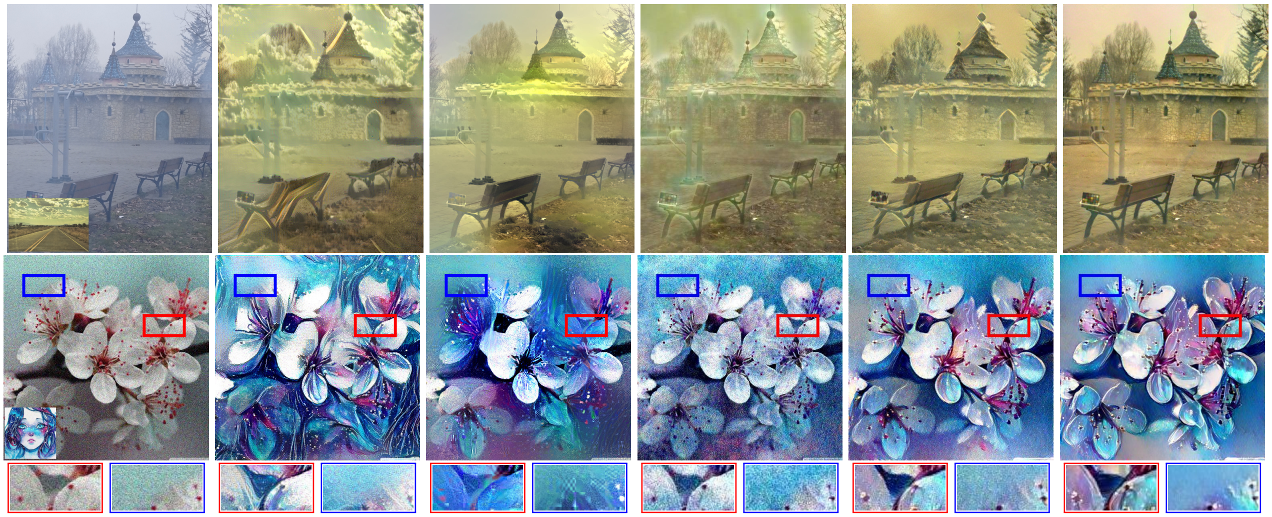
DILIE: Deep Internal Learning for Image Enhancement

Indra Deep Mastan[1], Shanmuganathan Raman[2], Prajwal Singh[2]
LNM Institute of Information Technology[1], Indian Institute of Technology Gandhinagar[2]
indradeep.mastan@lnmiit.ac.in[1], {shanmuga, singh_prajwal}@iitgn.ac.in[2]



(a) Content & Style (b) Neural style [10] (c) DPST [18] (d) WCT2 [40] (e) STROTSS [14] (f) DILIE (ours)

Figure 1: The figure shows that DILIE framework outputs images with better perceptual quality. The style image is shown at the left corner of content image. The first row shows that DILIE output image with minimum haze corruption for hazy image enhancement. The second row shows that DILIE output images with better clarity for noisy image enhancement.

Abstract

We consider the generic deep image enhancement problem where an input image is transformed into a perceptually better-looking image. The methods mostly fall into two categories: training with prior examples methods and training with no-prior examples methods. Recently, Deep Internal Learning solutions to image enhancement in training with no-prior examples setup are gaining attention. We perform image enhancement using a deep internal learning framework. Our Deep Internal Learning for Image Enhancement framework (DILIE) enhances content features and style features and preserves semantics in the enhanced image. To validate the results, we use structure similarity and perceptual error, which is efficient in measuring the unrealistic deformation present in the images. We show that DILIE framework outputs good quality images for hazy and noisy image enhancement tasks.

1. Introduction

Deep image enhancement is an ill-posed problem that aims to improve the perceptual quality of an image using a deep neural network [16, 39, 35, 37]. An image could be considered as the composition of content features and style features. The content features denote the objects, their structure, and their relative positions. The style features represent the color and the texture information of the objects. Deep image enhancement aims to improve the quality of the content and the style features.

Performing deep image enhancement without using prior examples of training data was proposed as an open problem [41]. The prior examples are related to supervised learning tasks that use clean and corrupted image pairs to perform image restoration. Recently, Deep Internal Learning (DIL) methods have shown how to use deep convolutional neural networks to perform image restoration and image synthesis without using prior examples [34, 9, 20, 30, 29]. DIL is different from training data-based methods that use prior examples to supervise the image enhancement task [23, 13].

Let us discuss an example of a deep image enhancement task. Suppose I denotes a hazy image. The haze particles degrade content features and style features. The content features are corrupted because haze particles reduce the clarity of the structure of the objects. The style features are corrupted due to gray and blueish patterns introduced by haze. The image enhancement task is to improve the perceptual quality of the hazy image I.

The enhancement of content and style features of hazy image I may draw inspiration from the image restoration and the style transfer methods. One strategy is to utilize the content features from I and transferring the photo-realistic features from a style image S. The interesting observation here is that maintaining the balance between the content feature and the style feature is challenging (Fig. 6).

We formulate a generic framework called Deep Internal Learning for Image Enhancement (DILIE) (Algorithm 1). It does not use prior examples of training data to perform image enhancement. Fig. 1 shows the hazy and noisy image enhancement tasks using DILIE. The good perceptual quality of DILIE framework is due to the ability of CNN to learn good quality image statistics from a single image [34, 9, 30, 21].

We illustrate DILIE framework in Fig. 2 for hazy image enhancement[1]. Given the degraded image I as input, the aim is to generate the enhanced image I^*. The *main idea* is to formulate the content feature enhancement (CFE) and the style feature enhancement (SFE) models separately for generalizability. Fig. 2 shows CFE decomposes the hazy image I into environmental haze layer H and haze-free image I^{cfe}. SFE transfers photo-realistic features from style image S to I^{cfe}.

The content feature enhancement is performed based on the type of corruption. CFE module for image dehazing is modeled using the image decomposition model. CFE module for image denoising is modeled using image reconstruction model. The image decomposition model performs joint optimization to separate the degraded image into clean and corrupted features. Image reconstruction generates a clean image with pixel-based reconstruction loss. Both these approaches rely upon the strong image prior captured by the encoder-decoder network [34].

The aim of SFE is to transform the input image (content) into a visually appealing output image by transferring style features from the style image. SFE is modeled based on the desired style specification, *i.e.*, photo-realistic style transfer [18, 40] or artistic style transfer [14]. Note that the distortions in the style transfer output lead to a lack of photo-realism. We measure the deformations using perceptual error Pieapp [26] computed between the content image and the output image. DILIE output images with low perceptual error and better visual quality (Table 1).

DILIE preserves the semantics of the input image using the contextual content loss denoted by $\mathcal{L}_{CL}$. The contextual content loss compares context vectors between input corrupted image and enhanced image. The context vectors represent high-level semantics information. The context vectors are extracted using pre-trained feature extractor VGG19 [23]. Fig. 2 illustrates $\mathcal{L}_{CL}$ is computed between I and I^{cfe} to preserve the contextual content features in I^{cfe}. We describe DILIE framework in Sec. 3.

Contributions. The major contributions are as follows.

- We propose a generic framework (DILIE) that addresses corruption-specific image enhancement using image reconstruction, image decomposition, and photo-realistic feature enhancement (Algorithm 1).
- We show image enhancement for the challenging scenario where photos were taken in hazy weather (Fig. 3 and Fig. 4). We also perform enhancement of the noisy images (Fig. 5).
- DILIE shows that utilization of contextual features improves image dehazing (Table 2). DILIE outputs images with good visual quality and lower perceptual error (Table 1 and Fig. 6). We also show the limitation of DILIE framework (Fig. 8).

2. Related Work

Deep Internal Learning. DIL aims to learn the internal patch distribution and utilize the deep image prior to perform image restoration [34, 9] and image synthesis [20, 30, 29] For simplicity, we divide DIL approaches into two categories: DIL methods that use Generative Adversarial Networks (GANs) [36, 20, 30, 29] and DIL methods that do not use GAN [34, 9]. DIL is similar to Zero-reference method (ZR) [11] that does not require any paired or unpaired training data. It is interesting to note that ZR [11] uses example images of different exposure levels to train DCE-Net for low-light image enhancement. DILIE does not use task-specific reference images for image enhancement. DILIE uses a pretrained feature extractor (VGG19) trained on image classification, which makes it different from other DIL methods [34, 9, 30, 29].

Content Feature Enhancement. CFE is performed using image reconstruction and image decomposition models. Image reconstruction models use an encoder-decoder network (ED) to perform denoising, super-resolution, and inpainting [34]. Dehazing is formulated as an image decomposition problem [9], where ED separates the image layer and haze layer. For simplicity, image dehazing could be

[1] We describe DILIE framework for noisy image enhancement in the supplementary material.

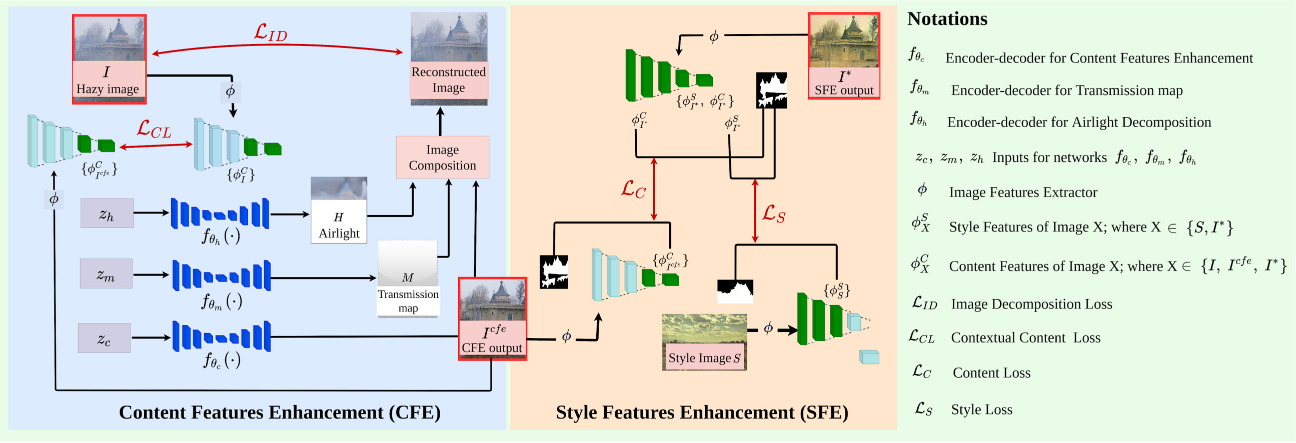

$\mathcal{L}_{ID}$
I
Hazy image
Reconstructed Image
ϕ
$\mathcal{L}_{CL}$
$\{\phi^C_{I^{cfe}}\}$
$\{\phi^C_I\}$
Image Composition
ϕ
z_h
$f_{\theta_h}(\cdot)$
H
Airlight
z_m
$f_{\theta_m}(\cdot)$
M
Transmission map
z_c
$f_{\theta_c}(\cdot)$
I^{cfe}
CFE output
Content Features Enhancement (CFE)
ϕ
$\{\phi^S_{I^*}, \phi^C_{I^*}\}$
$\phi^C_{I^*}$
$\phi^S_{I^*}$
I^*
SFE output
$\mathcal{L}_C$
$\mathcal{L}_S$
$\{\phi^C_{I^{cfe}}\}$
$\{\phi^S_S\}$
ϕ
Style Image S
ϕ
Style Features Enhancement (SFE)
Notations
f_{θ_c} Encoder-decoder for Content Features Enhancement
f_{θ_m} Encoder-decoder for Transmission map
f_{θ_h} Encoder-decoder for Airlight Decomposition
z_c, z_m, z_h Inputs for networks f_{θ_c}, f_{θ_m}, f_{θ_h}
ϕ Image Features Extractor
ϕ^S_X Style Features of Image X; where X $\in$ $\{S, I^*\}$
ϕ^C_X Content Features of Image X; where X $\in$ $\{I, I^{cfe}, I^*\}$
$\mathcal{L}_{ID}$ Image Decomposition Loss
$\mathcal{L}_{CL}$ Contextual Content Loss
$\mathcal{L}_C$ Content Loss
$\mathcal{L}_S$ Style Loss

Figure 2: **Hazy Image Enhancement.** The figure shows DILIE framework for hazy image enhancement. Hazy image I is transformed into an enhanced image I^*. The left side shows the content feature enhancement (CFE) and the right side shows the style feature enhancement (SFE). CFE performs image decomposition to output haze-free image I^{cfe}, transmission map M and haze layer H. VGG19 network ϕ is used to extract features to compute contextual content loss $\mathcal{L}_{CL}$, content loss $\mathcal{L}_C$, and style loss $\mathcal{L}_S$. Image decomposition loss $\mathcal{L}_{ID}$ is a pixel-based loss (Eq. 3). SFE improves style features using content loss $\mathcal{L}_C$ and style loss $\mathcal{L}_S$.

classified into classical [6, 8, 12, 33], supervised method using deep learning [15], and unsupervised methods [9].

Style Feature Enhancement. Gatys et al. proposed Neural style [10] for style feature enhancemnt. Luan et al. [18] improved Neural style [10] for photo-realism. WCT2 enhances photorealism using wavelet transforms [40]. STROTSS [14] uses optimal transport for more general style transfer.

Algorithm 1: Deep Internal Learning for Image Enhancement (DILIE).

```
DILIE ($I, f, S, \phi, \alpha, \beta$):
    $I^{cfe}$ = CFE(($I, f, \phi, \alpha$))
    if $S$ then
        $I^*$ = SFE($I^{cfe}, S, \phi, \beta$)
        output $I^*$
    else
        output $I^{cfe}$ as $I^*$

CFE ($I, f, \phi, \alpha$):
    if $\alpha = 1$ then
        $I^{cfe}$ = ID($I, f, \phi$)        ▷ Image Decomposition.
    else
        $I^{cfe}$ = IR($I, f$)              ▷ Image Reconstruction.
    return $I^{cfe}$

SFE ($I^{cfe}, S, \phi, \beta$):
    if $\beta = 1$ then
        $I^*$ = PE($I^{cfe}, S, \phi$)      ▷ Photorealistic Style.
    else
        $I^*$ = AE($I^{cfe}, S, \phi$)      ▷ Artistic Style.
    return $I^*$
```

3. Our Approach

DILIE is a unified framework to restore the content features and synthesizes new style features for image enhancement. We have provided the DILIE framework in Algorithm 1. For simplicity, DILIE framework is described as follows (Eq. 1).

$$I^* = \text{DILIE}(I, f, S, \phi, \alpha, \beta). \tag{1}$$

Here, I denotes the input image and I^* is the output enhanced image. The encoder-decoder network f is used for the reconstruction or decomposition of input I. The style image S is used to enhance the style features of image I. The VGG19 network ϕ is used for image context learning [19] and the style features enhancement [10, 14]. DILIE framework performs content feature enhancement (CFE) and style features enhancement (SFE) separately. α and β are the parameters used for CFE and SFE. CFE enhances content features by learning deep features using encoder-decoder f. SFE uses the style image S for photo-realistic and artistic feature enhancement.

Fig. 2 illustrate CFE and SFE procedures for hazy image enhancement. The input is a hazy image I and the output is an enhanced image I^* (Eq. 1). CFE aims to separate a hazy image I into its underlying haze-free content image I^{cfe} and haze layers H. SFE aims to improve the style features of computed content image I^{cfe}. Formally, CFE performs image decomposition to separate hazy image I into content image I^{cfe}, haze layer H, and transmission map M. Here,

M is a mask that is used to compute the reconstruction of I from I^{cfe} and H. The reconstruction of I is compared with I using image decomposition loss $\mathcal{L}_{ID}$ to preserve relationship between I^{cfe}, H, and M. Finally, SFE takes haze-free content image I^{cfe} as input to compute the final enhanced image as I^*.

Algorithm 1 highlights CFE and SFE procedures. CFE performs content features enhancement by image decomposition or image reconstruction models. SFE performs style feature enhancement using a photo-realist style or artistic style. We describe CFE in Sec. 3.1 and SFE in Sec. 3.2.

3.1. Content Feature Enhancement

CFE could be majorly performed in the following two ways: image reconstruction (IR) and image decomposition (ID). The formulation of content feature enhancement is given in Eq. 2.

$$I^{cfe} = \text{CFE}(I, f, \phi, \alpha). \tag{2}$$

Here, I^{cfe} denotes the output of the content feature enhancement. The structure of the encoder-decoder network f provides an implicit image prior for the restoration of image features [34]. The corruption-specific image prior enables diverse applications. For example, dark channel prior for the image dehazing [9, 12] and encoder-decoder without skip connections as denoising prior [34]. The VGG network ϕ is used to extract the contextual features to compute the contextual content loss for preserving the context of image I in CFE output [23]. The parameter α denotes whether CFE is used to model image decomposition ($\alpha = 1$) or reconstruction ($\alpha = 2$).

3.1.1 Image Decomposition

Image decomposition (ID) improves the quality of images by separating image features and corrupted features. Suppose an image I as a combination of image feature layer and environmental noise. ID separates I into the image features layer I^{cfe} and the image corruption layer D, where the separation is determined by a mask M. In the image dehazing (Sec. 4.1), the mask is a transmission map that determines image features I^{cfe} and airlight H (*i.e.*, corruption layer D). ID is defined in Eq. 3.

$$(\theta_c^*, \theta_d^*, \theta_m^*) = \underset{(\theta_c, \theta_d, \theta_m)}{\arg\min}\ \mathcal{L}_{ID}\big(I; f_{\theta_c}, f_{\theta_d}, f_{\theta_m}\big). \tag{3}$$

Here, $\mathcal{L}_{ID}$ denotes the image decomposition loss. f_{θ_c}, f_{θ_d}, and f_{θ_m} are the instances of encoder-decoder network. θ_c is the parameter of image content layer, θ_d is the parameter of distortion layer, and θ_m is the parameter of mask M. z_c, z_d, and z_m are random vectors that are the inputs for the networks. Formally, Eq. 3 models the joint optimization to compute $I^{cfe} = f_{\theta_c^*}(z_c)$, $D = f_{\theta_d^*}(z_d)$, and $M = f_{\theta_m^*}(z_m)$. We have shown $\mathcal{L}_{ID}$ in Eq. 4.

$$\begin{aligned}\mathcal{L}_{ID}\big(I; f_{\theta_c}, f_{\theta_d}, f_{\theta_m}\big) = \Big\| \big(f_{\theta_m}(z_m) \odot f_{\theta_c}(z_c) \\ +\ (1 - f_{\theta_m}(z_m)) \odot f_{\theta_d}(z_d)\big) \\ - I \Big\|.\end{aligned} \tag{4}$$

Here, $\odot$ denotes Hadamard product. Eq. 4 shows that the layer separation is achieved by composing image I from image features $I^{cfe} = f_{\theta_c^*}(z_c)$ and corruption layer $D = f_{\theta_d^*}(z_d)$, and then minimizing pixel-wise differences. We will discuss the image decomposition for hazy image enhancement in Sec. 4.

The image decomposition in Eq. 4 does not consider context of the input image. The abstract information of content features represents the context of the image, *i.e.*, objects and their relative positions. The features extractor ϕ (VGG19) is used to preserve the context of the image using contextual content loss. The content features are mostly present at the higher layers of feature extractor ϕ denoted by ϕ^C and the style features are mostly contained at the initial layers denoted by ϕ^S [10]. The contextual content loss $\mathcal{L}_{CL}$ is defined between the content features of I and $I^{cfe} = f_{\theta_c}(z_c)$ as given in Eq. 5.

$$\mathcal{L}_{CL}\big(I, \phi; f_{\theta_c}\big) = -\log CX\big(\phi^C(f_{\theta_c}(z_c)), \phi^C(I)\big). \tag{5}$$

Here, CX denotes the contextual similarity [23]. CX is computed by using the cosine distance between feature vectors. CX is computed by finding for each feature $\phi^C(f_{\theta_c}(z_c))_i$ of the image I^{cfe}, the contextually similar feature $\phi^C(I)_j$ of the corrupted image I, and then sum over all the features in $\phi^C(f_{\theta_c}(z_c))$. We call the strategy above as the contextual similarity criterion. The key observation is that high-level content information (image context) is similar in both I^{cfe} and I. $\mathcal{L}_{CL}$ maximizes the contextual similarity between I^{cfe} and I to improve performance.

3.1.2 Image Reconstruction

Image reconstruction model (IR) uses encoder-decoder (ED) denoted by f_{θ_r} to reconstruct the desired image, where θ_r is the network parameters. Suppose z_r is input to the network f_{θ_r}. IR model is described in Eq. 6.

$$\begin{aligned}\theta^* = \underset{\theta}{\arg\min}\ \mathcal{L}_{IR}(I; f_{\theta_r}), \\ \text{where } \mathcal{L}_{IR}(I; f_{\theta_r}) = \big\| f_\theta(z_r) - \mathcal{T}(I) \big\|.\end{aligned} \tag{6}$$

Here, $\mathcal{L}_{IR}$ is the reconstruction loss and $\mathcal{T}$ is the image transformation function. The output of CFE in image reconstruction is $I^{cfe} = f_{\theta_r^*}(z_r)$. Note that $\mathcal{T}$ varies based on the application under consideration. For example, $\mathcal{T}$ is an identity function for denoising and $\mathcal{T}$ is a downsampling function for super-resolution [34]. The encoder-decoder network f_{θ_r} in Eq. 6 is observed to provide an implicit prior for image feature enhancement [34].

3.2. Style Feature Enhancement

We described that CFE enhances the content features of I. SFE aims to improve style features and output the enhanced image I^* given the CFE output I^{cfe}. SFE transfer the style features to I^{cfe} using style image S. We define SFE in Eq. 7.

$$I^* = \text{SFE}(I^{cfe}, S, f, \phi, \beta). \tag{7}$$

Here, I^* is the enhanced image and S is the reference style image. β represents the type of feature enhancement, *i.e.*, photo-realistic ($\beta = 1$) or painting style artistic ($\beta = 2$).

The style features enhancement is performed using the content loss $\mathcal{L}_C$ and style loss $\mathcal{L}_S$. The content loss $\mathcal{L}_C$ is defined between the content feature representations $\phi^C(I^{cfe})$ extracted from I^{cfe} and the content feature representations $\phi^C(I^*)$ extracted from I^*. The content loss is given by $\mathcal{L}_C = \mathcal{L}\big(\phi^C(I^{cfe}), \phi^C(I^*)\big)$. The style loss $\mathcal{L}_S$ is computed between the style feature representations $\phi^S(S)$ extracted from S and the style feature representation $\phi^S(I^*)$ of I^*. Formally, $\mathcal{L}_S = \mathcal{L}\big(\phi^S(S), \phi^S(I^*)\big)$. We provide the detailed description of $\mathcal{L}_C$ and $\mathcal{L}_S$ in the supplementary material.

SFE could be considered as photo-realistic or artistic features enhancement. The photo-realistic feature enhancement (PE) is aimed to minimize the distortion of object boundaries and preserve photo-realism using loss $\mathcal{L}_{PE}$. In contrast, the artistic feature enhancement (AE) allows small deformations to achieve an artistic look using loss $\mathcal{L}_{AE}$.

3.2.1 Photo-realistic Feature Enhancement

The photo-realism characterization in the image is an unsolved problem [18]. The enhancement of the photo-realistic features is based on the observation that if the input image is photo-realistic, then those features could be retained [18]. The image with lower perceptual errors is observed to be more photo-realistic [26]. Therefore, the quality of photo-realism in the output I^* is measured by the perceptual error score PieAPP [26].

The total loss for PE is defined as $\mathcal{L}_{PE} = \mathcal{L}_m + \mu \times \mathcal{L}_C + \kappa \times \mathcal{L}_S$, where μ and κ are the coefficients for the content loss $\mathcal{L}_C$ and the style loss $\mathcal{L}_S$. The affine loss $\mathcal{L}_m$ preserves the object structure while transforming the style features. More specifically, affine loss uses Matting Laplacian $\mathcal{M}_{I^{cfe}}$ of the input I^{cfe} [18], where $\mathcal{M}_{I^{cfe}}$ represents the grayscale matte for the content features. Intuitively, the affine loss function transforms the color distribution of I^* while preserving the object structure.

3.2.2 Artistic Feature Enhancement

We described that small image feature deformation could be present in the artistic style transfer. Therefore, the strategy is to match the distribution of the style and the content features and do not use the affine loss to reduce deformations in I^*.

The total loss for AE is defined as $\mathcal{L}_{AE} = \mu \times \mathcal{L}_C + \kappa \times \mathcal{L}_S$, where μ and κ are the coefficients for the content loss $\mathcal{L}_C$ and the style loss $\mathcal{L}_S$. We use relaxed earth mover distance (EMD) to match the image feature distribution [14]. The EMD loss preserves the distance between all the pairs of features extracted from the VGG19 ϕ to allow pixel value modification for style features while preserving the structure of the objects.[2]

4. Applications

4.1. Hazy Image Enhancement

Pictures taken in the hazy weather may lack scene information such as contrast, colors, and object structure. The image degradation model [38] for the hazy image is shown in Eq. 8.

$$I(p) = \hat{I}(p) \times M(p) + H(p) \times (1 - M(p)). \tag{8}$$

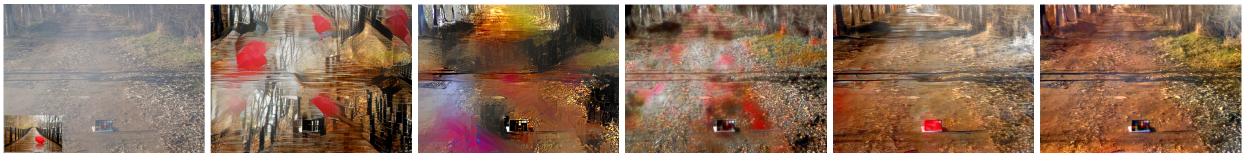
Here, p is the pixel location and I is the degraded observation. $\hat{I}$ is the haze-free image and M is the transmission map. Intuitively, the hazy image I could be considered as a haze layer H superimposed on the true scene content $\hat{I}$.

Image dehazing can be formulated as a layer decomposition problem to separates the hazy image (I) into a haze-free image layer (I^{cfe}) and a haze layer (H), where I^{cfe} is the approximation of haze-free image $\hat{I}$. We have discussed the generalized image decomposition framework for image enhancement in Eq. 3 (Sec. 3). We show its applicability for hazy image enhancement in Eq. 9 (Fig. 2).

$$\begin{aligned}(\theta_c^*, \theta_h^*, \theta_m^*) = \underset{(\theta_c, \theta_h, \theta_m)}{\arg\min}\ & \mathcal{L}_{ID}\big(I; f_{\theta_c}, f_{\theta_h}, f_{\theta_m}\big) \\ & + \mathcal{L}_{CL}\big(I, \phi; f_{\theta_c}\big).\end{aligned} \tag{9}$$

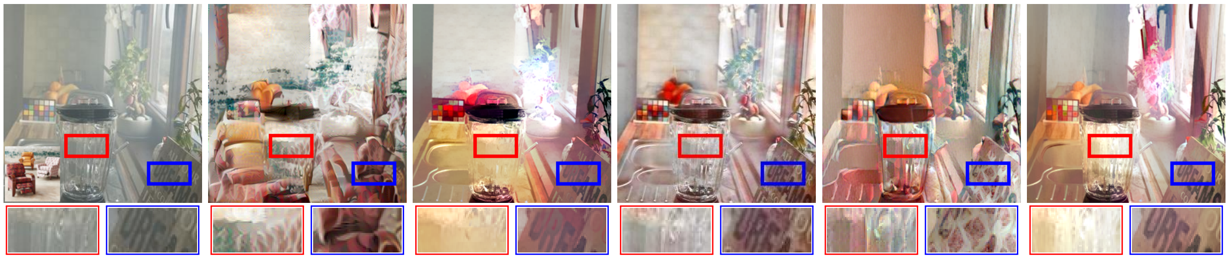
Here, $\mathcal{L}_{ID}$ is for image decomposition (Eq. 3) and $\mathcal{L}_{CL}$ is for preserving image context (Eq. 5). θ_h represents the parameters for haze layer. The transmission map $M = f_{\theta_m^*}(z_m)$ separates the haze-free image $I^{cfe} = f_{\theta_c^*}(z_c)$ and the atmospheric light $H = f_{\theta_h^*}(z_h)$. The joint framework is aimed to estimate $\hat{I}$ and H preserving their relations.

[2] We provide more details of DILIE framework in the supplementary material.



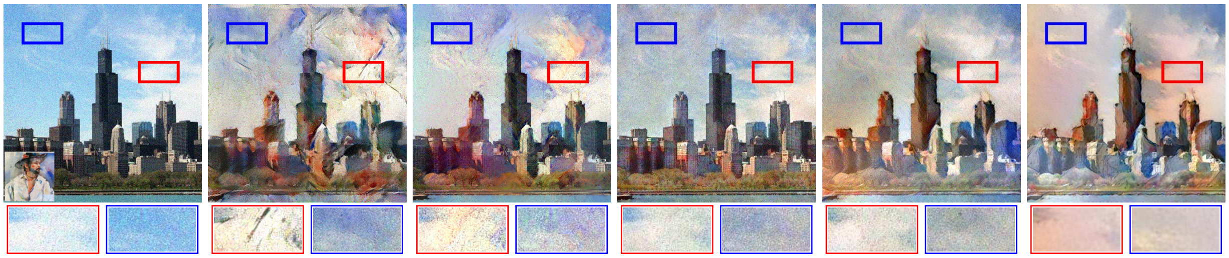
(a) Content & Style (b) Neural style [10] (c) DPST [18] (d) WCT2 [40] (e) STROTSS [14] (f) DILIE (ours)

Figure 3: **Hazy Image Enhancement (outdoor).** The content image contains haze and the style images are clear images (photo-realistic). Neural style [10] deforms the geometry of the objects. DPST [18] does not distribute image features well. WCT2 [40] output contains haze corruption, as shown by white spots. STROTSS [14] does not preserve fine image features details. It could be observed that DILIE (ours) output images with better visual quality.



(a) Content & Style (b) Neural style [10] (c) DPST [18] (d) WCT2 [40] (e) STROTSS [14] (f) DILIE (ours)

Figure 4: **Hazy Image Enhancement (indoor).** The figure shows the image feature enhancement of the indoor scene. It could be observed that DILIE outputs images with good quality (see the cropped images).



(a) Content & Style (b) Neural style [10] (c) DPST [18] (d) WCT2 [40] (e) STROTSS [14] (f) DILIE (ours)

Figure 5: **Noisy image enhancement.** The figure shows image enhancement for content images containing noise with strength $\sigma = 0.25$ using artistic style images.

	Neural [10]	**DPST [18]**	**WCT2 [40]**	**STROTSS [14]**	**DILIE**
I-Haze [2]	3.80	3.33	3.52	2.91	**2.78**
O-Haze [3]	3.00	2.71	2.88	2.81	**2.55**
Denoising 100	5.00	4.98	4.53	4.82	**4.27**

Table 1: The table shows that DILIE (ours) performs image enhancement with minimum perceptual error PieAPP [26].

	AODNet [15]	**MSCNN [27]**	**DcGAN [17]**	**GFN [28]**	**GCANet [4]**	**PFFNet [24]**	**DoubleDIP [9]**	**DILIE (ours)**
I-Haze [2]	0.732	0.755	0.733	0.751	0.719	0.740	0.691	**0.790**
O-Haze [3]	0.539	0.650	0.681	0.671	0.645	0.669	0.643	**0.705**

Table 2: The table shows SSIM comparison for dehazing of I-Haze and O-Haze dataset. DILIE outperforms other methods in comparison.

The main goal of Eq. 9 is to separate image features and haze features based on the semantics. The characteristics of haze particles in I are similar. Therefore, they accumulate into haze layer H. Similarly, the image features of I have similar characteristics and get separated into the haze-free image layer I^{cfe}. We have discussed contextual content

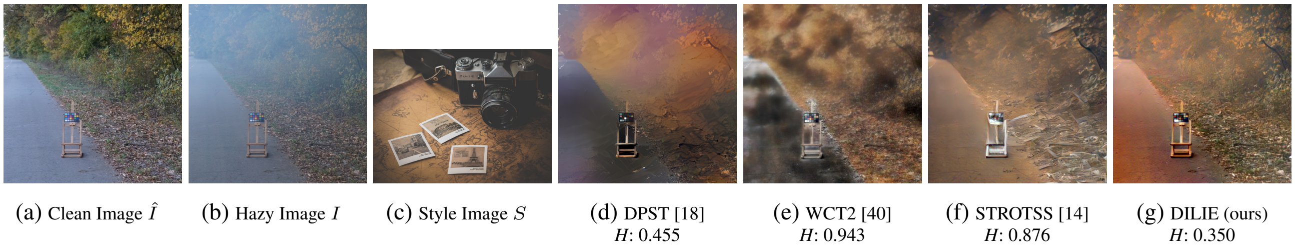


Figure 6: **Ablation Study (I).** The figure highlights the corruption of image features due to the haze in the enhanced output images. The style features (color information) of the outputs get affected by haze even when the input style image does not contain haze particles. H denotes the relative perceptual error due to haze computed using PieAPP [26]. DILIE output image with the minimum perceptual error and minimum effect from the haze.

loss $\mathcal{L}_{CL}$ given in Eq. 5 matches the contextual similarity between features. $\mathcal{L}_{CL}$ improves the performance of the layer decomposition framework.

Fig. 3 shows the image enhancement of the outdoor scene and Fig. 4 shows the enhancement of the indoor scene. The outdoor scenes mostly contain clouds and trees and the indoor images mostly contain objects present in the household. The hazy image enhancement improves the quality of image features of hazy outdoor and indoor scenes.

Table 1 shows that DILIE output images with better perceptual quality for hazy image enhancement[3]. Table 2 shows that DILIE achieves a good Structural Similarity Index (SSIM) for image dehazing. It is interesting to observe that the generalisability of DILIE (ours) allows good performance for both content feature enhancement (image dehazing) and style feature enhancement (hazy image enhancement).

Fig. 6 shows that if the input image contains haze particles, then the haze information gets incorporated into the output even when S does not include haze information. Ideally, the output should contain the content features from I and style features from S. The hazy image enhancement highlight that preserving a perceptually good balance between the style and the content features is challenging. CFE module removes haze features so that the final output I^* has less influence due to bad weather conditions.

4.2. Noisy Image Enhancement.

Denoising aims to recover a clean image from a noisy observation. The image degradation model for the noisy image is given as $I = \hat{I} + \epsilon$. Here, I the noisy image, $\hat{I}$ is the clean content image, and ϵ is the additive noise.

Image denoising is formulated as image reconstruction, where an encoder-decoder f reconstructs the clear image I^{cfe} from the noisy observation I. The network f provides a high impedance to noise and allows image features [34]. We have discussed the generalized framework for image reconstruction using transformation $\mathcal{T}$ in Eq. 6. Image denoising is performed by taking $\mathcal{T}$ to be identity function as given in Eq. 10.

$$I^{cfe} = f_\theta(z), \text{ where, } \theta^* = \arg\min_\theta \| f_\theta(z) - I \|. \quad (10)$$

Here, the restored image $I^{cfe} = f_\theta(z)$ is the approximation of $\hat{I}$. The reconstruction loss given in Eq. 10 is iteratively minimized, and early stopping is used to get the best possible outcome before the network over-learn the noisy features.

We make noisy image enhancement more challenging by using the style and the content images containing noise with the strength $\sigma = 0.25$. We show the output images in Fig. 5. It could be observed that DILIE gets a better distribution of features with better clarity (see cropped images). We have shown a quantitative comparison in Table 1. It can be observed that DILIE outperforms other methods in comparison.

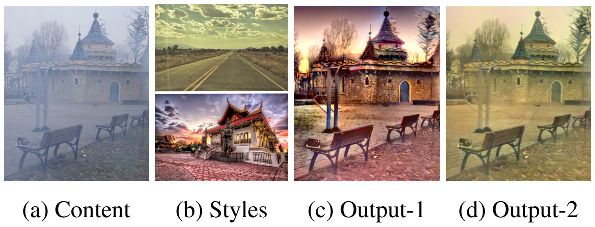
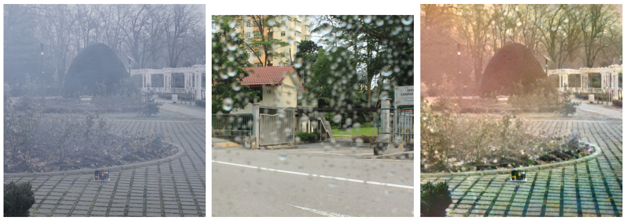


Figure 7: **Ablation Study (II).** The figure shows image enhancement of content image with two different style images using DILIE. Output-1 is with style image at the bottom and Output-2 is with style image at the top.

5. Ablation Studies

Fig. 6 illustrates that DILIE output images with less environmental noise. The quantitative comparison for haze corruption is described as follows. Consider the hazy image I, haze-free image $\hat{I}$, and the style image S (Fig. 6).

[3] We used implementation of Neural style provided in [32], Tensorflow implementation of DPS given in [1], contextual loss implementation in [22], STROTSS implementation in [25], and WCT2 implementation in [5]. We have provided more visual comparisons in the supplementary material.



(a) Content (b) Style (c) DILIE

Figure 8: **Limitation.** The figure shows weather change, where the content image is a hazy scene and the style image is a rainy scene.

The difference of image features between I and $\hat{I}$ is due to the haze. Let ST(y, z) denote the style transfer of content y using style z. Fig. 6 shows that when performing ST between I and S, the output image is observed to have haze corruption even when S does not have haze information.

To quantify haze corruption, let *E(*w*,* x*)* denote the perceptual error [26] between image w and image x. The relative error $H = \|E\big(\hat{I},\, \mathrm{ST}(\hat{I},\, S)\big) - E\,\big(\hat{I},\, \mathrm{ST}(I,\, S)\big)\|$ with reference to haze-free image $\hat{I}$ measures the deformations caused by haze in ST(I, S) by comparing ST output of the clean image $\hat{I}$ and the corrupted image I using perceptual error PieAPP [26].

Fig. 6 shows that DILIE output image with minimum perceptual error *H*. It could also be observed visually that in WCT2 [40] output contains haze corruption. DPST [18] and STROTSS [14] outputs also have haze effects when looking carefully. DILIE has the minimum haze effect [4].

Fig. 7 shows the results which are produced by the proposed framework with the same input but different reference images. Here, Output-1 is from the hazy content image and style-1 image (bottom). Output-2 is from the hazy content image and style-2 image (top). DILIE achieves image enhancement of the hazy image by incorporating a variety of style features.

6. Limitation

Fig. 8 shows the limitation of DILIE. The content image is a hazy image and the style image is a rainy image. The challenges are haze removal and improvising style features from a rainy scene with preserving object structure. The image features are not very clear in the output image. It could be because both style and content images have a high degree of corruption. We propose as future work to perform image enhancement when both content and style images are of bad weather conditions.

7. Implementation Details

DILIE uses instances of encoder-decoder networks (ED) for content feature enhancement. We used depth-5 encoder-decoder network (ED) in our experiments. Fig. 2 shows that hazy image enhancement uses three instances of ED networks for the separation of haze from image features. These ED networks do not share weights. ED uses convolution with strides for downsampling. For upsampling operation, we used bilinear upsampling and nearest neighbor upsampling.

We used pre-trained VGG19 as the feature extractor denoted by ϕ [10, 31]. VGG19 network extract style features, content features, and contextual features for image enhancement. It is interesting to note that VGG19 network is pre-trained for the image classification task on the ImageNet dataset [7]. We used the following layers of ϕ: conv1_2, conv2_2, conv3_2, and conv4_2. The content features are ϕ^C={conv4_2} and style features are ϕ^S={conv1_2, conv2_2, conv3_2} (Fig. 9).

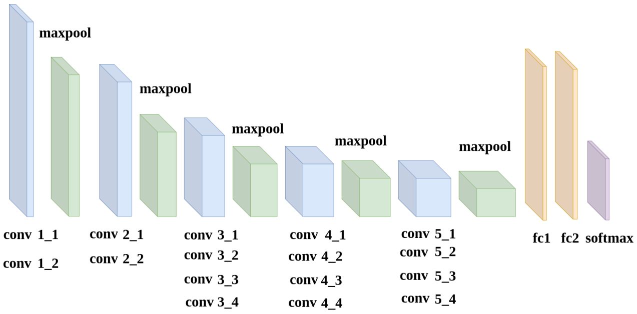


Figure 9: The figure illustrate layers of feature extractor (VGG19 Network ϕ) used in our experiments.

DILIE framework does not use a training dataset and uses only the corrupted input image and a reference image in the training process. The training process refers to iteratively minimizing the loss function. The corrupted image is used to train ED networks for content feature enhancement. The reference style image is used to provide style features extracted using a feature extractor.

8. Conclusion

We have discussed a deep internal learning framework for image enhancement (DILIE). The interesting challenge in image enhancement is that the degraded input image corrupts both style and content features. DILIE is a generic framework for content feature enhancement (CFE) and style feature enhancement (SFE). We show that CFE and SFE together lead to output images with a low perceptual error and good structure similarity. As future work, we propose to explore image enhancement for other image degradation models such as underwater scenes and snowfall.

[4] We discuss the ablation study more in the supplementary material.

References

[1] https://github.com/LouieYang/deep-photo-styletransfer-tf.

[2] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *arXiv:1804.05091v1*, 2018.

[3] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *IEEE Conference on Computer Vision and Pattern Recognition, NTIRE Workshop*, NTIRE CVPR'18, 2018.

[4] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383. IEEE, 2019.

[5] clovaai. https://github.com/clovaai/WCT2.

[6] Guang Deng. A generalized unsharp masking algorithm. *IEEE transactions on Image Processing*, 20(5):1249–1261, 2010.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[8] Raanan Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):1–9, 2008.

[9] Yossi Gandelsman, Assaf Shocher, and Michal Irani. Double-dip: Unsupervised image decomposition via coupled deep-image-priors. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 6, page 2, 2019.

[10] 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.

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020.

[12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[13] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[14] Nicholas Kolkin, Jason Salavon, and Gregory Shakhnarovich. Style transfer by relaxed optimal transport and self-similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10051–10060, 2019.

[15] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[16] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: a convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104:15–22, 2018.

[17] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[18] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6997–7005, 2017.

[19] Indra Deep Mastan and Shanmuganathan Raman. Dcil: Deep contextual internal learning for image restoration and image retargeting. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2366–2375, 2020.

[20] Indra Deep Mastan and Shanmuganathan Raman. Deepcfl: Deep contextual features learning from a single image. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2897–2906, 2021.

[21] Indra Deep Mastan and Shanmuganathan Raman. Deepobjstyle: deep object-based photo style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 711–719, 2021.

[22] Roey Mechrez. https://github.com/roimehrez/contextualLoss.

[23] Roey Mechrez, Itamar Talmi, and Lihi Zelnik-Manor. The contextual loss for image transformation with non-aligned data. *European Conference on Computer Vision (ECCV)*, 2018.

[24] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *Asian Conference on Computer Vision*, pages 203–215. Springer, 2018.

[25] Nkolkin13. https://github.com/nkolkin13/STROTSS.

[26] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1808–1817, 2018.

[27] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.

[28] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[29] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[30] Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. Ingan: Capturing and retargeting the “dna” of a natural image. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4491–4500. IEEE.

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[32] Cameron Smith. `https://github.com/cysmith/neural-style-tf`.

[33] Jean-Philippe Tarel and Nicolas Hautiere. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2201–2208. IEEE, 2009.

[34] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[35] Tianren Wang, Teng Zhang, and Brian C Lovell. Ebit: Weakly-supervised image translation with edge and boundary enhancement. *Pattern Recognition Letters*, 138:534–539, 2020.

[36] ECCV2020 Workshop. Deep internal learning: Training with no prior examples. `https://sites.google.com/view/deepinternallearning`.

[37] Junyi Xie, Hao Bian, Yuanhang Wu, Yu Zhao, Linmin Shan, and Shijie Hao. Semantically-guided low-light image enhancement. *Pattern Recognition Letters*, 138:308–314, 2020.

[38] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 702–717, 2018.

[39] Shibai Yin, Yibin Wang, and Yee-Hong Yang. A novel image-dehazing network with a parallel attention block. *Pattern Recognition*, 102:107255, 2020.

[40] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *International Conference on Computer Vision (ICCV)*, 2019.

[41] Lei Zhang and Wangmeng Zuo. Image restoration: From sparse and low-rank priors to deep priors [lecture notes]. *IEEE Signal Processing Magazine*, 34(5):172–179, 2017.