vect}_{\text{Patch}}(S_i), \quad j = 1, 2, ..., J \tag{3}$$

This approach ensures that our stroke generation respects the natural semantic boundaries that guide human artistic decision-making, rather than imposing arbitrary geometric divisions.

2.2 Artistic Stroke Parameter Estimation

Traditional stroke-based methods optimize parameters to minimize pixel-wise reconstruction error, often producing spatially incoherent results that bear little resemblance to natural painting processes. Instead, we derive stroke parameters directly from the geometric and semantic structure of vectorized regions, capturing the systematic approach characteristic of human artistic practice.

Stroke Construction: Following established parametric representations Liu et al. [2021], Zou et al. [2021], Hu et al. [2024], we model each brushstroke as $\{x, y, w, h, \theta, r, g, b\}$ where (x, y) defines position, (w, h) specify dimensions, θ determines orientation, and (r, g, b) encode color. However, rather than learning these through optimization, we extract them through geometric analysis of vectorized regions.

For each SVG region, we compute its polygon representation and derive stroke parameters through minimum-area rotated rectangle analysis. This geometric approach naturally captures the orientation and scale relationships that characterize coherent brushwork:

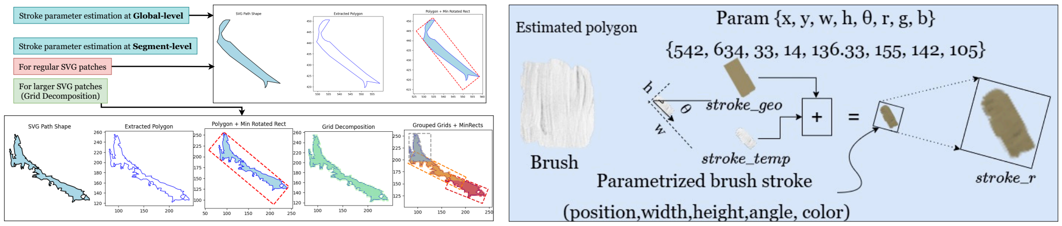
$$G^0_{P_n} = \text{Polygon}(V_G^{(0)}{}_j), \quad S^i_{P_n} = \text{Polygon}(V_S^{(i)}{}_j) \tag{4}$$

$$G^0_{B_p} = \text{ExtractParams}(G^0_{P_n}), \quad S^i_{B_p} = \text{ExtractParams}(S^i_{P_n}) \tag{5}$$

Adaptive Region Processing: At the global stage, each polygon corresponds to a single stroke unit, capturing the broad compositional elements. At the segment stage, we adaptively subdivide large polygons using grid decomposition when their area exceeds a threshold δ, enabling generation of fine-grained strokes for detailed regions while maintaining computational efficiency.

2.3 Process-Aware Stroke Sequencing

The temporal ordering of brushstrokes fundamentally distinguishes our approach from existing methods. Rather than random or optimization-driven sequences, we model the systematic progression that characterizes human artistic workflow.



(a) (b)

Figure 2: **From vector regions to artistic strokes: bridging geometric analysis and creative expression. (a) Stroke parameter extraction:** Our approach derives meaningful stroke parameters from geometric analysis rather than pixel-wise optimization. At the global level, each SVG region yields a single stroke capturing broad compositional elements. At the segment level, large regions undergo adaptive grid decomposition to generate fine-grained strokes for detailed refinement, mirroring how artists progress from coarse to fine detail. **(b) Process-aware rendering:** Template brush textures combine with geometric stroke properties through alpha blending, preserving the organic quality of painted surfaces while enabling systematic artistic reconstruction that reveals the creative process.

Global Sequencing: The hierarchical vectorization naturally provides stroke ordering for compositional elements, progressing from largest structural components to medium-scale details. This reflects how artists establish overall composition before refining specific areas.

Local Sequencing: Within semantic segments, we employ perceptual grouping principles to determine natural stroke progressions. Using proximity-based clustering combined with traveling salesman optimization, we generate sequences that maintain spatial coherence within regions:

$$B_p^t = \text{Seq}_g(G^0_{B_p}) \parallel \text{Seq}_s(S^i_{B_p}) \tag{6}$$

This dual-level sequencing ensures that the reconstructed painting process exhibits the systematic, region-aware progression characteristic of human artistic practice.

2.4 Progressive Artistic Reconstruction

The final painting reconstruction simulates the accumulative nature of artistic creation, where each stroke builds upon previous work to gradually reveal the complete artwork. Beginning with an empty canvas C_0, we render strokes sequentially, blending geometric stroke properties with brush texture through alpha compositing operations that preserve the organic quality of painted surfaces.

This process-aware reconstruction enables the creative applications that motivate our work: immersive experiences that reveal artistic creation in real-time, educational tools that demonstrate painting techniques, and collaborative platforms where AI can provide process-aware assistance to human artists.

3 Evaluation

We evaluate our approach across technical effectiveness and creative applicability. Our evaluation demonstrates that computationally modeling artistic workflow produces more coherent, semantically meaningful painting reconstructions than existing optimization-based approaches.

3.1 Experimental Setup

Datasets: We evaluate across WikiArt Saleh and Elgammal [2015] (portraits, landscapes, abstracts), METFace Karras et al. [2020] (facial artwork), DELAUNAY Gontier et al. [2022] (abstract compositions), and ImageNet Deng et al. [2009] (natural imagery).

Implementation: Our pipeline uses VTracer Pun and Tsang [2020] for image vectorization with two modes: *stacked* mode for layer-by-layer global analysis and *cutout* mode for patch-based segment processing. We employ SAM's Automatic Mask Generator (AMG) for segmentation, using

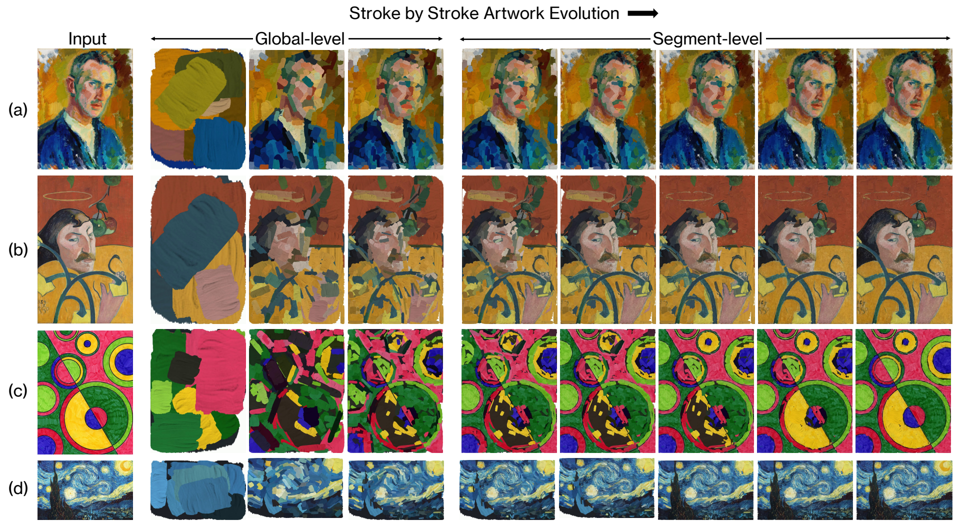


Figure 3: **Process-aware artistic reconstruction on WikiArt masterpieces.** Our method decomposes classic artworks to reveal a possible underlying creation process, demonstrating systematic hierarchical development from broad compositional elements to refined local details. The progression mirrors natural artistic workflow, progressing from abstract structural elements that establish overall composition to increasingly detailed region-specific refinement. This computational understanding of artistic process enables immersive experiences where viewers can witness how masterpieces evolved from initial conception to final form.

ViT-H model with parameters ranging from 2-8 points per side and IoU thresholds of 0.6-0.8. To handle overlapping masks, we sort by area and apply IoU-based filtering. Stroke parameters are extracted using Shapely's minimum-rotated rectangle algorithm Gillies and contributors [2007–present], followed by brush template resizing, rotation, and alpha blending for seamless stroke integration.

3.2 Evaluation on Datasets

Figure 3 demonstrates our method's ability to generate semantically coherent stroke sequences on WikiArt masterpieces. Unlike baseline methods that produce spatially random progressions, our approach exhibits clear hierarchical progressions, from compositional structure to refined details, closely mirroring natural artistic workflow. Figure 4 further demonstrates robust cross-domain performance across diverse visual domains: METFace portraits, abstract DELAUNAY compositions, style-transferred images, and natural ImageNet photography, confirming the generalizability of our process-aware approach. Quantitative evaluations are provided in Appendix A.2.

3.3 Extended Evaluations

Brush style significantly impacts rendered image quality, as illustrated through comparative examples in Appendix A.1. Detailed qualitative and quantitative comparisons against existing stroke-based rendering methods - including Stylized Neural Painting (SNP) Zou et al. [2021], region-based approaches such as Compositional Neural Painter (CNP) Hu et al. [2023], Sketch & Paint Prudviraj and Jamwal [2024], and Segmentation-Based Parametric Painting (SBPP) de Guevara et al. [2024] - are presented in Appendix A.3. Additionally, user survey results confirming human preference for our approach are provided in Appendix A.4.

Limitations: Our method faces several acknowledged limitations. The approach relies heavily on segmentation quality, and current segmentation methods may lack art-specific semantic understanding, potentially affecting region quality for highly stylized content. We model strokes as rectangular primitives, missing complex brush dynamics including pressure variation, color spread effects, and curved stroke trajectories. Additionally, we approximate complex splines through decomposition into straight atomic sub-strokes, which may not capture the fluidity of natural brushwork.

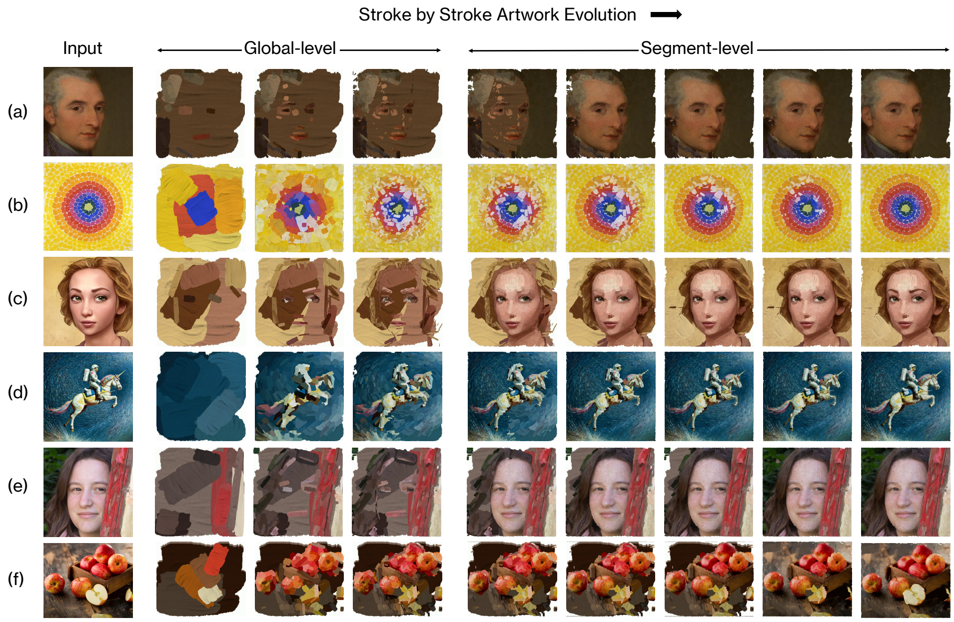


Figure 4: **Cross-domain generalization of computational artistic process modeling.** Our approach demonstrates robust performance across diverse visual domains: **(a)** METFace portraits reveal systematic facial feature development; **(b)** abstract DELAUNAY compositions show coherent geometric progression; **(c,d)** style-transferred images maintain natural artistic workflow; **(e,f)** natural ImageNet photography demonstrates applicability beyond traditional artwork. Across all domains, the method captures hierarchical, region-aware progression characteristic of systematic creative practice, enabling applications from art education to collaborative creation platforms.

4 Conclusion

This work advances computational creativity by demonstrating how AI systems can understand and reconstruct the temporal processes underlying artistic creation - capturing the essence of brushes in motion that brings art to life. While we **do not claim** the discovery of the original artistic process, **our key insight** is that meaningful artistic stroke sequences emerge from analyzing semantic and geometric structure rather than just pixel-wise optimization, capturing the systematic progression and natural flow that characterizes human creative practice.

By introducing process-aware stroke sequence generation that models the dynamic movement of brushes during creation, we enable new Creative AI applications.

Potential Applications: Our stroke sequences capture the motion and flow of artistic creation, potentially enabling diverse creative applications: (1) *Immersive museum experiences* where visitors witness masterpieces unfold stroke-by-stroke, experiencing the brushes in motion that originally created each work; (2) *Educational art tools* that reveal the systematic progression and movement patterns for pedagogical insight, helping students understand how professional brushwork flows and develops; (3) *Collaborative creation platforms* where AI provides process-aware assistance that respects the natural motion and rhythm of human artistic workflow.

This establishes foundations for Creative AI systems that engage with the temporal, dynamic aspects of creative practice - understanding not just what artists create, but how their brushes move to bring ideas into being. The fundamental insight - that creative processes exhibit discoverable motion patterns that can be computationally modeled - extends beyond visual art to other temporal creative domains where movement and flow are essential.

As Creative AI evolves, approaches that capture the motion and dynamics of creative practice will be essential for developing systems that truly complement human creativity by understanding not just creative outputs, but the rich, flowing processes of brushes in motion that generate them.

References

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-toend object detection with transformers. in eccv. *Springer*, 1(2):4, 2020.

Bowei Chen, Yifan Wang, Brian Curless, Ira Kemelmacher-Shlizerman, and Steven M Seitz. Inverse painting: Reconstructing the painting process. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024.

Manuel Ladron de Guevara, Matt Fisher, and Aaron Hertzmann. Segmentation-based parametric painting. In *2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2024.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Fofr. Style transfer - replicate, 2025. URL `https://replicate.com/fofr/style-transfer`.

Yaroslav Ganin, Tejas Kulkarni, Igor Babuschkin, SM Ali Eslami, and Oriol Vinyals. Synthesizing programs for images using reinforced adversarial learning. In *International Conference on Machine Learning*, pages 1666–1675. PMLR, 2018.

Sean Gillies and contributors. Shapely: manipulation and analysis of geometric objects. GitHub repository, 2007–present. URL `https://github.com/shapely/shapely`.

Camille Gontier, Jakob Jordan, and Mihai A Petrovici. Delaunay: a dataset of abstract art for psychophysical and machine learning research. *arXiv preprint arXiv:2201.12123*, 2022.

Bruce Gooch, Greg Coombe, and Peter Shirley. Artistic vision: painterly rendering using computer vision techniques. In *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, pages 83–ff, 2002.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

David Ha and Douglas Eck. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*, 2017.

Paul Haeberli. Paint by numbers: Abstract image representations. In *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 207–214, 1990.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Aaron Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 453–460, 1998.

Aaron Hertzmann. A survey of stroke-based rendering. Institute of Electrical and Electronics Engineers, 2003.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Teng Hu, Ran Yi, Haokun Zhu, Liang Liu, Jinlong Peng, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Stroke-based neural painting and stylization with dynamically predicted painting region. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7470–7480, 2023.

Zhangli Hu, Ye Chen, Zhongyin Zhao, Jinfan Liu, Bilian Ke, and Bingbing Ni. Towards artist-like painting agents with multi-granularity semantic alignment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10191–10199, 2024.

Zhewei Huang, Wen Heng, and Shuchang Zhou. Learning to paint with model-based deep reinforcement learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8709–8718, 2019.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33:12104–12114, 2020.

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

TP Lillicrap. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Peter Litwinowicz. Processing images and video for an impressionist effect. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 407–414, 1997.

Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. Paint transformer: Feed forward neural painting with stroke prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6598–6607, 2021.

Florian Nolte, Andrew Melnik, and Helge Ritter. Stroke-based rendering: From heuristics to deep learning. *arXiv preprint arXiv:2302.00595*, 2022.

Jeripothula Prudviraj and Vikram Jamwal. Sketch & paint: Stroke-by-stroke evolution of visual artworks. In *European Conference on Computer Vision*, pages 202–217. Springer, 2024.

Jeripothula Prudviraj and Vikram Jamwal. Sketch & paint: Stroke-by-stroke evolution of visual artworks. *arXiv preprint arXiv:2502.20119*, 2025.

Samuel Pun and Chris Tsang. Vtracer: A free and open-source vectorization tool for images. `https://github.com/visioncortex/vtracer`, 2020. Accessed: 2025-04-15.

Paul Rosin and John Collomosse. *Image and video-based artistic stylisation*, volume 42. Springer Science & Business Media, 2012.

Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

Jaskirat Singh and Liang Zheng. Combining semantic guidance and deep reinforcement learning for generating human level paintings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16387–16396, 2021.

Jaskirat Singh, Cameron Smith, Jose Echevarria, and Liang Zheng. Intelli-paint: Towards developing more human-intelligible painting agents. In *European Conference on Computer Vision*, pages 685–701. Springer, 2022.

Yiren Song, Shijie Huang, Chen Yao, Xiaojun Ye, Hai Ci, Jiaming Liu, Yuxuan Zhang, and Mike Zheng Shou. Processpainter: Learn painting process from sequence data. *arXiv preprint arXiv:2406.06062*, 2024.

Sara L Su, Ying-Qing Xu, Heung-Yeung Shum, and Falai Chen. Simulating artistic brushstrokes using interval splines. In *Proceedings of the 5th IASTED International Conference on Computer Graphics and Imaging*, pages 85–90, 2002.

Zhengyan Tong, Xiaohang Wang, Shengchao Yuan, Xuanhong Chen, Junjie Wang, and Xiangzhong Fang. Im2oil: stroke-based oil painting rendering with linearly controllable fineness via adaptive sampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1035–1046, 2022.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Zhengxia Zou, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. Stylized neural painting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15689–15698, 2021.

A Appendix: Additional Results and Discussion

A.1 Effect of Brush Styles on Rendering

We further examine brush type effects on rendered painting texture and morphology in Figure 5. This analysis reveals how different brush attributes-size, texture, and shape orientation-significantly influence visual style and aesthetic quality. Rectangular brushes produce smooth, uniform compositions, while angular, textured brushes introduce structural emphasis and artistic character. These findings underscore the importance of brush selection in achieving specific artistic effects and style replication.

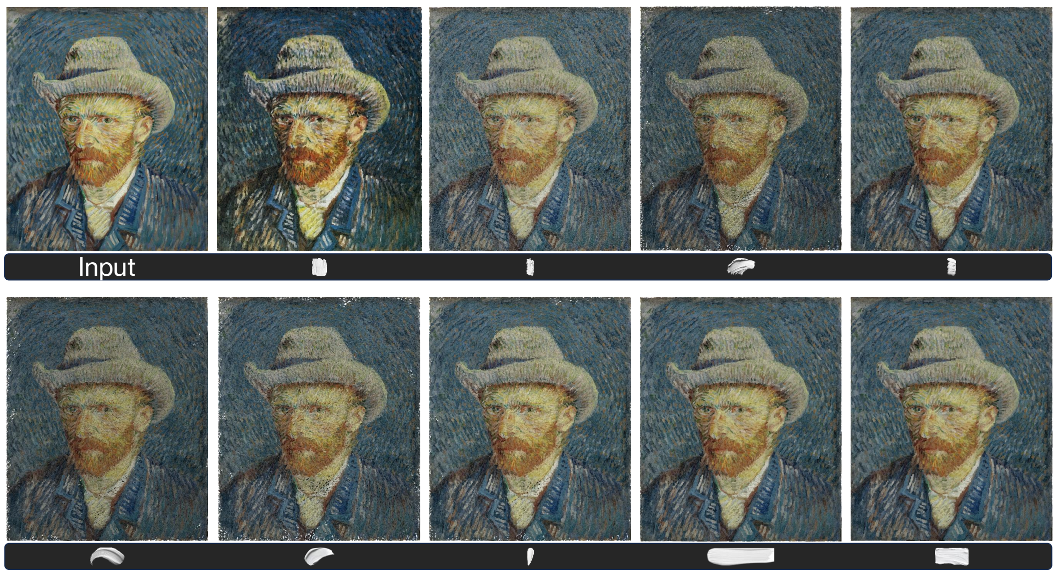


Figure 5: Different final rendered outputs of an input image (top left) based on various brush types [Zoom in to observe stroke details closely.].

A.2 Quantitative Results

Beyond creating artist-aligned region-by-region painting processes, our stroke construction model effectively recreates painterly marks that closely approximate original images, as evidenced in Figures 3, 4, and 5. To quantify this similarity, we compute perceptual metrics Zhang et al. [2018] and cosine similarity between input images and rendered outputs. Table 1 demonstrates that our model establishes semantically coherent painting processes while preserving intricate visual details of the source artwork.

After resizing images to 224×224, we employ perceptual metrics measuring deep feature similarity using ResNet101 He et al. [2016], where lower values indicate higher similarity. We also calculate pixel-level cosine similarity between inputs and outputs, with higher values (0-1 range) representing greater correspondence. Our results show perceptual similarity ranging from 0.04 to 0.20 and cosine similarity from 0.98 to 0.99 across WikiArt, METFace, Delaunay, and style-transferred samples. Table 1 reports only our method's results, as prior works lack comparable metrics on our evaluation datasets, precluding direct quantitative comparison.

Metric	WikiArt Saleh and Elgammal [2015]	METFace Karras et al. [2020]	Delaunay Gontier et al. [2022]	Style Fofr [2025]
CS	0.9959	0.9888	0.9934	0.9899
PM	0.0479	0.1395	0.0471	0.1123

Table 1: Similarity between input images and rendered outputs. CS: Cosine Similarity, PM: Perceptual Metric.

A.3 Comparison with Existing Methods

We evaluate our approach against three prominent stroke-based rendering methods: optimization-based Stylized Neural Painting (SNP) Zou et al. [2021], and region-based approaches including Compositional Neural Painter (CNP) Hu et al. [2023], Sketch & Paint Prudviraj and Jamwal [2024], and Segmentation-Based Parametric Painting (SBPP) de Guevara et al. [2024]. Figure 6 presents qualitative comparisons highlighting our method's advantages.

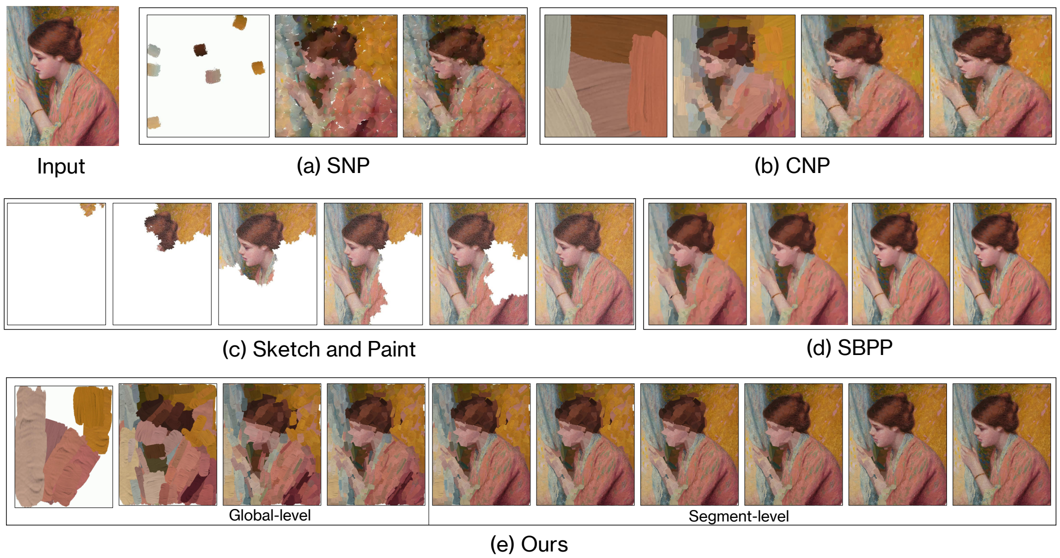


Figure 6: Qualitative comparison of our proposed method with (a) Stylized neural painting (SNP) Zou et al. [2021] (b) Compositional neural painter (CNP) Hu et al. [2023] (c) Sketch & paint Prudviraj and Jamwal [2025] (d) Segmentation-based parametric painting (SBPP) de Guevara et al. [2024]

Our approach uniquely supports variable resolutions and aspect ratios while maintaining artwork quality, unlike existing methods that often lack dimensional flexibility. SNP restricts inputs to square formats and frequently produces blurred outputs. Both CNP and SBPP generate strokes within rigid grid partitions, constraining natural semantic flow in region-guided stroke generation. Among compared methods, SBPP handles high-resolution inputs effectively and produces satisfactory strokes, though all generate limited stroke counts with substantial computational overhead.

Metric	SNP Zou et al. [2021]	PT Liu et al. [2021]	CNP Hu et al. [2023]	Ours
CS	0.9815	0.9892	0.9923	0.9959
PM	0.1686	0.1253	0.0931	0.0479

Table 2: Similarity between input images and rendered outputs on WikiArt against prior methods. CS: Cosine Similarity, PM: Perceptual Metric.

Method	PT Liu et al. [2021]	SRL Singh and Zheng [2021]	SNP Zou et al. [2021]	Im2oil Tong et al. [2022]	CNP Hu et al. [2023]	Ours
PM	0.1353	0.1950	0.1379	0.118	0.1026	0.1516
L_2	0.0128	0.0161	0.0081	0.0091	0.0046	0.0102

Table 3: Quantitative comparison of perceptual similarity (PM) and L_2 distance on ImageNet against prior methods.

Quantitative comparison with existing methods on WikiArt and ImageNet is presented in Table 2 and Table 3. From Table 2, we observe that our method outperforms prior approaches in overall performance. It yields marginally higher perceptual values compared to previous methods (Table 3); we attribute it largely to the inherent challenges of applying stroke-specific vectorization to ImageNet. Unlike artworks, ImageNet images generally exhibit limited color diversity and lower resolution, which complicates precise vector region extraction. Nevertheless, our method achieves comparable similarity scores even under these constraints.

Our vectorized approach renders 100 strokes per second and processes high-resolution $4K$ images with 80,000 vectorized regions in approximately 10 minutes, significantly outperforming current reinforcement learning-based methods. While Sketch & Paint employs SVG-based representations and sequencing but simply overlays vector patches without simulating brush strokes, leading to limited artistic rendering. Additionally, it treats the artwork as a single unified entity, without decomposing it into meaningful regions. In contrast, our method renders brush strokes progressively on a region-by-region basis.

A.4 User Study

Further, we conducted an initial user study with limited participants (5), each evaluating five images rendered using SNP Zou et al. [2021], PT Liu et al. [2021], CNP Hu et al. [2023], and our proposed method. Participants rated video renderings on naturalness, user engagement, stroke quality, and overall experience using a 1–10 scale. The survey assessed how each system emulates the natural stroke sequence, the level of engagement during interaction, the quality of stroke texture and tone, and overall satisfaction. The average ratings, presented in Figure 7, indicate that our method achieves more systematic stroke sequencing and offers a satisfying user experience compared to the other approaches.

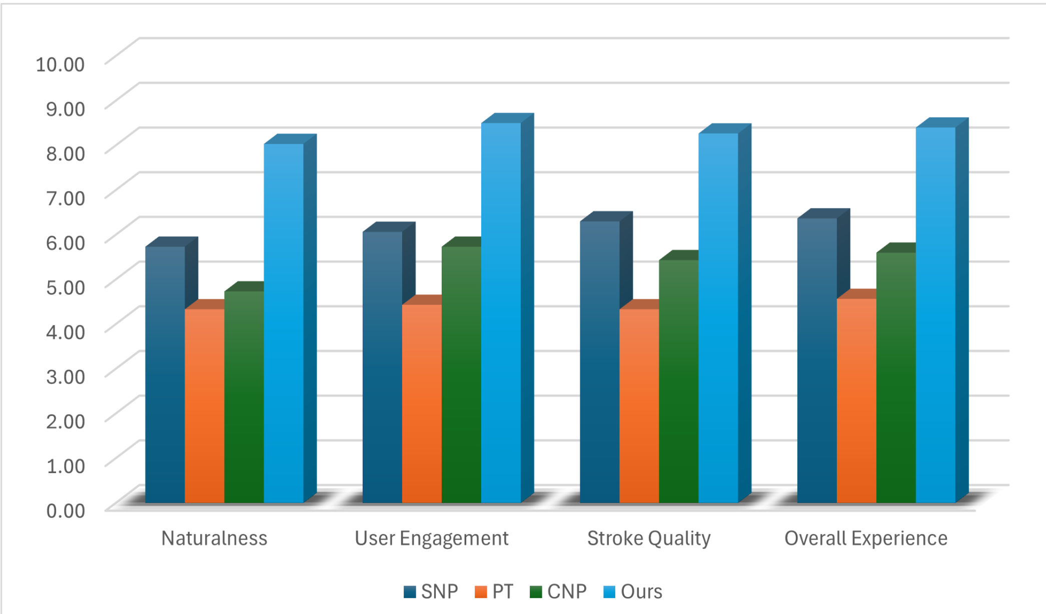


Figure 7: User survey results on generated renderings across different methods

B Appendix: Related Work

The fundamental challenge of determining *how to paint* and *where to paint* has been extensively explored within the Stroke-Based Rendering (SBR) framework, which aims to recreate non-photorealistic imagery by strategically calculating the placement and characteristics of discrete elements such as paint strokes or stipples Hertzmann [2003], Nolte et al. [2022], Song et al. [2024]. Unlike conventional image synthesis models such as VAEs Kingma [2013], GANs Goodfellow et al. [2014], and diffusion models Ho et al. [2020], which operate at the pixel level, SBR models align with human painting processes by treating brushstrokes as primary building blocks. These models reconstruct images stroke by stroke, simulating artistic creation and achieving superior fidelity to real artistic works Hu et al. [2023].

Classical SBR Methods. Early SBR algorithms explored procedural Haeberli [1990], rule-based systems Litwinowicz [1997], greedy search techniques Gooch et al. [2002], and user-guided methods Hertzmann [1998] to estimate stroke characteristics and placement. These methods typically rely on mathematical models such as cubic B-splines to define stroke primitives Hertzmann [1998], Su et al. [2002]. While effective for basic SBR tasks, they struggle to capture intricate painting details and face significant challenges in precise stroke placement. Optimization-based SBR approaches have introduced greater flexibility in stroke positioning and better alignment with artistic intent. For instance, EM-like packing algorithms Hertzmann [2003], Rosin and Collomosse [2012] based on heuristic optimization arrange non-overlapping stroke primitives. Similarly, Im2Oil Tong et al. [2022] employs texture-encapsulated meta brushstrokes as primitives and leverages adaptive sampling with greedy search guided by probability density maps to determine stroke parameters, producing visually appealing rendered paintings. However, its stroke sequences lack semantic alignment with image content, resulting in unnatural painting processes.

Deep Learning-Based SBR. Recent years have witnessed widespread adoption of deep learning models for sequential stroke generation. Early approaches Graves [2013], Ha and Eck [2017] utilized RNNs to decompose images into brushstrokes, but their reliance on manual annotations severely limited scalability and applicability across diverse artistic scenarios. Reinforcement learning-based painting agents Zou et al. [2021], Ganin et al. [2018], de Guevara et al. [2024], Chen et al. [2024], Huang et al. [2019], Liu et al. [2021], Song et al. [2024], Hu et al. [2024, 2023] have been introduced to synthesize stroke sequences, though these methods remain limited to rendering simple images such as basic sketches. Learning to Paint Huang et al. [2019] represents a pioneering approach combining deep deterministic policy gradient (DDPG) Lillicrap [2015] reinforcement learning with a WGAN-based reward system to achieve realistic painting. Although it reconstructs images using numerous tiny strokes with visually satisfying results, the method suffers from stylistic constraints and struggles with high-resolution images. To address the low training efficiency of RL-based methods, Paint Transformer Liu et al. [2021] employs a CNN-Transformer architecture that generates strokes in a feed-forward manner, improving both efficiency and quality of stroke generation. Zou *et al.* Zou et al. [2021] introduced a stroke optimization strategy that decomposes images into parameterized strokes and iteratively searches for optimal parameters. While achieving relatively good results in rendering realistic images, this method suffers from long optimization times.

Semantic Region-Based Approaches. To better mimic artistic painting processes, several methods have explored semantic region-based strategies. Intelli-Paint Singh et al. [2022] trains an RL model to predict strokes within enumerated semantic regions using attention windows and stroke parameters. However, its reliance on object detection results in block-style painting regions that lack alignment with image semantics. Hu *et al.* Hu et al. [2023] proposed a compositional neural painter that dynamically predicts the next painting region based on the current canvas, addressing boundary inconsistency artifacts caused by uniform block-dividing strategies. Recent segmentation-based methods have shown promise: de Guevara et al. [2024] employs DETR Carion et al. [2020] and dynamic attention to enable patch-based optimization across different image regions, effectively capturing both large-scale structures and fine details. Similarly, Hu *et al.* Hu et al. [2024] leveraged SAM Kirillov et al. [2023] and Depth Anything to implement a coarse-to-fine strategy using hierarchical painting regions, minimizing the mapping between pixel domain and stroke parameter search space. Recently, a vector region–based approach **?** was introduced for stroke rendering. However, it fails to incorporate a hierarchical stroke structure, ranging from global-level coarse details to segment-level finer strokes for refinement.

Limitations of Existing Approaches. Despite producing satisfying rendering results, existing methods fail to predict strokes region by region in a manner consistent with artistic strategy, where, for example, one segment is fully painted before transitioning to another. These methods struggle to capture stroke primitive features within sub-regions of segments, such as SVG-based vectorized regions. Vectorized regions effectively represent geometric and color features by tracing pixel values, which are crucial for detailed and accurate rendering. Recent work, Sketch & Paint Prudviraj and Jamwal [2025], explores SVG-based sequencing for art evolution but treats images globally, lacking region-by-region painting strategy. Additionally, it merely overlays SVG patches to complete paintings instead of utilizing brush-like strokes, resulting in a lack of painterly rendering quality.