Line Search-Based Feature Transformation for Fast, Stable, and Tunable Content-Style Control in Photorealistic Style Transfer

Tai-Yin Chiu
University of Texas at Austin
chiu.taiyin@utexas.edu

Danna Gurari
University of Colorado Boulder
Danna.Gurari@colorado.edu

Supplementary Materials

This document supplements the main paper with the following.

1. Experiments showing insufficiency of LST and DSTN for photorealistic style transfer (supplements Section 2 of the main paper).
2. Insufficiency of linear interpolation for content-style control (supplements Section 3 of the main paper).
3. Experiments which show that removing centralization and decentralization from AdaIN and ZCA leads to worse image quality (supplements Section 3.2 of the main paper).
4. Explanation for how centralization and decentralization support mean vector matching (supplements Section 3.2 of the main paper).
5. Derivation of Equation 7 in the main paper.
6. Proof of at least one positive solution to Equation 7 in the main paper.
7. Computation of the values of η searched by LS-FT (supplements Section 3.3 of the main paper).
8. Qualitative results demonstrating the effect of the content-style control knob α (supplements Section 3.3 of the main paper).
9. Convergence comparison between Modified IterFT and LS-FT on WCT^2, PhotoWCT, and PCA-d (supplements Section 4.1 of the main paper).
10. Qualitative results showing that our approaches fix the unreasonable results from IterFT (supplements Section 4.2 of the main paper).
11. Speed of transformations on PCA-d (supplements Section 4.3 of the main paper).

Insufficiency of LST for photorealistic style transfer

The prior work of LST [7] and DSTN [5] claim their autoencoder-based models can be used for photorealistic style transfer. However, they did not provide strong quantitative analysis to support the claim. Here we show that they are insufficient for photorealistic style transfer with the quantitative and qualitative evidence showing they preserve content poorly.

First, we notice from Fig. 1 that LST results in almost as bad content preservation as PhotoWCT with ZCA as the feature transformation and DSTN has an even worse content loss. Qualitatively, as exemplified in Fig. 2, compared to the results from WCT^2 [12], PhotoWCT [9], PhotoWCT^2 [2], and PCA-d [3] with our LS-FT as the feature transformation, the results from LST are prone to blurred boundaries (low sharpness) and dullness (low contrast) and the results from DSTN have many severe artifacts.

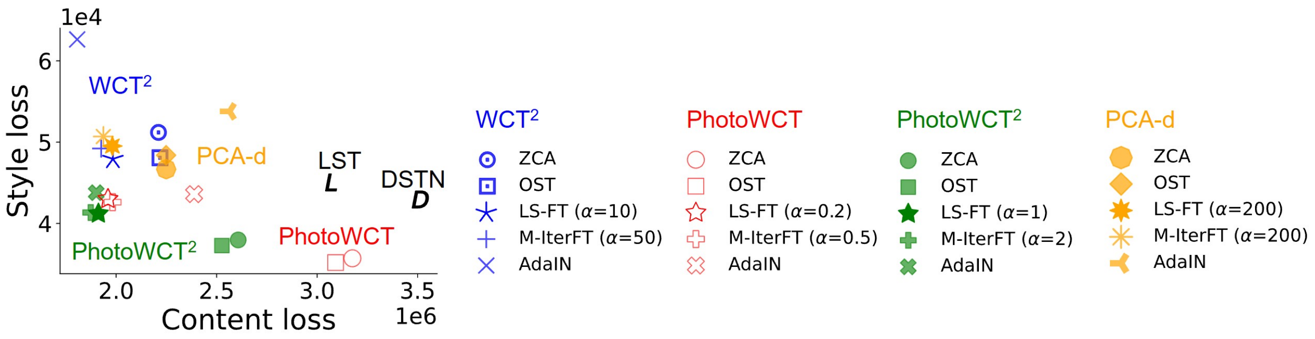

Figure 1: LST and DSTN do not preserve content and photorealism well compared to most transformation-model pairs.

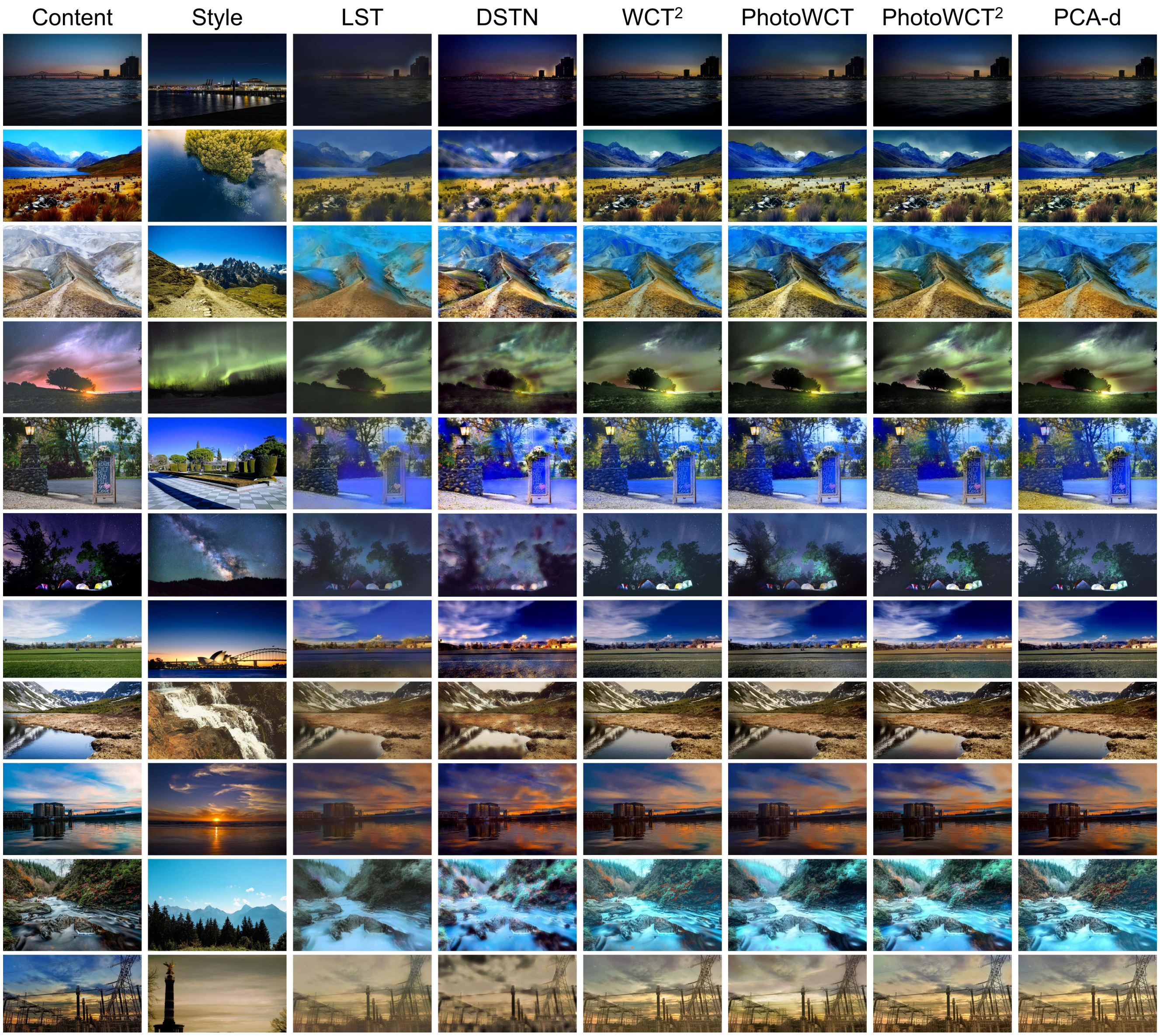

Figure 2: Qualitative comparison of LST [7] and DSTN [5] to WCT^2 [12], PhotoWCT [9], PhotoWCT^2 [2], and PCA-d [3] with our LS-FT as the feature transformation. The results show that LST and DSTN produce poor photorealism.

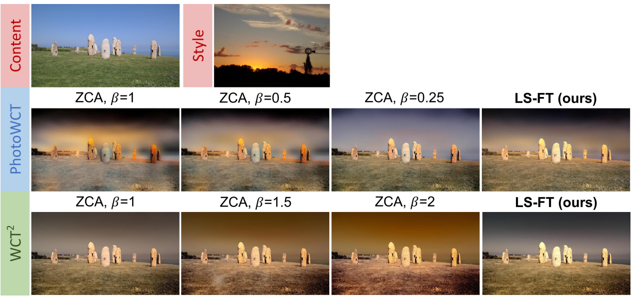

Figure 3: Insufficiency of linear interpolation for content-style control.

Insufficiency of linear interpolation for content-style control

In [8], where ZCA is proposed for style transfer, the authors also propose to use linear interpolation between the style feature $\mathbf{F}_s$ and the transformed feature $\mathbf{F}_t$ to realize content-style controllability. That is, the feature to be decoded is $\beta\mathbf{F}_t{+}(1{-}\beta)\mathbf{F}_s$, where β is the parameter that controls style strength. Mathematically, such a linear interpolation is insufficient since the content-style relation is a non-linear function as described by Eq. (1) in our main paper. Qualitatively, Fig. 3 shows that when trying to reduce reduce artifacts that come from strong stylization strength for PhotoWCT by reducing β, both style effects and artifacts disappear together. Alternatively, when trying to strengthen stylization strength for WCT^2 by increasing β, we can observe distorted style effects (Fig. 3). In contrast, our LS-FT can maintain the style effects well while controlling the balance between content and style.

Removing centralization and decentralization from AdaIN and ZCA leads to worse image quality

In Section 3.2 of the main paper, we add centralization and decentralization to stabilize the performance of IterFT [1]. Here we conduct an ablation study of removing centralization and decentralization from AdaIN [6] and ZCA [8] and test the resulting performance for the PhotoWCT^2 model [2]. The results are shown in Fig. 4. These reveal that the ablated AdaIN and the ablated ZCA may suffer from incomplete stylization, where parts of a content image receive limited to no stylization. Consequently, bad results may occur. This offers promising evidence that centralization and decentralization play an important role in synthesizing reasonable images.

Explanation for how centralization and decentralization support mean vector matching

While prior works [4, 10] focus on explaining the reason of matching second-order statistics between style and stylized features for style transfer, we conjecture that matching first-order mean vectors is also important, and this is supported by centralization and decentralization. We illustrate this here for the ZCA algorithm and the same argument applies to OST and AdaIN.

Let $\mathbf{F}_{c/s}$ be the content/style feature and $\mu_{c/s}$ and $\mathbf{C}_{c/s}{=}\bar{\mathbf{F}}_{c/s}\bar{\mathbf{F}}_{c/s}^{\mathrm{T}}$ be their mean vectors and covariance matrices. The ZCA transformed feature $\mathbf{F}_t$ is given by $\mathbf{C}_s^{\frac{1}{2}}\mathbf{C}_c^{\frac{-1}{2}}\bar{\mathbf{F}}_c{+}\mu_s$. It can be shown that the mean and the covariance of $\mathbf{F}_t$ are exactly those of $\mathbf{F}_s$. However, without centralization and decentralization and replacing $\mathbf{C}_{c/s}$ by the gram matrix $\mathbf{G}_{c/s}{=}\mathbf{F}_{c/s}\mathbf{F}_{c/s}^{\mathrm{T}}$, the transformed feature becomes $\mathbf{F}_g{=}\mathbf{G}_s^{\frac{1}{2}}\mathbf{G}_c^{\frac{-1}{2}}\mathbf{F}_c$. Now only the gram matrices of $\mathbf{F}_g$ and $\mathbf{F}_s$ match, while their mean vectors differ. This may explain the results in Fig. 4 here in the Supplementary Materials. An interesting area for future work is to establish *why* mean matching is important.

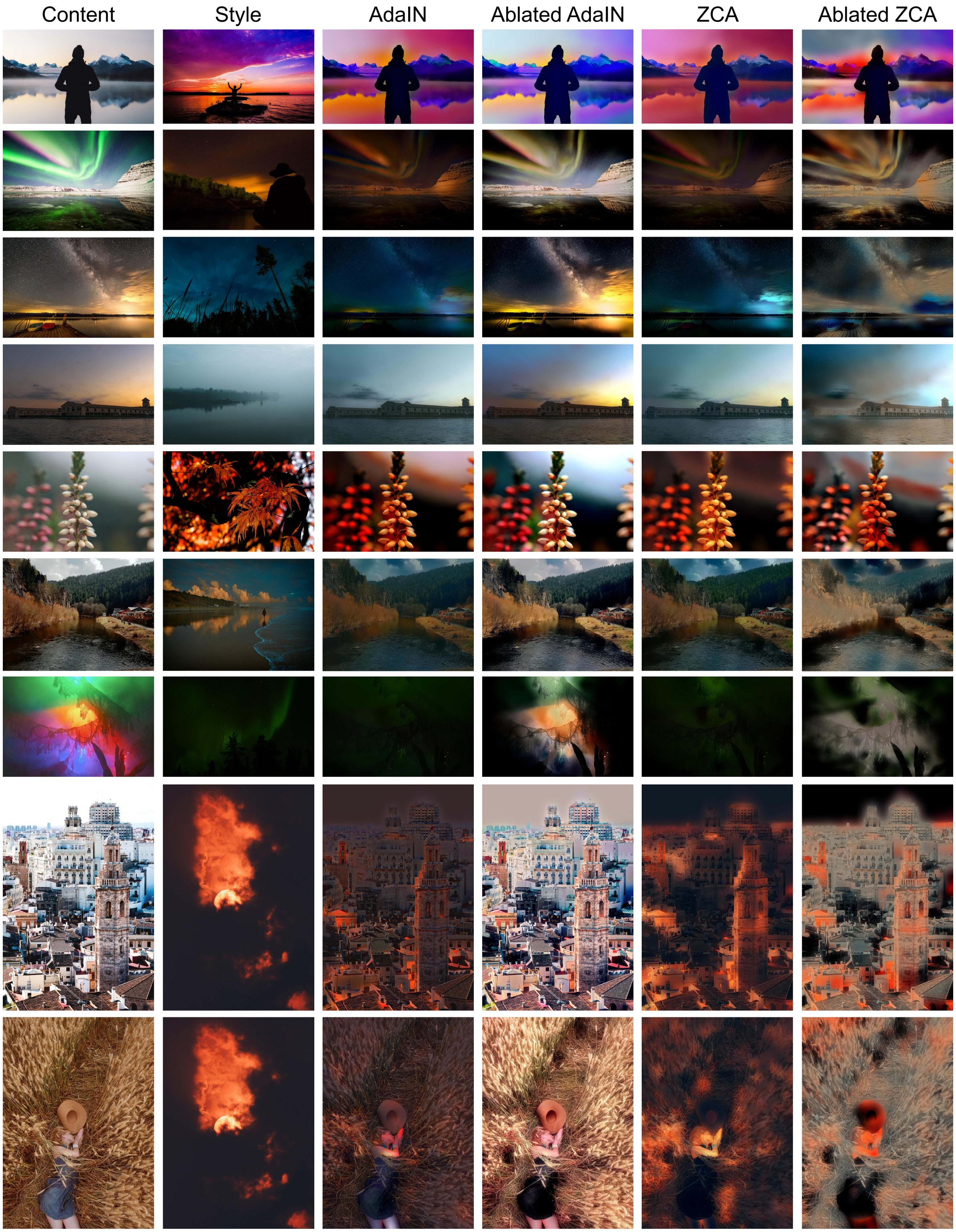

Figure 4: Ablation study of removing centralization and decentralization from AdaIN [6] and ZCA [8]. The results shows that centralization and decentralization play an important role in synthesizing reasonable images. The study is done with the PhotoWCT^2 model [2].

Derivation of Equation 7 in the main paper

We show the following optimization problem:

$$\min_{\eta} l(\bar{\mathbf{F}}_t - \eta \frac{\mathrm{d}l}{\mathrm{d}\bar{\mathbf{F}}_t}), \tag{1}$$

where:

$$l(\bar{\mathbf{F}}_t) = \underbrace{||\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c||_2^2}_{\text{content loss}} + \underbrace{\lambda\, ||\frac{1}{n_c}\bar{\mathbf{F}}_t\bar{\mathbf{F}}_t^{\mathrm{T}} - \frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}}||_2^2}_{\text{style loss}} \tag{2}$$

and:

$$\frac{\mathrm{d}l}{\mathrm{d}\bar{\mathbf{F}}_t} = 2(\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c) + \frac{4\lambda}{n_c}(\frac{1}{n_c}\bar{\mathbf{F}}_t\bar{\mathbf{F}}_t^{\mathrm{T}} - \frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}})\bar{\mathbf{F}}_t, \tag{3}$$

is equivalent to the following cubic equation:

$$a\eta^3 + b\eta^2 + c\eta + d = 0, \tag{4}$$

with the coefficients being:

$$a = \frac{2\lambda}{n_c^2}\mathrm{tr}[\mathbf{D_2D_2}],\ b = -\frac{6\lambda}{n_c^2}\mathrm{tr}[\mathbf{D_F D_2}],\ d = -\frac{1}{2}\mathrm{tr}[\mathbf{D_2}], \tag{5}$$

$$c = \mathrm{tr}[\mathbf{D_2}] + \frac{2\lambda}{n_c}\mathrm{tr}[\mathbf{D_2 S}] + \frac{2\lambda}{n_c^2}\big(\mathrm{tr}[\mathbf{D_F D_F}] + \mathrm{tr}[\mathbf{D_F D_F^{\mathrm{T}}}]\big), \tag{6}$$

where $\mathbf{D_2} \equiv \mathbf{DD}^{\mathrm{T}}$, $\mathbf{D_F} \equiv \mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}}$, $\mathbf{D} \equiv \frac{\mathrm{d}l}{\mathrm{d}\bar{\mathbf{F}}_t}$ and $\mathbf{S} \equiv \frac{1}{n_c}\bar{\mathbf{F}}_t\bar{\mathbf{F}}_t^{\mathrm{T}} - \frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}}$.

Proof. First, we plug $\bar{\mathbf{F}}_t - \eta\frac{\mathrm{d}l}{\mathrm{d}\bar{\mathbf{F}}_t} = \bar{\mathbf{F}}_t - \eta\mathbf{D}$ into the loss function in Eq. 2 and we get:

$$\begin{aligned} l(\bar{\mathbf{F}}_t - \eta\mathbf{D}) &= \underbrace{||\mathbf{A}||_2^2}_{\text{content loss}} + \lambda \underbrace{||\mathbf{B}||_2^2}_{\text{style loss}} \\ &= \mathrm{tr}[\mathbf{AA}^{\mathrm{T}}] + \lambda \cdot \mathrm{tr}[\mathbf{BB}^{\mathrm{T}}], \end{aligned} \tag{7}$$

where:

$$\begin{aligned} \mathbf{A} =&\ \bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c - \eta\mathbf{D}, \\ \mathbf{B} =&\ \frac{1}{n_c}(\bar{\mathbf{F}}_t - \eta\mathbf{D})(\bar{\mathbf{F}}_t - \eta\mathbf{D})^{\mathrm{T}} - \frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}} \\ =&\ (\frac{1}{n_c}\bar{\mathbf{F}}_t\bar{\mathbf{F}}_t^{\mathrm{T}} - \frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}}) - \frac{1}{n_c}\big(\eta(\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}})^{\mathrm{T}} + \eta\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}} - \eta^2\mathbf{DD}^{\mathrm{T}}\big) \\ =&\ \mathbf{S} - \frac{1}{n_c}\big(\eta(\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}})^{\mathrm{T}} + \eta\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}} - \eta^2\mathbf{DD}^{\mathrm{T}}\big) \\ =&\ \mathbf{S} - \frac{1}{n_c}\Big(\eta\mathbf{D_F^{\mathrm{T}}} + \eta\mathbf{D_F} - \eta^2\mathbf{D_2}\Big) \end{aligned} \tag{8}$$

To find the minimum, we differentiate $l(\bar{\mathbf{F}}_t - \eta\mathbf{D})$ with respect to η and solve the derivative equal to zero. Before the differentiation, we first derive an useful the identity that given a matrix $\mathbf{M}$ the derivative $\frac{\mathrm{d}(\mathrm{tr}[\mathbf{MM}^{\mathrm{T}}])}{\mathrm{d}p}$ with respect to a parameter p is equal to $2\mathrm{tr}[\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}p}\mathbf{M}^{\mathrm{T}}]$:

$$\begin{aligned} \frac{\mathrm{d}(\mathrm{tr}[\mathbf{MM}^{\mathrm{T}}])}{\mathrm{d}p} &= \mathrm{tr}\big[\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}p}\mathbf{M}^{\mathrm{T}}\big] + \mathrm{tr}[\mathbf{M}\frac{\mathrm{d}\mathbf{M}^{\mathrm{T}}}{\mathrm{d}p}] \\ &\overset{(*)}{=} \mathrm{tr}\big[\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}p}\mathbf{M}^{\mathrm{T}}\big] + \mathrm{tr}[\big(\mathbf{M}\frac{\mathrm{d}\mathbf{M}^{\mathrm{T}}}{\mathrm{d}p}\big)^{\mathrm{T}}] \\ &= 2\mathrm{tr}\big[\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}p}\mathbf{M}^{\mathrm{T}}\big], \end{aligned} \tag{9}$$

where in (*) we use the identity that the trace of a matrix is equal to the trace of its transpose. With the identity in Eq. 9, we have:

$$\frac{\mathrm{d}l(\bar{\mathbf{F}}_t - \eta\mathbf{D})}{\mathrm{d}\eta} = \cancel{2}\mathrm{tr}\big[\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}\eta}\mathbf{A}^{\mathrm{T}}\big] + \cancel{2}\lambda \cdot \mathrm{tr}\big[\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}\eta}\mathbf{B}^{\mathrm{T}}\big] = 0, \tag{10}$$

where the 2's are crossed out and the equality to zero still holds. In detail, we have:

$$\begin{aligned} &\mathrm{tr}\big[\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}\eta}\mathbf{A}^{\mathrm{T}}\big] \\ =&\ \mathrm{tr}[-\mathbf{D}(\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c - \eta\mathbf{D})^{\mathrm{T}}] \\ =&\ \mathrm{tr}[-\mathbf{D}(\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c)^{\mathrm{T}} + \eta\mathbf{D_2})] \end{aligned} \tag{11}$$

and:

$$\begin{aligned} \mathrm{tr}\big[\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}\eta}\mathbf{B}^{\mathrm{T}}\big] &= -\frac{1}{n_c}\mathrm{tr}\big[\left(\mathbf{D_F^{\mathrm{T}}} + \mathbf{D_F} - 2\eta\mathbf{D_2}\right)\mathbf{B}^{\mathrm{T}}\big] \\ &= -\frac{1}{n_c}\mathrm{tr}\big[\left(\mathbf{D_F^{\mathrm{T}}} + \mathbf{D_F} - 2\eta\mathbf{D_2}\right)\mathbf{B}\big], \end{aligned} \tag{12}$$

where the identity $\mathbf{B} = \mathbf{B}^{\mathrm{T}}$ is used. Due to the following equality:

$$\begin{aligned} \mathrm{tr}\big[(\mathbf{D_F^{\mathrm{T}}}\mathbf{B}\big] &\overset{(a)}{=} \mathrm{tr}\big[\mathbf{B}^{\mathrm{T}}\mathbf{D_F}\big] \\ &\overset{(b)}{=} \mathrm{tr}\big[\mathbf{BD_F}\big] \\ &\overset{(c)}{=} \mathrm{tr}\big[\mathbf{D_F B}\big], \end{aligned} \tag{13}$$

where we use in (a) the identity $\mathrm{tr}[\mathbf{M}^{\mathrm{T}}]$ = $\mathrm{tr}[\mathbf{M}]$ for any square matrix $\mathbf{M}$, in (b) the identity $\mathbf{B} = \mathbf{B}^{\mathrm{T}}$, and in (c) the identity $\mathrm{tr}[\mathbf{M}_1\mathbf{M}_2]$ = $\mathrm{tr}[\mathbf{M}_2\mathbf{M}_1]$ for any square matrix that can be decomposed into the product of two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, Eq. 12 can be further written as:

$$\begin{aligned} &\mathrm{tr}\big[\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}\eta}\mathbf{B}^{\mathrm{T}}\big] \\ =& -\frac{2}{n_c}\mathrm{tr}\big[(\mathbf{D_F} - \eta\mathbf{D_2})\mathbf{B}\big] \\ =& -\frac{2}{n_c}\mathrm{tr}\left[\left(\mathbf{D_F} - \eta\mathbf{D_2}\right)\left(\mathbf{S} - \frac{1}{n_c}\left(\eta\mathbf{D_F^{\mathrm{T}}} + \eta\mathbf{D_F} - \eta^2\mathbf{D_2}\right)\right)\right]. \end{aligned} \tag{14}$$

If we substitute Eq. 11 and Eq. 14 into Eq. 10 and group the terms by the order of η, arranging them into the format

$a\eta^3 + b\eta^2 + c\eta + d = 0$, we have the coefficients being:

$$a = \frac{2\lambda}{n_c^2}\mathrm{tr}[\mathbf{D_2D_2}], \tag{15}$$

$$b = -\frac{2\lambda}{n_c^2}\mathrm{tr}\left[\mathbf{D_2D_F^T} + \mathbf{D_2D_F} + \mathbf{D_FD_2}\right], \tag{16}$$

$$c = \ \mathrm{tr}[\mathbf{D_2}] + \frac{2\lambda}{n_c}\mathrm{tr}[\mathbf{D_2S}] + \frac{2\lambda}{n_c^2}\big(\mathrm{tr}[\mathbf{D_FD_F}] + \mathrm{tr}[\mathbf{D_FD_F^T}]\big), \tag{17}$$

$$d = -\mathrm{tr}[\mathbf{D}(\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c)^{\mathrm{T}}] - \frac{2\lambda}{n_c}\mathrm{tr}[\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}}\mathbf{S}]. \tag{18}$$

Since $\mathrm{tr}[\mathbf{D_2D_F^T}]$ is equal to $\mathrm{tr}[\mathbf{D_FD_2}]$ due to the aforementioned trace identity $\mathrm{tr}[\mathbf{M}] = \mathrm{tr}[\mathbf{M}^{\mathrm{T}}]$, and $\mathrm{tr}[\mathbf{D_FD_2}]$ is equal to $\mathrm{tr}[\mathbf{D_2D_F}]$ due to the aforementioned trace identity $\mathrm{tr}[\mathbf{M}_1\mathbf{M}_2]$ = $\mathrm{tr}[\mathbf{M}_2\mathbf{M}_1]$, Eq. 16 can be further simplified as:

$$b = -\frac{6\lambda}{n_c^2}\mathrm{tr}\big[\mathbf{D_FD_2}\big]. \tag{19}$$

In addition, d in Eq. 18 can be further simplified as follows:

$$\begin{aligned} d &= -\mathrm{tr}[\mathbf{D}(\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c)^{\mathrm{T}}] - \frac{2\lambda}{n_c}\mathrm{tr}[\mathbf{D}\bar{\mathbf{F}}_t^{\mathrm{T}}\mathbf{S}] \\ &= -\mathrm{tr}\big[\mathbf{D}\big((\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c)^{\mathrm{T}} + \frac{2\lambda}{n_c}(\mathbf{S}^{\mathrm{T}}\bar{\mathbf{F}}_t)^{\mathrm{T}}\big)\big] \\ &\overset{(a)}{=} -\mathrm{tr}\big[\mathbf{D}\big((\bar{\mathbf{F}}_t - \bar{\mathbf{F}}_c)^{\mathrm{T}} + \frac{2\lambda}{n_c}(\mathbf{S}\bar{\mathbf{F}}_t)^{\mathrm{T}}\big)\big] \\ &\overset{(b)}{=} -\frac{1}{2}\mathrm{tr}[\mathbf{D}\mathbf{D}^{\mathrm{T}}] \\ &= -\frac{1}{2}\mathrm{tr}[\mathbf{D_2}], \end{aligned} \tag{20}$$

where in (a) we use $\mathbf{S}^{\mathrm{T}} = \mathbf{S}$ and in (b) we introduce the equality in Eq. 3. We conclude the derivation with the Equations (15), (19), (17), and (20). ■

Proof of at least one positive solution to Equation 7 in the main paper

To comply with the constraint $\eta > 0$ in line search, we have to ensure that there is at least one positive solution to Equation 7 (Eq. 4 in this document). We prove this in the following.

Proof. For a cubic equation, there are two possible sets of solutions: either three real roots or one real root with two complex conjugated roots. From the relation between roots and coefficients, the product of three roots of Eq. 4 is equal to $-\frac{d}{a} = \frac{n_c^2}{4\lambda}\frac{\mathrm{tr}[\mathbf{D}\mathbf{D}^{\mathrm{T}}]}{\mathrm{tr}[\mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{D}^{\mathrm{T}}]} = \frac{n_c^2}{4\lambda}\frac{||\mathbf{D}||_{\mathrm{F}}^2}{||\mathbf{D}\mathbf{D}^{\mathrm{T}}||_{\mathrm{F}}^2} > 0$, where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm. If the solutions to Eq. 4 are three real roots, since the product of three real roots is positive, at least one of them must be positive. If the solutions are one real root and two complex conjugated roots, since the product of three roots is positive and the product of two complex conjugated roots is the squared 2-norm of the conjugated roots, which is positive, the remaining real root should be positive. Therefore, there is at least one positive solution to Eq. 4. ■

Values of η searched by LS-FT

To illustrate how the values of line-searched η differ from the constant η equal to 0.01 used in IterFT [1], we embed LS-FT in PhotoWCT^2 [2] and computed the values of line-searched η at each *reluN_1* layer using the PST dataset [11]. As shown in Fig. 5, depending on the layer where LS-FT line-searches η and the input images, the value of a line-searched η can be as small as 0.04 and as large as 0.8.

Effect of the content-style control knob α

We introduce a content-style control knob α in Section 3.3 in the main paper to adjust the value of λ in the unit of $\frac{||\bar{\mathbf{F}}_c||_2^2}{||\frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}}||_2^2}$ ($\lambda = \alpha\frac{||\bar{\mathbf{F}}_c||_2^2}{||\frac{1}{n_s}\bar{\mathbf{F}}_s\bar{\mathbf{F}}_s^{\mathrm{T}}||_2^2}$) such that we can tune the value of α to realize content-style controllability. As explained in Section 3.3 in the main paper, we want different values of α to boost the stylization strength of WCT^2 [12], to boost the content preservation for PhotoWCT [9] and PCA-d [3], and to balance content preservation and stylization strength for PhotoWCT^2 [2].

Fig. 6(a) shows the performance of WCT^2, PhotoWCT, and PhotoWCT^2 with LS-FT of different α values, while Fig. 6(e) shows the performance of Modified IterFT on the same models. The other panels in Fig. 6 are the zoomed plots for different models. To boost the stylization strength of WCT^2, we observe in Fig. 6(b,f) that for LS-FT and Modified IterFT the style loss marginally decreases when α is larger than 10 and 50, respectively. Therefore, we set α equal to 10 and 50 for LS-FT and Modified IterFT, respectively, when they are used in the WCT^2 model. To boost the content preservation ability of PhotoWCT, we observe in Fig. 6(c,g) that for LS-FT and Modified IterFT the content loss marginally decreases when α is smaller than 0.2 and 0.5, respectively. Therefore, we set α equal to 0.2 and 0.5 for LS-FT and Modified IterFT, respectively, when they are used in the PhotoWCT model. To balance content preservation ability and stylization strength of PhotoWCT^2, we first recall in the Figure 5(d) in the main paper that the style loss of AdaIN is around 4.37e^4. Thus, to have stronger stylization strength than AdaIN and better content preservation ability than ZCA, we want the style losses of LS-FT and Modified IterFT smaller than 4.37e^4 and the content losses of LS-FT and Modified IterFT as small as possible. We observe in Fig. 6(d,h) that $\alpha = 1$ for LS-FT and $\alpha = 2$ for Modified IterFT best fit this criterion, and so we set α as such values.

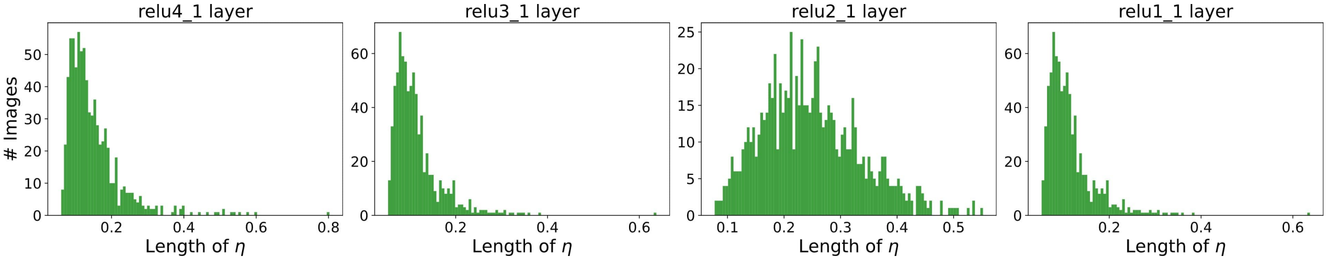

Figure 5: Histograms of the values of line-searched η at the *reluN_1* layers. This test is done using PhotoWCT^2 [2] and the PST dataset [11].

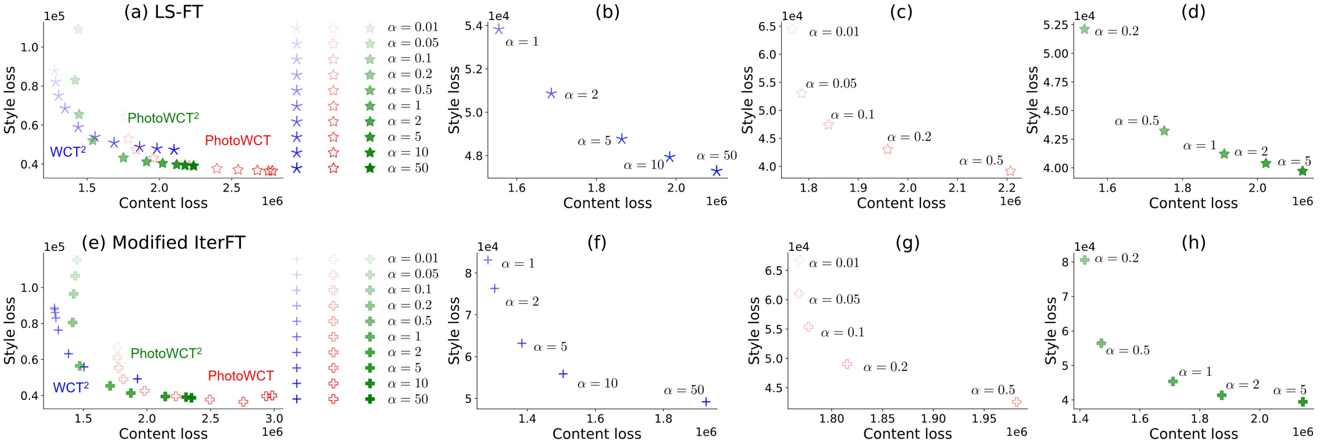

Figure 6: Effect of the content-style control knob α on WCT^2 [12], PhotoWCT [9], and PhotoWCT^2 [2].

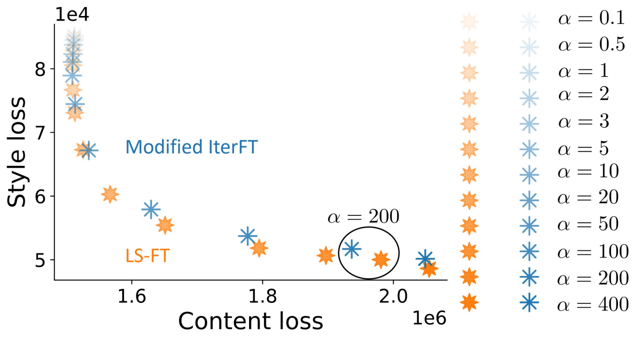

Figure 7: Effect of the content-style control knob α on PCA-d [3].

Fig. 7 shows the performance of PCA-d with LS-FT and Modified IterFT of different α values. We observe that for both LS-FT and Modified IterFT, when α is equal to 200, the content and style losses of PCA-d are very close to those from the aforementioned settings for WCT^2, PhotoWCT, and PhotoWCT^2. Therefore, we set α to 200 here.

Convergence comparison between Modified IterFT and LS-FT on WCT^2, PhotoWCT, and PCA-d

We show in Figure 4 in the main paper the convergence comparison between Modified IterFT and LS-FT on PhotoWCT^2 [2]. For completeness, we show in Fig. 8 the convergence comparison between Modified IterFT and LS-FT on WCT^2 [12], PhotoWCT [9], and PCA-d [3]. The result shows that LS-FT needs only one iteration at each transformation layer of each model to outperform Modified IterFT, which is the same as the result from PhotoWCT^2.

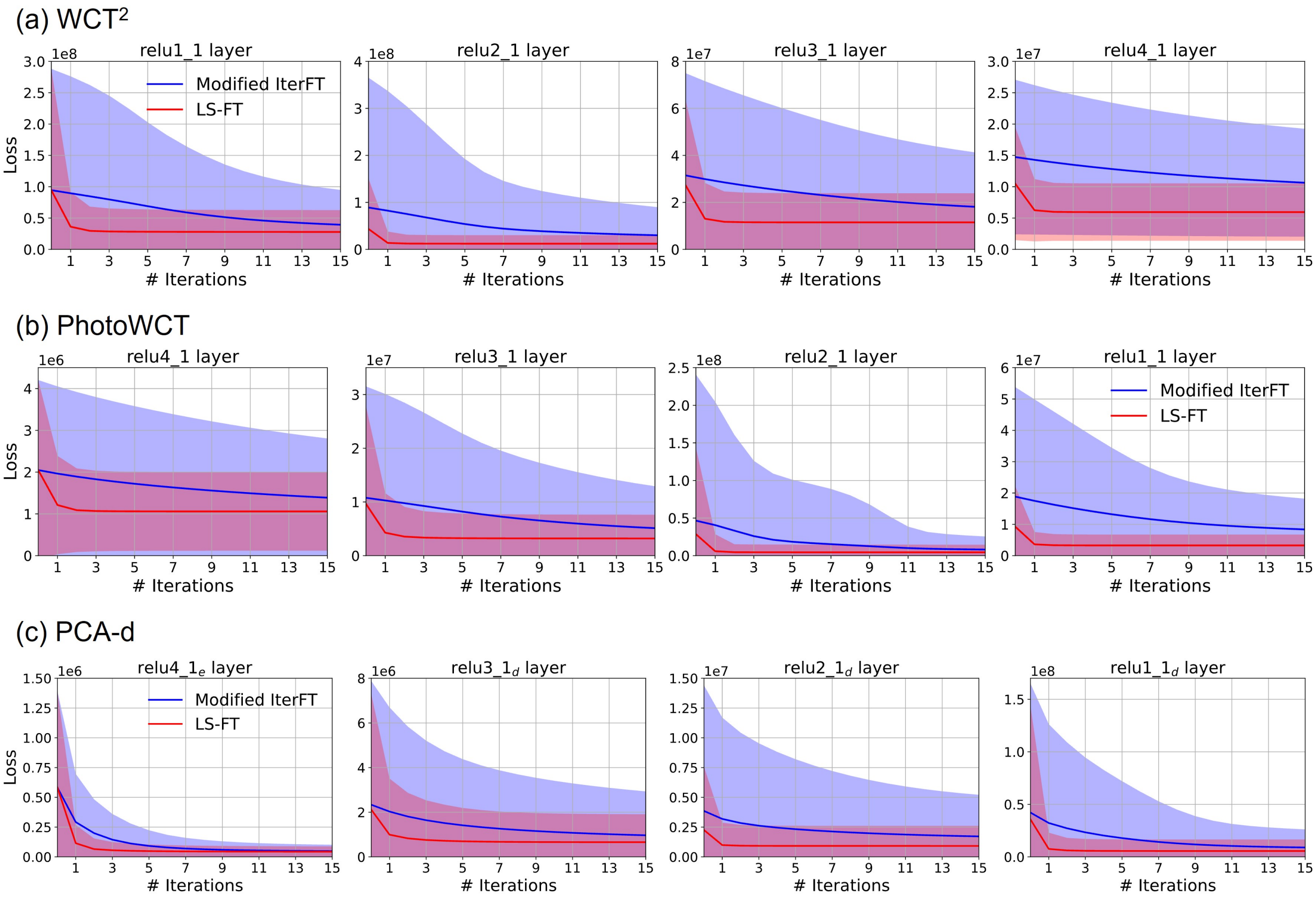

Figure 8: Convergence comparison between Modified IterFT and LS-FT on WCT^2 [12], PhotoWCT [9], and PCA-d [3]. The result shows that LS-FT needs only one iteration to outperform Modified IterFT at each transformation layer of each model, which is the same as the case of PhotoWCT^2 [2].

PCA-d	Not Tunable				Tunable	
	ZCA	OST	AdaIN	MAST	M-IterFT	LS-FT
HD	0.024	0.032	0.048	**0.005**	0.190	0.052	**0.031**
FHD	0.042	0.036	0.051	**0.006**	0.211	0.093	**0.032**
QHD	0.070	0.064	0.072	**0.013**	0.368	0.170	**0.064**
UHD	0.160	0.108	0.111	**0.027**	0.572	0.374	**0.120**

Table 1: The speeds for stylization of images of PCA-d using different transformations. Unit: Second.

Unreasonable results from IterFT are fixed with our Modified IterFT and LS-FT

We exclude IterFT [1] from the experiment in Section 4.2 in the main paper, since it results in dozens of unreasonable results from the PST dataset [11]. Here we show some failures from IterFT in Fig. 9 and how they are corrected with our Modified IterFT and LS-FT.

Speed of transformations on PCA-d

We show the speed of different transformations on WCT^2, PhotoWCT, and PhotoWCT^2 in Table 2 in the main paper. Here we show the speed of transformations on PCA-d in Table 1. We observe a similar result: our LS-FT is faster or comparably fast to ZCA.

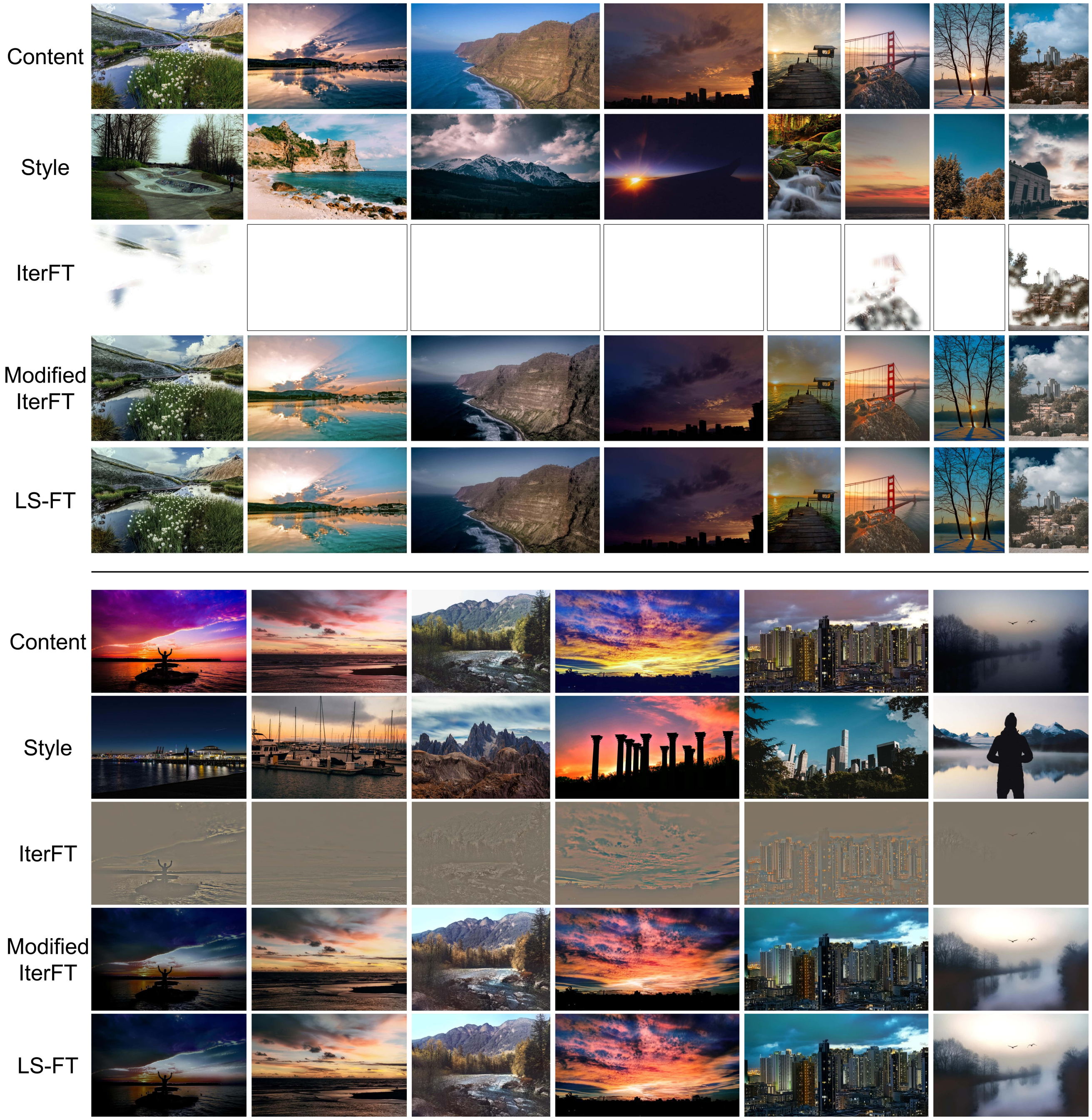

Figure 9: The failures of IterFT [1] can be corrected with our Modified IterFT and LS-FT.

References

[1] Tai-Yin Chiu and Danna Gurari. Iterative feature transformation for fast and versatile universal style transfer. In *European Conference on Computer Vision*, pages 169–184. Springer, 2020.

[2] Tai-Yin Chiu and Danna Gurari. Photowct2: Compact autoencoder for photorealistic style transfer resulting from blockwise training and skip connections of high-frequency residuals. *arXiv preprint arXiv:2110.11995*, 2021.

[3] Tai-Yin Chiu and Danna Gurari. Pca-based knowledge distillation towards lightweight and content-style balanced photorealistic style transfer models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7844–7853, 2022.

[4] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In

Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2414–2423, 2016.

[5] Kibeom Hong, Seogkyu Jeon, Huan Yang, Jianlong Fu, and Hyeran Byun. Domain-aware universal style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14609–14617, 2021.

[6] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[7] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3809–3817, 2019.

[8] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In *Advances in neural information processing systems*, pages 386–396, 2017.

[9] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 453–468, 2018.

[10] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. Demystifying neural style transfer. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 2230–2236, 2017.

[11] Xide Xia, Meng Zhang, Tianfan Xue, Zheng Sun, Hui Fang, Brian Kulis, and Jiawen Chen. Joint bilateral learning for real-time universal photorealistic style transfer. *arXiv preprint arXiv:2004.10955*, 2020.

[12] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9036–9045, 2019.