COAT: Correspondence-driven Appearance Transfer

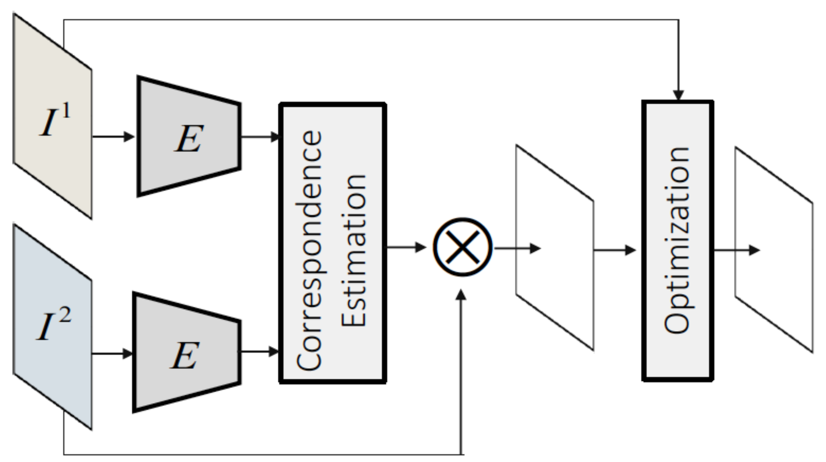
Sangryul Jeon[1], Zhifei Zhang[2], Zhe Lin[2], Scott Cohen[2], Zhihong Ding[2], and Kwanghoon Sohn[1]
[1]UC Berkeley, [2] Adobe Research, [3]Yonsei University

1. Introduction

Our goal : Re-rendering the structure of the original content image and transfer the arbitrary appearance of the exemplar style image

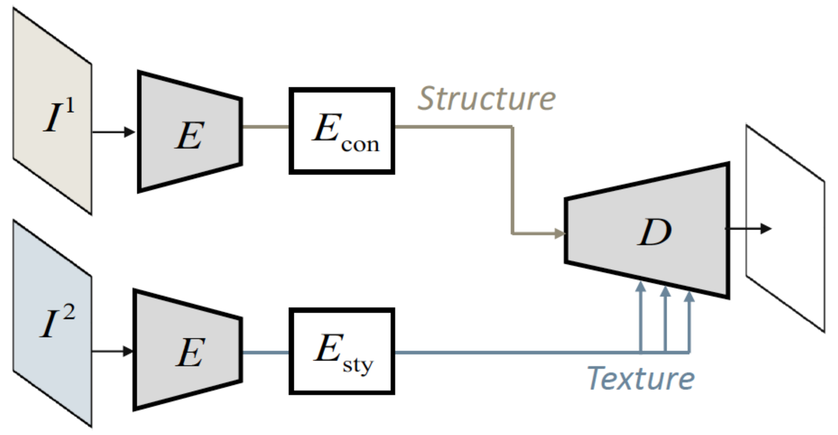
Content Style Transferred image

2. Previous Approaches



Based on *image warping*

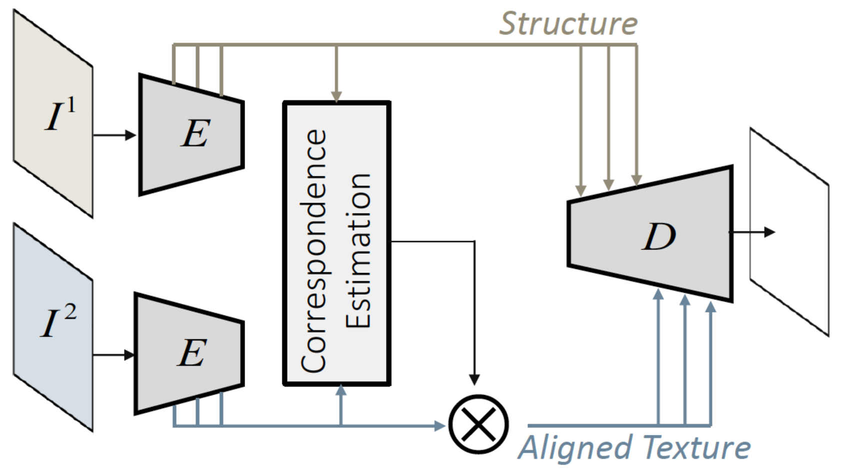
- First the content and style images are aligned by dense correspondences before generating the transferred image
- However, Its performance heavily relies on the quality of dense correspondence, often suffering from large appearance geometry variations



Based on *latent swapping*

- Disentangling the latent code into structure and texture components through autoencoder architecture
- However, the high level texture component mostly embeds the global style rather than capturing rich details from the style image.

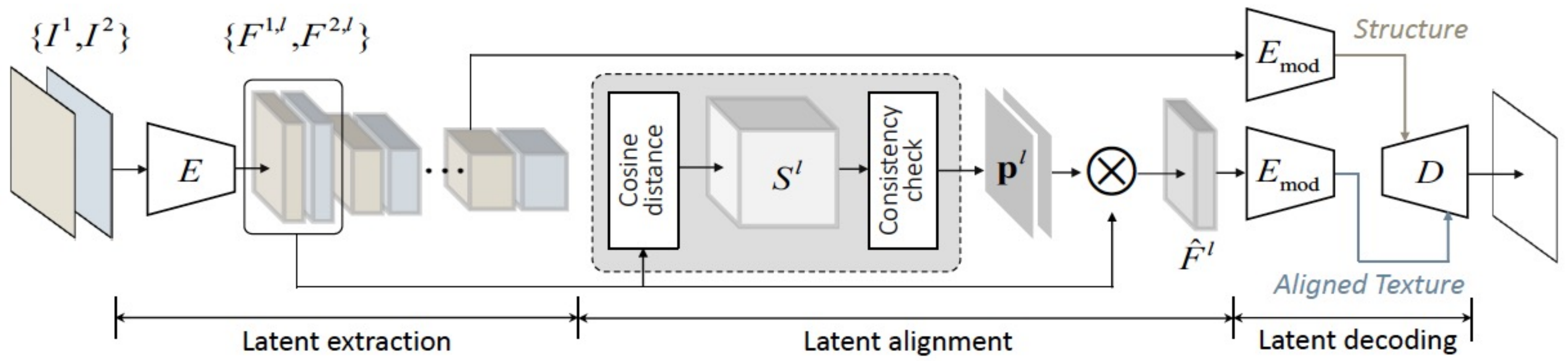
3. Proposed Method



Based on *multiscale latent alignment*

- Multi-scale features from the style image are aligned to obtain aligned texture features
- Providing more fine-grained textures without distorting the spatial structure from the content image

4. Network Architecture



Latent extraction

- Extracting the latents from the given image pair leveraging the hierarchy of the CNN-based encoder, obtaining multi-scale feature maps

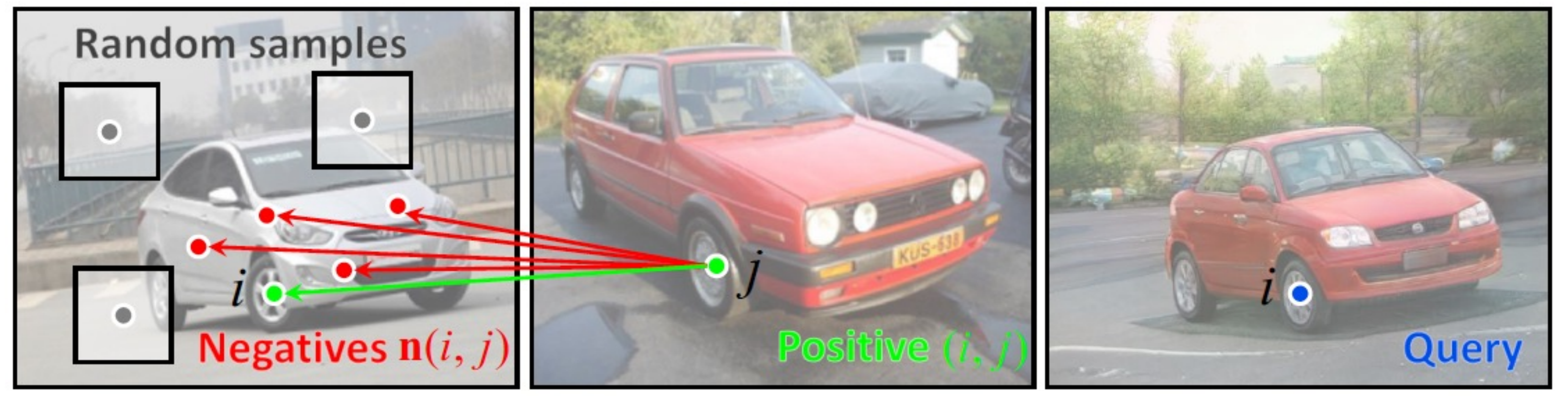
Latent alignment

- Spatially align the style feature maps to the content feature maps
- By measuring consistency of all pair-wise distances, only confident matches are used

Latent decoding

- StyleGAN-based architecture
- To preserve the structure from the content image, the content feature maps are fed to the first n modulation layers of the decoder
- To transfer high-fidelity textures from the style image, the rest layers are modulated by the aligned style feature maps.

5. Loss Functions



(a) Content image I^1 (b) Style image I^2 (c) Transfer result $I^{2\rightarrow 1}$

Correspondence contrastive loss

- Associates the patches that have a similar structure to each other, while disassociating them from other patches although with similar textures
- Positive and negative samples for a given query patch are determined based on the correspondences

Reconstruction loss

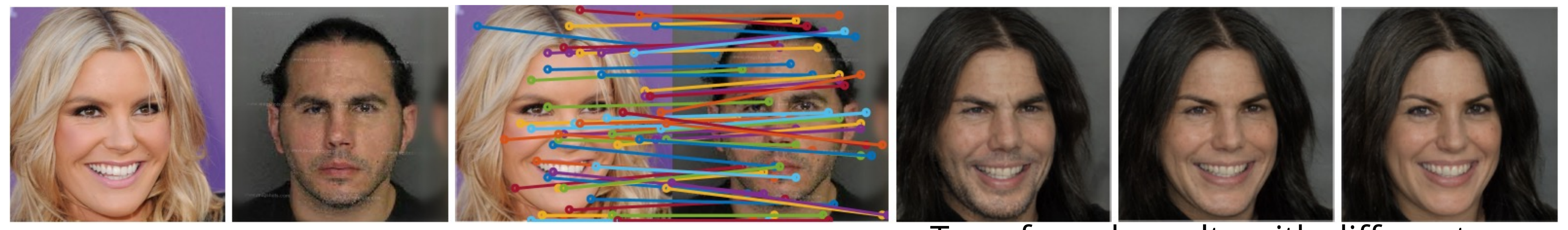
- To keep the consistency between the original image and predicted one

Regularization loss

- To keep the extracted latent vectors to be closer to the average latent space

6. Experiments

Ablation study



Content Style Correspondence Transferred results with different degree of appearance

Results on human face (CelebA) and animal face (AFHQ)

Content Style STROTSS DST StarGAN2 SAE Ours

Results on car (Stanford Car) and horse (LSUN-horse)

Content Style WCT STROTSS DST WST Ours