Dear Editors and Reviewers:

Thank you for your letter and for the reviewers’ comments concerning our manuscript entitled “Bridging the Metrics Gap in Image Style Transfer: A Comprehensive Survey of Models and Criteria” (ID: NEUCOM-D-24-08078). Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval. Furthermore, we would like to show the details as follows:

**Reviewer 1#:**

Thank you for your positive recognition of our work. Your approval has greatly encouraged us. We have carefully considered the revision suggestions you provided and have made the following changes in response to them.

1. *The discussions on interpretability and controllability need further clarification (Sec. 5.2). Current research on style transfer develops limited research on explicit interpretability [1-4]. Some diffusion model, use visual programming based approaches to achieve controllability [5-6], however, the transferring process is not transparent (This topic can be included in Sec. 5.5 in human-computer interaction). These works should be discussed, and highlight the future directions.*

**The authors’ answer:** We have reconsidered Section 6 "Frontiers and Challenges." As you mentioned, the original manuscript lacked sufficient discussion on interpretability and controllability, which resulted in a reduced practical value of the paper. Therefore, we have rewritten certain subsections of Section 6.

Specifically, regarding interpretability, after carefully reviewing the references you provided, we have expanded Section 6.2 "Interpretability." Following the structure of the sentences you suggested, we discuss the shortcomings of current results in terms of interpretability for different network architectures, such as attention mechanisms, diffusion models, and GANs. Due to space limitations, it is not convenient to paste the content here, so please refer to the latest version of our manuscript, Section 6.2 "Interpretability," for more detailed information.

As for controllability, we have carefully considered its relationship with the practical applications of style transfer. We believe that low controllability affects the real-world use of style transfer results, and hence, we added a new subsection 6.5 "Applications." This subsection discusses the barriers to applying current style transfer results in real-world scenarios from three perspectives: controllability, efficiency, and multimodality. Especially in terms of controllability, we have highlighted the consequences of low controllability, which hinders the adoption of style transfer in fields such as medical image processing, AR, VR, and assisted design. We also suggest potential techniques that could help improve controllability, aiming to provide clearer guidance for researchers.

1. *While the current paper covers a lot of specific works, a more general/high-level discussions of different methods, and future direction is encouraged. This can help the community target more specific goals.*

**The authors’ answer:** We have added Section 6.6 "Frontiers," where we introduce the potential development directions of style transfer from a higher-level perspective. Unlike the main body of the paper, which presents the development of the field in a chronological order, in this subsection, we classify the current major research directions in style transfer and briefly introduce each of them. We also highlight the challenges these research directions face and their potential future developments. Additionally, we discuss the advantages and disadvantages of using different backbone networks for style transfer tasks, aiming to reduce the understanding costs for researchers.

**Reviewer 2#:**

Thank you for your recognition of our work. We have carefully reviewed the revision suggestions you provided and have made the following changes in response:

1. *The introduction lacks clear motivation for why this survey is needed now.*

**The authors’ answer:** Indeed, we did not explicitly state the motivation of our paper in the Introduction or Abstract, and we apologize for any confusion this may have caused. To address this, we have added a new paragraph in the Introduction to explain our motivation. This paragraph introduces recent style transfer reviews, summarizes the strengths and weaknesses of these papers, and concludes with the motivation of our work: to revisit the current style transfer results in chronological order, and more importantly, to examine the lack of unified objective evaluation standards in the field of style transfer**.**.

1. *Missing discussion of several recent works on the narrow topic, such as: [StyleFormer: Real-time Arbitrary Style Transfer...] and [Stylerf: Zero-shot 3d style transfer of neural radiance...]*

**The authors’ answer:** We have carefully considered your comments and believe that your suggestions are indeed necessary. Although our work primarily focuses on image style transfer, we recognize that current style transfer reviews often overlook these narrow topics. However, to avoid disrupting the main thread of our introduction to image style transfer, we have added a new section, Section 5 "From 2D to Multimodal Exploration," to cover the achievements in these narrower fields. This section classifies recent works in these areas and focuses on introducing 21 papers in the fields of video style transfer and 3D style transfer.

1. *Limited coverage of real-world applications and practical challenges*

**The authors’ answer:** We have carefully considered your comments. As style transfer is a research field aimed at end-users, the challenges related to its applications are significant and should not be overlooked in a style transfer review. To address this, we have redesigned Section 6 "Frontiers and Challenges" and added a subsection, Section 6.5 "Applications," to specifically discuss the difficulties in applying style transfer results in real-world scenarios. This subsection introduces challenges from three perspectives: controllability, efficiency, and multimodality. We analyze the shortcomings of current style transfer results in these areas, using practical application scenarios as examples, with the goal of providing clearer guidance for researchers.

1. *Should include at least some quantitative comparison/tables of different methods*

**The authors’ answer:** We have carefully considered your comments, but we still believe that providing a quantitative comparison or table of different methods is a challenging task. We are well aware that there is still considerable controversy regarding objective evaluation standards in the field of style transfer. Different research methods use various evaluation metrics, and no unified standard has been established, making it difficult to provide a complete and representative table for quantitative comparison. Moreover, some methods may perform poorly on certain objective metrics, but the generated images might exhibit more distinctive style features, which is often difficult to fully capture with a single numerical indicator. Relying on such data may mislead researchers into overlooking the actual performance of the models. Therefore, in this study, we focus more on the systematic analysis of the methods and a deeper discussion of current evaluation standards to help readers gain a more comprehensive understanding of the effectiveness and challenges of style transfer. Nevertheless, we have still modified Table 1: Evaluation Metrics Used in Papers by replacing the original 0/1 with quantitative metrics, thus providing a table with some degree of quantitative comparison.

1. *Insufficient coverage of domain adaptation connections*

**The authors’ answer:** Thank you for your valuable suggestions. Regarding the issue of "insufficient discussion on the relevance to domain adaptation," it is true that domain adaptation is a broader task than style transfer, and style transfer can be considered a special case of domain adaptation. We believe that the core of style transfer lies in combining the content of the source domain with the style of the target domain, which is essentially a form of domain migration. Therefore, there is an inherent connection between style transfer and domain adaptation. In Section 6.5.1 "Controllability Issues," we have already discussed the relationship between the two and pointed out that style transfer shares similarities with domain adaptation tasks, especially in handling the differences between the source and target domains. However, since the research on domain adaptation covers a wider range of issues, including various migration problems between the source and target domains, we have limited our in-depth discussion of domain adaptation in this study.

1. *Need more discussion of style transfer for 3D content*

**The authors’ answer:** Thank you for your valuable suggestions. As we mentioned in our response to comment 2 "Missing discussion of several recent works on the narrow topic," we have added a section, Section 5 "From 2D to Multimodal Exploration," where we discuss 21 papers on video style transfer and 3D style transfer. Since these topics deviate somewhat from the main focus of our paper, "image style transfer," we did not include these papers in the previous sections.

1. *Currently the survey missed a lot of recent works from 2022-2024 in wider area of visual reasoning and image transfer method.*

**The authors’ answer** Thank you for your valuable suggestions. In response to this issue, we have referenced the articles you provided and added a total of 17 recent studies from 2023-2024 in the field of image style transfer to enrich our paper. Additionally, we have discussed the relationship between style transfer and other broader visual tasks in Section 6.5.1 "Controllability Issues," particularly highlighting the potential contributions of image segmentation and object detection to the field of style transfer.

Yours, Sincerely,

Xiaotong Zhou, Yuhui Zheng.

14 Dec, 2024

Nanjing University of Information Science and Technology.