Dear editors and reviewers:

Thank you all for your contribution to my manuscript entitled “Bridging the Metrics Gap in Image Style Transfer: A Comprehensive Survey of Models and Criteria” (ID: NEUCOM-D-24-08078). We highly appreciate your comments and suggestions. We have studied all your comments and suggestions one by one carefully and have made necessary corrections and clarifications. The following are our responses to the editors and reviewers.

**Reviewer 1#**

***Question 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.*

***Answer 1:*** 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. We think you have given us a good comment. Therefore, we have reorganized 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. Please kindly refer to the latest version of Section 6.2 "Interpretability" for more detailed information.

As for controllability, we 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 new 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 now 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 suggested potential techniques that could help improve the controllability, aiming to provide clearer guidance for researchers.

***Question 2:*** *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.*

***Answer 2:***Good idea! We have now added Section 6.6 "Frontiers”, in which we introduced 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 classified the current major research directions in style transfer and briefly introduced each of them. We also highlighted the challenges these research directions face and their potential future developments. Additionally, we discussed the advantages and disadvantages of using different backbone networks for style transfer tasks so that readers can understand them better.

**Reviewer 2#:**

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

***Answer 1:***Good idea! Thank you so much for your comment. To address this, we have added a new paragraph in the Introduction section to explain our motivation. In this paragraph, we introduced recent style transfer reviews, summarized the strengths and weaknesses of these papers, and pointed out the motivation of our work: to rediscuss the current style transfer progress in a chronological order, and more importantly, to examine the lack of unified objective evaluation standards in the field of style transfer.

***Question 2:*** *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...]*

***Answer 2:*** Good suggestion! Although our work primarily focused on image style transfer, we recognized that current style transfer reviews often overlook these narrow topics/research fields. 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 summarize the achievements in these narrower fields. In this section, we classified recent works in these areas and focused on introducing 21 papers in the fields of video style transfer and 3D style transfer.

***Questions 3:*** *Limited coverage of real-world applications and practical challenges*

***Answer 3:*** Thank you for your comments. As style transfer is a research field aiming 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 new subsection, Section 6.5 "Applications" to specifically discuss the difficulties in applying style transfer results in real-world scenarios. This subsection introduced challenges from three perspectives: controllability, efficiency, and multimodality. We analyzed 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.

***Question 4:*** *Should include at least some quantitative comparison/tables of different methods*

***Answer 4:*** Thank you for your comment. Providing a quantitative comparison or table of different methods is a challenging task. In this revised version of our manuscript, we have 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.

To be honest, there are considerable controversies regarding the 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.

***Question 5:*** *Insufficient coverage of domain adaptation connections*

***Answer 5:*** Thank you for your valuable suggestions. 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 this revised version, we added a new subsubsection: Section 6.5.1 "Controllability Issues". In this subsubsection, we analyzed 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 didn’t provide an in-depth discussion of domain adaptation.

***Question 6:*** *Need more discussion of style transfer for 3D content*

***Answer 6:*** Thank you for your valuable suggestions. As we mentioned in our response to *Question 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 discussed 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.

***Question 7:*** *Currently the survey missed a lot of recent works from 2022-2024 in wider area of visual reasoning and image transfer method.*

***Answer 7:*** Thank you for your valuable suggestions. We have now cited 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 also discussed the relationship between style transfer and other broader visual tasks in Section 6.5.1 "Controllability Issues" and particularly highlighted the potential contributions of image segmentation and object detection to the field of style transfer.

Dear editors and reviewers, thank you again for your efforts made to our manuscript. We wish our improvements can now satisfy you and our manuscript can be accepted for publication in your distinguished journal. We will certainly cooperate with you if you have further comments.

Yours, Sincerely,

Xiaotong Zhou, Yuhui Zheng.

Dec. 14, 2024

Nanjing University of Information Science and Technology.