Painterly Image Harmonization via Bi-Transformation with Dynamic Kernels

Zhangliang Sun Hui Zhang
Tsinghua University

BMVC 2024

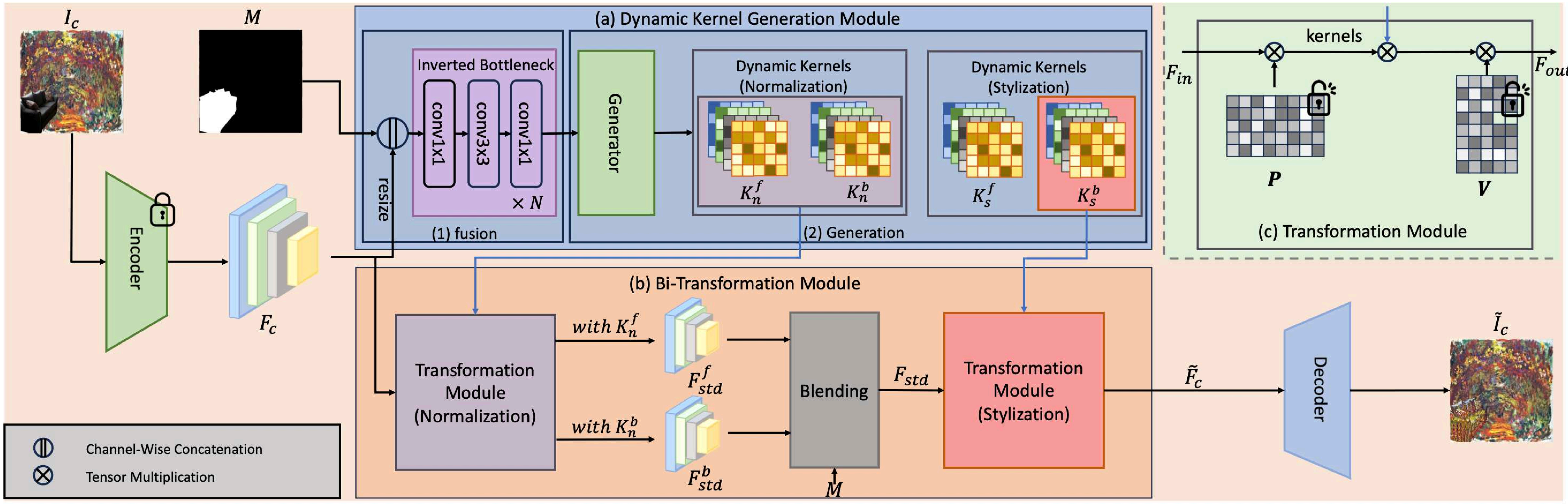
Motivation

- To achieve a more harmonized style fusion while retaining finer content details.
- To achieve an even more flexible alignment of foreground and background features which implemented with dynamic kernels in image harmonization tasks.

Conclusion

- DKTNet, the proposed model, introduces a learnable bidirectional transformation module to align content features with style features
- The model outperforms other methods in balancing style and content, as evidenced by both qualitative and quantitative comparisons

Method



1. Encoder-Decoder
2. Kernel Generator
3. Dynamic-Kernel Based Transformation

- We introduce a classifier-like generator to generate four types of dynamic kernels $\mathbf{K}$ to represent the relationships between different styles.
- A transformation module is designed to facilitate the migration between different feature maps.

$$F_{out} = \tau(F_{in}) = (F_{in} \cdot \mathbf{P} \cdot \mathbf{K_c} + \mathbf{K_b}) \cdot \mathbf{V}$$

- We assume that there is a specific standard style; the corresponding feature maps are named F_{std}.
- The specific standard feature maps can be obtained by combining feature maps that have been transformed with different dynamic kernels to 'standardize' the foreground and background parts.

$$F_{std}^f = \tau_n^f(F_c),\ F_{std}^b = \tau_n^b(F_c),\ F_{std} = F_{std}^f \circ M + F_{std}^b \circ (1 - M)$$

- The harmonized feature maps can be calculated by transferring the 'standard' feature maps using dynamic kernels $\tilde{F}_c = \tau_s(F_{std})$.
- Use the Encoder-Decoder Architecture to obtain the generated image $F_c = \mathcal{E}(I_c),\ \tilde{I}_c = \mathcal{D}(\tilde{F}_c)$.
- Utilize various loss functions to train our DKTNet.

$$\mathcal{L}_G = \mathcal{L}_s + \lambda_c \mathcal{L}_c + \lambda_{adv} \mathcal{L}_{adv} + \lambda_{cons} \mathcal{L}_{cons} + \mathcal{L}_{recons}$$

Result

	BT↑	FID↓	Params↓
Composite	-	28.469	-
PHDiffusion [24]	0.581	23.122	~900MB
PHDNet [2]	1.247	26.759	83MB
Ours	**1.379**	**11.000**	**34MB**

Table 1: Comparison results with other painterly image harmonization methods.

Version	BT↑
w/o $\mathcal{L}_{cons}$	0.669
w/o Bi-Trans	2.084
w/o KD	0.225
Full	**3.186**

Table 2: Results of ablation experiments for our model.

- Provides visual examples of harmonized images with more similar strokes and patterns to the background.
- User study demonstrates better preservation of content and style consistency compared to previous methods.
- The lower FID score indicates that our DKTNet is able to generate images that are more consistent with the style of the background.