IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 9, 2021, accepted May 19, 2021, date of publication May 27, 2021, date of current version June 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084339

Diverse and Adjustable Versatile Image Enhancer

WOOJAE KIM[1], ANH-DUC NGUYEN[1], JINWOO KIM[1], JONGYOO KIM[2], HEESEOK OH[3], AND SANGHOON LEE[1,4], (Senior Member, IEEE)

[1]Department of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, South Korea
[2]Microsoft Research Asia, Beijing 100080, China
[3]Division of IT Convergence Engineering, Hansung University, Seoul 02876, South Korea
[4]Department of Radiology, College of Medicine, Yonsei University, Seoul 120-749, South Korea

Corresponding author: Sanghoon Lee (slee@yonsei.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) through the Korea Government (Ministry of Science and ICT, MSIT) under Grant NRF-2020R1A2C3011697, and in part by the Yonsei University Research Fund of 2021 under Grant 2021-22-0001.

ABSTRACT Enhancing the quality of photographs is a highly subjective process and depends on users' preferences. Hence, it is often more desired to let users choose their own best from a set of diverse and adjustable enhanced images with astounding quality. However, a system that can satisfy this requirement has not yet been established. While classical algorithms blindly enhance an image by filtering, recent intelligent enhancement systems can only do it with limited styles through learning from a set of single expert-retouched (ER) images. To fill this void, we propose a novel framework, Diverse and adjustable Versatile Image Enhancer (DaVIE), that learns from multiple ER images simultaneously. Thereby, it can output diverse results without being bound to a specific enhancement style while allowing users to freely adjust the level of enhancement. For ease of diversity, we adopt a variational auto-encoder (VAE) that learns stochastic distribution of enhancement styles. By using the VAE, the proposed model provides diversely enhanced images. To establish better control in terms of enhancement level, we propose a more general form of adaptive instance normalization and loss functions, which can afford even extreme image editing. Through rigorous experiments, we demonstrate that the proposed DaVIE framework yields visually pleasing and diverse results. We also show the proposed model quantitatively outperforms existing methods on the MIT-Adobe-5K dataset. Furthermore, through a strict user-study, we show that the users consider the qualities of ER images and machine-retouched images to be similar, with about 35% selection probability for DaVIE enhanced images.

INDEX TERMS Diverse image enhancement, adjustable learning, automatic photo enhancement, variational autoencoder, adaptive-instance normalization.

I. INTRODUCTION

Digital photography has been a revolutionary advancements in human artistic expression. With the emergence of innovative digital imaging technology, users can enjoy taking photos more easily regardless of time and place. However, individual users have their own preferences or purposes to capture a given real-world scenario [2]–[4]. So, users frequently want to retouch the captured photograph for their satisfaction. Thus, it becomes more and more important to provide user-friendly image manipulation tools and photograph enhancement techniques.

For many years, semi-automatic photo-retouching software such as *Lightroom* and *Photoshop* have been dominating this field. These software provide traditional image processing tools such as histogram equalization, denoising, deblurring, contrast enhancement, adaptive adjustments, and color mapping [5]–[9]. However, the quality of editing requires user's aesthetic outlook, software familiarity, and a deeper understanding of image processing algorithms, which causes high barriers to entry. Therefore, most users spend a long time familiarizing themselves with these tools. With the ongoing technological advancements, users should have access to a simple yet powerful image enhancement system that can satisfy a wide range of preferences.

Deep-learning techniques [10] have been successfully applied in photograph enhancers that allow users to enhance

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

FIGURE 1. Examples of the original image, the five expert-retouched images (experts A to E), and the six diversely enhanced images by DaVIE (DaVIE 1 to 6) with their enhancement level adjusted. The control parameters λ for the three enhancement levels *Original*, *Natural*, and *Strong* are set to 0, 1, and 3, respectively.

TABLE 1. Results of subjective rank test on MIT-Adobe-5K. The average rank score and standard deviation (std.) are reported for each expert-retouched set. Twenty-one subjects assessed the rank (Best (5) to Worst (1)) for randomly selected 100 sets (each set includes *five ER images*) on MIT-Adobe-5K.

Method	Expert A	Expert B	Expert C	Expert D	Expert E
Rank Score↑	3.29	2.44	3.07	2.82	**3.38**
(std.)	(1.67)	(1.46)	(1.40)	(1.19)	(1.15)

images to expert-level images automatically [11]–[14]. The MIT-Adobe-5K dataset [1] having 5K natural images and five corresponding expert-retouched (ER) images have been widely used for training them. However, most studies focused on only one type of ER images (mostly by expert C) while ignoring the rest; thus the learned style is unavoidably biased towards a specific preference for an expert. As depicted in Fig. 1, the five retouched styles (ER A to E) vary significantly. Based on this, we first performed a user-study that measures the users' subjective ranks of the five expert styles. Table 1 provides the result of the subjective rank test on the MIT-Adobe-5K dataset. As shown, the statistical users' preference is not biased towards a specific expert, even expert E showed slightly higher preference. This implies that winner-takes-all strategy is not suitable to reflect the actual preference distribution (The detailed protocol will be presented in Section IV). *Beauty* is in the eye of the beholder, so enhanced images should satisfy a wide range of viewer preferences. Ideally, the enhancement algorithm should be able to produce variously enhanced images so that users can choose their preferred styles. Furthermore, to better accommodate users' satisfaction, the system should let users adjust the degree of enhancement once they find the best style.

Motivated by these observations, we propose a novel framework termed *–Diverse and adjustable Versatile Image Enhancer (DaVIE)–* that presents diversely enhanced candidate images while the users' control the level of enhancement. To the best of our knowledge, this is the first work that offers diverse and manually adjustable image enhancement guided by multiple ER images. Fig. 1 shows six diverse results from DaVIE at three adjusted enhancement levels. In the figure, The DaVIE results show diverse styles compared to the original image and even the ER images. Moreover, the results at the *Natural* level demonstrate an similar level of enhancement to the ER images. On the other hand, when the level is *Original*, the results are similar to the original image with a slight enhancement effect. In contrast, when the enhancement level is *Strong* (extrapolation), the editing style is noticeably emphasized but still visually pleasing. The extrapolation outside the convex hull is challenging due to a lack of appropriate supervision signals.

To achieve this, we train a variational auto-encoder (VAE) based diverse image-to-image translation network. Moreover, to adjust the level of enhancement, we propose a generalized version of adaptive instance normalization (AdaIN) that provides control over the enhancement level. Hence, the users can freely choose their preferred image while adjusting the degree of enhancement according to their preference and can also produce excessive enhancement. Inspired by [15], we impose a constraint on the latent space to force the network to learn a locally linear manifold that can offer adjustable enhancement levels beyond what the network sees in training. By an extensive benchmark of various image enhancers, we demonstrate that DaVIE can provide multiple preferable candidates offering state-of-the-art performance. Moreover, from the user study, we demonstrate that the proposed scheme can be used practically.

The main contributions of our study are:

- A novel framework that learns a diverse image-to-image-translation model for user-oriented image enhancement that allows users choose image of their preference.

- A new adjustable enhancement scheme, which can be achieved by adjusting the translation of the original image toward the ER images in the latent space. To realize this, we propose a generalized version of AdaIN, called AdaIN_a and its necessary regularization terms. This scheme results in high-quality output, even for extreme-enhancement level.

The remainder of this paper is organized as follows. Section II discusses recent studies that used image enhancement techniques and diverse image generation models. Section III describes the architecture of the DaVIE framework and, the implementation of VAE based diverse and adjustable learning. Section IV discusses the experimental results of DaVIE and presents model for visualization and analysis. Finally, Section V concludes the paper.

II. RELATED WORK

A. DIGITAL IMAGE ENHANCEMENT

Image enhancement has been widely studied in the field of signal processing and computer vision. Several previous studies focused on low-level processing such as contrast adjustment [16], detail sharpening [17], and image denoising [18]. Recently, as data-driven approaches have become popular, high-level approaches are being studied extensively. Bychkovsky *et al.* [1] created a large-scale image-retouching database, which includes pairs of the original and the corresponding five ER images and trained machine learning models to predict the adjustment parameters. Hwang *et al.* [19] proposed context-aware image enhancement for automated processing. More recently, with the development in deep-learning, Yan *et al.* [20] proposed an automatic image adjustment method using a simple deep neural network. Ignatov *et al.* [21] proposed a CNN-based quality enhancement model. Hu *et al.* [22] developed a global retouching curve prediction model using a white-box framework. Gharbi *et al.* [23] attempted to learn a locally-affine model in bilateral space for real-time enhancement. Chen *et al.* [11] recently proposed the deep photo enhancer (DPE) technique that introduces unpaired generative adversarial networks (GANs) for automatic enhancement. Another stream is engaging in the enhancement of underexposed images. Chen *et al.* [24] made use of short- and long-exposure mapping, while Wang *et al.* [14] estimated a scaling illumination map. Park *et al.* presented dual autoencoder based low-light image enhancement [25], while Guo *et al.* also proposed a pipeline neural network for this task [26]. Kim *et al.* introduced patch-based principal energy analysis [27], and they also proposed maximal diffusion values based low-light image enhancement [28]. Ni *et al.* [29] presented unsupervised image enhancement method using GAN.

By using a reinforcement learning, Park *et al.* [30] proposed a global image modification model. To account for local adjustments, Moran *et al.* [13] trained several local parametric functions. Similarly, Kim *et al.* [12] proposed a two-stage approach that includes a channel-wise intensity transformation and local refinement network. Zeng *et al.* [31] proposed a learning framework for rapid image enhancement based on image-adaptive three-dimensional lookup tables (3D LUTs). However, most of the studies in this stream only considered one image pair (*i.e.,* a single pair of original-GT, with expert C in the MIT-Adobe-5K). This, of course, limits individual preference in applications.

In connection with personal-preference, several latent searching algorithms have been proposed to determine user-oriented target [32], [33]. For image enhancement, Kang *et al.* [34] proposed an enhancement module that observes user preferences and conducts personalized enhancement of unseen images. Similarly, Caicedo *et al.* [35] introduced a collaborative filtering method to discover clusters of user preferences for automatic enhancement. Recently, Bianco *et al.* [36] proposed personalized image enhancement using neural spline color transforms. Also, Kim *et al.* [37] introduced the concept of personalized image enhancement. They gathered users' preferred image sets and trained the enhancement module by embedding the estimated preference. However, it is still difficult to characterize the users' preferences precisely, which can vary depending on the users' momentary emotion or situation.

B. DIVERSE IMAGE GENERATION

Diverse image generative models are mainly classified into one of two categories: noise-to-image generation and image-to-image translation. The most popular approaches to noise-to-image generation are well-known VAE [38], [39] and GAN [40]. Especially, VAE learns high-dimensional distributions as a variational inference problem. Moreover, the learned latent distribution can be directly used to generate diverse and creative results while providing more stable training than GAN. For the image-to-image translation, in [41], they developed a conditional adversarial network as a general-purpose solution to image-to-image translation. However, since the generator relies on the target domain, the style of the generated image is still limited. More recently, AdaIN has been widely used in the encoder-decoder network [42]. In [43], [44], the content and style codes of the target were utilized and translated into those of another domain for diverse image-to-image translation. In [45], they applied a deep network interpolation for various image translation tasks.

Our approach takes advantage of both ways. While our model learns to translate the original image to an enhanced image, we realize diverse styles of enhancement by explicitly deploying a learned distribution of various styles using a variational approach. Moreover, we propose an adjustable learning scheme that provides a method to control the level of enhancement as shown in the examples of Fig. 1.

III. PROPOSED APPROACH

Fig. 2 presents an overview of the proposed framework for diverse and adjustable image enhancement. Suppose we have a set of training images: a low-quality original image I_o and

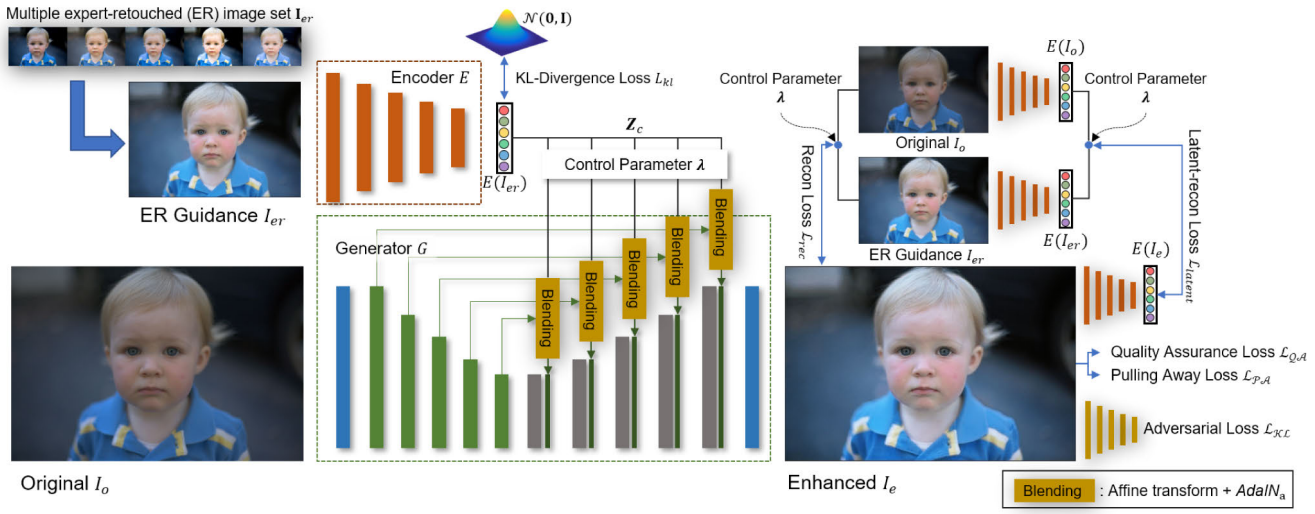


FIGURE 2. The network architecture of DaVIE. The structure mainly includes the encoder *E* and generator *G* modules while the blending module is responsible for modulating the instance guidance space to the enhanced image space.

its multiple ER image sets $\mathbf{I}_{er} = \{I_{er}^1, I_{er}^2, \ldots, I_{er}^N\}$ as instance guidance, where N is the total number of experts. From the given training set, the enhanced images I_e are produced while the entire pipeline is trained in an end-to-end manner.

Firstly, from the instance guidance set $\mathbf{I}_{er}$, the guidance image I_{er} is randomly selected and projected into a low-dimensional manifold $\mathcal{F}$ by the encoder. From the manifold, the generator obtains a blended feature and learns to transfer the instance space into the enhanced space.

To control the level of enhancement, we introduce two possible approaches: simple linear extrapolation and linear weighting into AdaIN when the model transfers the low-dimensional latent into the generator. Figs. 3(a) and (b) show the results of each approach with respect to the linear weight parameter λ. As shown in the figure, a simple extrapolating process is not sufficient to guarantee a satisfactory result. Therefore, we newly propose a generalized version of AdaIN called AdaIN_a that enables adjustable enhancement. AdaIN_a performs linearly approximated embedding over the manifold by using a control parameter λ.

Generally, the variational network suffers from a posterior collapse in the diverse generative scenario, which results in weak diversity of enhancement. To address this, our network also uses a pulling away (PA) loss that maximizes the distance of diversely enhanced images. Moreover, to assure the local quality of excessive enhancement (for $\lambda > 1$), we also utilize a modified version of structural similarity (SSIM) as a quality assurance (QA) loss.

A. DIVERSE AND ADJUSTABLE FRAMEWORK

1) PROBABILISTIC FRAMEWORK FOR DIVERSE ENHANCEMENT

To obtain a distribution of diverse enhancement samples, we take an amortized variational inference (AVI) approach that learns a global inference network to predict the parameters of the per-sample latent distribution, which is similar to the encoder of a VAE. AVI involves the minimization of the Kullback-Leibler divergence (KLD) between the variational distribution and the ground-truth (GT) posterior distribution. As optimizing directly is not feasible in practice, we follow the approach in [38], [46] and minimize a variational lower bound instead, as follows:

$$\begin{aligned}\log p(I_e|I_o) \geq\ & -\mathcal{KL}(f_\phi(Z_c|I_{er})||f_\psi(Z_c)) \\ & + \mathbb{E}_{Z_c\sim f_\phi(Z_c|I_{er})}[\log, g_\theta(I_e|Z_c, I_o)] \end{aligned} \quad (1)$$

where I_e, I_{er} and I_o are the enhanced, ER, and original images, respectively. Z_c is the latent style vector of instance guidance space. Further details on the derivation of the lower bound can be found in [38], [46]. f_ϕ, f_ψ, and g_θ are the posterior, prior, and likelihood parametrized by deep neural network parameters ϕ, ψ, θ, respectively. Here, the prior is set as $f_\psi = \mathcal{N}(0, \text{I})$. The first term ($\mathcal{KL}(\cdot)$) is used to ensure new enhancement styles to be drawn from a unit Gaussian, while the second term ($\mathbb{E}_{Z_c}(\cdot)$) denotes an auto-encoder reconstruction loss.

2) ADJUSTABLE ENHANCEMENT FRAMEWORK

An adjustable model can be learned the most intuitive by controlling the latent vector before passing it into the generator. For this, we use the concept of AdaIN [42] and further develop the module to incorporate adjustable property. The mathematical expression of AdaIN can be summarized as:

$$\text{AdaIN}(u, v) = \sigma(v)\frac{u - \mu(u)}{\sigma(u)} + \mu(v), \quad (2)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are the spatial *first-* and *second*-order statistics of the given two feature maps u and v. In our framework, u and v are the feature maps from the original and enhanced images, respectively. If we assume a locally linear manifold, we can easily draw new samples with given

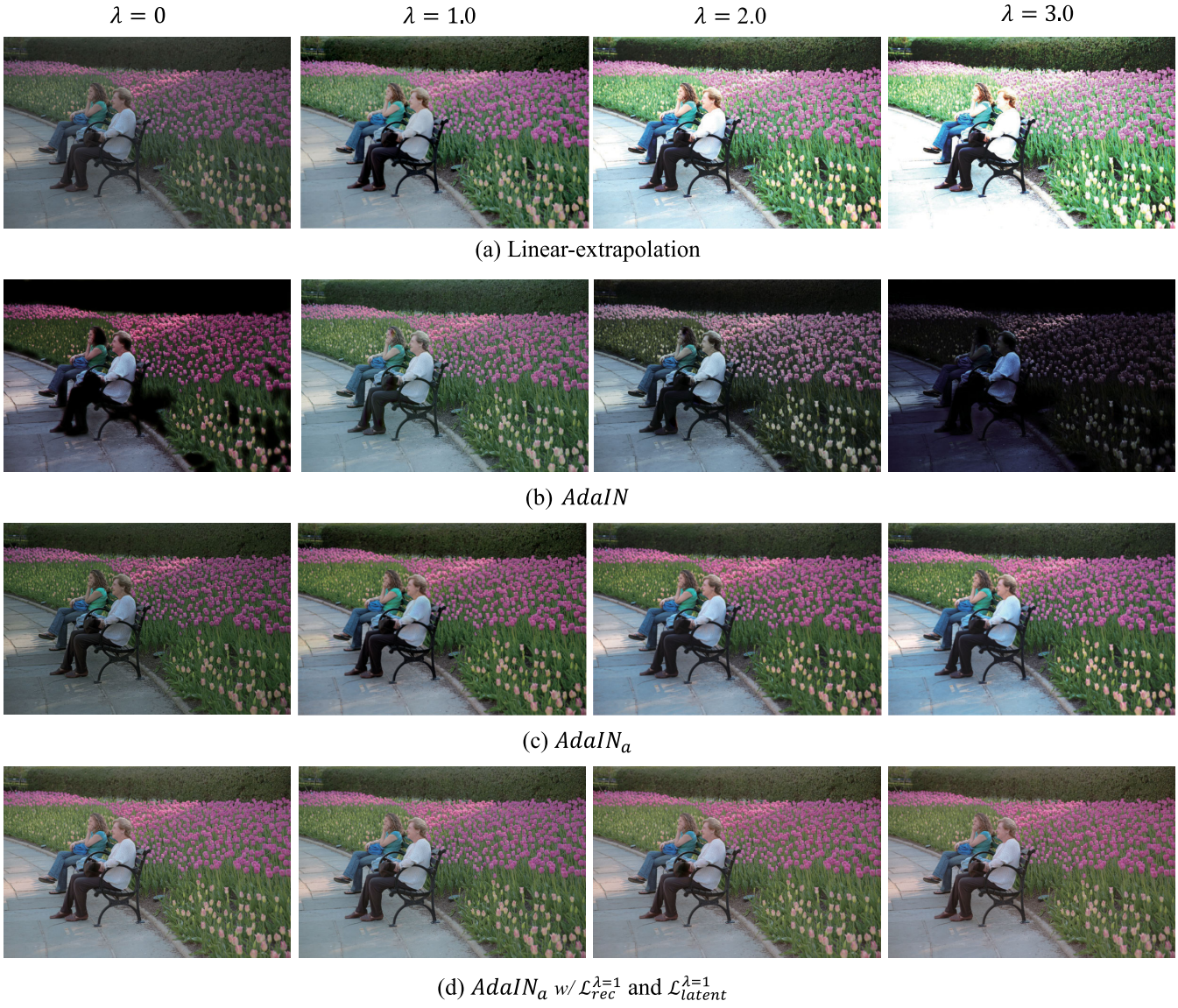


FIGURE 3. Comparisons of four different extrapolating schemes according to the control parameter λ: (a) simple linear extrapolation using the original image (λ=0) and the expert-C retouched image (λ=1) in the image space, (b) original AdaIN based DaVIE, (c) proposed AdaIN$_a$ based DaVIE, and (d) DaVIE *w/* $\mathcal{L}_{rec}^{\lambda=1}$ and $\mathcal{L}_{latent}^{\lambda=1}$, which is equivalent to a simple L_1 loss. Note that the subfigures (a)-(d) are described throughout Section III.

$\mu(u)$, $\mu(v)$, $\sigma(u)$, and $\sigma(v)$. Here, when the modulation parameters $\mu(v)$ and $\sigma(v)$ are linearly weighted, the result can be adjusted along the true manifold. However, without explicit regularization, the manifold may be highly non-linear. In this case, the linear sampling scheme offers a poor approximation and may lead to undesirable enhancements. For example, Fig. 3(b) shows an example when a linear weight is applied to the original AdaIN which leads to an unpredictable change when the control (weight) parameter increases.

To address this problem, we propose a generalized version of AdaIN termed AdaIN$_a$, which imposes a locally linear constraint for the neighborhood of the original image I_o and the enhanced image I_{er}.

We assume that if the similar levels of enhancement are close to each other in the image space, there also exists a latent space such that the latent representations of these images are also close to each other. From this assumption, we constrain the manifold by a locally linear rule

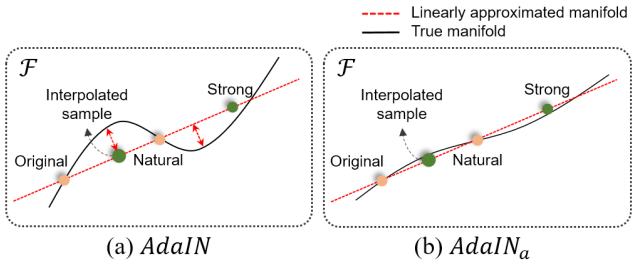


FIGURE 4. A schematic example of the true manifolds induced by AdaIN and AdaIN$_a$. (a) shows the manifold space learned by the original AdaIN and (b) shows the linearly approximated manifold space learned by proposed AdaIN$_a$. In (b), the linearly generated latent codes are distributed nearer to the true manifold (green samples).

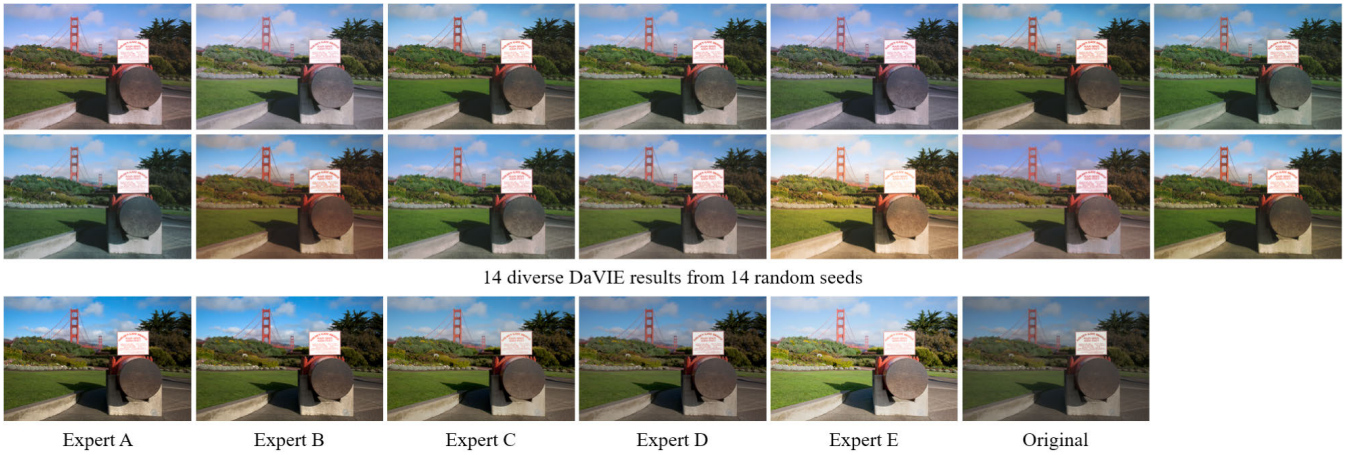


FIGURE 5. Examples of 14 diversely enhanced images (the control parameter λ is set to 3), five expert-retouched images (experts A to E), and the original image. Note that DaVIE results are enhanced from the random seeds.

dictated by a control parameter λ. Hence, we can approximate the locality of the true manifold by a linearly approximated manifold in the latent space $\mathcal{F}$. AdaIN_a is defined as follows:

$$\text{AdaIN}_\text{a}(u, v) = \frac{\sigma(u) + \lambda \cdot \vec{\sigma}(u, v)}{\sigma(u)}(u - \mu(u)) + \mu(u) + \lambda \cdot \vec{\mu}(u, v), \tag{3}$$

where $\mu(\cdot)$, $\sigma(\cdot)$, and λ are the spatial *first-* and *second-*order statistics of the input feature maps u and v, and the control parameter ($\lambda > 0$). $\vec{\mu}(\cdot)$ and $\vec{\sigma}(\cdot)$ denote the statistical direction of u and v, that is, $\vec{\mu}(u, v) = \mu(v) - \mu(u)$ and $\vec{\sigma}(u, v) = \sigma(v) - \sigma(u)$, respectively. Using this linear interpolation scheme, we implicitly force the network to produce plausible samples for the regions near the data point. As a result, the manifold learned by the network becomes smoother. Fig. 4 shows a difference between (a) AdaIN and (b) AdaIN_a over the interpolated samples between *Original*, *Natural*, and *Strong* (extrapolated) in the true manifold. As illustrated, AdaIN constrains the function at only the data point, and the distribution of manifold space shows high curvature. Thus, when we linearly sample from that manifold in (a), the distance between the sampled results and true manifold is large, which may not produce satisfactory outcomes. In contrast, when we employ AdaIN_a, the manifold is smoothened and becomes more suitable for our linear sampling strategy.

For clearer visualization, Figs. 3 (b) and (c) show the comparison of DaVIE trained with the original AdaIN and AdaIN_a. As can be seen, the adjusted images in Fig. 3(b) show high distortion as λ varies. However, as shown in Fig. 3(c), this embedding is intuitive for adjusting the level of enhancement, even for extrapolation ($\lambda > 1$). Nonetheless, trivially applying AdaIN_a does not immediately translate to high-quality adjusted results. To make this work properly, we introduce several new losses, as will be discussed in Section III-C.

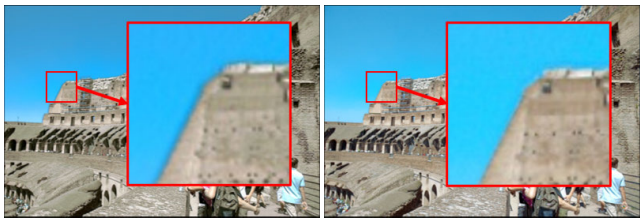
B. ARCHITECTURE

Our framework includes two separate modules: the encoder module E ($f_\phi(Z_c|I_{er})$) that projects the instance guidance space into low-dimensional manifold; the generator module G ($g_\theta(I_e|Z_c, I_o)$) which generates the diversely enhanced image I_e from the given original image I_o and the latent vector Z_c. The encoder module E includes six convolutional layers with down-samplings and ends with two dense layers that estimate the per-sample statistical mean and variance of the proposal distribution f_ϕ. Our generator G is based on the U-Net structure [47] which has a set of down- and up-sampling convolutional layers. To modulate the re-sampled latent vector into the generator, we use a blending block which contains affine transformation and AdaIN_a. In this block, the activation outputs (skips) of down-sampling layers, the latent vector Z_c, and the control parameter λ are fed into AdaIN_a for diverse and adjustable enhancement. For the encoder module and the down-sampling layers in the generator, we use batch-normalization [48]. During inference time, the latent vector is randomly sampled from a normal distribution instead of the ER guidance images. Therefore, DaVIE can provide multiple enhanced candidates from an infinite number of trials. Fig. 5 shows 14 diversely enhanced examples from a randomly generated latent vector. It can be seen that DaVIE provides diverse and visually pleasing results beyond the ER images.

Following the current trend, to produce better-looking output, we optionally employ an adversarial loss [11]. For the discriminator network D, we adopt a multi-scale discriminator [49] to differentiate real and fake.

C. ADJUSTABLE LOSS TERMS

Unlike previous works, the proposed model is based on AdaIN_a, which imposes a locally linear constraint on the manifold to adjust the level of enhancement. However, when using a usual L_1-norm as reconstruction loss, the adjustable effect is easily washed away during training. Fig. 3 (d) shows



(a) *w/o* $\mathcal{L}_{\mathcal{QA}}$ (b) *w/* $\mathcal{L}_{\mathcal{QA}}$

FIGURE 6. Effectiveness of the quality assurance loss for extrapolated results (λ is set to 2): (a) the result without $\mathcal{L}_{\mathcal{QA}}$ and (b) the result with $\mathcal{L}_{\mathcal{QA}}$. The artifacts on the edge lines are clearly removed in (b) (*w/* - with, *w/o* - without).

an example when DaVIE is trained using simple L_1 loss. As can be seen that the adjusted images with the control parameters $\lambda \in \{1, 2, 3\}$ still show similar effects. In other words, since we have only a single signal to supervise all the levels of outputs, if we use a simple L_1 loss, the network ignores λ. Therefore, to avoid the washing-away problem, we take into account the control parameter λ in the objective function explicitly. Intuitively, when the control parameter λ is closer to zero, the reconstruction loss equals the error between the enhanced image I_e and the original image I_o. In contrast, when λ is near to one, the loss function becomes closer to the error between the enhanced image I_e and the ER image I_{er}. Toward this, we define two locally linear embedded reconstruction losses: the content reconstruction loss and the latent reconstruction loss.

1) CONTENT RECONSTRUCTION LOSS

The content reconstruction loss $\mathcal{L}_{rec}$ is defined as an interpolated L_1-norm between the original image I_o and the ER image I_{er} as

$$\begin{aligned}\mathcal{L}_{rec} = &\max(0, (1-\lambda)) \cdot ||I_o - G(Z_c, I_o; \lambda)||_1 \\ &+ \lambda \cdot ||I_{er} - G(Z_c, I_o; \lambda)||_1,\end{aligned} \tag{4}$$

where $G(\cdot)$ is the generator module which outputs the enhanced image I_e and λ is the control parameter ($\lambda > 0$). Here, when $(1-\lambda)$ is negative, we manually set it to zero using $\max(0, (1-\lambda))$.

2) LATENT RECONSTRUCTION LOSS

Likewise, in the low-dimensional manifold, the latent reconstruction loss $\mathcal{L}_{latent}$ is also defined by an interpolated L_1-norm between the original image latent $E(I_o)$, the ER image latent $E(I_{er})$ as

$$\begin{aligned}\mathcal{L}_{latent} = &\max(0, (1-\lambda)) \cdot ||E(I_o) - E(I_e)||_1 \\ &+ \lambda \cdot ||E(I_{er}) - E(I_e)||_1,\end{aligned} \tag{5}$$

where $E(\cdot)$ and λ are, respectively, the encoder module that extracts the latent vector of each image and the control parameter.

3) QUALITY ASSURANCE LOSS

Since the proposed framework performs extrapolation ($\lambda > 1$), the strongly enhanced image unavoidably suffers from unexpected artifacts due to the non-linearity of the manifold. Fig. 6 (a) shows an example of visual artifacts when the control parameter is set to $\lambda = 2$. As can be seen, a structural degradation is distributed along the edge region. To address this problem, we maximize the structural consistency using SSIM [50], but we take only the structural term among the luminance, contrast, and structural similarities. This is because the other terms may also adversely affect our adjustable learning, which is the reason for using (4) and (5) instead of simple L_1 loss. The QA loss is defined as follows:

$$\mathcal{L}_{\mathcal{QA}} = \text{SSIM}_\text{s}(I_e, I_{er}), \tag{6}$$

where SSIM_s is the structural term of SSIM. As demonstrated in Fig. 6 (b), when our model is trained with $\mathcal{L}_{\mathcal{QA}}$, the artifacts are largely mitigated even the control parameter is extremely large.

D. DIVERSITY TERMS

1) KL DIVERGENCE LOSS

To achieve diverse enhancement, we minimize the typical interpretation of the KLD loss $\mathcal{L}_{\mathcal{KL}}$ between the encoded latent space $E(Z_c|I_{er})$ and multivariate normal distribution space $\mathcal{N}(0, \mathbf{I})$. The KLD loss is defined as

$$\mathcal{L}_{\mathcal{KL}} = \mathcal{KL}(E(Z_c|I_{er})||\mathcal{N}(0, \mathbf{I})). \tag{7}$$

2) PULLING AWAY LOSS

As mentioned above, VAE easily suffers the posterior-collapse problem. To solve this, we additionally maximize the diversity of the enhanced images $G(Z_c^1, I_o)$ and $G(Z_c^2, I_o)$ for certain latent codes Z_c^1 and Z_c^2. The pulling away constraint is defined as in [51]:

$$\mathcal{L}_{\mathcal{PA}} = \frac{||G(Z_c^1, I_o) - G(Z_c^2, I_o)||_1}{||Z_c^1 - Z_c^2||_1 + C}, \tag{8}$$

where Z_c^1 and Z_c^2 are two randomly sampled latent vectors, $G(Z_c^1, I_o)$ and $G(Z_c^2, I_o)$ are their enhanced images from the generator G, and C is a constant for numerical stability.

3) ADVERSARIAL LOSS

In our framework, the diversity relies only on the ER images. Nevertheless, we confirmed that the spatial characteristic of the enhanced image slightly follows the adversarial true set similar to DPE [11]. The discriminator loss $\mathcal{L}_{adv}$ is based on LSGAN [52]. As mentioned before, this term can be optionally used. The detailed experimental results will be introduced in Section IV.

E. TOTAL LOSS

Finally, the total loss function utilized for the training process is given by:

$$\begin{aligned}\mathcal{L}_{total} = &\gamma_{rec}\mathcal{L}_{rec} + \gamma_{latent}\mathcal{L}_{latent} - \gamma_{\mathcal{QA}}\mathcal{L}_{\mathcal{QA}} \\ &\times \gamma_{\mathcal{KL}}\mathcal{L}_{\mathcal{KL}} - \gamma_{\mathcal{PA}}\mathcal{L}_{\mathcal{PA}} + \gamma_{adv}\mathcal{L}_{adv},\end{aligned} \tag{9}$$

where γ_{rec}, γ_{latent}, $\gamma_{\mathcal{QA}}$, $\gamma_{\mathcal{KL}}$, $\gamma_{\mathcal{PA}}$, and γ_{adv} are the hyper-parameters to weight each term.

TABLE 2. Quantitative comparison with the state-of-the-art approaches on the MIT-Adobe-5K-*ER-C* (test set is same as in UPE [14]). ↑ indicates higher is better while ↓ implies smaller is better.

Methods	U-Net [47]	HDRNet [23]	White-Box [22]	DPE [11]	Distort-and-Recover [30]	UPE [14]	DLPF [13]	Kim *et al.** [12]	Zeng *et al.* [31]	DaVIE
SSIM↑	0.850	0.866	0.701	0.850	0.841	0.893	0.887	0.925	0.920	**0.915**
LPIPS↓	–	–	–	–	–	0.158	0.103	–	–	**0.091**
PSNR↑	22.24	21.96	18.57	22.15	20.97	23.04	24.48	25.88	25.06	**25.26**

* indicates numbers are taken from the original paper.

IV. EXPERIMENTAL RESULTS

A. IMPLEMENTATION DETAILS

Because the proposed DaVIE is guided by multiple ER image pairs, the experiments are mainly conducted on the MIT-Adobe-5K dataset [1]. Unlike existing studies, we fully utilize five ER image sets in the dataset as instance guidance and GT. We randomly divided the dataset into two separate training sets (2,250 images for training, 2,250 images for discriminator) and one test set (500 images), as same for their multiple ER images. The images are scaled to 256×256 resolution in the training session. In the experiment, the dimension of the latent vector was set as 256. We set the hyper-parameters as $\gamma_{rec} = 100$, $\gamma_{latent} = 1$, $\gamma_{\mathcal{QA}} = 10$, $\gamma_{\mathcal{KL}} = 1$, $\gamma_{\mathcal{PA}} = 1$, and $\gamma_{adv} = 1$, respectively. The range of the control parameter λ was set to $\lambda \in [0, 2]$ in the training. For optimization, we used the Adam optimizer [53] with ($\beta_1 = 0.0$, $\beta_2 = 0.99$) for the encoder, generator, and discriminator modules. We set the learning rate as 0.001 and decreased it by 0.95 for each of the 20 epoches. The batch size was set to 10, and the training was powered by three GPUs (RTX 2080 Ti 11GB).

B. QUANTITATIVE COMPARISON

We first compared the quantitative quality of DaVIE with existing image methods on the dataset. Since this is an initiative that diversely enhances images, it is required to select a specific sample that is the most similar to preference from the diverse results. Most of the previous studies trained only on a specific ER set (mostly expert C) and obtained only one solution. For fairness, we also benchmarked on widely compared ER-C. However, the proposed DaVIE outputs multiple enhanced candidates. Therefore, even if the quality of the diversely enhanced image is pleasing, it is likely to be different from the target ER image. To address this, we first enhanced each test image to a set of 30 images from the random seeds of three control parameters ($\lambda \in \{1.0, 2.0, 3.0\}$). From the enhanced image set, an image that is most quantitatively similar to the ER-C image selected for the comparison. Here, we used the mean squared error (MSE) for their similarity check (*i.e., similarity between 30 diverse images and ER-C*). In the experiment, a training/testing split follows UPE [14], and nine recent state-of-the-art image enhancement methods were benchmarked: U-Net based model [47], HDRNet [23], White-Box [22], DPE [11], Distort-and-Recover [30], UPE [14], DLPF [13], Kim *et al.* [12], and Zeng *et al.* [31].

TABLE 3. PSNR and SSIM comparison over five *ER* sets, and the *ER-C* results without $\mathcal{L}_{\mathcal{QA}}$ loss on MIT-Adobe-5K. *w/* and *w/o* indicate *with* and *without*, respectively.

Methods	Experts	PSNR↑	SSIM↑
DaVIE	*ER-A*	22.50	0.896
	ER-B	25.01	**0.936**
	ER-C	**25.26**	0.915
	ER-D	23.08	0.918
	ER-E	21.48	0.905
DaVIE *w/o* $\mathcal{L}_{\mathcal{QA}}$	*ER-C*	23.61	0.886
DaVIE-AdaIN		20.41	0.862
DaVIE *w/* $\mathcal{L}_{rec}^{\lambda=1}$, *w/* $\mathcal{L}_{latent}^{\lambda=1}$		23.86	0.867

We employed three well-utilized metrics for quantitative comparison, namely: SSIM [50], LPIPS [54], and PSNR, between the enhanced images and GTs. Table 2 tabulates the performance comparison over the tested models on MIT-Adobe-5K-*ER-C*. As shown in the table, DaVIE delivers higher quantitative scores than most of the benchmarked models except a recent work by Kim *et al.* [12]. Nonetheless, the proposed model provides more diverse enhancement. For clearer demonstration, Table 3 lists the individual performance comparison over the five ER sets as GTs in the MIT-Adobe-5K dataset. As may be seen, proposed DaVIE delivers reliable scores on the most ER sets, even though these are tested by a single trained model.

C. VISUALIZATION RESULTS

Fig. 7 reports examples of our diverse and adjustable enhancement with their multiple ER images. Fig. 8 presents the enhanced results of DaVIE in comparison to those of the existing image enhancers (DPE, UPE, DLPF) over five ER images. Note that all of benchmarked models are reproduced by pretrained models. As shown in Fig. 7, when the control parameter λ is equal to zero, the results are closer to the original image. Conversely, as λ increases, the level of enhancement proportionally increases towards the GTs, and it is still visually pleasing (as in Fig. 1). Specifically, the results from DaVIE A to E are highly similar to five ERs (experts A to E) while DaVIE F provides more diverse and pleasing results that go beyond the diversity of ER images (experts A-E). Furthermore, when DaVIE is compared to existing methods, it suggests more diverse candidates while the benchmarking methods are mostly close to expert C in Fig. 8.

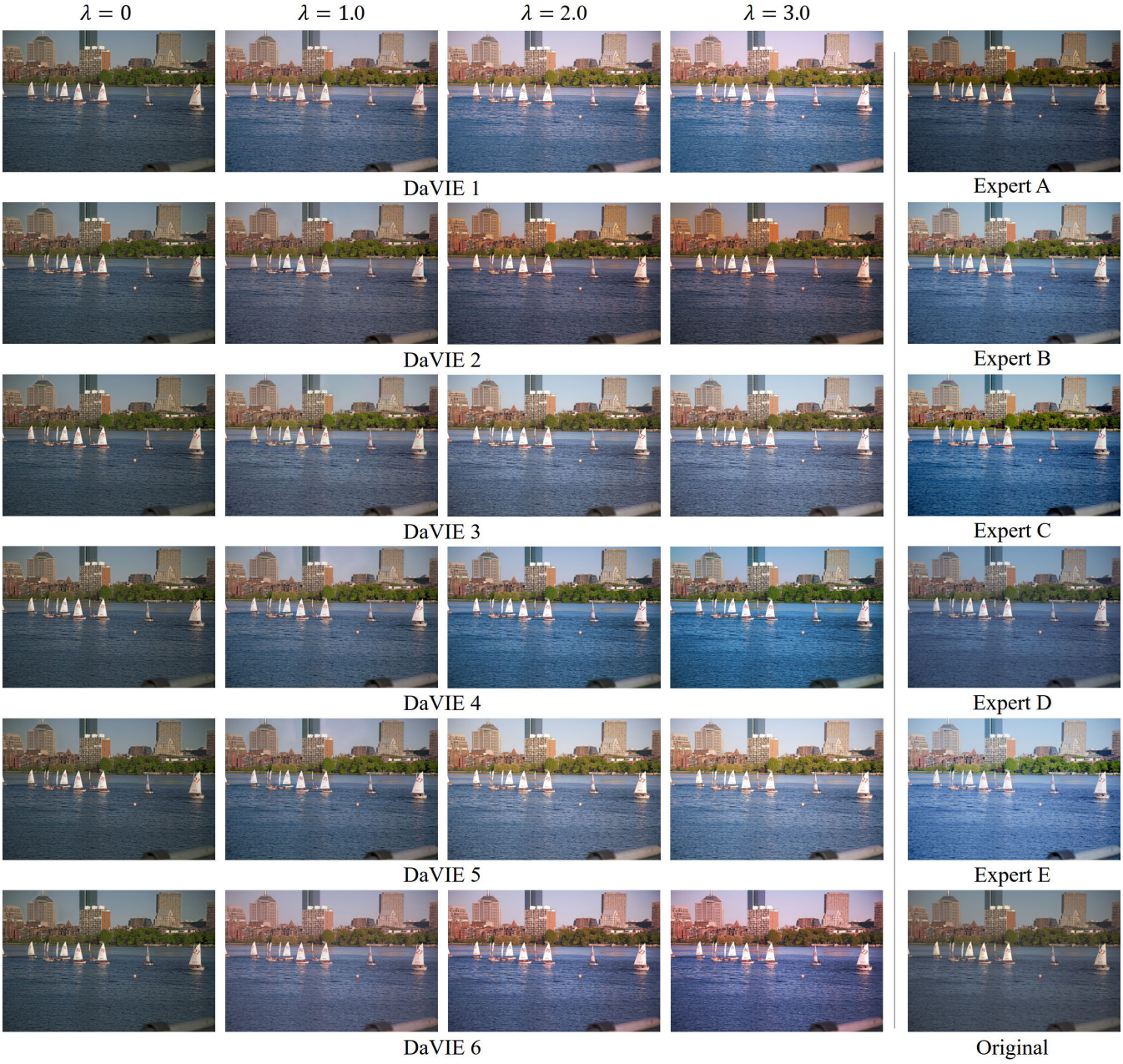


FIGURE 7. Examples of six diversely enhanced images (DaVIE 1 to 6) with four control parameters ($\lambda \in \{0, 1.0, 2.0, 3.0\}$), five expert-retouched images (experts A to E), and the original image. Note that DaVIE 1-6 results are enhanced from the random seeds.

D. QUANTITATIVE DIVERSITY

We also show the quantitative diversity via the LPIPS metric score [54]. As in [43], [56], an average LPIPS score is a well-utilized indicator of diversity quantification. Here, we use the entire test set, and twenty-diversely enhanced image pairs per input were utilized, which amounts to 10K pairs. Table 4 reports the average LPIPS scores *w.r.t.* λ. As may be seen, DaVIE gradually increases the diversity scores when λ increases. This tendency also can be clearly shown in Figs. 1 and 7. Moreover, to validate the diversity when using $\mathcal{L}_{\mathcal{PA}}$, we report the ablation results of $\mathcal{L}_{\mathcal{PA}}$. When DaVIE trained without $\mathcal{L}_{\mathcal{PA}}$, the LPIPS score decreases by nearly half.

TABLE 4. Quantitative diversity comparison with different control parameters and without $\mathcal{L}_{\mathcal{PA}}$ loss on MIT-Adobe-5K-*ER*-*C*. The diversity score is the average LPIPS metric.

Methods	LPIPS↑ (λ=0.0)	LPIPS↑ (λ=1.0)	LPIPS↑ (λ=2.0)	LPIPS↑ (λ=3.0)	LPIPS↑ (*all*)
DaVIE	0.0057	0.0128	0.0185	0.0248	0.0223
DaVIE *w/o* $\mathcal{L}_{\mathcal{PA}}$	0.0028	0.0046	0.0092	0.0134	0.0103

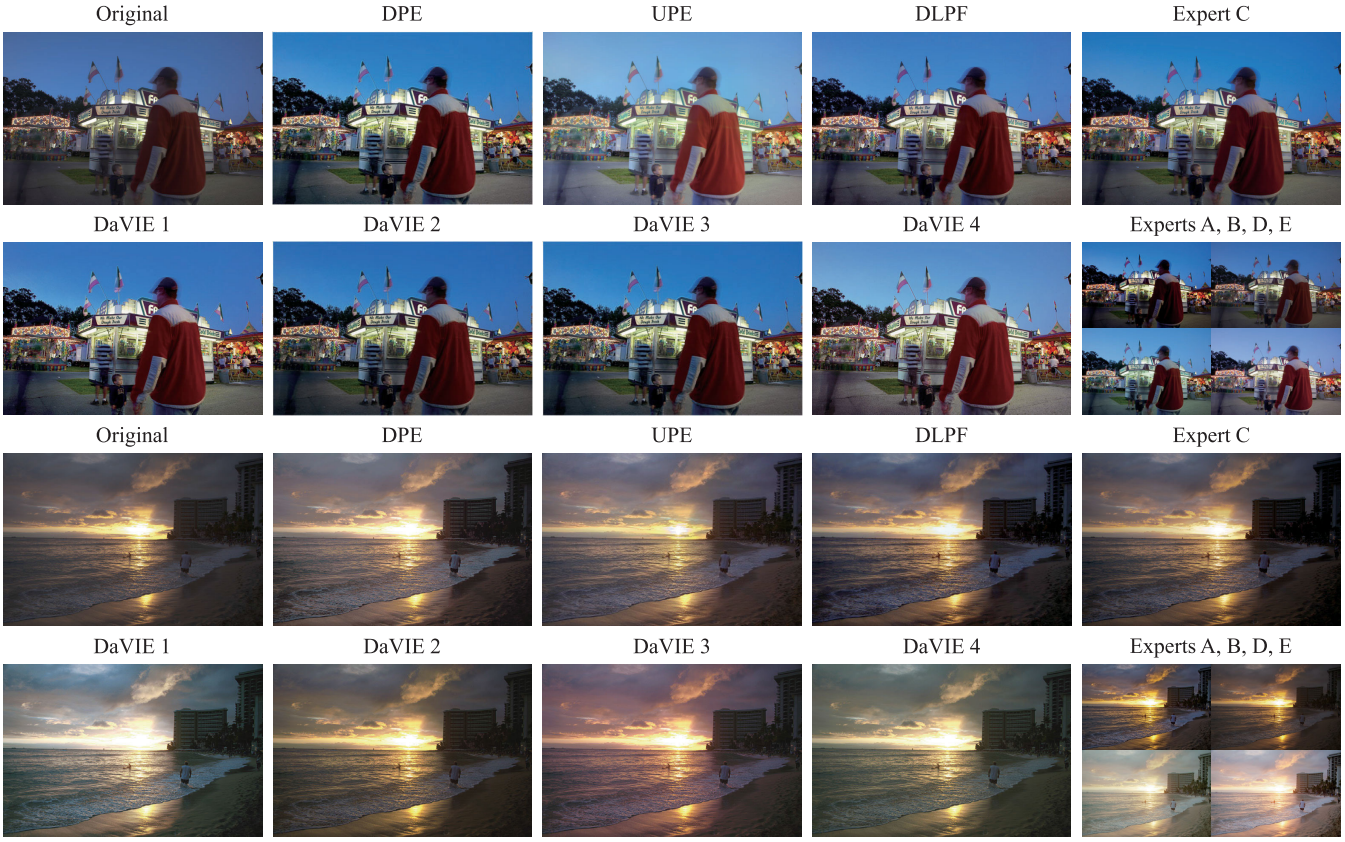


FIGURE 8. Comparisons of four diversely enhanced DaVIE results (DaVIE 1-4) with three existing methods (DPE [11], UPE [14], and DLPF [13]) and five ER images.

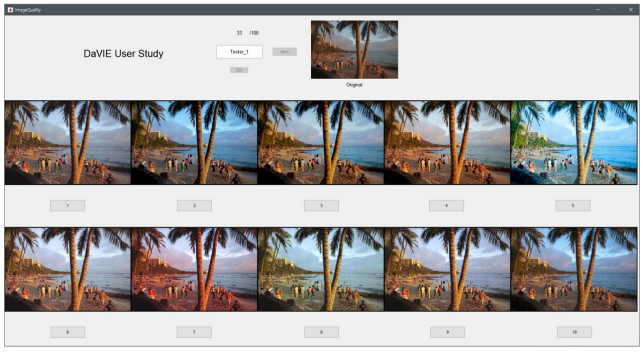


FIGURE 9. Screen-shot of the user interface for the preference user study.

E. ABLATION STUDY

We further investigate the effectiveness of our contributions one by one. Here, we quantitatively compare three versions: DaVIE without QA loss $\mathcal{L}_{\mathcal{QA}}$, DaVIE with AdaIN, and DaVIE with L_1 loss. At the bottom of Table 3, we report the results of three versions tested on the MIT-Adobe-5K-*ER-C*. As mentioned in Section III-C, it can also be seen that removing $\mathcal{L}_{\mathcal{QA}}$ decreases the performance compared to DaVIE-*ER-C*. Also, when we trained the model with original AdaIN, the performance was the poorest as predicted in Section III-A. In addition, when DaVIE is trained with the usual L_1 loss, the performance is relatively lower than that of DaVIE using the proposed interpolation-based reconstruction losses. Overall, when all the loss terms are used together, DaVIE exhibits the best performance.

F. USER STUDY

1) PROTOCOL

In our user study, 21 subjects participated satisfying the subject criteria recommended in ITU-R BT. 500 [57], [58]. Their ages ranged from 21 to 38 years, and they had normal vision (all of the subjects were screened for normal visual acuity on the Landolt chart). A 65-inch monitor with UHD resolution (Samsung UN65JU7500F) was used to display images. In our subjective examination, the following two sessions were conducted: (1) ranking five ER images in the MIT-Adobe-5K dataset and (2) selecting the most preferred image from the 10 randomly shuffled images that contain five ER images and five enhanced images by DaVIE. Before each session, we informed each subject individually to help them understand the goal and procedure of the tests.

2) SUBJECTIVE RANK TEST

To clarify the users' preferences, we first conducted the subjective rank test. For this test, we randomly selected 100 images from MIT-Adobe-5K. Each testing set included the original image and five ER images. Each subject viewed them at the same time, and was asked to rank the five ER

FIGURE 10. In-the-wild test examples on the Flickr30K [55] dataset. Each example set includes four DaVIE enhanced results (DaVIE 1-4) with three control parameters ($\lambda \in \{0, 1.0, 3.0\}$) and the original image (Original 1-4).

images in order of preference. From the first to the last ranks, we scored each from 5 to 1, then averaged the scores for each ER set as shown in Table 1. As can be seen in the table, all ER images seem to have similar preference scores. This implies that it is necessary to take into account various styles of enhancement to reflect the real-world user preference rather than depending on only one style (as done in most studies) [11], [13].

3) PREFERENCE USER STUDY

We designed a *'Turing-test'* based user-study so that half of the images were ER images and the other half were the DaVIE enhanced images (five human-enhanced vs. five machine-enhanced images). It means that an ideal selection probability of 50% indicates that the users are totally confused between expert- and machine-retouched images [59].

Similar to the rank test, we also randomly selected 100 image sets from the MIT-Adobe-5K dataset. We first generated five diverse images from the random seeds with a fixed control parameter of $\lambda = 3$. Then, 11 images (the original image, five DaVIE results, and five corresponding ER images) were shown to each subject using our user-interface as shown in Fig. 9. To ensure a fair comparison, it was not specified whether the images were DaVIE enhanced or ER. Then, each user selected the top-five preferred images for each testing set. After subjects assessed 100 image sets, the overall selected ratios between DaVIE results and ER images were obtained.

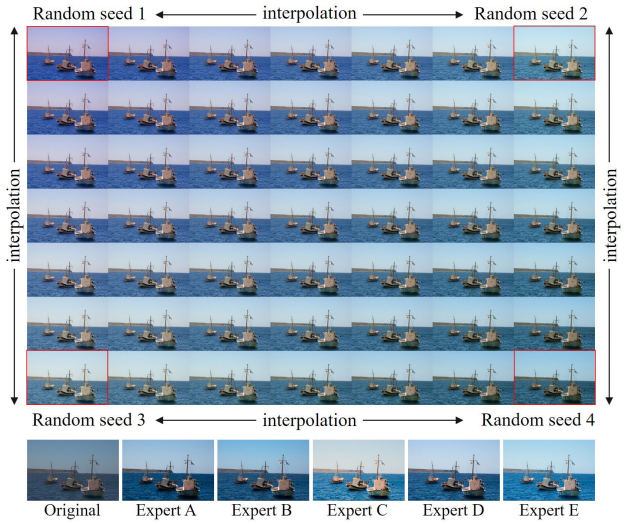


FIGURE 11. Samples enhanced from interpolated latent representations of four independent latent vector. The control parameter λ was set to 3.

In the experiment, the subjects selected DaVIE enhanced images with a probability of 34.53%. The result shows that for more than *one-third of the time* users may mistake our DaVIE results with real ER images, or in other words, more than one-third of our DaVIE results were of the highest standard. Even though 34.5% is not reaching the ideal 50% yet, DaVIE is significantly competitive in comparison with five human experts' results. The selected percentage of individual experts

TABLE 5. Additional results of a subjective user study*.

Ratio	Expert A	Expert B	Expert C	Expert D	Expert E
% ↑	13.35%	12.34%	13.14%	12.73%	13.91%

*the ratio of selecting preferred images from five DaVIE enhanced images and five ER images on the MIT-Adobe-5K dataset. Each ratio indicates how much users preferred each result.

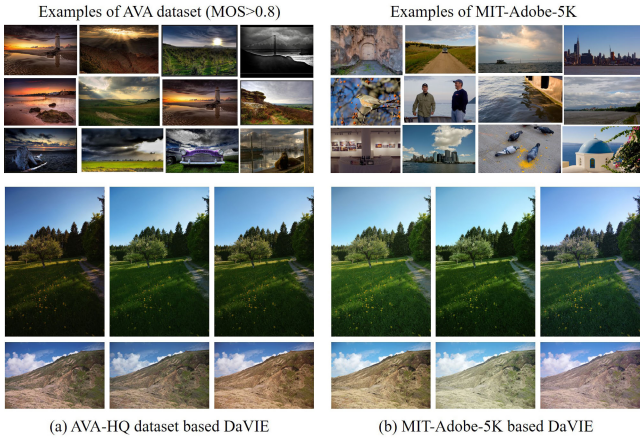


FIGURE 12. Effects of two different adversarial true-sets: (a) AVA-HQ based DaVIE, (b) MIT-Adobe-5K based DaVIE.

is shown in Table 5, where each ER image set was selected with about 13% individual preference. This again suggests that the user's preferences are also not biased toward one expert style, and it can be seen that DaVIE can replace them with a 34.53% probability. Thus, the proposed DaVIE is highly competitive in terms of subjective quality or the preference.

G. IN-THE-WILD TEST

We also visualized the results of in-the-wild images on the Flickr30K dataset [55]. Fig. 10 depicts four examples of sets using DaVIE from the Flickr30K dataset. Each image is enhanced diversely from four random seeds with three control parameters $\lambda \in \{0, 1, 3\}$. It can be seen that the results are diversely enhanced and visually pleasing. Moreover, as the enhancement level increases, each style is emphasized.

H. LATENT INTERPOLATION

The latent representation from the feature extraction process was analyzed. The proposed model generates enhanced images from the random seed. To further analyze this, we conducted a latent-interpolation experiment using bi-linear interpolation. We first randomly chose four latent representations with the control parameter λ of 3. Then, we synthesized a convex collection of 49 latent codes using bilinear interpolation. Finally, we enhanced the input image from the interpolated latent codes and the results are reported in a 7×7 grid as shown in Fig. 11. As can be seen, DaVIE smoothly interpolates between four enhancement styles, which implies that the learned manifold is also locally smooth.

I. DIFFERENT ADVERSARIAL TRUE-SET

As previously mentioned, the adversarial loss can be optionally used in our framework. Since our model is based on paired-learning, the adversarial true-set does not significantly affect the enhanced results; however, we observed that the spatial characteristic was slightly changed *w.r.t.* the adversarial true-set. Fig. 12 shows the examples of two different adversarial true-sets for DaVIE: AVA-HQ [60] and MIT-Adobe-5K dataset [1]. For the AVA-HQ dataset, we selected 1000 high-aesthetic quality images with higher mean opinion scores (MOSs) (MOS > 0.8) in the AVA dataset [60]. As can be seen, the overall images in AVA-HQ are dark, and DaVIE results with the AVA-HQ adversarial true set show also very dark but visually pleasing. In contrast, when we used the MIT-Adobe-5K dataset as an adversarial true-set, the results are much brighter and visually pleasing.

V. CONCLUSION

Unlike existing *winner-take-all* approaches, in this study, we explored a diverse and adjustable image enhancer by utilizing the AVI approach and locally linear embedding based AdaIN_a. Through a diverse and adjustable scheme, the proposed DaVIE provides diversely enhanced multiple-images so that users can choose their own preferred images. The diversely enhanced images are favorable among all the benchmarking models and sometimes even makes viewer confuse with professionally retouched images. In particular, the proposed AdaIN_a can be effectively applied to other image-to-image translation tasks for controlling the level of transfer. In the future, we plan to customize the proposed framework to an individual-preference-considered enhancement according to user satisfaction with personal devices.

REFERENCES

[1] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 97–104.
[2] H. Kim, S. Ahn, W. Kim, and S. Lee, "Visual preference assessment on ultra-high-definition images," *IEEE Trans. Broadcast.*, vol. 62, no. 4, pp. 757–769, Dec. 2016.
[3] W. Kim, S. Ahn, A.-D. Nguyen, J. Kim, J. Kim, H. Oh, and S. Lee, "Modern trends on quality of experience assessment and future work," *APSIPA Trans. Signal Inf. Process.*, vol. 8, p. e23, Oct. 2019.
[4] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87884–87917, 2020.
[5] H. Xu, G. Zhai, X. Wu, and X. Yang, "Generalized equalization model for image enhancement," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 68–82, Jan. 2014.
[6] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
[7] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.
[8] J. Lim, M. Heo, C. Lee, and C.-S. Kim, "Contrast enhancement of noisy low-light images based on structure-texture-noise decomposition," *J. Vis. Commun. Image Represent.*, vol. 45, pp. 107–121, May 2017.
[9] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 6306–6314.

[12] H.-U. Kim, Y. J. Koh, and C.-S. Kim, "Global and local enhancement networks for paired and unpaired image enhancement," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2020, pp. 339–354.

[13] S. Moran, P. Marza, S. McDonagh, S. Parisot, and G. Slabaugh, "DeepLPF: Deep local parametric filters for image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12826–12835.

[14] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6849–6857.

[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[16] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 889–896, May 2000.

[17] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast local Laplacian filters: Theory and applications," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 1–14, Sep. 2014.

[18] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 60–65.

[19] S. J. Hwang, A. Kapoor, and S. B. Kang, "Context-based automatic local image enhancement," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Florence, Italy: Springer, 2012, pp. 569–582.

[20] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu, "Automatic photo adjustment using deep neural networks," *ACM Trans. Graph.*, vol. 35, no. 2, pp. 1–15, May 2016.

[21] A. Ignatov, N. Kobyshev, R. Timofte, and K. Vanhoey, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3277–3285.

[22] Y. Hu, H. He, C. Xu, B. Wang, and S. Lin, "Exposure: A white-box photo post-processing framework," *ACM Trans. Graph.*, vol. 37, no. 2, pp. 1–17, Jul. 2018.

[23] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.

[24] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3291–3300.

[25] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22084–22093, 2018.

[26] Y. Guo, X. Ke, J. Ma, and J. Zhang, "A pipeline neural network for low-light image enhancement," *IEEE Access*, vol. 7, pp. 13737–13744, 2019.

[27] W. Kim, "Image enhancement using patch-based principal energy analysis," *IEEE Access*, vol. 6, pp. 72620–72628, 2018.

[28] W. Kim, R. Lee, M. Park, and S.-H. Lee, "Low-light image enhancement based on maximal diffusion values," *IEEE Access*, vol. 7, pp. 129150–129163, 2019.

[29] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Trans. Image Process.*, vol. 29, pp. 9140–9151, 2020.

[30] J. Park, J.-Y. Lee, D. Yoo, and I. S. Kweon, "Distort-and-recover: Color enhancement using deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 5928–5936.

[31] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 25, 2020, doi: 10.1109/TPAMI.2020.3026740.

[32] C.-H. Chiu, Y. Koyama, Y.-C. Lai, T. Igarashi, and Y. Yue, "Human-in-the-loop differential subspace search in high-dimensional latent space," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 1–85, Jul. 2020.

[33] Y. Koyama, I. Sato, D. Sakamoto, and T. Igarashi, "Sequential line search for efficient visual design optimization by crowds," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, Jul. 2017.

[34] S. B. Kang, A. Kapoor, and D. Lischinski, "Personalization of image enhancement," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1799–1806.

[35] J. C. Caicedo, A. Kapoor, and S. B. Kang, "Collaborative personalization of image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 249–256.

[36] S. Bianco, C. Cusano, F. Piccoli, and R. Schettini, "Personalized image enhancement using neural spline color transforms," *IEEE Trans. Image Process.*, vol. 29, pp. 6223–6236, 2020.

[37] H.-U. Kim, Y. J. Koh, and C.-S. Kim, "PieNet: Personalized image enhancement network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2020, pp. 374–390.

[38] D. P Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: http://arxiv.org/abs/1312.6114

[39] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 3483–3491.

[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[42] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1501–1510.

[43] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 172–189.

[44] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 35–51.

[45] X. Wang, K. Yu, C. Dong, X. Tang, and C. C. Loy, "Deep network interpolation for continuous imagery effect transition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1692–1701.

[46] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *J. Mach. Learn. Res.*, vol. 38, no. 4, pp. 361–369, 2012. [Online]. Available: http://jmlr.org/papers/v14/hoffman13a.html and http://arxiv.org/abs/1206.7051

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.

[48] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[49] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8798–8807.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[51] D. Yang, S. Hong, Y. Jang, T. Zhao, and H. Lee, "Diversity-sensitive conditional generative adversarial networks," 2019, *arXiv:1901.09024*. [Online]. Available: http://arxiv.org/abs/1901.09024

[52] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[54] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 465–476.

[55] B. A. Plummer, L. Wang, C. M. Cervantes, J. C. Caicedo, J. Hockenmaier, and S. Lazebnik, "Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2641–2649.

[56] L. Zhao, Q. Mo, S. Lin, Z. Wang, Z. Zuo, H. Chen, W. Xing, and D. Lu, "UCTGAN: Diverse image inpainting based on unsupervised cross-space translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5741–5750.

[57] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document RIR BT.500, 2002.
[58] W. Kim, S. Lee, and A. C. Bovik, "VR sickness versus VR presence: A statistical prediction model," *IEEE Trans. Image Process.*, vol. 30, pp. 559–571, 2021.
[59] J. H. Moor, "An analysis of the Turing test," *Philos. Stud.*, vol. 30, no. 4, pp. 249–257, 1976.
[60] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2408–2415.

WOOJAE KIM received the B.S. degree in electronic engineering from Soongsil University, Seoul, South Korea, in 2015. He is currently pursuing the M.S. and Ph.D. degrees with the Multidimensional Insight Laboratory, Yonsei University.

He was a Research Assistant with the Laboratory for School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore, in 2018, under the guidance of Prof. Weisi Lin. His research interests include image and video processing based on the human visual systems, image/video quality assessment, computer vision, and machine learning.

ANH-DUC NGUYEN received the B.Eng. degree in automatic control from the Hanoi University of Science and Technology, Vietnam, in 2015. He is currently pursuing the joint M.Sc. and Ph.D. degrees with Yonsei University, South Korea.

His research interests include image/video analysis, geometric computer vision, and deep learning.

JINWOO KIM received the B.S. degree in electrical and electronic from Hongik University, South Korea, in 2016. He is currently pursuing the M.S. and Ph.D. degrees with the Multidimensional Insight Laboratory, Yonsei University.

His research interests include quality assessment, computer vision, and machine learning.

JONGYOO KIM received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2011, 2013, and 2018, respectively.

He joined Microsoft Research Asia, in 2018. His current research interests include 2-D/3-D image and video processing based on human visual systems, quality assessment of 2-D/3-D image and video, 3-D computer vision, and deep learning. He was a recipient of the Global Ph.D. Fellowship from the National Research Foundation of Korea, from 2011 to 2016.

HEESEOK OH received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2010, 2012, and 2017, respectively.

He was a Senior Engineer with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. He is currently an Assistant Professor with Hansung University, Seoul. His research interests include 2D/3D image and video processing based on human visual systems, computer vision, extended reality, and deep generative networks.

SANGHOON LEE (Senior Member, IEEE) received the B.S. degree from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1991, and the Ph.D. degree from The University of Texas at Austin, TX, USA, in 2000.

From 1991 to 1996, he was with Korea Telecom, South Korea. From 1999 to 2002, he was with Lucent Technologies, NJ, USA. In 2003, he joined the EE Department, Yonsei University, as a Faculty Member, where he is currently a Full Professor. His current research interests include image/video processing, computer vision, and graphics. He received the 2015 Yonsei Academic Award from Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society. He was the General Chair of the 2013 IEEE IVMSP Workshop. He served in steering committees for IEEE and APISPA Conferences. He has been serving as the Chair for the IEEE P3333.1 Quality Assessment Working Group, since 2011. He was the IVM Technical Committee Chair of APSIPA, from 2018 to 2019, and is a Board of Governors Member of APSIPA, in 2020. He was with the IEEE IVMSP Technical Committee, from 2014 to 2019, and has been with the IEEE MMSP Technical Committee, since 2016. He also served as an Editor for the *Journal of Communications and Networks*, from 2009 to 2015, an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2010 to 2014, and a Guest Editor for the Special Issue of the IEEE TRANSACTIONS ON IMAGE PROCESSING, in 2013.