Transparency-Aware Style Transfer via Soft Alpha-Guided Feature Propagation

Anonymous Submission

Abstract

Neural style transfer (NST) aims to synthesize images that retain the structural content of one image while adopting the visual style of another. While recent methods have advanced stylization quality and speed, they typically assume fully opaque inputs and disregard the alpha channel commonly present in RGBA images. However, pixel-wise transparency is essential in many practical settings, including compositional graphics, UI layering, and even adversarial attacks where hidden content is encoded in alpha channels. We introduce the task of alpha-aware neural style transfer (A-NST), which generalizes conventional NST to handle partially visible inputs. To address this challenge, we propose a novel framework called Soft Alpha-Guided Feature Propagation (SAFP), where the alpha channel serves as a continuous visibility prior to guide feature modulation throughout the stylization process. To evaluate A-NST, we construct AlphaStyle, a new dataset comprising RGBA-style pairs with varying transparency levels, derived from MSCOCO and WikiArt via segmentation and compositing. Experiments demonstrate that our method significantly outperforms RGB-only baselines in preserving structural fidelity and transparency alignment. To our knowledge, this is the first formal study to frame neural style transfer in the RGBA domain.

Introduction

Neural style transfer (NST) reimagines visual content by synthesizing the semantics of one image with the artistic identity of another using deep neural networks. Since the seminal work of Gatys et al. (Gatys, Ecker, and Bethge 2016), which framed style transfer as an optimization problem using the Gram matrix of VGG features, the field has advanced significantly toward real-time stylization (Huang and Belongie 2017), high-fidelity generation (Kwon et al. 2024), and multimodal controllability (Ahn et al. 2024).

However, a critical limitation persists across almost all existing neural style transfer (NST) models: the implicit assumption that input images are fully opaque. This design paradigm effectively excludes the alpha channel—an integral component of digital imagery that encodes per-pixel transparency. The alpha channel is central to layered graphics, UI composition, and image compositing workflows, where transparency governs how visual elements interact across spatial hierarchies. By ignoring this dimension, conventional NST models fail to account for transparency semantics, often resulting in stylization artifacts such as edge bleeding, style leakage into transparent regions, and compositional inconsistencies during downstream integration.

Unlike RGB channels that define visual appearance, the alpha channel governs perceptual visibility. In the context of style transfer, alpha values modulate stylistic salience—pixels with high alpha demand stronger stylization fidelity, while those with low or zero alpha suppress stylistic rendering. This transparency signal encodes important spatial semantics that are entirely discarded in RGB-only NST pipelines.

To address this gap, we introduce **alpha-aware neural style transfer**—a novel NST paradigm that models transparency semantics throughout the stylization process. Our end-to-end framework accepts RGBA inputs, applies alpha-guided feature computation, and outputs stylized RGBA images that faithfully preserve both artistic identity and transparency structure. To our knowledge, this is the first NST method to treat soft alpha information as a first-class modeling component.

While our study focuses on the alpha channel as a prominent transparency cue, we view our design as a first step toward modeling more general per-pixel semantics such as depth, surface normals, or material attributes. We leave these extensions for future work.

Our contributions are summarized as follows:

- Problem formulation: We define alpha-aware neural style transfer as a new task setting that extends traditional RGB-only NST to RGBA-space modeling with explicit transparency semantics.
- Soft Alpha-Guided Feature Propagation (SAFP): We introduce a transparency-aware feature routing mechanism that integrates the alpha channel as a continuous visibility prior throughout the network.
- Alpha-guided perceptual loss: We design an alpha-aware perceptual loss that modulates stylization supervision based on transparency, encouraging semantic consistency between visibility and stylized output.
- AlphaStyle Dataset: We construct AlphaStyle, the first large-scale RGBA dataset for NST, derived from WikiArt

and MSCOCO using Segment Anything (Kirillov et al. 2023) to extract irregular masks. The dataset contains over 20,000 high-quality RGBA image pairs in lossless PNG format.

Realted Works

Neural Style Transfer

Neural Style Transfer (NST) aims to synthesize an image that preserves the semantic content of a source image while reflecting the artistic style of another. The field was initiated by Gatys et al. (Gatys, Ecker, and Bethge 2016), who formulated NST as an optimization problem by minimizing a style-content loss defined on the Gram matrices of deep feature maps extracted by a pretrained VGG network (Simonyan and Zisserman 2014). This seminal work demonstrated the capability of convolutional features to separate and recombine style and content.

However, optimization-based methods (Li and Wand 2016; Berger and Memisevic 2016; Risser, Wilmot, and Barnes 2017) are computationally expensive, requiring a forward and backward pass for each output image. To address efficiency, feed-forward approaches were proposed, training a generative model to approximate the style transformation. Early works (Johnson, Alahi, and Fei-Fei 2016; Ulyanov et al. 2016) trained one model per style, while later extensions (Dumoulin, Shlens, and Kudlur 2016; Chen et al. 2017) enabled multiple styles per model. More recent methods (Huang and Belongie 2017; Xu et al. 2021; Kwon et al. 2024) support arbitrary style transfer via feature statistic alignment or adaptive modulation.

Despite substantial progress in efficiency, fidelity, and control, all aforementioned methods operate under a common assumption: input images are fully opaque and represented solely in RGB space. This assumption simplifies modeling but ignores transparency semantics crucial to layered visual content, rendering these methods ineffective when dealing with partial visibility or alpha-masked regions.

Transparency Modeling in Vision Tasks

While transparency is largely ignored in current neural style transfer methods, it plays a significant role in several other vision tasks. For instance, in image matting (Yao et al. 2024), the alpha channel is a central component used to separate foreground from background, guiding the generation of high-fidelity composite images. Recent works such as Matte Anything (Yao et al. 2024) leverage large vision models to predict continuous alpha mattes, demonstrating that soft transparency cues provide essential structure for downstream editing.

Beyond traditional compositing, transparency has also been used as a medium for adversarial attack design. Xia et al. (Xia and Chen 2025) introduced AlphaDog, an adversarial example framework that manipulates the alpha channel to fool Large Language Models (LLMs) equipped with vision encoders. They show that many vision-language models rely solely on RGB content, neglecting alpha information. As a result, alpha-perturbed images are perceptually unchanged to humans but lead to erroneous model predictions—highlighting a critical semantic gap between human and model perception.

These applications affirm that the alpha channel encodes nontrivial semantics relevant to content structure and visibility. Yet, to the best of our knowledge, such transparency cues have never been explicitly modeled in neural style transfer. Our work fills this gap by incorporating soft alpha information into both feature learning and stylization loss functions.

Visibility-Aware Feature Propagation

Stylization under partial visibility presents unique challenges: not all pixels contribute equally to the perceptual structure or aesthetic of an image. In traditional computer vision tasks such as inpainting, techniques like partial convolution (Liu et al. 2018) and gated convolution (Yu et al. 2019) were introduced to address irregular missing regions by modulating feature propagation with spatial masks. These methods condition the convolution operation on binary or learned soft masks to restrict computation to valid areas, enabling reconstruction in structurally incomplete inputs.

Some recent approaches also extend the use of (soft) masks to guide feature modulation. For instance, segmentation-aware stylization (Zhao et al. 2020; Yu, Wang, and Li 2024; Ko and Kim 2023) employs semantic (soft) masks to preserve object-level consistency across style domains. These methods treat masks as external priors—typically reflecting semantic class identity or object boundaries—but do not address the notion of visibility as a first-class semantic cue.

Our approach differs in both motivation and formulation. Instead of treating masks as auxiliary spatial signals, we regard the alpha channel as a native representation of pixel-level visibility semantics. It directly encodes how strongly each pixel should participate in stylization, offering a more principled foundation for visibility-aware feature routing. To realize this, we implement a *soft alpha-guided feature propagation* mechanism, where alpha values modulate both intermediate features and perceptual losses. This extends prior masked-feature methods beyond spatial reconstruction, enabling stylization to be conditioned on natural transparency cues embedded in RGBA data.

To our knowledge, this is the first attempt to integrate visibility-aware guidance—derived from alpha transparency—into the stylization process in an end-to-end learnable framework.

Method

Problem Formulation: Alpha-aware Neural Style Transfer

We define **Alpha-aware Neural Style Transfer (A-NST)** as an extension of conventional neural style transfer from the RGB space to the RGBA domain, where the alpha channel encodes per-pixel transparency. Unlike standard NST, which assumes full visibility across all pixels, A-NST explicitly models *visibility semantics*—that is, the spatial extent and intensity of stylization—guided by the input alpha map.

Formally, let $I_c \in \mathbb{R}^{H\times W\times 4}$ and $I_s \in \mathbb{R}^{H\times W\times 4}$ denote the content and style images in RGBA format, respectively. The goal is to synthesize a stylized output $I_t \in \mathbb{R}^{H\times W\times 4}$ such that:

- The RGB channels of I_t preserve the structural content of I_c while reflecting the style patterns of I_s;
- The alpha channel of I_t conveys semantically meaningful transparency, consistent with I_c and/or I_s depending on the stylization objective.

This problem formulation introduces a new dimension to neural stylization: transparency-aware modeling. It demands the network to account not only for appearance transformation, but also for selective feature propagation conditioned on pixel-wise visibility. To address this, we propose an end-to-end framework that incorporates soft alpha-guided feature routing and transparency-aware loss functions to produce stylized RGBA outputs that preserve both artistic identity and alpha structure.

Soft Alpha-Guided Feature Propagation

We propose **Soft Alpha-Guided Feature Propagation (SAFP)**, a transparency-aware mechanism that integrates pixel-wise visibility semantics into the stylization process. Instead of treating the alpha channel as an auxiliary mask applied post-hoc, we embed it directly into the convolutional backbone, allowing alpha transparency to dynamically modulate feature propagation throughout the network.

Given an RGBA input image $I \in \mathbb{R}^{H\times W\times 4}$, we decompose it into an RGB tensor $I_{RGB} \in \mathbb{R}^{H\times W\times 3}$ and a normalized alpha map $M_\alpha \in [0,1]^{H\times W}$. The alpha map is interpreted as a soft visibility prior: higher values indicate greater relevance in stylization, while lower values correspond to transparent or visually suppressed regions.

To incorporate M_α into the feature extraction pipeline, we introduce *Soft Partial Convolution (SoftPConv)*, an extension of Partial Convolution (Liu et al. 2018) to continuous masks. The output feature Y at pixel (i,j) is computed as:

$$Y_{i,j} = \begin{cases} \dfrac{\sum\limits_{(u,v)\in\Omega_{i,j}} W_{u,v}\cdot I_{RGB}^{(u,v)}\cdot M_\alpha^{(u,v)}}{\sum\limits_{(u,v)\in\Omega_{i,j}} M_\alpha^{(u,v)}} + b, & \text{if } \sum M_\alpha > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Here, $\Omega_{i,j}$ denotes the local kernel window centered at pixel (i,j), W is the convolution kernel, and b is the bias term.

Following (Liu et al. 2018), we update the alpha map after each layer to reflect the effective visibility at the next stage:

$$M'_\alpha(i,j) = \frac{1}{|\Omega|}\sum_{(u,v)\in\Omega_{i,j}} M_\alpha(u,v) \tag{2}$$

We apply SoftPConv throughout the encoder by replacing all vanilla convolutions. This ensures that visibility-aware feature propagation is maintained at every spatial scale, enabling the network to stylize partially visible regions while respecting transparency boundaries and semantic layering.

Importantly, SoftPConv is implemented as a drop-in replacement for conventional convolutional layers and does not require architectural re-design. As such, it can be seamlessly integrated into existing convolutional architectures, making it a lightweight and modular enhancement for transparency-aware vision tasks.

Alpha-Aware Network Architecture

SoftPConv layers, introduced in the previous section as the implementation of SAFP, replace all vanilla convolutions in the encoder to enable transparency-aware stylization.To validate its effectiveness, we integrate SoftPConv into a recent state-of-the-art architecture for arbitrary style transfer: *AesFA* (Kwon et al. 2024).

Our design goal is to extend style transfer models (here is AesFA) to operate in the RGBA domain while preserving its expressive capacity and training stability. We therefore adopt a minimally invasive strategy: only convolutional components are modified to support alpha guidance, while higher-level design such as feature adaptation remains unchanged.

Input and Output. The model receives two RGBA images as input: a content image $I_c \in \mathbb{R}^{H\times W\times 4}$ and a style image $I_s \in \mathbb{R}^{H\times W\times 4}$. The output is a stylized image $I_t \in \mathbb{R}^{H\times W\times 4}$. We decompose each input into RGB and alpha components. The RGB parts are used to encode content and style features. The alpha channel of the *style image* is passed into the network as a soft visibility prior via SoftPConv, modulating feature propagation. In contrast, the *content alpha* is not processed during stylization but preserved for post-processing to maintain downstream compositional fidelity.

Module Replacement Strategy. All vanilla convolutions in the encoder of AesFA are replaced with SoftPConv modules, allowing the style alpha map to influence activations throughout the hierarchy. These modified layers serve as drop-in replacements and require no change to the surrounding network topology or loss functions. This design ensures compatibility with pretrained weights and minimizes engineering overhead.

Alpha Preprocessing and Postprocessing. Before feature extraction, the style alpha channel is normalized and injected as a soft mask into SoftPConv layers. During decoding, the network reconstructs a stylized RGB image, to which the original content alpha is appended to form the RGBA output. This ensures that style propagation respects stylistic visibility while maintaining structural transparency from the content.

Generalization and Compatibility. Although our implementation is instantiated on AesFA, the proposed integration strategy is model-agnostic. Any encoder-decoder-based NST architecture can benefit from alpha-aware modulation by substituting convolutional layers with SoftPConv. This highlights the plug-and-play nature of SoftPConv as a modular enhancement for RGBA-compatible stylization networks.

Alpha-Aware Loss Functions

To ensure that our training objective is consistent with the transparency-aware design of Soft Alpha-Guided Feature Propagation (SAFP), we adapt the loss framework of AesFA (Kwon et al. 2024) by incorporating visibility guidance through the alpha channel. Rather than redesigning all loss terms from scratch, we selectively apply alpha-aware weighting where it is most critical, while keeping the other objectives unchanged to preserve stable baselines.

Content and Style Losses. We retain the original content and style reconstruction losses of Gatys et al. (Gatys, Ecker, and Bethge 2016) without modification. These objectives rely on global feature statistics and are largely agnostic to local visibility variations, making them less sensitive to partial transparency. Preserving their formulation also allows for a cleaner comparison with existing RGB-only baselines.

Alpha-aware EFDM Contrastive Loss. The perceptual contrastive loss in AesFA, which is built on Exact Feature Distribution Matching (EFDM) (Zhang et al. 2022), is highly sensitive to local feature activations and therefore benefits most from explicit visibility control. We extend EFDM by incorporating alpha masks M_α when matching positive and negative samples: features at transparent or occluded pixels are excluded from the sorted feature distributions before distance computation. A subtle challenge arises in this setting: the effective number of valid pixels may differ between the output O_i and the positive sample $S_{\text{pos},i}$ (or negative samples $S_{\text{neg},j}$) due to different alpha masks. Because EFDM relies on comparing sorted feature distributions, unequal sample sizes would result in biased or unstable matching. To address this, we resample the larger set via linear interpolation so that both distributions have the same number of elements before sorting. This step ensures that the alpha-aware EFDM remains a fair and consistent measure of feature alignment.

Formally, the alpha-aware contrastive loss is defined as:

$$\mathcal{L}^{\alpha}_{\text{contrastive}} = \sum_{i=1}^{N} \frac{\left\| F_l(O_i) - \text{EFDM}\left(F_l(O_i), F_l(S_{\text{pos},i}); M_\alpha\right) \right\|_2}{\sum_{j=1}^{k} \left\| F_l(O_i) - \text{EFDM}\left(F_l(O_i), F_l(S_{\text{neg},j}); M_\alpha\right) \right\|_2} \tag{3}$$

where $F_l(\cdot)$ denotes the features extracted at layer l, $S_{\text{pos},i}$ and $S_{\text{neg},j}$ are the positive and negative style samples, and $\text{EFDM}(\cdot; M_\alpha)$ represents masked feature matching with resampling.

Total Loss. Our final training objective is a weighted combination of the three terms:

$$\mathcal{L}_{\text{total}} = \lambda_c \mathcal{L}_{\text{content}} + \lambda_s \mathcal{L}_{\text{style}} + \lambda_{\text{con}} \mathcal{L}^{\alpha}_{\text{contrastive}} \tag{4}$$

where λ_c, λ_s, and λ_{con} are scalar weights controlling the relative importance of each loss.

Discussion. By making the EFDM-based contrastive loss alpha-aware while keeping the content and style losses intact, we achieve two goals: (i) the most visibility-sensitive component of the objective is explicitly aligned with transparency semantics, and (ii) the overall formulation remains plug-and-play and compatible with existing NST frameworks. This minimal yet effective modification closes the loop for SAFP, ensuring that both feature propagation and supervision consistently respect the alpha channel.

References

Ahn, N.; Lee, J.; Lee, C.; Kim, K.; Kim, D.; Nam, S.-H.; and Hong, K. 2024. Dreamstyler: Paint by style inversion with text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 674–681.

Berger, G.; and Memisevic, R. 2016. Incorporating long-range consistency in cnn-based texture generation. *arXiv preprint arXiv:1606.01286*.

Chen, D.; Yuan, L.; Liao, J.; Yu, N.; and Hua, G. 2017. Stylebank: An explicit representation for neural image style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1897–1906.

Deng, Y.; Tang, F.; Dong, W.; Ma, C.; Pan, X.; Wang, L.; and Xu, C. 2022. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11326–11336.

Dumoulin, V.; Shlens, J.; and Kudlur, M. 2016. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*.

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.

Huang, S.; An, J.; Wei, D.; Luo, J.; and Pfister, H. 2023. Quantart: Quantizing image style transfer towards high visual fidelity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5947–5956.

Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510.

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.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026.

Ko, K.; and Kim, C.-S. 2023. Continuously masked transformer for image inpainting. In *Proceedings of the IEEE/CVF international conference on computer vision*, 13169–13178.

Kwon, J.; Kim, S.; Lin, Y.; Yoo, S.; and Cha, J. 2024. Aesfa: an aesthetic feature-aware arbitrary neural style transfer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 13310–13319.

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.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13*, 740–755. Springer.

Liu, G.; Reda, F. A.; Shih, K. J.; Wang, T.-C.; Tao, A.; and Catanzaro, B. 2018. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European conference on computer vision (ECCV)*, 85–100.

Phillips, F.; and Mackintosh, B. 2011. Wiki art gallery, inc.: A case for critical thinking. *Issues in Accounting Education*, 26(3): 593–608.

Risser, E.; Wilmot, P.; and Barnes, C. 2017. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv preprint arXiv:1701.08893*.

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

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*.

Xia, Q.; and Chen, Q. 2025. AlphaDog: No-Box Camouflage Attacks via Alpha Channel Oversight. In *NDSS*.

Xu, W.; Long, C.; Wang, R.; and Wang, G. 2021. Drb-gan: A dynamic resblock generative adversarial network for artistic style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6383–6392.

Yao, J.; Wang, X.; Ye, L.; and Liu, W. 2024. Matte anything: Interactive natural image matting with segment anything model. *Image and Vision Computing*, 147: 105067.

Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; and Huang, T. S. 2019. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4471–4480.

Yu, Y.; Wang, J.; and Li, N. 2024. Foreground and background separated image style transfer with a single text condition. *Image and Vision Computing*, 143: 104956.

Zhang, C.; Xu, X.; Wang, L.; Dai, Z.; and Yang, J. 2024. S2wat: Image style transfer via hierarchical vision transformer using strips window attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 7024–7032.

Zhang, Y.; Li, M.; Li, R.; Jia, K.; and Zhang, L. 2022. Exact feature distribution matching for arbitrary style transfer and domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8035–8045.

Zhao, H.-H.; Rosin, P. L.; Lai, Y.-K.; and Wang, Y.-N. 2020. Automatic semantic style transfer using deep convolutional neural networks and soft masks. *The Visual Computer*, 36(7): 1307–1324.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.