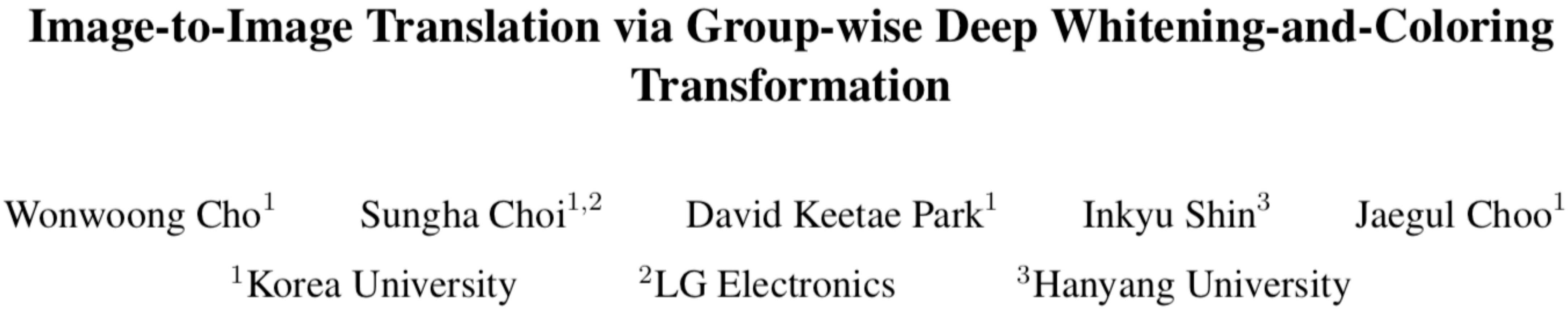


Image-to-Image Translation via Group-wise Deep Whitening-and-Coloring Transformation

Wonwoong Cho[1] Sungha Choi[1,2] David Keetae Park[1] Inkyu Shin[3] Jaegul Choo[1]

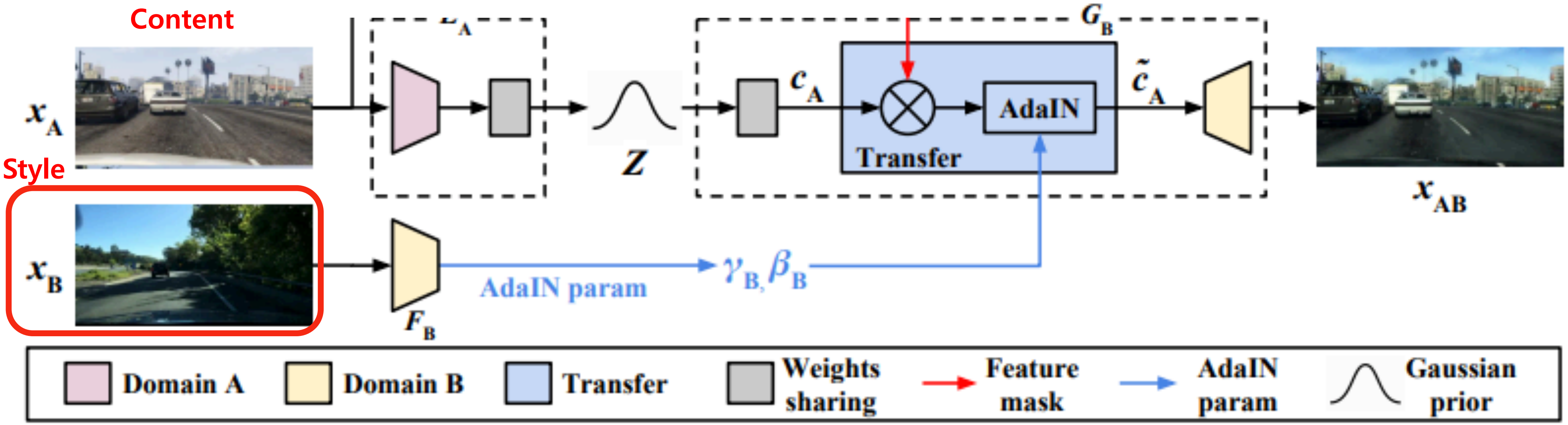
[1]Korea University [2]LG Electronics [3]Hanyang University

Presented by : Kangyeol Kim

DAVIAN Lab, Korea University

Background

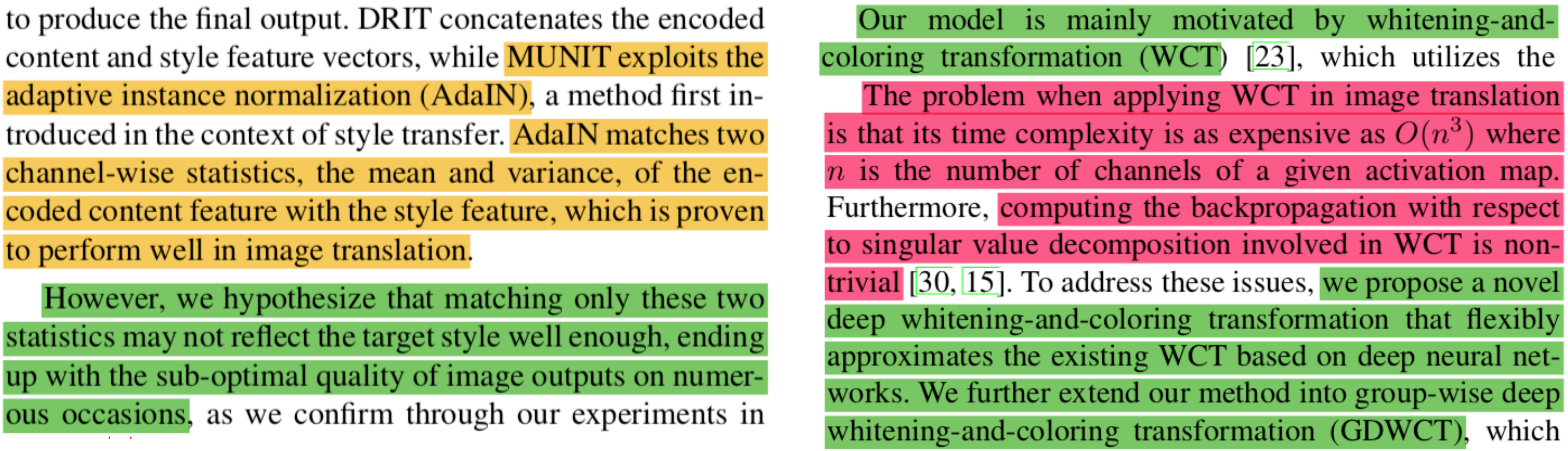
Toward Various outputs in image-to-image translation



- To tackle unimodal problem in image-to-image translation as CycleGAN, DRIT and MUNIT propose new architecture to generate **multi-modal outputs from a single input image**

- We can use a **random noise** from gaussian distribution or an **exemplar image** as depicted above.

Motivation



to produce the final output. DRIT concatenates the encoded content and style feature vectors, while MUNIT exploits the adaptive instance normalization (AdaIN), a method first introduced in the context of style transfer. AdaIN matches two channel-wise statistics, the mean and variance, of the encoded content feature with the style feature, which is proven to perform well in image translation.

However, we hypothesize that matching only these two statistics may not reflect the target style well enough, ending up with the sub-optimal quality of image outputs on numerous occasions, as we confirm through our experiments in

Our model is mainly motivated by whitening-and-coloring transformation (WCT) [23], which utilizes the

The problem when applying WCT in image translation is that its time complexity is as expensive as $O(n^3)$ where n is the number of channels of a given activation map. Furthermore, computing the backpropagation with respect to singular value decomposition involved in WCT is non-trivial [30, 15]. To address these issues, we propose a novel deep whitening-and-coloring transformation that flexibly approximates the existing WCT based on deep neural networks. We further extend our method into group-wise deep whitening-and-coloring transformation (GDWCT), which

Proposed methods (1/6)

Overall architecture

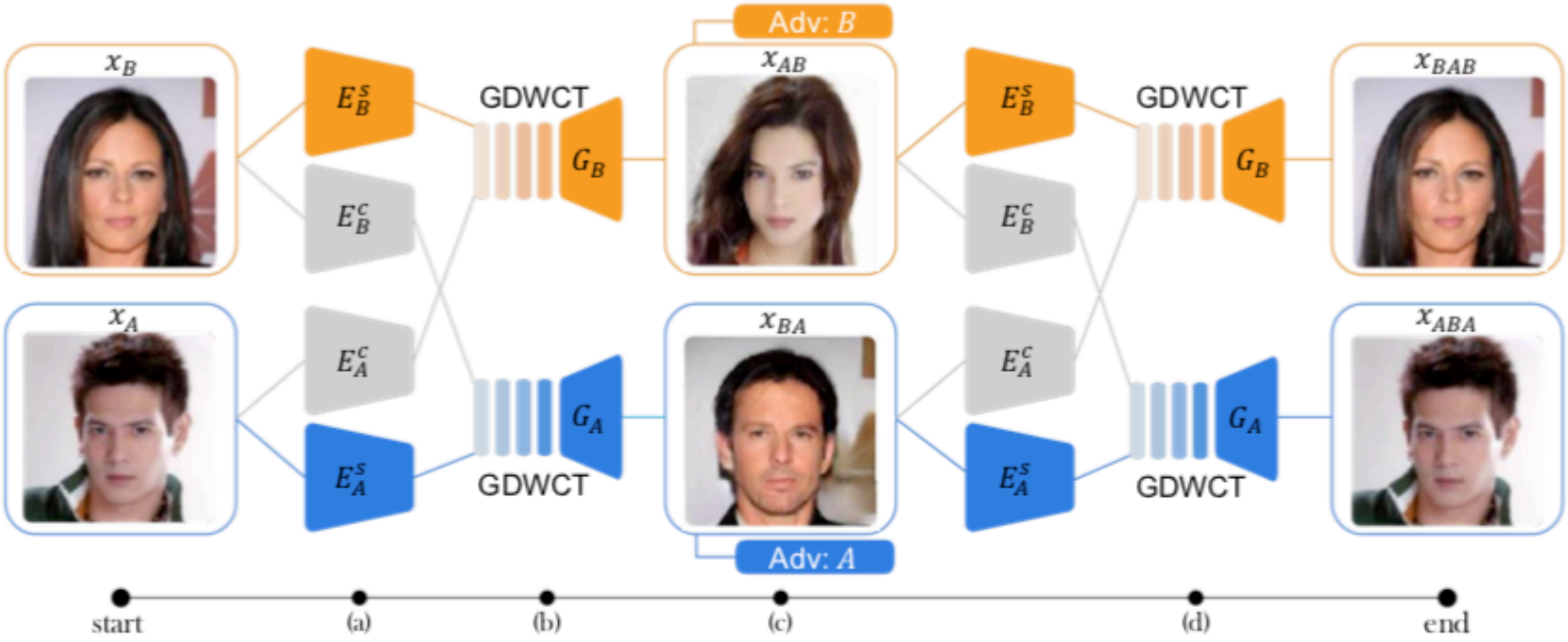


Figure 1: Overview of our model. (a) To translate from $\mathcal{A} \rightarrow \mathcal{B}$, we first extract the content feature $c_{\mathcal{A}}$ from the image $x_{\mathcal{A}}$ (i.e., $c_{\mathcal{A}} = E^c_{\mathcal{A}}(x_{\mathcal{A}})$) and the style feature $s_{\mathcal{B}}$ from the image $x_{\mathcal{B}}$ (i.e., $s_{\mathcal{B}} = E^s_{\mathcal{B}}(x_{\mathcal{B}})$). (b) The obtained features are combined in our GDWCT module while forwarded through the generator $G_{\mathcal{B}}$. (c) The discriminator $D_{\mathcal{B}}$ classifies whether the input $x_{\mathcal{AB}}$ is a real image of the domain $\mathcal{B}$ or not. (d) Similar to the procedures from (a) to (c), the generator $G_{\mathcal{B}}$ generates the reconstructed image $x_{\mathcal{BAB}}$ by combining the content feature $c_{\mathcal{BA}}$ and the style feature $s_{\mathcal{AB}}$.

Proposed methods (2/6)

Losses

- **Style consistency loss** – Styles of both style image and translated image are similar

$$\mathcal{L}_s^{\mathcal{A}\to\mathcal{B}} = \mathbb{E}_{x_{\mathcal{A}\to\mathcal{B}},x_{\mathcal{B}}}\left[\|E_{\mathcal{B}}^s(x_{\mathcal{A}\to\mathcal{B}}) - E_{\mathcal{B}}^s(x_{\mathcal{B}})\|_1\right]$$

- **Content consistency loss** – To maintain the content feature of the input image

$$\mathcal{L}_c^{\mathcal{A}\to\mathcal{B}} = \mathbb{E}_{x_{\mathcal{A}\to\mathcal{B}},x_{\mathcal{A}}}\left[\|E_{\mathcal{B}}^c(x_{\mathcal{A}\to\mathcal{B}}) - E_{\mathcal{A}}^c(x_{\mathcal{A}})\|_1\right]$$

- **Cycle consistency loss** and **identity loss** – To obtain high-quality images

$$\begin{aligned}\mathcal{L}_{cyc}^{\mathcal{A}\to\mathcal{B}\to\mathcal{A}} &= \mathbb{E}_{x_{\mathcal{A}}}\left[\|x_{\mathcal{A}\to\mathcal{B}\to\mathcal{A}} - x_{\mathcal{A}}\|_1\right] \\ \mathcal{L}_i^{\mathcal{A}\to\mathcal{A}} &= \mathbb{E}_{x_{\mathcal{A}}}\left[\|x_{\mathcal{A}\to\mathcal{A}} - x_{\mathcal{A}}\|_1\right].\end{aligned}$$

- **Adversarial Loss** – To minimize the discrepancy between dist(real) and dist(generated)

$$\begin{aligned}\mathcal{L}_{D_{adv}}^{\mathcal{B}} &= \tfrac{1}{2}\,\mathbb{E}_{x_{\mathcal{B}}}[(D(x_{\mathcal{B}}) - 1)^2] + \frac{1}{2}\,\mathbb{E}_{x_{\mathcal{A}\to\mathcal{B}}}[(D(x_{\mathcal{A}\to\mathcal{B}}))^2] \\ \mathcal{L}_{G_{adv}}^{\mathcal{B}} &= \tfrac{1}{2}\,\mathbb{E}_{x_{\mathcal{A}\to\mathcal{B}}}[(D(x_{\mathcal{A}\to\mathcal{B}}) - 1)^2]\end{aligned}$$

Proposed methods (3/6)

Whitening transform

Recall whitening procedure is to transform content feature map f_c to $\widehat{f_c}$ where $\Sigma_c = f_c f_c^T$, $\widehat{f_c}\ \widehat{f_c^T} = I$. However, eigendecomposition is not only computationally intensive but also difficult to backpropagate the gradient signal

The authors propose deep whitening transformation(DWT):

$$R_w = E\left[||\Sigma_c\ - I||\right]$$

Then,

$$\widehat{f_c} = f_c - \mu_c$$

However, performing DWT to entire channels may excessively throw away the content feature. Thus, the authors propose group-DWT:

$$f_{ci} \in R^{G\times\left(\frac{C}{G}\right)\times BHW} \Rightarrow \Sigma_{ci} \in \mathrm{R}^{\mathrm{G}\times\left(\frac{\mathrm{C}}{\mathrm{G}}\right)\times\left(\frac{C}{G}\right)}$$

Proposed methods (4/6)

Coloring transform

Recall coloring procedure is to transform whitend feature map $\widehat{f_c}$ to $\widehat{f_{cs}}$ where $\widehat{f_{cs}}\,\widehat{f_{cs}^T} = f_s f_s^T, \widehat{f_{cs}} = E_s D_s^{\frac{1}{2}} E_s^T \widehat{f_c}$. However, considering same reasons as whitening transform, The authors propose deep coloring transformation(DCT):

$$MLP^{CT}(f_s) = S = UD = E_s D_s^{\frac{1}{2}}$$

where $U \in R^{C \times C}$,orthonormal matrix, $D \in R^{C \times C}$ is diagonal matrix whose diagonal entries correspond to the L_2 norm of each column vector of S. Additionally, add regularization term to ensure orthogonality of U.

$$R_c = E\left[||U^T U \ - I||\right]$$

Then,

$$\widehat{f_{cs}} \ = UDU^T \widehat{f_c}$$

Due to expensive computational cost, the authors propose group-DCT:

$$Compute\ set\ of\ \{UDU^T\}_i \in R^{\left(\frac{C}{G}\right) \times \left(\frac{C}{G}\right)}$$

Proposed methods (5/6)

GDWCT operations

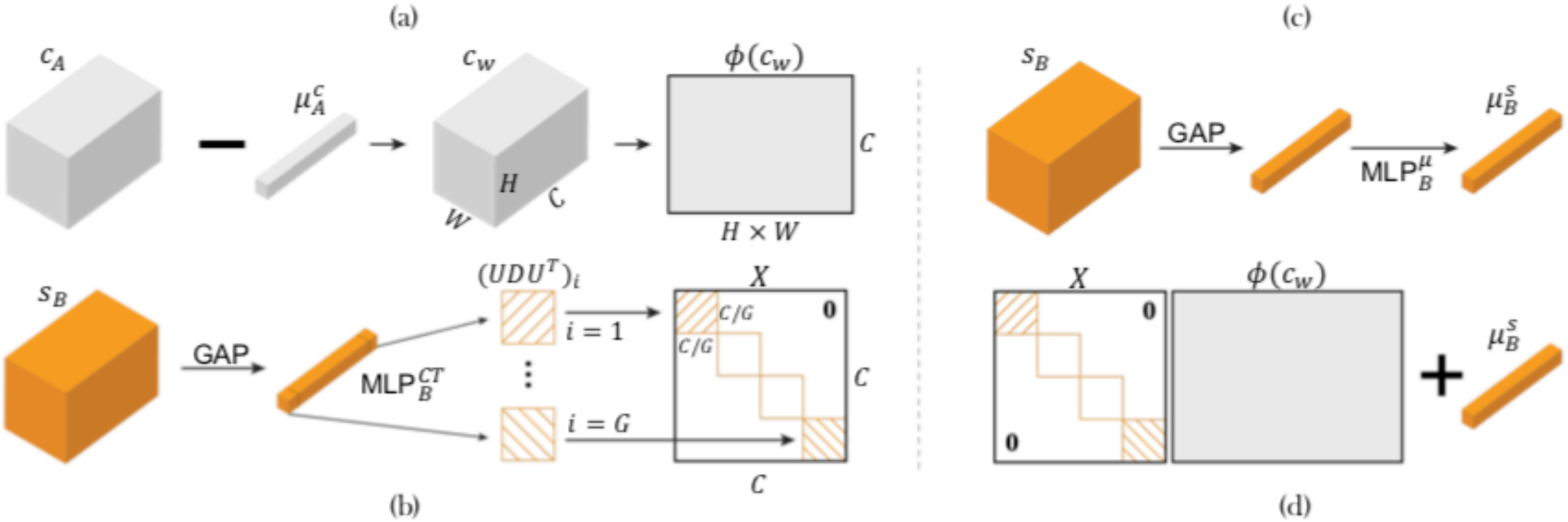


Figure 3: Details on the proposed GDWCT module. (a) The process for obtaining the whitened feature. Because the regularization term (Eq. (2)) encourages the zero-mean content feature $\bar{c}$ to be the whitened feature c_w, we just subtract the mean of the content feature μ_A^c from c_A. (b) The procedure of approximating the coloring transformation matrix (Section 3.3). (c) We obtain the mean of the style feature μ_B^s by forwarding it to the MLP layer MLP_B^{μ}. (d) Our module first multiply the whitened feature c_w with the group-wise coloring transformation matrix GDCT. We then add it with the mean of the style μ_B^s.

Proposed methods (6/6)

GDWCT overview

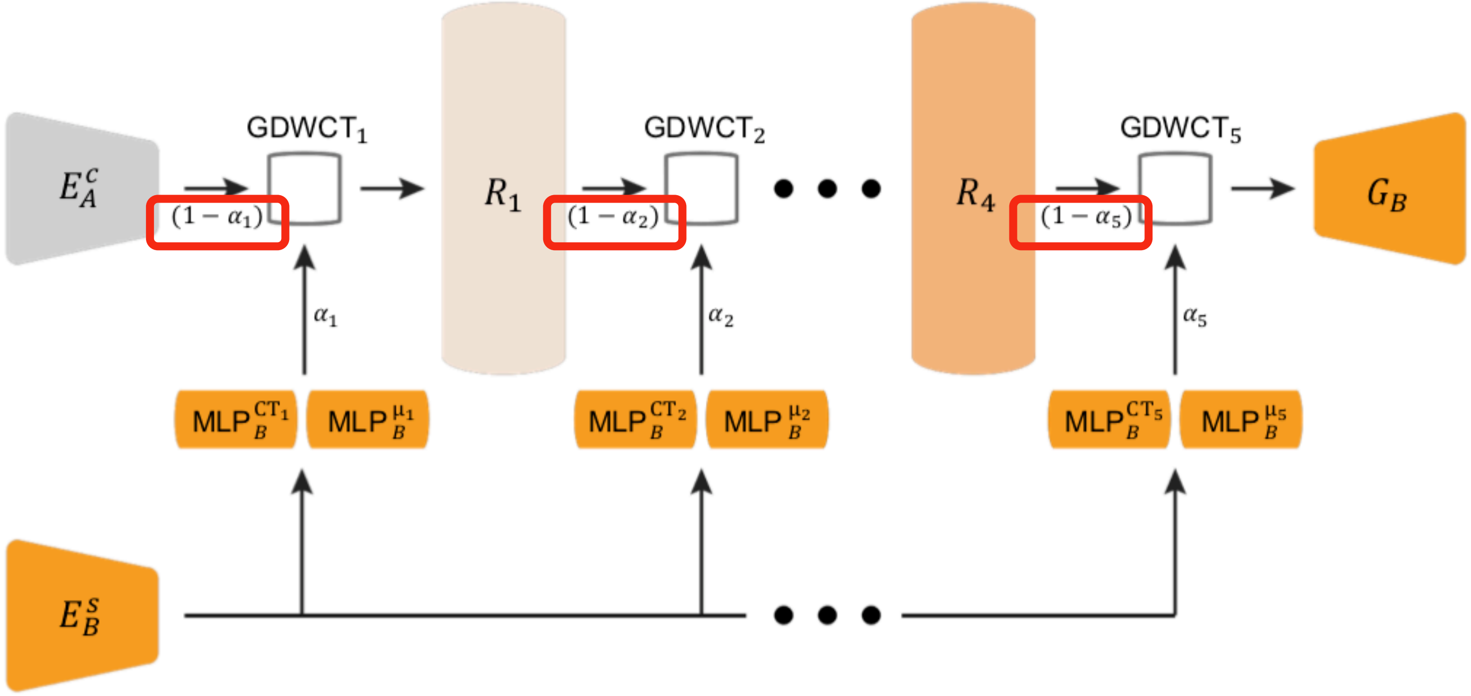
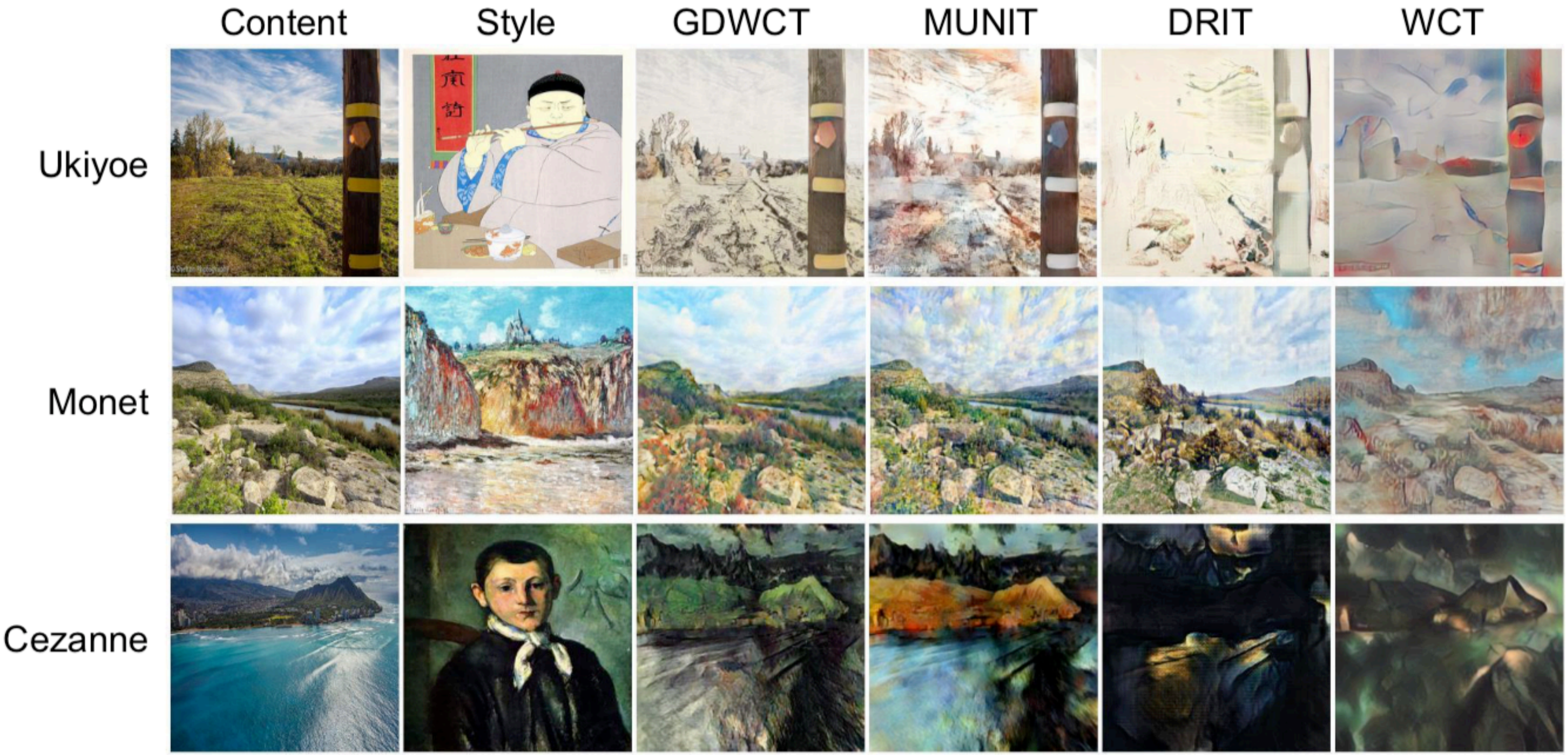


Figure 2: Image translation via the proposed GDWCT. We apply the style via multiple hops to apply the style from the low-level feature to the high-level feature.

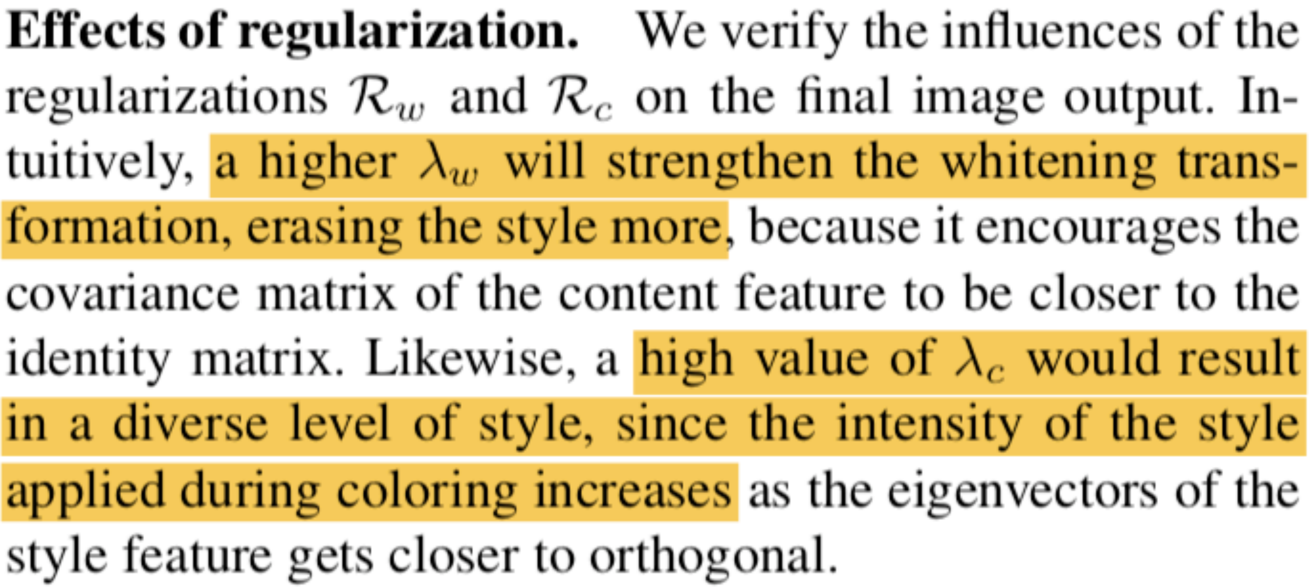
Experiments (1/2)

Style transfer.



Experiments (2/2)

Regularizations effects



Effects of regularization. We verify the influences of the regularizations $\mathcal{R}_w$ and $\mathcal{R}_c$ on the final image output. Intuitively, a higher λ_w will strengthen the whitening transformation, erasing the style more, because it encourages the covariance matrix of the content feature to be closer to the identity matrix. Likewise, a high value of λ_c would result in a diverse level of style, since the intensity of the style applied during coloring increases as the eigenvectors of the style feature gets closer to orthogonal.

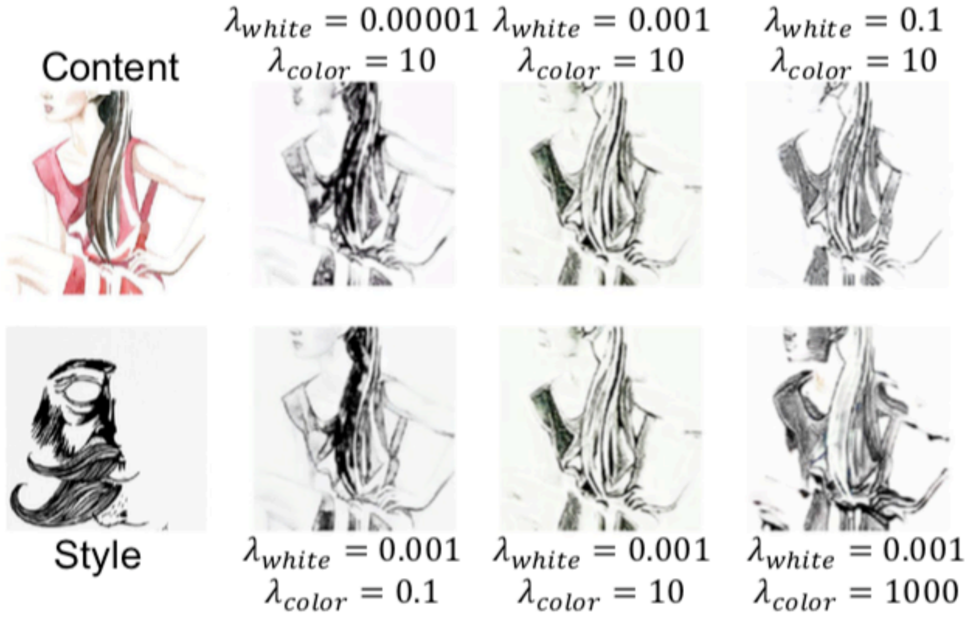

Content
$\lambda_{white} = 0.00001$ $\lambda_{color} = 10$
$\lambda_{white} = 0.001$ $\lambda_{color} = 10$
$\lambda_{white} = 0.1$ $\lambda_{color} = 10$
Style
$\lambda_{white} = 0.001$ $\lambda_{color} = 0.1$
$\lambda_{white} = 0.001$ $\lambda_{color} = 10$
$\lambda_{white} = 0.001$ $\lambda_{color} = 1000$

Figure 6: Visualization of the regularization influences.

Thank You