AAAI Press Formatting Instructions for Authors Using LaTeX — A Guide

Anonymous Submission

Abstract

AAAI creates proceedings, working notes, and technical reports directly from electronic source furnished by the authors. To ensure that all papers in the publication have a uniform appearance, authors must adhere to the following instructions.

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 underlies almost all existing NST models: the assumption that input images are fully opaque. This design paradigm excludes the alpha channel—a key component in digital imaging that encodes per-pixel transparency and is widely used in UI composition, image compositing, and layered artistic workflows. As a result, standard NST methods ignore transparency semantics, which often leads to undesirable stylization artifacts such as edge bleeding, occluded detail transfer, and inconsistency in downstream composition.

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 Partial Convolution: We propose a soft partial convolution module that generalizes conventional partial convolutions to handle continuous-valued alpha masks, enabling effective feature propagation under partial visibility.
- 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 traditional neural style transfer from the RGB domain to the RGBA space, where the alpha channel provides pixel-level transparency signals. In contrast to conventional NST, which assumes full visibility across all pixels, A-NST explicitly incorporates *visibility semantics*—i.e., where and to what extent style should be rendered—guided by the alpha channel.

Formally, let $I_c \in \mathbb{R}^{H\times W\times 4}$ and $I_s \in \mathbb{R}^{H\times W\times 4}$ be the content and style images in RGBA format, respectively. The objective 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 layout of I_c while reflecting the style characteristics of I_s;
- The alpha channel of I_t conveys meaningful transparency, inherited from I_c and/or I_s depending on task constraints.

This problem formulation introduces a new modeling dimension in NST: transparency-aware processing. To this end, we propose a framework that integrates soft alpha-guided feature propagation and visibility-weighted loss computation, ensuring style synthesis aligns with the perceptual transparency structure of the input.

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.

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, 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.

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.

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.

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.