

PieNet: Personalized Image Enhancement Network

Han-Ul Kim¹[0000-0001-7450-6600], Young Jun Koh²[0000-0003-1805-2960], and
Chang-Su Kim¹[0000-0002-4276-1831]

¹ Korea University, Seoul, Korea
hanulkim@mcl.korea.ac.kr, changsukim@korea.ac.kr

² Chungnam National University, Daejeon, Korea
yjkoh@cnu.ac.kr

Abstract. Image enhancement is an inherently subjective process since people have diverse preferences for image aesthetics. However, most enhancement techniques pay less attention to the personalization issue despite its importance. In this paper, we propose the first deep learning approach to personalized image enhancement, which can enhance new images for a new user, by asking him or her to select about 10~20 preferred images from a random set of images. First, we represent various users' preferences for enhancement as feature vectors in an embedding space, called preference vectors. We construct the embedding space based on metric learning. Then, we develop the personalized image enhancement network (PieNet) to enhance images adaptively using each user's preference vector. Experimental results demonstrate that the proposed algorithm is capable of achieving personalization successfully, as well as outperforming conventional general image enhancement algorithms significantly. The source codes and trained models are available at <https://github.com/hukim1124/PieNet>.

Keywords: Image enhancement, personalization, metric learning

1 Introduction

Nowadays, people take photographs casually but are often unsatisfied with them. Photos may be noisy because of limited camera sensors. Also, photos taken in uncontrolled environments may suffer from low dynamic ranges or distorted color tones [25,32]. Thus, image enhancement is required, which post-processes and edits its photographs to satisfy user preferences. Professional softwares provide many tools to support manual image enhancement. However, the results of manual enhancement depend on users' skills and experience. Moreover, the manual process requires lots of efforts.

Many researches have been carried out to perform image enhancement automatically. But, image enhancement is a non-trivial problem partly due to the non-linear relationship between input and output images. Furthermore, it makes enhancement even more challenging that people have different preferences for

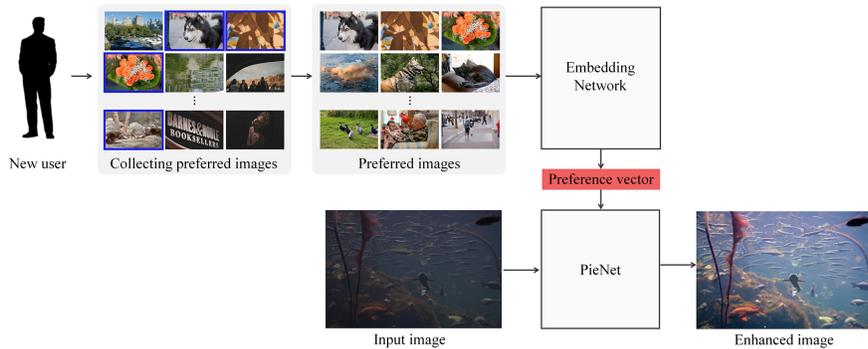


Fig. 1: Illustration of personalization for a new user. A user selects only about 10~20 preferred images from a random set of images. Then, the proposed algorithm analyzes the user’s preference and enhances a new image according to the preferred style. Please see the supplemental video for this demonstration.

images; image enhancement is a subjective process. In this regard, the deep-learning-based enhancement algorithms in [9, 19, 33, 35, 42, 45] have the common limitation that they cannot handle various user preferences. In [7, 24], personalized image enhancement systems have been developed. However, they are based on traditional enhancement techniques such as gamma-correction and S-curve. They may not yield output qualities that are as high as those of professionally enhanced images by experts.

We propose a deep learning algorithm for personalized image enhancement. First, we model diverse user preferences for image enhancement as feature vectors, called preference vectors, in an embedding space. More specifically, we perform metric learning to learn the embedding space, in which a preference vector conveys the preferred enhancement style of the corresponding user. Next, we propose a novel image enhancement network, referred to as PieNet, which employs the preference vector to achieve personalized enhancement. The proposed PieNet has an encoder-decoder architecture. The encoder part yields multi-scale features, representing local and global information for image enhancement. The decoder part includes personalized up-sample blocks, which employ the preference vector to produce personalized results. Experimental results demonstrate that the proposed algorithm outperforms the conventional deep learning algorithms [9, 14, 19, 35, 42] for general (*i.e.* non-personalized) image enhancement on the MIT-Adobe 5K dataset [5]. Moreover, it is shown that the proposed algorithm achieves personalization successfully. In particular, it is shown that the proposed algorithm achieves the personalization for a new user with the minimal effort of selecting only a few preferred images, as illustrated in Fig. 1.

This paper has three main contributions:

1. Development of PieNet to tackle the personalization issue in image enhancement, which is the *first* deep learning approach to the best of our knowledge.

2. Remarkable general image enhancement performance on MIT-Adobe 5K.
3. Excellent scalability of PieNet to achieve the personalization for a new user with the minimal effort of selecting only 10~20 preferred images. Please see the supplemental video for personalization demos.

2 Related Work

Image enhancement: Early studies on image enhancement focused on improving image contrast. Histogram equalization [15] and its variants [2, 26, 29–31, 41, 44] modify the histogram of an image to improve its limited dynamic range. Also, retinex methods [6, 12, 13, 16, 20, 21, 43, 47] regard an image as the product of reflectance and illumination [28], and alter the illumination to enhance a poorly lit image. However, these methods may not reconstruct the complex mapping function between an image and its professionally enhanced version, edited by an expert.

An alternative approach to image enhancement is the data-driven one to learn the mapping between input and enhanced images from a large dataset. Bychkovsky *et al.* [5] introduced the MIT-Adobe 5K dataset, composed of 5,000 input and expert-retouched image pairs. They used the dataset to estimate mapping functions, based on regression schemes for predicting user adjustments. However, their method still may fail to reconstruct highly non-linear mapping functions between input and enhanced images.

Recently, motivated by the success of deep learning, several deep neural networks have been developed to deal with the non-linear image enhancement. Yan *et al.* [45] proposed a deep learning scheme, which uses image descriptors to predict a color mapping for each pixel. Lore *et al.* [33] developed a deep autoencoder to enhance low-light images. Gharbi *et al.* [14] proposed deep bilateral learning for real-time enhancement, which predicts local affine transforms in the bilateral space. Based on the retinex theory, Wang *et al.* [42] designed a deep network to predict an image-to-illumination mapping function instead of a direct image-to-image function. Also, to achieve unpaired learning for image enhancement, Park *et al.* [35] introduced the distort-and-recover approach that degraded high-quality images to generate pseudo paired data. Chen *et al.* [9] used two-way generative adversarial networks (GANs) for stable training. Deng *et al.* [11] developed an aesthetic-driven image enhancement algorithm. In [19, 46], an adversarial loss is integrated into reinforcement learning to learn to generate a sequence of enhancement operations. These deep learning algorithms provide promising performances, but are limited in that they do not consider different users’ diverse preferences.

Personalization: Joshi *et al.* [23] proposed a personal photo enhancement algorithm, which uses a person’s favorite photographs as examples to perform several tasks, including deblurring, super-resolution, and white-balancing. The proposed algorithm is, however, more related to the Kang *et al.*’s personalization system [24] for image enhancement. Their system asks a user to enhance

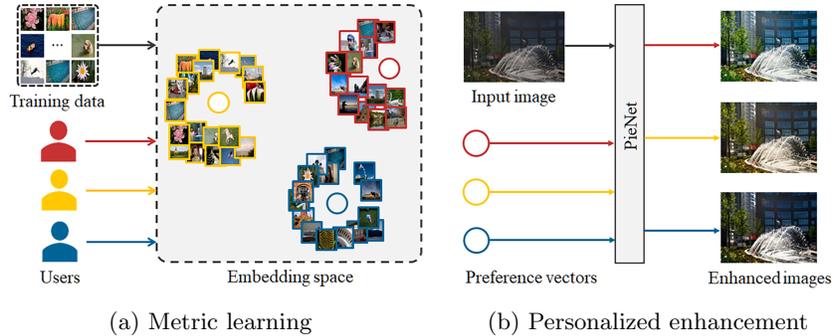


Fig. 2: Overview of the proposed system, which has two stages. (a) Metric learning is performed to discover an embedding space, in which the characteristics of each user’s preferred images are encoded into the preference vector. (b) PieNet employs the user’s preference vector to yield personalized enhancement results.

25 representative images by controlling a set of parameters. Given a new image, it finds the most ‘similar’ representative image. Then, it uses the corresponding set of parameters to enhance the new image. They adopted a metric learning scheme to define the ‘similarity’ between images such that it correlates well with the enhancement parameters. Caicedo *et al.* [7] extended the Kang *et al.*’s system to consider the enhancement results of other users with similar preferences based on collaborative filtering.

Compared to [7, 24], the proposed algorithm has noticeable differences. First, while they demand a user to enhance training images by controlling parameters, the proposed algorithm requires a user to select only a few preferred images from a set of candidate images. Thus, the proposed algorithm needs much less user efforts. Second, we consider more complex mappings between input and enhanced images, by developing the *first* deep learning algorithm for personalized image enhancement.

Metric learning: The objective of metric learning is to learn an embedding space, in which the distance between similar objects is shorter than the distance between dissimilar ones. In [4, 10, 17], a contrastive loss was employed to minimize the distances between objects in the same class, while constraining the distances between inter-class objects to be larger than a margin. Schroff *et al.* [37] proposed a triplet loss to encourage the distance between anchor and positive objects to be smaller than that between anchor and negative objects. To overcome slow convergence of the triplet loss, many extensions have been proposed [8, 38–40].

Notice that, whereas the conventional personalized algorithms in [7, 24] use metric learning to embed images with similar enhancement parameters tightly, the proposed algorithm performs it to directly embed preferred images of each user tightly and thus obtain the user’s preference vector in the embedding space.

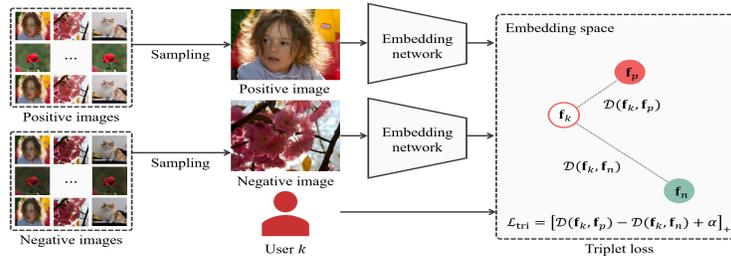


Fig. 3: Metric learning for the embedding space of preference vectors.

3 Proposed Algorithm

Fig. 2 shows an overview of the proposed system for personalized image enhancement. First, we do metric learning to determine an embedding space, in which a user’s preference for enhancing images is represented by a feature vector. We refer to this feature vector as the preference vector. Second, we develop PieNet that employs the preference vector to produce personalized enhanced results adaptively for the specific user.

3.1 Preference vector

Let us consider an embedding space, in which a preference vector represents a user’s preferred style for enhancement. We learn this embedding space to yield the preference vectors for multiple users based on metric learning. Each user provides two sets of preferred (positive) and non-preferred (negative) images. More specifically, we assume that there are N training images annotated by K users: $\{(I_i, \mathbf{y}_i)\}_{i=1}^N$, where I_i is the i th image and $\mathbf{y}_i = [y_{i1}, \dots, y_{iK}]^T$ is its label vector. The k th element y_{ik} in \mathbf{y}_i is 1 if user k likes I_i , and 0 otherwise.

Fig. 3 illustrates how to determine the embedding space. We feed a pair of positive and negative images for each user into two identical embedding networks. Specifically, at each training iteration, we sample a triplet (k, I_p, I_n) , where I_p and I_n are positive and negative images for user k , whose labels are $y_{p,k} = 1$ and $y_{n,k} = 0$. Each embedding network produces a 512-dimensional feature vector from an RGB color image of size 256×256 . We employ ResNet-18 [18] as the twin embedding networks and perform L_2 normalization to each output feature vector.

Let \mathbf{f}_p and \mathbf{f}_n denote such feature vectors for positive and negative images, respectively, and \mathbf{f}_k be the preference vector for user k . Then, we learn the embedding space, where the preference vector \mathbf{f}_k is similar to the positive feature vector but dissimilar from the negative one. To this end, we compute the triplet loss [37], given by

$$\mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n) = \left[\mathcal{D}(\mathbf{f}_k, \mathbf{f}_p) - \mathcal{D}(\mathbf{f}_k, \mathbf{f}_n) + \alpha \right]_+ \quad (1)$$

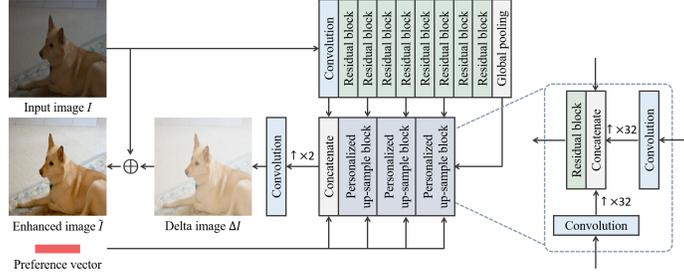


Fig. 4: Architecture of PieNet, composed of an encoder (top part) and a decoder (bottom part).

where $\mathcal{D}(\cdot)$ is the squared Euclidean distance, $[\cdot]_+$ is the rectifier, and α is a margin to be enforced between positive and negative images.

In the training phase, we initialize each preference vector with random values and the embedding networks with the pre-trained weights on ImageNet [36]. Then, we compute gradients to minimize the triplet loss,

$$\frac{\partial \mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n)}{\partial \mathbf{f}_p} = 2(\mathbf{f}_p - \mathbf{f}_k) \cdot \mathbb{I}(\mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n) > 0), \quad (2)$$

$$\frac{\partial \mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n)}{\partial \mathbf{f}_n} = 2(\mathbf{f}_k - \mathbf{f}_n) \cdot \mathbb{I}(\mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n) > 0), \quad (3)$$

$$\frac{\partial \mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n)}{\partial \mathbf{f}_k} = 2(\mathbf{f}_n - \mathbf{f}_p) \cdot \mathbb{I}(\mathcal{L}_{\text{tri}}(\mathbf{f}_k, \mathbf{f}_p, \mathbf{f}_n) > 0), \quad (4)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Note that these gradients are back-propagated to update the preference vector \mathbf{f}_k and the weight parameters in the embedding networks simultaneously. In this way, we determine the embedding space to yield the preference vectors for the K users, by employing the training images and their label vectors.

3.2 PieNet architecture

Using the preference vectors, we perform personalized image enhancement. Fig. 4 is the architecture of PieNet, which has an encoder and a decoder. The encoder takes an RGB color image as input. Its spatial resolution is 512×512 . We also employ ResNet-18 to implement the encoder, which consists of one convolution layer, eight residual blocks, and one average pooling layer. The encoder extracts five multi-scale features from the convolution layer, the 2nd, 4th, 6th residual blocks, and the pooling layer, respectively. The intermediate features from the convolution layer and the residual blocks preserve detailed local information, while the global feature, extracted from the pooling layer, contains high-level information such as global brightness and scene category of the input image.

From the extracted features of the encoder, the decoder reconstructs a delta image ΔI , which is added to the input image I to enhance its quality. For the

decoder, we develop the personalized up-sample block (PUB) to consider user preferences. In Fig. 4, each PUB takes three inputs: 1) the preference vector, 2) the output of the previous block, and 3) the intermediate feature of the encoder. It makes the first two inputs have the same size as the third one through convolution and up-sampling, and then concatenates the three data along the channel dimension. Then, the residual block in the PUB produces the output. By feeding the preference vector to every PUB, the decoder can satisfy the preferred style of the specific user. The output of the last PUB is concatenated with the output of the convolution layer in encoder. Then, the concatenated signal is up-sampled to be the same size as the input image and fed into the last convolution layer. The last convolution layer yields the delta image ΔI .

Finally, the enhanced image \tilde{I} is obtained by adding the delta image ΔI to the input image I , given by

$$\tilde{I} = I + \Delta I. \quad (5)$$

Note that we predict the delta image instead of the enhanced image directly. This is because the down-sampling process in the network may lose image details. Even though the delta image loses some details, the enhanced image can restore those details from the input image.

3.3 PieNet training

For user k , suppose that the preference vector \mathbf{f}_k and an image pair (I, I_k^*) are available. Here, I_k^* is the ground-truth enhanced image that user k prefers to obtain from image I . We train PieNet using the preference vector and the image pair. Note that PieNet estimates a delta image ΔI_k and produces a personalized enhanced image \tilde{I}_k via (5). We compare the estimated result with the ground-truth to train PieNet, by employing the loss function

$$\mathcal{L}(I, \Delta I_k, \tilde{I}_k, I_k^*) = \mathcal{L}_c(\tilde{I}_k, I_k^*) + \lambda_p \mathcal{L}_p(\tilde{I}_k, I_k^*) + \lambda_t \mathcal{L}_t(I, \Delta I_k) \quad (6)$$

where \mathcal{L}_c , \mathcal{L}_p , and \mathcal{L}_t are color, perceptual, and total variation losses, respectively, and λ_p and λ_t are balancing parameters.

The color loss penalizes the mean absolute error between the predicted and ground-truth enhanced images, given by $\mathcal{L}_c(\tilde{I}_k, I_k^*) = \|\tilde{I}_k - I_k^*\|_1$. The perceptual loss [22] encourages the enhanced image and the ground-truth image to have similar features. Specifically, it is defined as $\mathcal{L}_p(\tilde{I}_k, I_k^*) = \|\hat{\mathbf{f}}_k - \mathbf{f}_k^*\|_1$, where $\hat{\mathbf{f}}_k$ and \mathbf{f}_k^* denote the features for the estimated and ground-truth enhanced images, respectively, extracted from the embedding network in Section 3.1. Notice that the embedding network attempts to construct the embedding space, where the features of ground-truth enhanced images are compactly distributed near the preference vector. Hence, the perceptual loss constrains that the feature of the enhanced image should be near the preference vector.

Also, we use the total variation loss [1] to enforce the spatial smoothness of the delta image. To constrain neighboring pixels to exhibit similar delta values, the total variation loss is defined as

$$\mathcal{L}_t(I, \Delta I_k) = \|W_x \otimes \nabla_x(\Delta I_k)\|_1 + \|W_y \otimes \nabla_y(\Delta I_k)\|_1 \quad (7)$$

where \otimes is the element-wise multiplication, and ∇_x and ∇_y denote the partial derivatives in the horizontal and vertical directions. Also, $W_x = \exp(-|\nabla_x I|)$ and $W_y = \exp(-|\nabla_y I|)$ are weight maps, which have large values in smooth regions in the input image. Thus, the total variation loss \mathcal{L}_t imposes large penalties when neighboring pixels in smooth regions are assigned quite different delta values. On the contrary, near edges or complicated texture in the input image, delta values may be dissimilar from one another without causing large penalties.

3.4 Personalization for new users

A critical issue in personalized enhancement is ‘scalability,’ which means the capability of accommodating the preference of a new user with minimal efforts. A straightforward approach is to repeat the entire training process, *i.e.* performing metric learning and training PieNet to consider a new user as well as the existing users. However, the fine-tuning of the embedding space and PieNet is a time-consuming process. Therefore, we decide the preference vector for the new user within the pre-trained embedding space and also use the pre-trained PieNet to produce personalized results. We consider two schemes to determine the preference vector for the new user.

The first scheme assumes that a new user provides two sets of positive and negative images. Then, given the pre-trained embedding space, the gradient in (4) is back-propagated to update the preference vector, while the other gradients in (2) and (3) are not back-propagated. Thus, we fix the embedding space while determining the preference vector. Then, using this preference vector, the pre-trained PieNet yields personalized enhanced images for the new user.

In the second scheme (which is computationally much simpler and is thus adopted in the default mode), a new user provides preferred images only. The pre-trained embedding network encodes these preferred images into feature vectors. Then, we determine the preference vector, by averaging the feature vectors. Note that the new user need not provide non-preferred images. Thus, the second scheme demands less user effort than the first scheme does. Moreover, it is shown in Section 4 that about 10 preferred images are sufficient for PieNet to yield desirable personalization results.

4 Experiments

4.1 Evaluation on MIT-Adobe 5K

Dataset and metrics: We assess the proposed algorithm on the MIT-Adobe 5K dataset [5]. It consists of 5,000 input images, each of which was manually enhanced by five different photographers (A/B/C/D/E). Thus, there are five sets of 5,000 pairs of input and enhanced images, one set for each photographer. Among the 5,000 images, we randomly select 500 images to compose the test set as done in [9, 42], and use the remaining 4,500 images as the training set.

Table 1: Comparison of the proposed algorithm with the conventional algorithms on MIT-Adobe 5K. For the ‘single user’ test, we use the photographer C’s retouched images as the ground-truth. For the ‘multiple users’ test, we use the retouched images by the five photographers A/B/C/D/E as the ground-truth.

Method	Single user		Multiple users	
	PSNR	SSIM	mPSNR	mSSIM
WB [19]	18.36	0.810	17.83	0.799
D&R [35]	20.97	0.841	18.65	0.834
HDR [14]	23.44	0.882	21.64	0.872
DPE [9]	22.34	0.873	21.09	0.858
DUPE [42]	23.61	0.887	21.74	0.881
Proposed	25.28	0.908	24.28	0.907

For quantitative assessment, we employ PSNR and SSIM, which measure, respectively, color and structural similarity between predicted and ground-truth enhanced images.

Implementation details: We jointly train the embedding network and the preference vectors for photographers A/B/C/D/E using the training set in MIT-Adobe 5K. For instance, when the preference vector for photographer A is trained, we regard his or her retouched images as the positive set, but the input images and the other photographers’ retouched images as the negative set. This process is carried out for the other photographers similarly. We minimize the triplet loss in (1) using the Adam optimizer [27] with a learning rate of 1.0×10^{-4} . The training is iterated for 25,000 mini-batches, each of which includes 64 triplets. For data augmentation, we randomly rotate image pairs by multiples of 90 degrees. The margin α in (1) is set to 0.2.

To train PieNet, we use all image pairs in the training set, *i.e.* the five sets of image pairs for photographers A/B/C/D/E. In other words, we train PieNet for all five photographers using their preference vectors and image pairs. We also use the Adam optimizer to minimize the loss function in (6) with a learning rate of 1.0×10^{-4} for 100,000 mini-batches. The mini-batch size is 8. We randomly rotate images by multiples of 90 degrees. Also, we randomly perturb the preference vectors to make PieNet insensitive to small perturbations. Specifically, we add noise \mathbf{n} to the preference vectors, where \mathbf{n} is sampled from the hypersphere, $\|\mathbf{n}\|_2 = 0.1$. The parameters λ_p and λ_t in (6) are fixed to 0.4 and 0.01, respectively.

Experimental results: Table 1 compares the proposed algorithm with the recent state-of-the-art algorithms in [9, 14, 19, 35, 42]. We obtain the results of the conventional algorithms using the source codes and parameters, provided by the respective authors. Note that these conventional algorithms are for general (*i.e.* non-personalized) image enhancement. Specifically, they attempt to mimic the retouching of photographer C only. In contrast, the proposed PieNet can provide enhanced images in five different styles using the preference vectors of photographers A/B/C/D/E.

In the ‘single user’ test, we compare the proposed algorithm with the conventional algorithms using the photographer C’s enhanced images as the ground-truth. For this test, we only use the training images, retouched by C, to train the proposed algorithm for a fair comparison. WB [19] and D&R [35] provide poor performance than the other algorithms, since they use unpaired images for training. The proposed algorithm significantly outperforms the conventional algorithms [9, 14, 42], which conduct supervised learning directly using pairs of input and enhanced images. The proposed algorithm provides excellent performance due to two main factors. First, we adopt the effective network architecture. Second, combining different losses in (6) further improves the performance. Especially, we find that the perceptual loss \mathcal{L}_p , based on the embedding network in Section 3.1, leads to a notable PSNR improvement.

The ‘multiple users’ test analyzes the personalization performance. The proposed algorithm can produce differently enhanced results according to the preference vectors. In this test, it yields personalized enhanced results for photographers A/B/C/D/E. We compare the personalized results with the corresponding ground-truth to compute PSNR and SSIM for each photographer, and then compute the average PSNR (mPSNR) and average SSIM (mSSIM) over the five photographers. In contrast, since the conventional algorithms provide only one enhanced result for an input image, we use the same enhanced result to compute PSNR and SSIM for each photographer.

By comparing the ‘single user’ and ‘multiple users’ tests, we see that the conventional algorithms experience significant degradation in the performance. This is because the conventional algorithms are designed to mimic the retouching style of photographer C only. Thus, they provide low PSNR and SSIM scores when compared with the other photographers’ ground-truth. In contrast, the proposed algorithm experiences only minor degradation. It is worth pointing out that the proposed personalization performances in ‘multiple users’ even surpass the performances of all conventional algorithms in ‘single user,’ whose scores are computed for C only.

Notice that the proposed algorithm can produce personalized results for all five photographers without any additional training of PieNet, by changing only the preference vectors. On the contrary, for the conventional algorithms to provide reliable results for photographers A/B/D/E, they should retrain their networks four more times for the adaptation. The proposed algorithm hence achieves personalization more efficiently than the conventional algorithms.

Fig. 5 compares the proposed algorithm with HDR, DPE, and DUPE qualitatively. In Fig. 5(b)~(d), these conventional algorithms provide reasonable results. They are designed to yield similar color tones to the photographer C’s retouched image in Fig. 5(h). However, image enhancement is subjective, and photographer B prefers a different output in Fig. 5(f). As shown in Fig. 5(e) and (g), the proposed algorithm adaptively produces output images in B and C’s styles, respectively. More experiments are available in the supplementary materials.

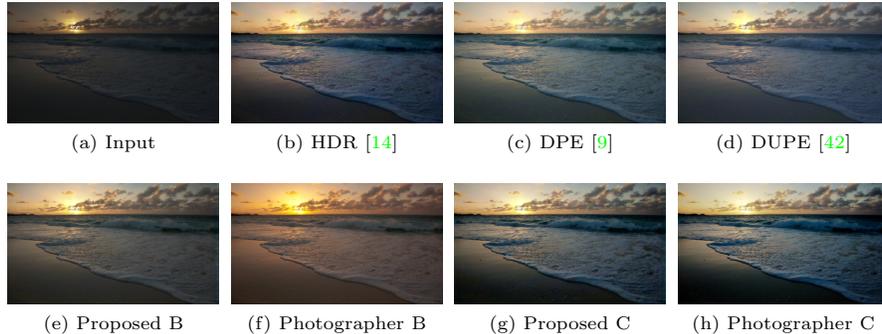


Fig. 5: Qualitative comparison of enhanced images.

4.2 Personalization

Dataset and metrics: We expand the MIT-Adobe 5K dataset to evaluate the personalization performances of the proposed algorithm for new users other than photographers A/B/C/D/E. However, it is an expensive task to collect 5,000 ground-truth images that are manually retouched by each new user. Therefore, for the expanded dataset, we enhance images using 28 conventional methods (including photographers A/B/C/D/E in [5] and predefined settings in Adobe Lightroom), instead of employing people to enhance images manually. These methods are regarded as users. Then, we divide the expanded dataset into the training and test sets:

- **Training (20 users):** 11 presets in Adobe Lightroom, 5 conventional methods [2, 3, 6, 12, 43], Photographers A/B/C/D [5]
- **Test (8 users):** 4 presets in Adobe Lightroom, 3 conventional methods [13, 16, 30], Photographer E [5]

Note that the enhancement methods in the training and test sets do not overlap. For each method in the training set, there are 4,500 pairs of input and enhanced images for training the embedding space and PieNet. On the other hand, for each method in the test set, there are 500 pairs of input and enhanced images, which are used to assess the personalization performances of the proposed algorithm.

Implementation details: We train the embedding space and PieNet using the pairs of input and enhanced images in the training set. We use the same training settings in Section 4.1. In the test phase, we regard the methods in the test set as new users. Then, we compute the preference vector for each new user using the two schemes in Section 3.4: 1) triplet loss minimization and 2) feature vector average. For the first scheme, each method regards its enhanced images as positive or preferred images, while considering enhanced images of the other methods as negative images. For the second schemes, each method uses its enhanced images as preferred images.

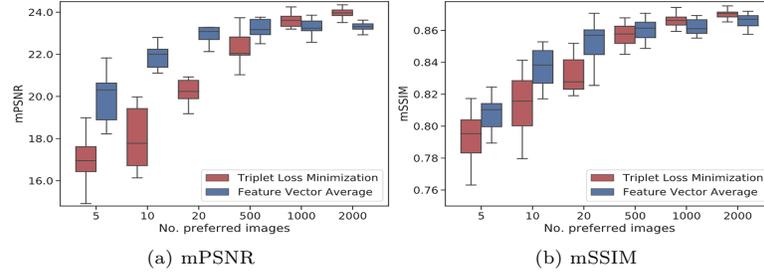


Fig. 6: Bar plots of the mPSNR and mSSIM scores for the 8 new users in the expanded MIT-Adobe 5K dataset. In each plot, we repeat ten experiments.



Fig. 7: Qualitative comparison of personalized enhancement results of the proposed algorithm with the ground-truth.

Experimental results: Fig. 6 shows the mPSNR and mSSIM scores for the 8 new users according to the number N_{pref} of preferred images, which are used to determine preference vectors. For the first scheme ‘triplet loss minimization,’ we use three times as many negative images as preferred images. The preferred images and negative images are randomly selected for plotting the bar graphs. Given sufficiently many preferred images ($N_{\text{pref}} = 2,000$) to compute the preference vectors, the first scheme outperforms the second scheme. However, when there are only a limited number of preferred images, it experiences severe performance degradation. In contrast, the second scheme provides reliable personalization performances, even when only 20 preferred images are available. Therefore, in the following tests, we use the second scheme with $N_{\text{pref}} = 20$.

Fig. 7 shows personalization results. The top row presents personalized enhancement results, while the bottom row shows the corresponding ground-truth generated by the four enhancement methods in the test set. Note that these four ground-truth images are considerably different from one another. For instance, photographer E in Fig. 7(h) boosts the brightness of the image, while

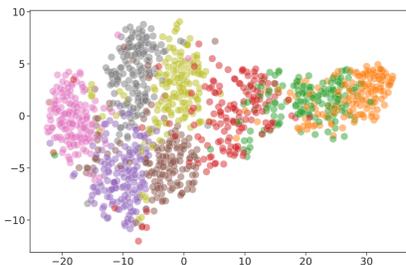


Fig. 8: Visualization of the embedding space.

the RedLiftMatte (RLM) preset in Lightroom in Fig. 7(g) renders the floor in different color tones. Nevertheless, the proposed algorithm successfully reflects the tendency of each method in Fig. 7(e)~(h). The supplementary materials provide more personalization results.

Fig. 8 visualizes the embedding space, in which the feature vectors of ground-truth images of the 8 test users are depicted as dots in different colors. For each user, 100 ground-truth images are randomly selected. The t-SNE technique [34] is employed for this visualization. We see that the feature vectors for the images are well clustered according to the users, which indicates that the embedding space is suitable for representing the preferred image styles of users. For the same reason, in this embedding space, we can easily construct the preference vector of a new user by averaging the feature vectors of only a few preferred images.

4.3 User study

We conducted a user study with 10 participants to assess the personalization performance of the proposed algorithm for real people. It was designed as follows.

1. Each participant selected 20 preferred images from various pre-enhanced images in the test set of the expanded MIT-Adobe 5K dataset.
2. The proposed algorithm generated the preference vector for each participant using those preferred images.
3. Each participant was presented with six enhanced results of the same photograph, obtained by the proposed algorithm and the conventional algorithms [9, 14, 19, 35, 42], and was asked to vote for the algorithm yielding the most pleasing result. The photograph was also selected from the test set, but not used to generate the preference vector. This was repeated for 10 photographs.

For this user study, we pre-trained the proposed algorithm using the training set of the expanded dataset. All conventional algorithms were trained using the photographer E’s retouched images, since it was the most often selected by the participants as their preferred method. Table 2 summarizes the voting results. The proposed algorithm gets the most votes by providing personalized results to each participant effectively.

Table 2: User study results. N_{vote} is the number of votes that each method gets.

	HDR [14]	DPE [9]	WB [19]	D&R [35]	DUPE [42]	Proposed
N_{vote}	101	64	17	2	71	245

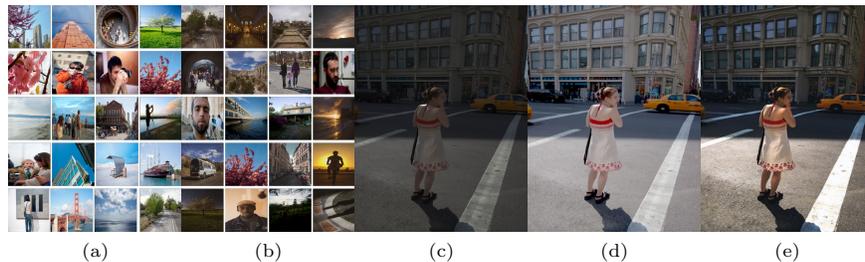


Fig. 9: Personalization results for two participants P1 and P2 in the user study: (a) P1’s preferred images, (b) P2’s preferred images, (c) input image, (d) enhanced image for P1, and (e) enhanced image for P2.

Fig. 9 shows personalization results for two participants (P1 and P2) in the user study. P1 and P2, respectively, selected preferred images in (a) and (b). These preferred styles are represented by their preference vectors. PieNet enhances an input image in (c) using the preference vectors. We see that the enhancement results in (d) and (e) faithfully reflect their preferred styles.

5 Conclusions

We addressed the personalization issue in image enhancement. We trained an embedding space to obtain preference vectors based on metric learning, and developed PieNet to produce personalized results using the preference vectors. Experiments demonstrated that the proposed algorithm significantly outperforms the state-of-the-art algorithms on the MIT-Adobe 5K dataset. Also, it was demonstrated that a new user can obtain reliable results by providing only 10~20 preferred images to the proposed enhancement system.

Acknowledgements

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00464) supervised by the IITP (Institute for Information & communications Technology Promotion), in part by the National Research Foundation of Korea (NRF) through the Korea Government (MSIP) under Grant NRF-2018R1A2B3003896, and in part by the research fund of Chungnam National University.

References

1. Aly, H.A., Dubois, E.: Image up-sampling using total-variation regularization with a new observation model. *IEEE Trans. Image Process.* **14**(10), 1647–1659 (2005) [7](#)
2. Arici, T., Dikbas, S., Altunbasak, Y.: A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. Image Process.* **18**(9), 1921–1935 (2009) [3](#), [11](#)
3. Aubry, M., Paris, S., Hasinoff, S.W., Kautz, J., Durand, F.: Fast local Laplacian filters: Theory and applications. *ACM Trans. Graphics* **33**(5), 167 (2014) [11](#)
4. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a “Siamese” time delay neural network. In: *NIPS* (1994) [4](#)
5. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: *CVPR* (2011) [2](#), [3](#), [8](#), [11](#)
6. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: *ICCV* (2017) [3](#), [11](#)
7. Caicedo, J.C., Kapoor, A., Kang, S.B.: Collaborative personalization of image enhancement. In: *CVPR* (2011) [2](#), [4](#)
8. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: A deep quadruplet network for person re-identification. In: *CVPR* (2017) [4](#)
9. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. In: *CVPR* (2018) [2](#), [3](#), [8](#), [9](#), [10](#), [11](#), [13](#), [14](#)
10. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: *CVPR* (2005) [4](#)
11. Deng, Y., Loy, C.C., Tang, X.: Aesthetic-driven image enhancement by adversarial learning. In: *ACM MM* (2018) [3](#)
12. Fu, X., Liao, Y., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Trans. Image Process.* **24**(12), 4965–4977 (2015) [3](#), [11](#)
13. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: *CVPR* (2016) [3](#), [11](#), [12](#)
14. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. *ACM Trans. Graphics* **36**(4), 118 (2017) [2](#), [3](#), [9](#), [10](#), [11](#), [13](#), [14](#)
15. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing* (4th Edition). Pearson (2018) [3](#)
16. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2016) [3](#), [11](#)
17. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: *CVPR* (2006) [4](#)
18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR* (2016) [5](#)
19. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. *ACM Trans. Graphics* **37**(2), 26 (2018) [2](#), [3](#), [9](#), [10](#), [13](#), [14](#)
20. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **6**(7), 965–976 (1997) [3](#)

21. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **6**(3), 451–462 (1997) [3](#)
22. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *ECCV* (2016) [7](#)
23. Joshi, N., Matusik, W., Adelson, E.H., Kriegman, D.J.: Personal photo enhancement using example images. *ACM Trans. Graphics* **29**(2), 12–1 (2010) [3](#)
24. Kang, S.B., Kapoor, A., Lischinski, D.: Personalization of image enhancement. In: *CVPR* (2010) [2](#), [3](#), [4](#)
25. Kim, J.H., Jang, W.D., Sim, J.Y., Kim, C.S.: Optimized contrast enhancement for real-time image and video dehazing. *J. Vis. Commun. Image. Represent.* **24**, 410–425 (Apr 2013) [1](#)
26. Kim, Y.T.: Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Trans. Consum. Electro.* **43**(1), 1–8 (1997) [3](#)
27. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *ICLR* (2014) [9](#)
28. Land, E.H.: The retinex theory of color vision. *Scientific American* **237**(6), 108–129 (1977) [3](#)
29. Lee, C., Kim, J.H., Lee, C., Kim, C.S.: Optimized brightness compensation and contrast enhancement for transmissive liquid crystal displays. *IEEE Trans. Circuits Syst. Video Technol.* **24**, 576–590 (Apr 2014) [3](#)
30. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans. Image Process.* **22**(12), 5372–5384 (2013) [3](#), [11](#), [12](#)
31. Lee, C., Lee, C., Lee, Y.Y., Kim, C.S.: Power-constrained contrast enhancement for emissive displays based on histogram equalization. *IEEE Trans. Image Process.* **21**(1), 80–93 (2011) [3](#)
32. Lim, J., Heo, M., Lee, C., Kim, C.S.: Contrast enhancement of noisy low-light images based on structure-texture-noise decomposition. *J. Vis. Commun. Image. Represent.* **45**, 107–121 (May 2017) [1](#)
33. Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: a deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **61**, 650–662 (2017) [2](#), [3](#)
34. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (Nov 2008) [13](#)
35. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: *CVPR* (2018) [2](#), [3](#), [9](#), [10](#), [13](#), [14](#)
36. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**(3), 211–252 (2015) [6](#)
37. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In: *CVPR* (2015) [4](#), [5](#)
38. Sohn, K.: Improved deep metric learning with multi-class N-pair loss objective. In: *NIPS* (2016) [4](#)
39. Song, H.O., Jegelka, S., Rathod, V., Murphy, K.: Deep metric learning via facility location. In: *CVPR* (2017) [4](#)
40. Song, H.O., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: *CVPR* (2016) [4](#)
41. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **9**(5), 889–896 (2000) [3](#)

42. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: CVPR (2019) [2](#), [3](#), [8](#), [9](#), [10](#), [11](#), [13](#), [14](#)
43. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* **22**(9), 3538–3548 (2013) [3](#), [11](#)
44. Wang, Y., Chen, Q., Zhang, B.: Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Trans. Consum. Electro.* **45**(1), 68–75 (1999) [3](#)
45. Yan, Z., Zhang, H., Wang, B., Paris, S., Yu, Y.: Automatic photo adjustment using deep neural networks. *ACM Trans. Graphics* **35**(2), 11 (2016) [2](#), [3](#)
46. Yu, R., Liu, W., Zhang, Y., Qu, Z., Zhao, D., Zhang, B.: Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In: NIPS (2018) [3](#)
47. Yue, H., Yang, J., Sun, X., Wu, F., Hou, C.: Contrast enhancement based on intrinsic image decomposition. *IEEE Trans. Image Process.* **26**(8), 3981–3994 (2017) [3](#)