Stylization-Based Architecture for Fast Deep Exemplar Colorization

ZHONGYOU XU, TINGTING WANG, FAMING FANG, YUN SHENG, GUIXU ZHANG

Stylization-Based Architecture for Fast Deep Exemplar Colorization

Zhongyou Xu[†], Tingting Wang[†], Faming Fang[†]*, Yun Sheng[§], Guixu Zhang[†]
[†]Shanghai Key Laboratory of Multidimensional Information Processing,
and the School of Computer Science and Technology, East China Normal University
[§]Liverpool John Moores University

Abstract

Exemplar-based colorization aims to add colors to a grayscale image guided by a content related reference image. Existing methods are either sensitive to the selection of reference images (content, position) or extremely time and resource consuming, which limits their practical application. To tackle these problems, we propose a deep exemplar colorization architecture inspired by the characteristics of stylization in feature extracting and blending. Our coarse-to-fine architecture consists of two parts: a fast transfer sub-net and a robust colorization sub-net. The transfer sub-net obtains a coarse chrominance map via matching basic feature statistics of the input pairs in a progressive way. The colorization sub-net refines the map to generate the final results. The proposed end-to-end network can jointly learn faithful colorization with a related reference and plausible color prediction with unrelated reference. Extensive experimental validation demonstrates that our approach outperforms the state-of-the-art methods in less time whether in exemplar-based colorization or image stylization tasks.

tion often plays an important role in the previous colorization works. Some papers [18, 39, 25] propose to manually add color scribbles over a grayscale image carefully, then propagate these known colors to the whole image, which is, however a challenging task for an untrained user who lacks professional skills and artistic sensitivity. Other researches [23, 3] try to alleviate such empirical problem by replacing the color scribbles with a closely related reference image. Credible correspondences are established between the target gray image and the reference image for propagating and coloring directly. The fragile model usually generates an inferior result for the dissimilarity caused by lighting, viewpoints and content differences. In recent years, deep learning-based colorization techniques [17, 2, 40] have achieved remarkable results. The colorization network is trained on a large number of image dataset to learn the relationship between grayscale image and its corresponding color version. Once the net parameters are determined, the colorization result can be easily obtained. Although such automatic colorization models make a huge success, the colorized result is sort of uncontrollable without any user intervention. More recent works [31, 41, 8] attempt to combine the controllability from interaction and robustness

Introduction

Colorization is a classic task in computer vision which aims at adding colors to a gray image.

1. Manually add color scribbles over a grayscale image, then propagate to the whole image
   - A challenging task for an untrained user who lacks professional skills and artistic sensitivity
2. Replacing the color scribbles with a closely related reference image
   - Lighting, viewpoints and content differences affect the result.
3. Automatic colorization by neural network
   - Uncontrollable without any user intervention
4. Combine the controllability from interaction and robustness from learning

Introduction

A novel deep learning approach for fast exemplar-based colorization

- The proposed architecture consists of two parts:
  - Transfer sub-net
  - Colorization sub-net
- Jointly learn faithful colorization with a related reference and plausible color prediction with an unrelated one.
- Better colorization in less time compared with other exemplar-based methods
- Extend the transfer sub-set to photorealistic image stylization without any additional modification.

Related Work

Colorization

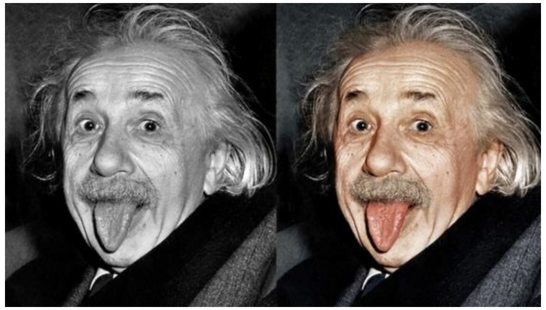
- Scribble-based colorization
  - Manually add color scribbles over a grayscale image, then propagate to the whole image
  - A challenging task for an untrained user who lacks professional skills and artistic sensitivity

- Exemplar-based colorization
  - Colorize a gray image guided by a reference one.

Related Work

Colorization

- Learning-based colorization
  - Automatic colorization by training an end-to-end network to reconstruct an image by predicting every pixel of the target image.
  - Only produce a single plausible result for each grayscale image.
  - Uncontrollable without any user interaction



- Hybrid colorization
  - Combine the controllability from interaction and robustness from learning to achieve more promising colorization.

Related Work

Photorealistic image stylization

- Photorealistic image stylization is evolved from traditional style transfer
  - The most discrepancy is that the output of the former should still maintain the original edge structure clearly.
- DSPT adds a regularization term to the loss function of the neural style algorithm to suppress distortions appearing in stylization results.
- Li *et al*. present a modified Whitening and Coloring Transforms (PhotoWCT) model that utilizes unpooling layers and additional smooth operations
  - Poor robustness and cause artifacts and blurriness
- We propose a transfer sub-net which can not only applied to image stylization but also generate an appropriate chrominance map for gray input.

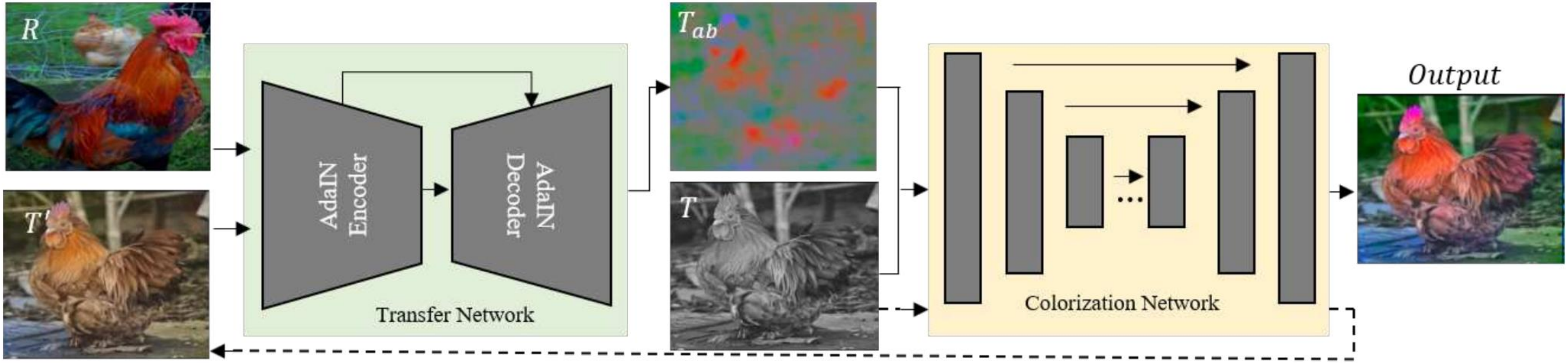


Figure 1. **System pipline.** Our model consists of two sub-nets: the transfer sub-net aims at generating coarse T_{ab} for target image that blends the color information of the reference; the colorization sub-net outputs the final result by refining T_{ab} and refers meaningful colors of the objects that are unrelated with the reference.

Stylization-based Colorization

Stylization-based Colorization

Transfer sub-net

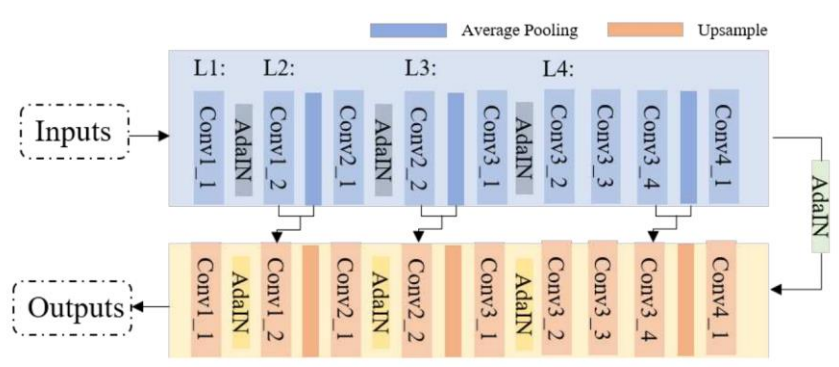


Figure 2. Detailed architecture of transfer sub-net.

- We adopt the pretrained VGG19 modules (from $conv1\_1$ layer to $conv4\_1$ layer) as the encoder and a symmetrical decoder for image reconstruction.
- To accelerate feature matching and blending, the fast Adaptive Instance Normalization (AdaIN) is utilized

$$AdaIN(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

- Concatenate the upsampled pooling feature and its previous convolution feature in each level of the encoder net for reconstruction in symmetrical module of the decoder by using a skip connection
- Use the original color VGG-19 by utilizing a pre-trained colorization network to give the gray target image pre-color.

Stylization-based Colorization

Colorization sub-net

- A colorization sub-net which takes luminance T along with chrominance T_{ab} as input.
  - No exact ground truth in exemplar-based colorization
  - We expect the sub-net can not only propagate "right" colors to "right" regions based on semantic features, but also refer meaningful colors of the objects when a reliable reference is unavailable.
- We take color images with randomly sampled ab channels as input and expect the network to learn their complete ab information.
- Adopts an analogous U-Net architecture which is formed by ten feature blocks and one output block.

$$L_c = L_h((1+\lambda M)\cdot F_c(x), (1+\lambda M)\cdot y)$$

  - Where L_h is Huber Loss function, F_c is colorization sub-net, x and y represent the input and output, M is the binary mask indicating the location of sampled ab channels

Experiments

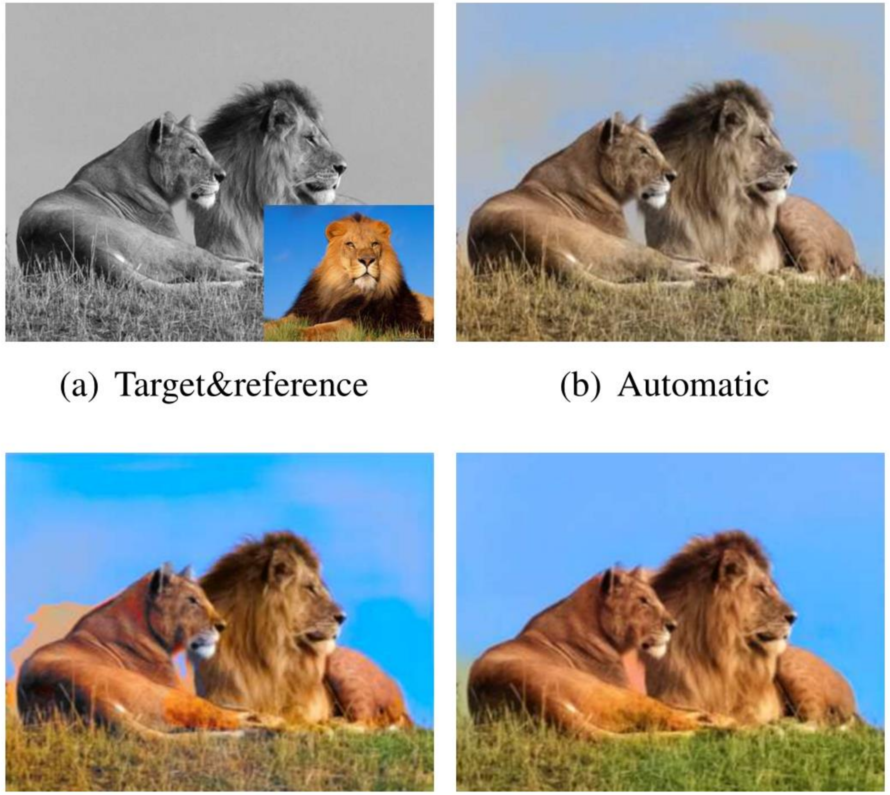
Implementation details

- The decoder of transfer sub-net is trained on the Microsoft COCO dataset
  - minimizing the sum of the L2 reconstruction loss with weight 0.8
  - perceptual loss with weight 0.2
  - the learning rate of which is initially set to 0.0001 and then decreased by a factor of 2 every 5 epochs
- The colorization sub-net is trained on the ImageNet dataset
  - weight coefficient is 10
  - ADAM optimizer
  - the learning rate of which is set to 0.00001 and then decreased by 10 after 10 epoch

Experiments

Ablation study

- First, we feed the target image with an empty ab map to the coloration sub-net and obtain the pre-colorized result in Fig. 4(b). Then, we feed the pre-colorized target image along with the reference to the transfer sub-net and show coarse colorization result in Fig. 4(c).



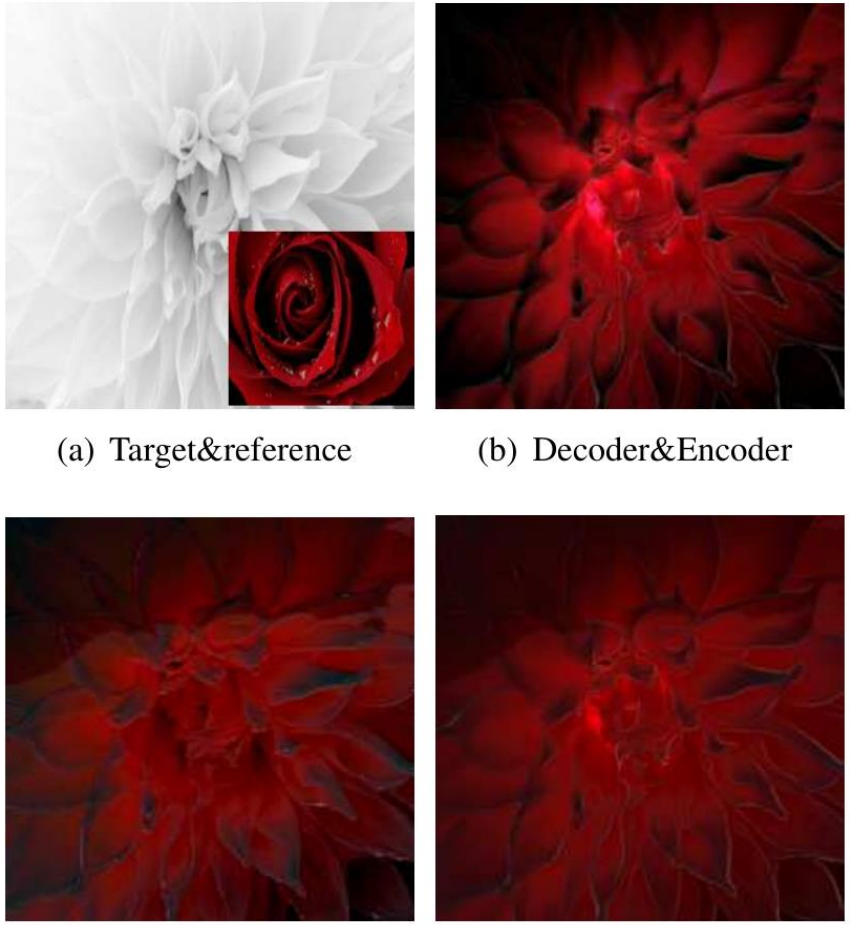
(a) Target&reference (b) Automatic

(c) Transfer (d) Transfer&Colorize

Figure 4. Ablation study about our model architecture.

Experiments

Ablation study



(a) Target&reference (b) Decoder&Encoder

(c) Decoder (d) Encoder

Figure 5. Results of AdaIN operations added in different parts.

Experiments

Comparison with colorization methods

Visual comparison

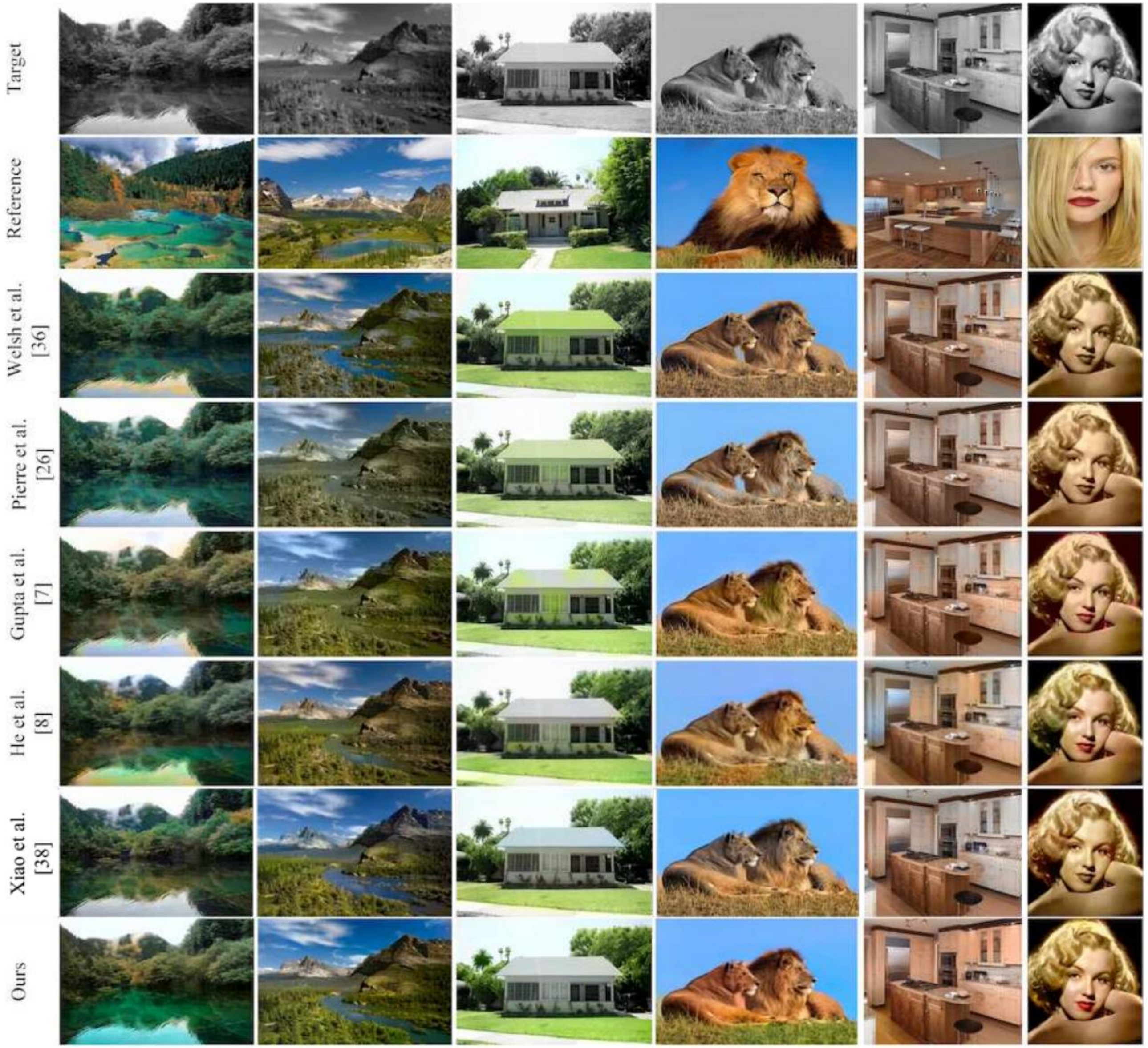
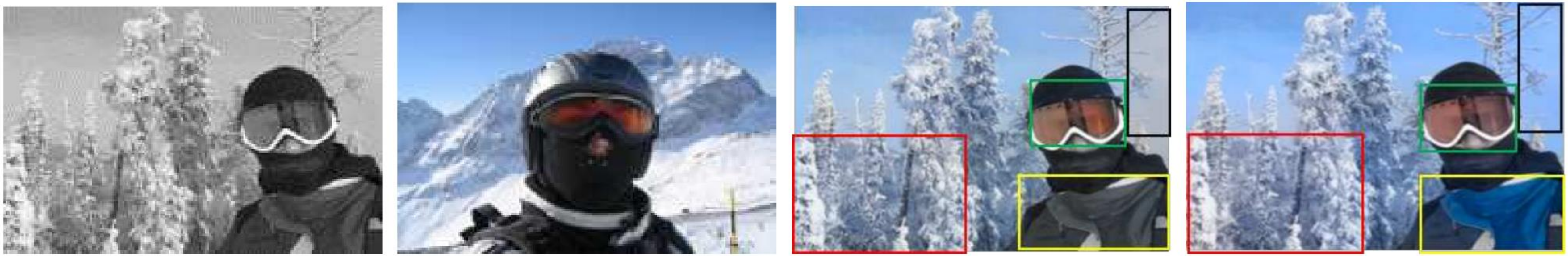


Figure 6. Comparison with example-based colorization methods. The first two rows are input target-reference image pairs. The last six rows are corresponding colorization results generated by [36, 26, 7, 8, 38] and the proposed method, respectively.



(a) Target (b) Reference (c) He *et al.* [8] (d) Ours

Figure 7. Detailed comparison between He *et al.* [8] and ours.

Experiments

Comparison with colorization methods

Visual comparison

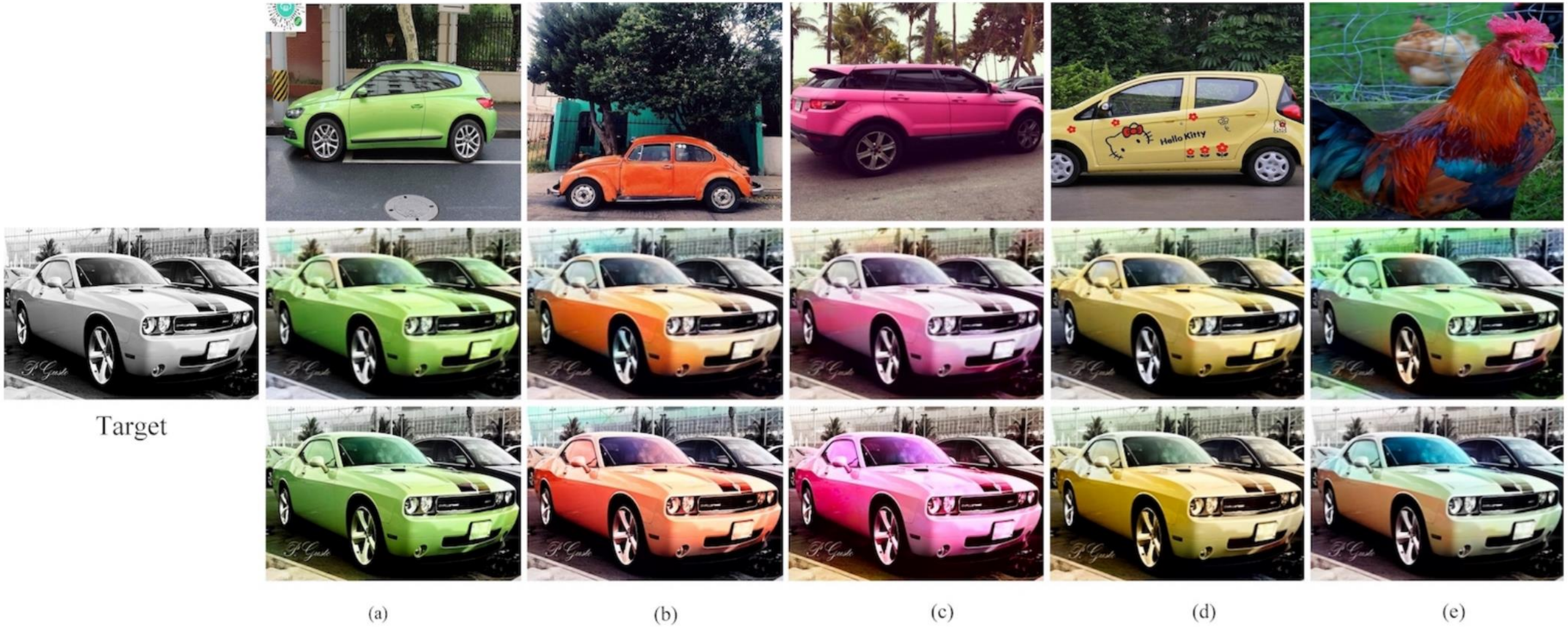


Figure 8. Results of colorization with different reference images. The first row shows five different references. The second and third row are their correspinding colorization results by He *et al.* [8] and the proposed method, respectively.

Experiments

Comparison with colorization methods

Visual comparison

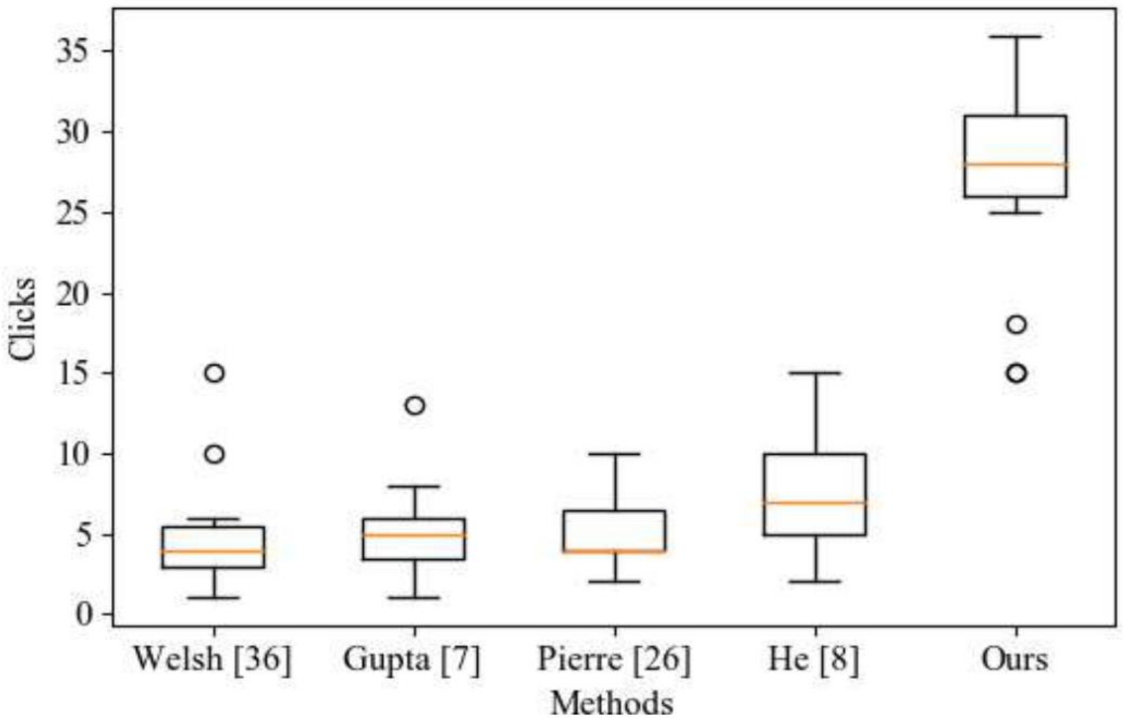


Figure 9. Boxplots of user preferences for different methods, showing the mean (yellow line), quartiles, and extremes (black lines) of the distributions.

Experiments

Comparison with colorization methods

User evaluation

Table 1. Runtime comparison on colorization tasks.

Image Size	256 x 256	512 x 512	1024 x 1024
Welsh *et al.* [36]	3.51	13.52	53.14
Pierre *et al.* [26]	3.69	14.12	66.38
Gupta *et al.* [7]	114.51	446.47	1784.58
He *et al.* [8]	7.25+1.14	33.69+1.30	49.96+OMM
Xiao *et al.* [38]	0.48	1.4	3.96
Ours	0.04+0.08x2	0.14+0.12x2	0.59+0.28x2

Experiments

Comparison with colorization methods

Runtime comparison

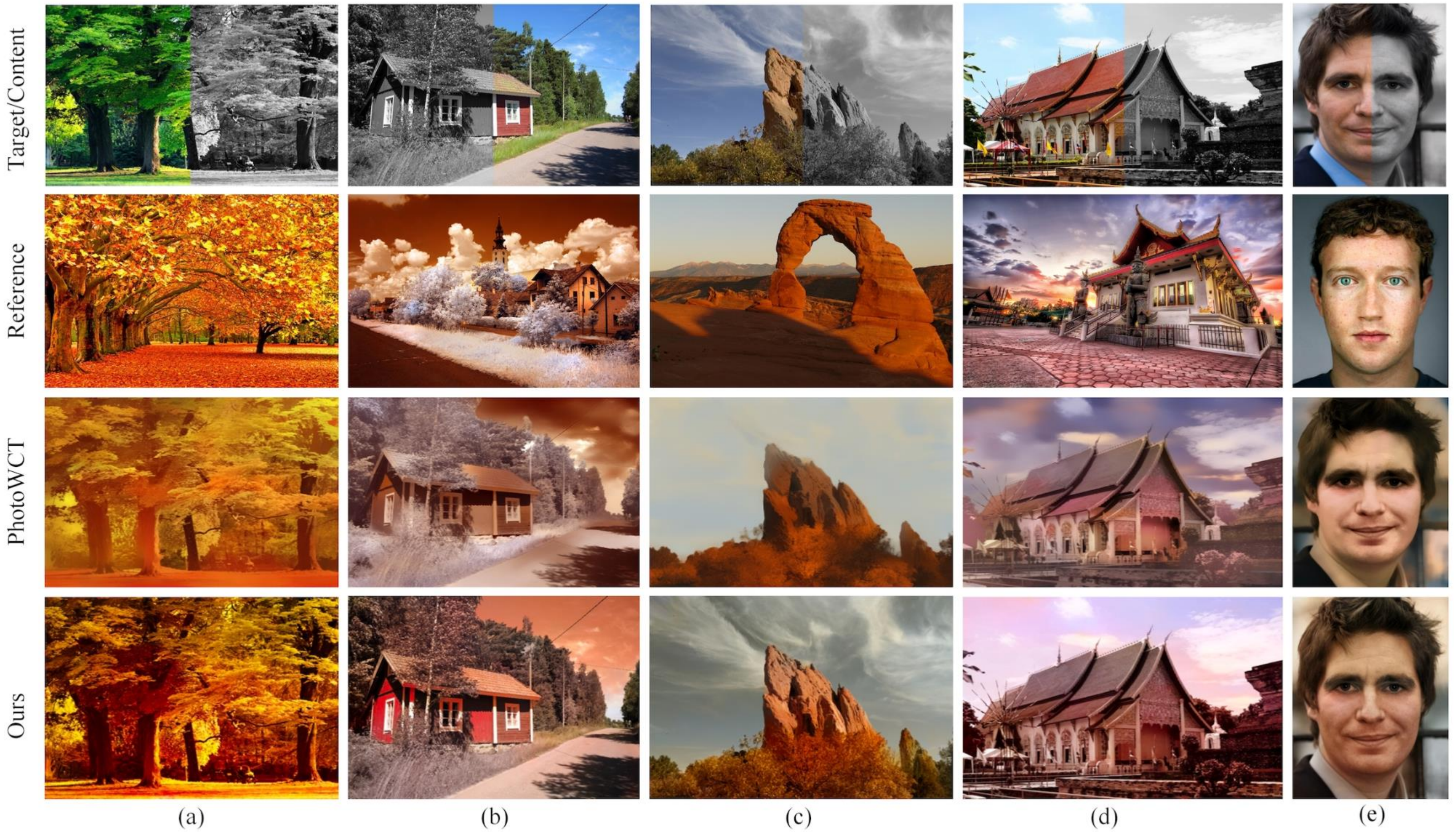


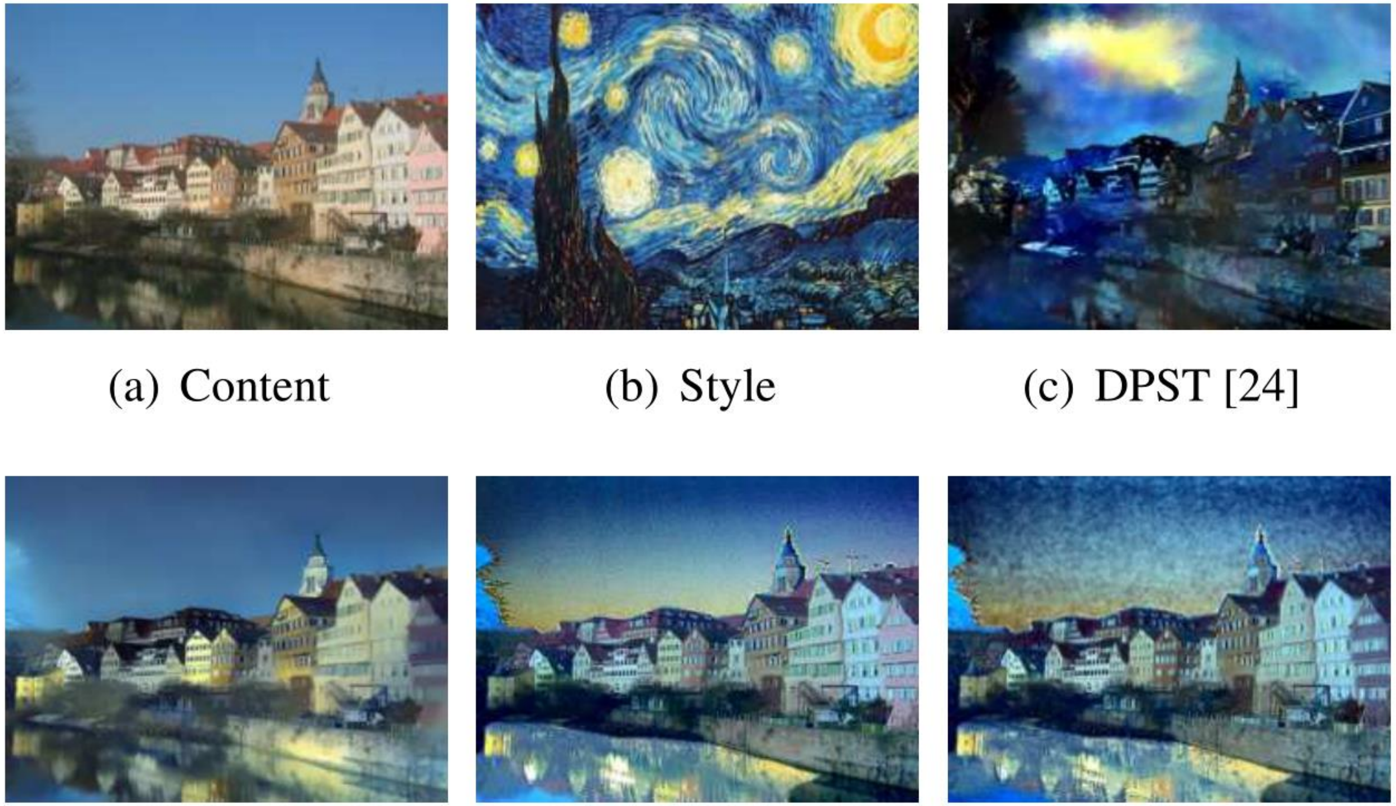
Figure 10. Comparison with PhotoWCT [20] on photo dataset used in photorealistic image stylization. The first two rows show the input image pairs and the last two rows show the result of the method, respectively.

Experiments

Comparison with Photorealistic Image Stylization Methods

Experiments

Comparison with Stylization Methods



(a) Content (b) Style (c) DPST [24]

(d) PhotoWCT [20] (e) Ours-AdaIN (f) Ours-WCT

Figure 11. Results of stylization methods [24, 20] and our transfer sub-net with WCT and AdaIN respectively.