

# Supplementary Materials: The Name of the Title is Hope

Anonymous Authors

## 1 OVERVIEW

In this supplementary material, more model implementation details, experimental results and analysis of experimental results are provided, which are organized as follows:

- Section 2 elucidates which specific layers within the TANet model are chosen to represent aesthetic features.
- Section 3 provides additional instances that illustrate the aesthetic assumption bias in AesUST, which assumes that every image from WikiArt embodies aesthetic quality that aligns with human perception.
- Section 4 delineates the intricate structure of the newly proposed aesthetic discriminator  $\mathcal{D}_{aes}$ .
- Section 5 presents a more comprehensive explanation regarding the implementation of the deception rate, along with an in-depth analysis of its implications and effectiveness.
- Section 6 provides an in-depth explanation of how the three user studies in the main paper are conducted.
- Section 7 presents an additional straightforward experiment to further validate the efficacy of the proposed aesthetic attention mechanism within USAesA and the novel Universal Aesthetic Codebook.
- Section 8 visualizes how the UAC retrieves universal aesthetic features to enhance global aesthetic attributes of style-specific aesthetic features.
- Section 9 provides more experimental results to further illustrate the versatility and robustness of our method across various content and styles.

## 2 IMPLEMENTATION DETAILS OF AESTHETIC ASSESSMENT MODULE

In Section 3 of the main paper, we provide an overview of our use of TANet [2] as the aesthetic assessment module within AesStyler. This section offers a more comprehensive explanation regarding the implementation details of this module.

TANet [2] is pre-trained on the extensive TAD66k [2] dataset, which is annotated with human-assessed aesthetic scores. This foundational training assures that TANet adeptly encapsulates aesthetic characteristics in alignment with human aesthetic predilections. To verify this, we present aesthetic score results predicted by TANet on some sample images from the TAD66k dataset in Fig. 1, demonstrating the proficiency of TANet in accurately forecasting these aesthetic scores.

In the training phase of AesStyler, we principally utilize the *Aesthetic Perceiving Network* (APNet) within TANet to extract aesthetic features. The architecture of APNet is primarily built upon MobileNetV2 [5], comprising 17 InvertedResidual blocks. These blocks are each composed of three convolution layers and an equal number of batch-normalization layers, employing ReLU as the activation function.

To be specific, following the idea of dividing the model into blocks, we extract feature maps from *Inverted Residual-57*, *Inverted*

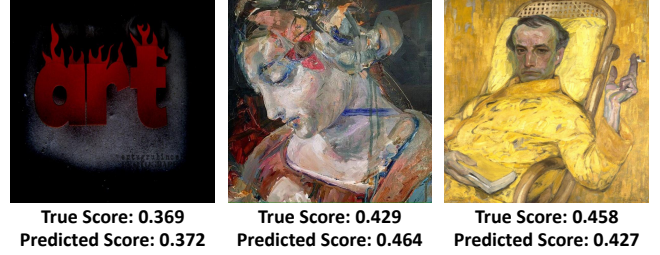


Figure 1: Images from TAD66k dataset with true scores and TANet predicted scores.

*Residual-93*, *Inverted Residual-120*, *Inverted Residual-147*, and *Inverted Residual-156* layers within TANet, designating them as aesthetic features. These feature maps are then utilized in the USAesA module to guide the style transfer process. The dimensions of these feature maps, represented as  $[channels, width, height]$ , are  $[32, 28, 28]$ ,  $[64, 14, 14]$ ,  $[96, 14, 14]$ ,  $[160, 7, 7]$ , and  $[320, 7, 7]$  respectively.

## 3 AESTHETIC ASSUMPTION BIAS

In Section 1 of the main paper, we explore the problem of Aesthetic Assumption Bias, a notable issue in previous aesthetic-aware style transfer methods, e.g. AesUST. This section is dedicated to presenting a more detailed analysis of this particular problem. To elucidate this problem in greater detail, we initially present the aesthetic score distribution of the WikiArt dataset [3] in Fig. 2, with the aesthetic scores evaluated by TANet [2].

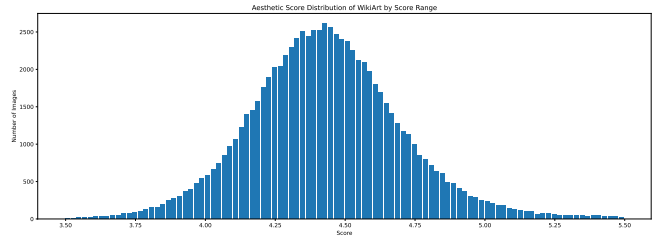


Figure 2: Aesthetic score distribution of WikiArt dataset.

From Fig. 2, it can be seen that the aesthetic score distribution of the WikiArt dataset resembles a Gaussian distribution, a reasonable observation given the extensive collection of images in the WikiArt dataset. Given this distribution, it appears illogical for AesUST [7] to assume that all images from the WikiArt dataset are highly aesthetic, because apparently some images have inferior aesthetics. This inherent bias in the aesthetic assumption ultimately results in the skewed extraction of aesthetic features, causing some of these so-called aesthetic features extracted by AesUST to focus on aspects unrelated to aesthetics.

To visually underscore the issue of aesthetic assumption bias, we also display some sample images from the WikiArt dataset that



Figure 3: Images from WikiArt dataset with high aesthetics and low aesthetics.

have high aesthetic scores alongside images with low aesthetic scores in Fig. 3, where the aesthetic scores are obtained by TAnet.

From Fig. 3, the first row showcases images from the WikiArt dataset with high aesthetic scores. These images are noticeably well-balanced in structure and harmonious in color, contributing to their aesthetic appeal. The second and third rows, however, exhibit overly simplistic content and styles, primarily characterized by pure colors

and minimal patterns, clearly lacking in aesthetic appeal, which is in accordance with their lower aesthetic scores. In contrast, the images in the last two rows are marked by overly complicated and cluttered patterns, and their colors are also excessively disordered, which diminishes their aesthetic appeal to humans and results in lower aesthetic scores.

**Table 1: Architecture of aesthetic discriminator.**

Part	Layer	In Channel	Out Channel	Kernel Size	Stride	Padding	Negative Slope
Disc	Conv	3	32	4	2	1	-
	InstanceNorm	32	32	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
	Conv	32	64	4	2	1	-
	InstanceNorm	64	64	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
	Conv	64	128	4	2	1	-
	InstanceNorm	128	128	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
	Conv	128	256	4	2	1	-
	InstanceNorm	256	256	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
	Conv	256	512	4	2	1	-
	InstanceNorm	512	512	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
	Conv	512	1024	4	2	1	-
	InstanceNorm	1024	1024	-	-	-	-
	LeakyReLU	-	-	-	-	-	0.2
Classifier	Conv	1024	1	4	1	0	-

## 4 ARCHITECTURE OF AESTHETIC DISCRIMINATOR

In Section 3 of the main paper, we provide a concise introduction to the newly proposed aesthetic discriminator  $\mathcal{D}_{aes}$ . This section provides a more detailed explanation regarding the specific structure of the aesthetic discriminator  $\mathcal{D}_{aes}$ . The detailed architecture of our newly proposed aesthetic discriminator  $\mathcal{D}_{aes}$  is shown in Table 1. The aesthetic discriminator plays the min-max game of discriminating between real artworks in WikiArt dataset [3] and style transfer results along with the generator to avoid the appearance of strange textures. We adapt the architecture from [1].

## 5 DETAILS OF DECEPTION RATE

In Section 4 of the main paper, we introduce the use of the deception rate to evaluate the quality of style transfer results. This section presents a more comprehensive explanation regarding the implementation of the deception rate, along with an in-depth analysis of its implications and effectiveness. Following [4], we train a VGG-16 network [6] using the WikiArt dataset [3] to classify artworks based on the annotated artist labels. We carefully curated the WikiArt dataset by selecting only those artist categories who created more than 30 artworks, thereby distilling the dataset down to 664 artist categories. We employ the Adam optimizer with an initial learning rate of 0.0001, and each batch comprises 64 images. We achieve a final accuracy of 0.4985 on the validation set of the WikiArt dataset.

In assessing style transfer results, the deception rate is calculated as the proportion of generated images that the network identifies as artworks of the artist who created the style image. Put simply, successful style transfer results deceive the model into classifying them as genuine works of the original artist, indicating that a higher deception rate corresponds to a greater resemblance of the style

transfer results to the actual artworks. This metric is calculated using a set of 18 style images, each representative of a different artist, and these style images are combined with 300 content images. This combination yields a total of 5,400 different style transfer results for thorough evaluation. Furthermore, we showcase sample instances of these successful deceptions, as produced by our AesStyler, to further demonstrate this concept in Fig. 4.

## 6 IMPLEMENTATION DETAILS OF USER STUDIES

In Section 4 of the main paper, we conduct three user studies focused on style transfer quality comparison, aesthetic comparison, and ablation studies. This section offers a more detailed explanation of the implementation aspects of these three user studies.

In each user study, the questionnaire comprises several ranking tasks. For each question, participants are presented with figures similar to those shown in Fig. 5 (note that the user study for aesthetic comparison does not display content and style images in the first row). The last two rows of each figure show six style transfer results produced by six distinct UST methods, with their order randomized. Participants are tasked with ranking the Top-3 images out of these six results. Scores are assigned as follows: 3 points for the first choice, 2 points for the second, 1 point for the third, and 0 for the remaining. We then calculate the average scores across all questions for each user study, thus deriving the final scores for the six different UST methods. As a result, a method achieves the full score of 3 only if the results of this particular method always rank the 1st in every question.

## 7 FURTHER ANALYSIS OF USAESA AND UAC

In Section 3 of the main paper, we introduce the innovative Universal and Style-specific Aesthetic-Guided Attention (USAesA) module





Figure 4: Examples of successful deceptions generated by our method.





Figure 5: Example of user study questions.

and the Universal Aesthetic Codebook (UAC) module. In Section 4 of the main paper, we have conducted thorough ablation studies to demonstrate their effectiveness. In this section, we will present an additional straightforward experiment to further validate the efficacy of these two modules.

In the **Phase 1 Universal Aesthetic Enhancement** of the USAesA module (main paper Fig. 4 left), we use the universal aesthetic feature  $F_u$  from UAC to enhance the global aesthetic attributes of the style-specific aesthetic feature  $F_a$ . This enhancement is tailored according to the channel distribution.

To validate the efficacy of both the aesthetic attention mechanism and the UAC, we applied the aesthetic enhancement (Phase 1 in USAesA) in the aesthetic assessment model TANet. Specifically, we randomly select 1,000 images from the WikiArt dataset and evaluate their aesthetic scores utilizing TANet. Notably, at the *Inverted Residual-147* layer in TANet, we deploy the aesthetic attention mechanism. This mechanism leverages the universal aesthetic feature  $F_u$  from the UAC, enhancing the style-specific aesthetic feature  $F_a$  as defined in Equations 3, 4, 5, and 6 in the main paper. The enhanced aesthetic features  $F_{ua}^x$  are subsequently processed through TANet. The results are presented in Table 2.

Table 2: Aesthetic scores of enhanced and original aesthetic features.

	Enhanced	Original
Aesthetic Score $\uparrow$	<b>0.4843</b>	0.4415

From the results presented in Table 2, it is evident that the newly proposed aesthetic enhancement mechanism (Phase 1) within the USAesA module, in conjunction with the novel Universal Aesthetic

Codebook (UAC), successfully enhances the aesthetic attributes of style images.

## 8 VISUALIZATION OF UNIVERSAL AESTHETIC CODEBOOK

In Section 3 of the main paper, we propose a novel Universal Aesthetic Codebook (UAC) and describe its role in guiding the style transfer process, particularly in terms of universal aesthetic attributes. In this section, we will provide a visual illustration to detail the function mechanism of the UAC in the style transfer process.

In the practical implementation of the UAC, it stores feature maps. However, for the purpose of visual illustration in this section, we will display the corresponding images of these feature maps. We exhibit several pairs of images in Fig. 6, where each pair consists of a style image and its corresponding universal aesthetic image retrieved from the UAC.

From the qualitative results presented in Fig. 6, it can be seen that the UAC effectively retrieves images with a similar style and higher (universal) aesthetics. Thus, we are able to employ the aesthetic feature retrieved from UAC, along with style-specific aesthetic features, to guide the model towards generating more universally appealing results.

## 9 MORE EXPERIMENT RESULTS

In Section 4 of the main paper, we present qualitative comparisons with existing state-of-the-art universal style transfer methods. In this section, to further illustrate the versatility and robustness of our method across various content and styles, we showcase stylization results from pair-wise combinations of 8 content images and 6 style images, as depicted in Fig. 7. These results demonstrate that our AesStyler method can consistently achieve aesthetically pleasing style transfer results across a diverse range of content-style pairings.

## REFERENCES

- [1] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. 2017. Improved training of wasserstein gans. *Advances in neural information processing systems* 30 (2017).
- [2] Shuai He, Yongchang Zhang, Rui Xie, Dongxiang Jiang, and Anlong Ming. 2022. Rethinking image aesthetics assessment: Models, datasets and benchmarks. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. 942–948.
- [3] Fred Phillips and Brandy Mackintosh. 2011. Wiki Art Gallery, Inc.: A Case for Critical Thinking. *Issues in Accounting Education* (Aug 2011), 593–608. <https://doi.org/10.2308/iace-50038>
- [4] Artsiom Sanakoyeu, Dmytro Kotovenko, Sabine Lang, and Bjorn Ommer. 2018. A style-aware content loss for real-time HD style transfer. In *proceedings of the European conference on computer vision (ECCV)*. 698–714.
- [5] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.
- [6] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations, International Conference on Learning Representations* (Jan 2015).
- [7] Zhizhong Wang, Zhanjie Zhang, Lei Zhao, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. 2022. AesUST: towards aesthetic-enhanced universal style transfer. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1095–1106.



Score: 0.3496



Score: 0.5493



Score: 0.3751



Score: 0.5425



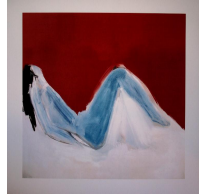
Score: 0.3702



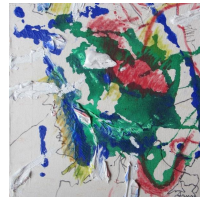
Score: 0.5540



Score: 0.3753



Score: 0.5581



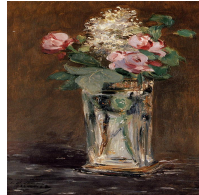
Score: 0.3771



Score: 0.5317



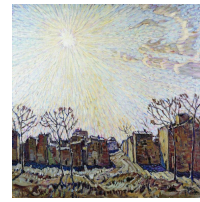
Score: 0.3663



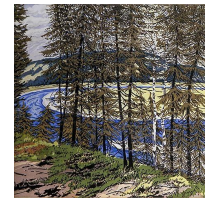
Score: 0.5266

**Style  
Image**

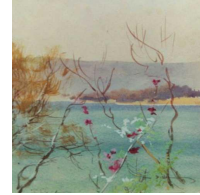
**Universal  
Aesthetic Image**



Score: 0.3792



Score: 0.5467



Score: 0.3741



Score: 0.5207



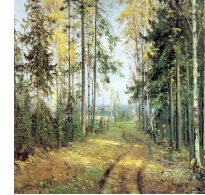
Score: 0.3617



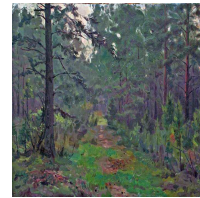
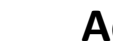
Score: 0.5399



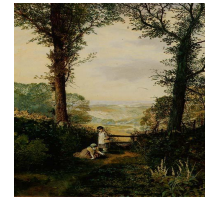
Score: 0.3593



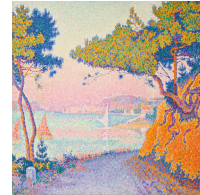
Score: 0.5468



Score: 0.3778



Score: 0.5253



Score: 0.3586



Score: 0.5388

**Style  
Image**

**Universal  
Aesthetic Image**

Figure 6: Visualization of the Universal Aesthetic Codebook.



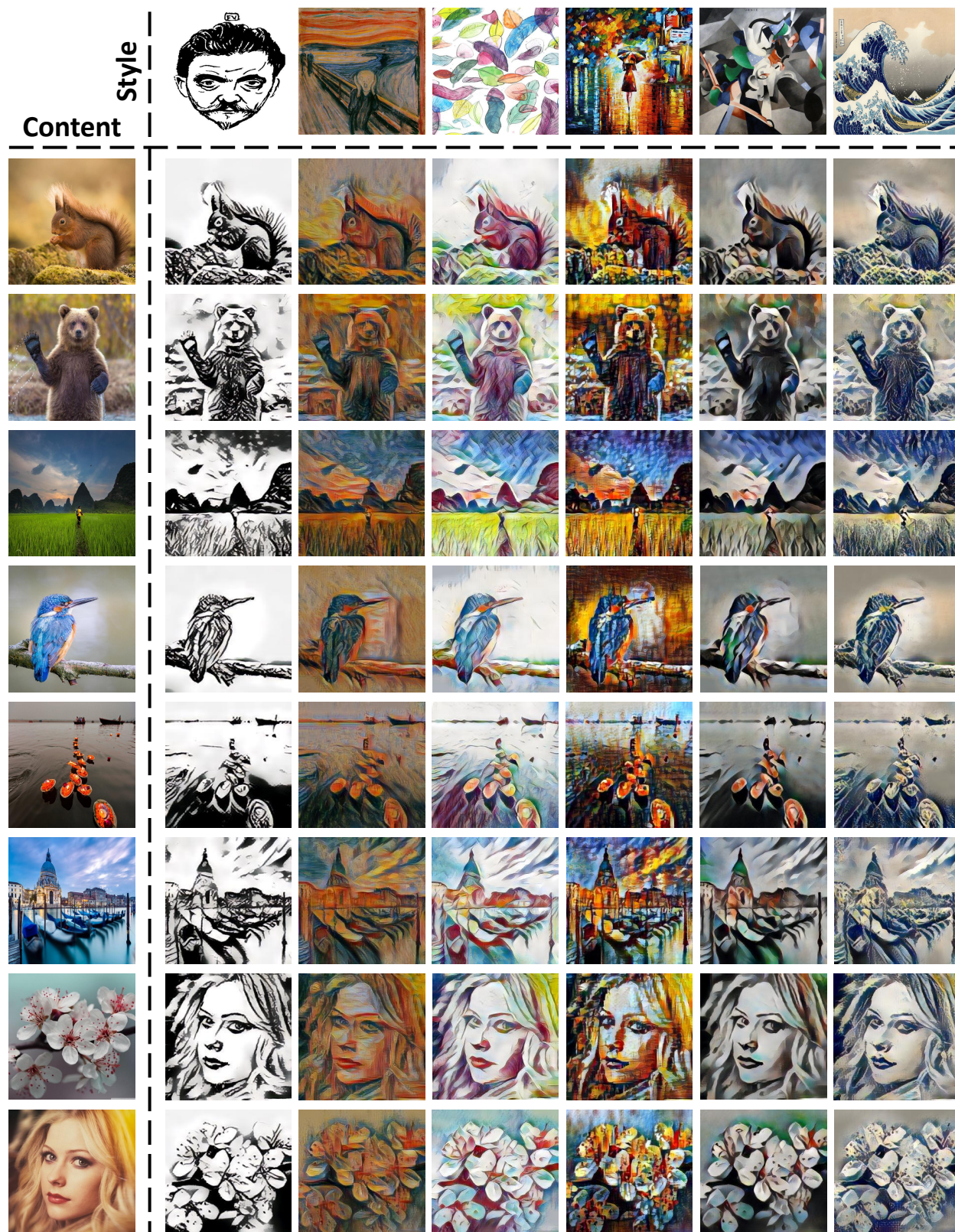


Figure 7: More image style transfer results of our AesStyler.