Style Transfer Dataset And Insights for Future Stylization Improvements

Mikhail Akhtyrchenko

Moscow state university

Course: My first scientific paper (Practice)/Group 417

Expert: V. V. Kitov

2024

Introduction

- Style transfer involves redrawing content images in the style of another image.
- Current methods lack standardization, making experiments irreproducible.
- A new public dataset with 10,000 stylizations and ratings is introduced to address these issues.
- Key goals:
  - Train models predicting stylization quality.
  - Recommend optimal styles for given content.
  - Analyze factors affecting stylization success.

Goal of research

Find objective features of a stylized image that affect subjective human perception of the image.

One-slide talk

Motivation

Explore drawbacks of current stylization models for future studies

- Manual data markup
- Make assumptions about the influence of various factors on the quality of stylization
- Figure out how to test these assumptions and verify them

Goal

Find dependencies between scores and some characteristics of stylizations and evaluate their influence on the final quality of the stylization

Problem statement

The concept of stylization does not have a clear constructive definition. There are several formulations of the stylization problem, each of which produces different results.

Therefore, it is proposed to create a marked dataset with subjective evaluations of stylizations created by the modern model and to analyze the image characteristics that influence human perception of stylization.

Each of them relies on some heuristic observation made on the basis of subjective human understanding of stylization

Hypothesis and quality criterions

Since each person's perception is subjective, we do not aim to obtain a dataset with highly correlated scores. We assume that there are image properties that are on average well correlated with human preferences, i.e., that do not depend on the preferences of a particular person.
Therefore, if we can identify some invariant properties, this will be the result of our work

Solution

Therefore, it is proposed to create a marked dataset with subjective evaluations of stylizations created by the modern model and to analyze the image characteristics that influence human perception of stylization.

Our work is divided into 2 phases:

- Dataset partitioning
- Analysis of stylization quality scores

Dataset description

$$Dataset = \{I_{(style, content, size)} \mid$$
$$(style, content, size) \in Contents \times Styles \times Sizes)\}$$
$$|Contents| = 50, \ |Styles| = 50,$$
$$Sizes = \{150, 300, 500, 700\}, \ labels = \overline{1, 10}$$

Scores consistency

	A1	A2	A3
A1	1.0	0.39459	0.346416
A2	0.394590	1.0	0.412650
A3	0.346416	0.41265	1.0

Table: Kendall tau-c rank correlation coefficient between scores of different annotators

	A1	A2	A3
A1	1.0	0.517117	0.461340
A2	0.517117	1.0	0.541933
A3	0.461340	0.541933	1.0

Table: Spearman rank correlation coefficient between scores of different annotators

Scores distribution

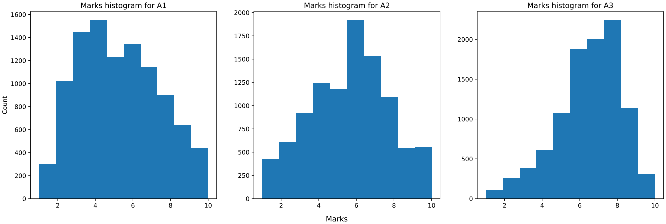

Figure: Scores distributions for different annotators

Key observations

After marking up, we came up with a few hypotheses.

- Images with low contrast are rated lower
- Images with low diversity of colours are rated lower
- Images with high sharpness are rated lower
- The appearance of borders on smooth surfaces after stylization reduces scores
- The more faces on content image are altered by stylization, the lower the score

Correlation tests for Color features (theory)

For colors we use CIELAB format:

$$I_{(content,style,size)} \in \mathbb{R}^{N\times M\times 3},\ L = (I_{ij1})_{i=1,j=1}^{N,M} \tag{1}$$

$$a = (I_{ij2})_{i=1,j=1}^{N,M},\ b = (I_{ij3})_{i=1,j=1}^{N,M}, \tag{2}$$

$$\tag{3}$$

$$L_{mean} = \mu_L, \tag{4}$$

$$C_{RMS} = s_L, \tag{5}$$

$$C_{webber} = \frac{\max(L)}{\max(\min(L), 1)} - 1, \tag{6}$$

$$CD_{ab\_rms} = \sqrt{s_a^2 + s_b^2} \tag{7}$$

$$\tag{8}$$

Correlation tests for Color features (results)

	Kendall tau-c			Spearman		
	A1	A2	A3	A1	A2	A3
L_{mean}	-0.015	-0.061	-0.049	-0.022	-0.090	-0.073
C_{RMS}	0.237	0.184	0.167	0.349	0.273	0.245
C_{webber}	0.074	0.031	-0.000	0.107	0.043	-0.001
CD_{ab_rms}	0.103	0.081	0.111	0.154	0.120	0.166

Table: Kendal tau-c and Spearman correlation for color metrics

Style Size Impact

- Smaller style sizes introduce noise and artifacts.
- Larger sizes provide better texture integration but may miss fine details.
- Optimal size range: 300^2–500^2 pixels.
- Examples:
  - High-resolution textures improve detail.
  - Oversized styles may reduce coherence.

Correlation tests for Face embeddings distance

The cosine distance between face embeddings on the content and on the stylised image was used to assess face similarity. Embeddings were obtained by applying the VGG-face model from the DeepFace library.

$$Sim = Cosine(emb_{content}, emb_{stylization})$$

	Kendall tau-c			Spearman		
	A1	A2	A3	A1	A2	A3
Sim	-0.373	-0.399	-0.45	-0.53	-0.565	-0.624

Table: Kendal tau-c and Spearman correlation between Sim and scores

Conclusion

- A standardized dataset for style transfer is introduced.
- Key factors for successful stylizations are identified.
- Recommendations for future algorithm development are provided.
- Future work:
  - Enhance attribute diversity in datasets.
  - Develop models addressing style and content mismatch.

Recommendations for Improvement

- Use style images with sufficient resolution and diverse textures.
- Avoid oversimplified or excessively complex textures.
- Ensure color diversity in style images.
- Implement adaptive algorithms for varying content and style sizes.

Future Work

- Expand dataset with additional styles and contents.
- Explore advanced models for adaptive style-texture integration.
- Address shortcomings in style size compatibility.
- Develop open-source tools for reproducibility in style transfer research.