IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 August 2022, accepted 13 September 2022, date of publication 16 September 2022,
date of current version 23 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3207299

RESEARCH ARTICLE

Edge-Aware Interactive Contrast Enhancement

KEUNSOO KO, (Student Member, IEEE), AND CHANG-SU KIM, (Senior Member, IEEE)
School of Electrical Engineering, Korea University, Seoul 136-701, South Korea

Corresponding author: Chang-Su Kim (changsukim@korea.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant through the Korea Government (MSIP) under Grant NRF-2021R1A4A1031864 and Grant NRF-2022R1A2B5B03002310.

ABSTRACT Contrast enhancement is required in many applications. Many studies have been conducted to perform contrast enhancement automatically, but most of them do not consider various personal preferences for contrast. We propose an edge-aware interactive contrast enhancement algorithm to enable a user to adjust image contrast easily according to his or her preference. A user provides a parameter for controlling the global brightness and two types of scribbles to darken or brighten local regions in an image. Then, the proposed algorithm generates an edge-aware mask by propagating the scribbles to nearby regions and restores an enhanced image through a neural network, called e-IceNet. The user can provide annotations iteratively until he or she obtains a desired image. We train e-IceNet on guidance images to yield reliable results for diverse input images. We also propose two differentiable losses to train e-IceNet effectively and reliably. Extensive experiments demonstrate that the proposed e-IceNet is capable of allowing users to enhance images satisfactorily with simple scribbles, as well as producing enhanced images automatically.

INDEX TERMS Interactive contrast enhancement, personalized contrast enhancement, convolutional neural networks, multiple random walkers, guidance images.

I. INTRODUCTION

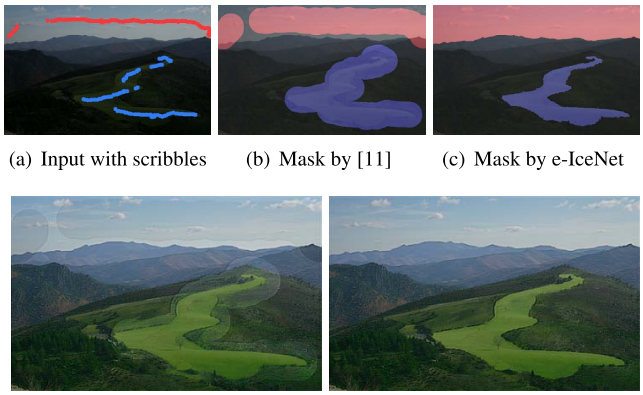
Nowadays people take many digital photographs casually, but uncontrolled environments often cause photographs with low dynamic ranges. Especially, abnormal lighting conditions would distort colors and textures considerably, thereby degrading the visual experience. Contrast enhancement (CE) can alleviate these problems, and its performance can be improved if user interactions are allowed. For example, Photoshop provides interactive tools to adjust contrast according to personal preferences, but using such tools takes much effort.

A lot of researches have been carried out to perform CE automatically. Recently, with the success of convolutional neural networks (CNNs) in the field of low-level vision [1], [2], [3], [4], various CNN-based CE algorithms [5], [6], [7], [8], [9], [10] have been developed. They learn mappings from low contrast images to high contrast ones using big training datasets. However, CE is a non-trivial task, partly due to the non-linear relationship between input and output images. Furthermore, it makes CE even more challenging that people have different preferences for images; CE is a subjective process.

In this regard, the conventional algorithms [5], [6], [7], [8], [9], [10] have the limitation that they cannot satisfy various personal preferences. To overcome it, Ko and Kim [11] recently developed the IceNet algorithm, which enhances image contrast after accepting user annotations. However, their algorithm demands meticulous interactions: users should pay attention to the boundaries of regions for controlling brightness. Otherwise, unwanted results may be obtained as illustrated in Figures 1(b) and 1(d).

To overcome this problem, we develop an edge-aware interaction system, which allows a user to specify desired regions roughly without painstaking annotations but can generate accurate masks, as shown in Figures 1(c) and 1(e). Specifically, a user provides a parameter for controlling the global brightness and two types of scribbles to darken or brighten local regions in an image. Since the scribbles represent only rough locations for controlling brightness, we propagate them to nearby regions in an edge-aware manner, by employing multiple random walkers (MRW) [12], and generate accurate masks. Then, we restore an enhanced image through the proposed edge-aware interactive CE network

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



(a) Input with scribbles (b) Mask by [11] (c) Mask by e-IceNet

(d) Enhanced image of [11] (e) Enhanced image of e-IceNet

FIGURE 1. Comparison of e-IceNet with the conventional algorithm [11]. An image with scribbles in (a) is used as input: red and blue scribbles are for darkening and brightening ones, respectively. Using the scribbles, in (b) and (c), the masks are generated by [11] and e-IceNet. Then, in (d) and (e), the enhanced images are obtained by [11] and e-IceNet.

(e-IceNet), which consists of the multi-scale feature fusion (MFF), local feature extraction (LFE), and gamma estimation (GME) modules. For interactive CE, we construct guidance images from low contrast images. By employing the guidance images, we can provide more reliable results than conventional interactive CE methods. In other words, we train e-IceNet using those guidance images to yield reliable results for diverse images. It is worth pointing out that e-IceNet also can produce an enhanced image automatically, which can serve as a basis for further interactive adjustments if so desired. Extensive experiments show that the proposed e-IceNet not only provides users with satisfactory images but also outperforms the state-of-the-art algorithms qualitatively and quantitatively. In particular, the proposed algorithm is preferred to existing algorithms three times more frequently in subjective tests. It is strongly recommended to watch the accompanying video for a real-time demo of e-IceNet.

To summarize, this work has three main contributions:

- We develop the edge-aware interactive CE system to enable users to obtain satisfactory images without requiring meticulous annotations.
- We propose an effective mask generation scheme that converts rough user scribbles into accurate masks based on the MRW simulation.
- We also propose e-IceNet, composed of the MFF, LFE, and GME modules, and train it on guidance images to yield reliable results for diverse input images.

II. RELATED WORK

The objectives of enhancement are closely related but different between CE, color enhancement [13], dehazing [14], [15], and detail enhancement [16], [17]. This section briefly reviews CE techniques.

Early CE methods improve the contrast of images using transformation functions or based on retinex theory. There are various transformation functions [18], [19], [20], [21], [22], among which gamma correction and logarithmic mapping are well-known parametric curves for mapping input pixel values to output ones. On the other hand, some CE algorithms [23], [24], [25], [26] have been developed based on retinex theory [27]. These conventional methods produce promising results. However, their performance usually depends on careful parameter tuning. It is difficult to find reliable parameters effectively for diverse input images. To address this problem, algorithms based on edge-preserving filters [28], [29], [30] have been proposed, but they take a long processing time to construct optimized filters [28].

The release of big paired datasets, containing pairs of low and high contrast images, has enabled CNN-based CE methods [5], [6], [7], [8], [9], [10] to yield impressive results. In the paired datasets, each pair is captured from the same scene, or a low contrast image is synthesized from a high contrast one. However, it is hard to capture the same scene twice because of moving objects, or synthesized low contrast images may not be photo-realistic. Therefore, by employing adversarial losses [1], some methods [31], [32], [33] train the CE networks using unpaired datasets, consisting of low and high contrast images of different scenes, but they should select unpaired images carefully. To avoid this cumbersome selection process, Guo *et al.* [34] proposed a self-supervised learning scheme that needs only low contrast images for training. However, all these CNN methods are not adaptive, so they cannot satisfy diverse user preferences.

To meet various preferences, professional software provides CE tools, but using these tools takes a lot of effort and training. To reduce such effort, simple interactive methods have been developed. Stoel *et al.* [35] proposed an interactive histogram equalization scheme for medical images. Their method allows a user to specify a region of interest (RoI) and applies the equalization to the region. Grundland and Dodgson [36] proposed an interactive tone adjustment method. When a user selects key tones on an image, it preserves those tones but adjusts the other tones, while maintaining the overall tonal balance. Lischinski *et al.* [37] and Dodgson *et al.* [38] also proposed interactive methods, in which a user specifies RoIs for local CE. Since these interactive methods are based on traditional CE methods, their performance depends on careful parameter tuning. Recently, Ko and Kim [11] proposed the first CNN-based interactive CE algorithm allowing a user to scribble local regions for CE, but their algorithm demands meticulous interactions.

III. PROPOSED ALGORITHM

The proposed algorithm yields an enhanced image according to simple user annotations, as shown in Figure 2. By inspecting an image $\mathbf{I}$, a user provides an exposure level η for controlling the global brightness and two types of scribbles: blue and red scribbles, respectively, mean that the user wants to brighten and darken the corresponding local regions. Then, the proposed algorithm generates an edge-aware mask M and reconstructs an enhanced image $\mathbf{J}$ through e-IceNet.

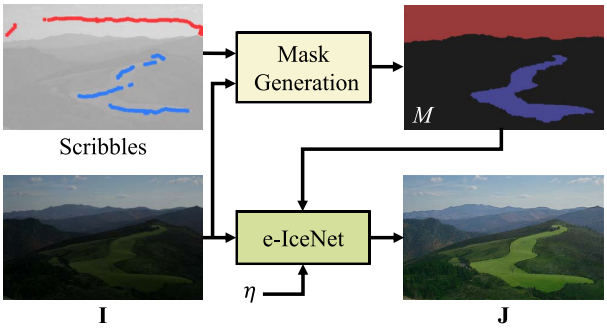


FIGURE 2. Overview of the proposed algorithm. A user provides an exposure level η and scribbles: red or blue scribbles are for darkening or brightening local regions in an image **I**, respectively. Given the annotations, the proposed algorithm generates an edge-aware mask ***M*** and then reconstructs an image **J** through e-IceNet. Please see the supplemental video for a real-time demo of the interactive enhancement.

A. MASK GENERATION

We generate an edge-aware mask by assigning each pixel to one label $l \in \{l_\mathrm{b}, l_\mathrm{d}, l_\mathrm{u}\}$: pixels labeled with l_b, l_d and l_u are to be brightened, darkened, and unchanged, respectively. Some may want to brighten a certain region, while others may want to darken the same region or leave it as it is. To satisfy such different preferences, we perform the labeling based on user scribbles. However, for users' convenience, we do not require the scribbles to be accurate to the pixel level. Instead, we need only rough scribbles for l_b and l_d. Then, we propagate them in an edge-aware manner by adopting the MRW system [12], which simulates the interactions of multiple agents on a graph. Note that MRW was originally developed for unsupervised segmentation. We extend it to accommodate user scribbles.

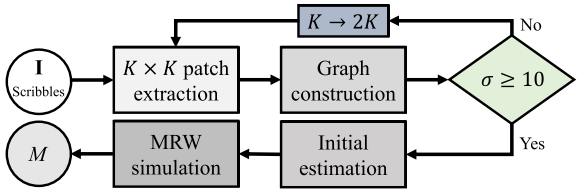


FIGURE 3. Flowchart for the edge-aware mask generation.

Figure 3 is the flowchart for the edge-aware mask generation. Instead of an entire image, we extract a $K \times K$ patch around each scribble to reduce the computational complexity. Then, we divide the patch into superpixels. We compute the standard deviation of each of R, G, B channels of all superpixels and calculate the average standard deviation σ over the three channels. If σ is smaller than a threshold ($= 10$), we double the patch size to consider a broader region. Initially, we set $K = 32$. Next, we estimate initial distributions of the three agents. Finally, we perform the MRW simulation, yielding an edge-aware mask.

TABLE 1. Distance metrics in the similarity function s.

l	Feature	Feature encoding	λ_l	Dimension	Distance
1	RGB	Super-pixel mean	2.5	3	Euclidean
2	LAB	Super-pixel mean	1.0	3	Euclidean
3	MFF	Super-pixel mean	10^3	32	Euclidean

1) GRAPH CONSTRUCTION

First, we over-segment an image into N superpixels [39]. The number of superpixels is set to $N = \min\{4K, 250\}$. We construct the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \ldots, v_N\}$ is the set of nodes (or superpixels) and $\mathcal{E} = \{e_{ij}\}$ is the set of edges. Each edge e_{ij}, connecting neighboring nodes v_i and v_j, is assigned a weight

$$w_{ij} = \exp\Big(-\sum_l \lambda_l d_l(v_i, v_j)\Big) \tag{1}$$

where d_l is the distance metrics of node features. Table 1 summarizes the distance metrics. Here, the MFF feature is extracted from the proposed MFF module in Figure 5.

2) INITIAL ESTIMATION

We employ three agents for the three classes in $\{l_\mathrm{b}, l_\mathrm{d}, l_\mathrm{u}\}$: b-agent, d-agent, and u-agent. We set the initial distributions of b-agent and d-agent uniformly along the nodes overlapping with brightening and darkening scribbles, respectively. Also, the initial distribution of u-agent is set uniformly along the boundary nodes, excluding those similar to the scribbled nodes. Specifically, we allocate uniform probabilities to the nodes v at the image boundaries that satisfy the conditions

$$s(v, v_\mathrm{b}) < \kappa, \quad \forall v_\mathrm{b} \text{ and } s(v, v_\mathrm{d}) < \kappa, \quad \forall v_\mathrm{d}, \tag{2}$$

where $\kappa = 10^{-4}$. Here, v_b and v_d are the nodes overlapping with the brightening and darkening scribbles, respectively.

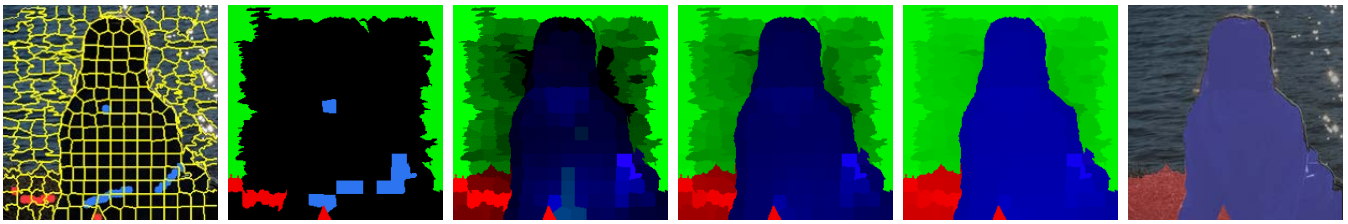
3) MRW SIMULATION

The three initial distributions are refined by MRW iterations. In MRW, each agent travels on the graph $\mathcal{G}$ according to the transition probability a_{ij} to move from node v_j to node v_i. We obtain a_{ij} by normalizing w_{ij} in (1), $a_{ij} = w_{ij}/\sum_k(w_{kj})$. We then construct the transition matrix $\mathbf{A} = [a_{ij}]$.

For simplicity, we describe the MRW process from the viewpoint of b-agent. The others, d-agent and u-agent, are processed in the same manner. Let $\mathbf{p}_\mathrm{b}^\theta = \left[p_{\mathrm{b},1}^\theta, \ldots, p_{\mathrm{b},N}^\theta\right]^\mathrm{T}$ be the distribution of b-agent, in which $p_{\mathrm{b},i}^\theta$ is the probability that b-agent is found at node v_i at iteration θ. The random movement of b-agent is determined recursively by

$$\mathbf{p}_\mathrm{b}^{\theta+1} = (1-\epsilon)\mathbf{A}\mathbf{p}_\mathrm{b}^\theta + \epsilon\mathbf{r}_\mathrm{b}^\theta \tag{3}$$

where $\mathbf{r}_\mathrm{b}^\theta = \left[r_{\mathrm{b},1}^\theta, \ldots, r_{\mathrm{b},N}^\theta\right]^\mathrm{T}$ is the restart distribution. With probability $1-\epsilon$, b-agent moves according to the transition matrix $\mathbf{A}$. On the other hand, with probability ϵ, it is forced to restart with the distribution $\mathbf{r}_\mathrm{b}^\theta$. To make the three agents



(a) Input image (b) $\theta = 0$ (c) $\theta = 10$ (d) $\theta = 20$ (e) $\theta = 30$ (f) Edge-aware mask

FIGURE 4. An example of the MRW simulation: (a) an input image with scribbles and the node set $\mathcal{V}$, (b)~(e) the evolution of the distributions of b-agent (blue), d-agent (red), and u-agent (green), and (f) the generated edge-aware mask.

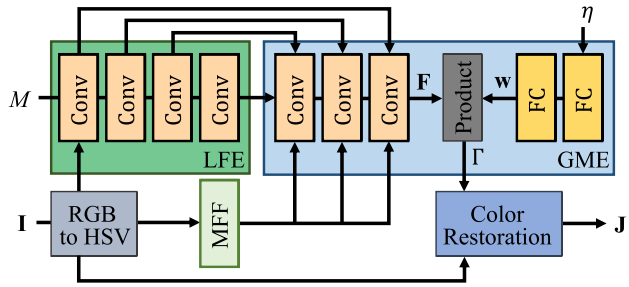


FIGURE 5. The network structure of e-IceNet, which includes the multi-scale feature fusion (MFF), local feature extraction (LFE), and gamma estimation (GME) modules.

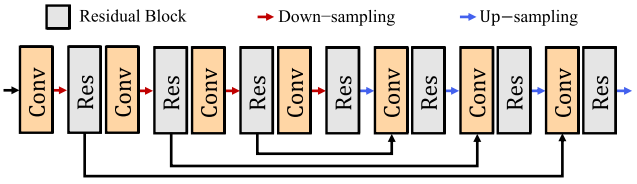


FIGURE 6. The multi-scale feature fusion (MFF) module.

interact with one another, we propose the restart rule, which determines the restart distributions $\mathbf{r}_{\mathrm{b}}^{\theta}$, $\mathbf{r}_{\mathrm{d}}^{\theta}$, and $\mathbf{r}_{\mathrm{u}}^{\theta}$ by considering $\mathbf{p}_{\mathrm{b}}^{\theta}$, $\mathbf{p}_{\mathrm{d}}^{\theta}$, and $\mathbf{p}_{\mathrm{u}}^{\theta}$ jointly. Specifically, we set

$$r_{\mathrm{b},i}^{\theta} = (1-\delta) r_{\mathrm{b},i}^{(\theta-1)} + \delta \frac{p_{\mathrm{b},i}^{\theta} + p_{\mathrm{b},i}^{0}}{\sum_{\mathrm{k}\in\{\mathrm{b,d,u}\}} (p_{\mathrm{k},i}^{\theta} + p_{\mathrm{k},i}^{0})} \tag{4}$$

where δ is a cooling factor [12] and $\mathbf{r}_{\mathrm{b}}^{0} = \mathbf{p}_{\mathrm{b}}^{0}$ is set. Here, we increase the probability that b-agent stays at its scribbled nodes, by adding the initial probability $p_{\mathrm{b},i}^{0}$ at iteration 0. Also, b-agent is enforced to restart with higher likelihoods at nodes in which it is more probable than the other two agents. This makes the three agents repel one another and form their own clusters.

We perform the MRW iterations until the agents yield the stationary distributions $\boldsymbol{\pi}_{\mathrm{b}}$, $\boldsymbol{\pi}_{\mathrm{d}}$, and $\boldsymbol{\pi}_{\mathrm{u}}$. Figure 4 visualizes the evolvement of the three distributions. As the iteration goes on, the agents yield more mutually exclusive probabilities. Eventually, we obtain the stationary distributions that delineate the target regions effectively. Finally, we generate a mask value m_i at node v_i, given by

$$m_i = \begin{cases} 1 & \text{if } \pi_{\mathrm{b},i} > \max\{\pi_{\mathrm{d},i}, \pi_{\mathrm{u},i}\}, \\ -1 & \text{if } \pi_{\mathrm{d},i} > \max\{\pi_{\mathrm{b},i}, \pi_{\mathrm{u},i}\}, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

These mask values form the edge-aware mask M.

B. E-IcENEt

Figure 5 shows the structure of e-IceNet. First, we convert an RGB image $\mathbf{I}$ into the HSV space. Next, given an edge-aware mask M and an exposure level η, we estimate a gamma map Γ for pixel-wise gamma correction. Finally, an enhanced image $\mathbf{J}$ is obtained through the color restoration.

1) GAMMA ESTIMATION

Gamma correction is widely used for CE [11], [22], [40], [41]. It is important to select an appropriate gamma value by considering personal preferences as well as contextual information in an image. We hence determine a gamma value for each pixel in the image $\mathbf{I}$ using the MFF, LFE, and GME modules in Figure 5.

In general, coarse-scale feature maps provide global contexts, whereas fine-scale ones convey detailed local contexts. Because both global and local contexts are important for CE [10], we extract multi-scale contexts through the MFF module in Figure 6. Based on the U-Net architecture [42], MFF includes seven residual blocks [43] and convolutional layers. Each residual block consists of two convolutional layers and a residual connection.

In addition to the contextual information extracted by MFF, we obtain user-specific information for controlling local brightness by feeding the HSV image and the edge-aware mask into the LFE module. Then, in the GME module, we mix the contextual information and the user-specific information using three convolutional layers, yielding a feature map $\mathbf{F}$. Also, we produce a brightness vector $\mathbf{w}$ from η via two fully-connected layers. Note that the brightness vector $\mathbf{w}$ encodes the user preference for the global brightness. Finally, based on the brightness vector $\mathbf{w}$, we convert the feature vector $\mathbf{F}(\mathbf{x})$ at pixel $\mathbf{x}$ to a gamma value $\Gamma(\mathbf{x})$, given by

$$\Gamma(\mathbf{x}) = 10\varphi(\langle \mathbf{F}(\mathbf{x}), \mathbf{w} \rangle) \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is the inner product and $\varphi(\cdot)$ is the sigmoid function. Hence, we have $0 < \Gamma(\mathbf{x}) < 10$.

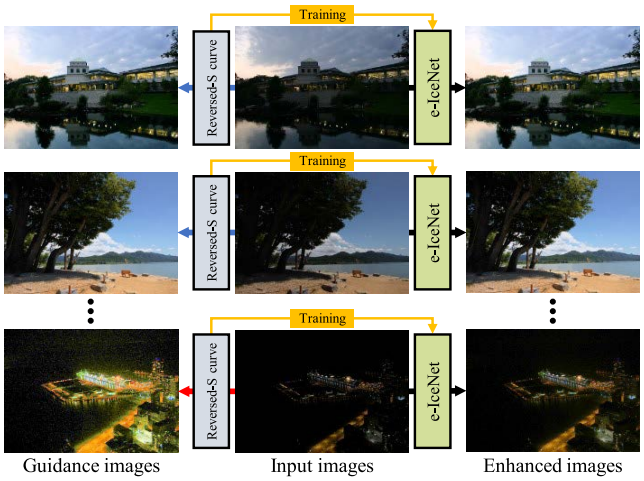


FIGURE 7. Examples of guidance images. Although e-IceNet is trained on both reliable (blue) and unreliable (red) guidance images, it yields robust and effective CE results in all cases.

2) COLOR RESTORATION

In the HSV space, the color image can be enhanced by processing only the value component (V), which corresponds to the luminance image, while preserving the hue and saturation components [22]. Hence, we obtain the gamma-corrected luminance image W by transforming each pixel in V using its gamma value in (6),

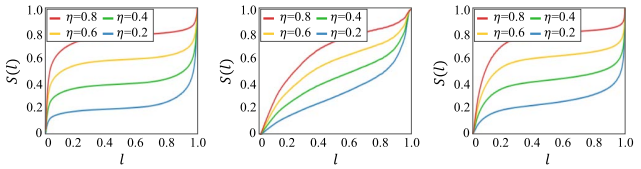
$$W(\mathbf{x}) = V_{\max}\left(\frac{V(\mathbf{x})}{V_{\max}}\right)^{\Gamma(\mathbf{x})} \tag{7}$$

where $V_{\max}$ is the maximum intensity (typically 255). To reconstruct the enhanced image $\mathbf{J}$, we transform the HSV image into the RGB image after replacing V with W.

3) GUIDANCE IMAGES

The CNN-based CE methods in [5], [6], [7], [8], [9], and [10] learn mappings from low contrast images to high contrast ones. Despite their promising CE results, they cannot satisfy various user preferences. On the other hand, the traditional CE methods in [18], [19], [20], [21], and [22] can provide personalized images, satisfying user preferences, by adjusting transformation functions. However, they may fail to yield reliable results for diverse input images.

To take advantage of both approaches, we construct guidance images by adapting traditional transformation functions to meet various requirements. Then, we use those guidance images to train the proposed e-IceNet, yielding reliable results for diverse input images. Since guidance images are obtained based on traditional transformation functions, they may be unreliable in some cases, such as the bottom one in Figure 7. However, e-IceNet is trained using not only unreliable guidance images but also (more numerous) reliable ones. Therefore, it enhances diverse input images effectively and robustly, even though some of them may have poor guidance images.



(a) HE (b) LDR (c) AGCWD

FIGURE 8. The reversed-S curves for various exposure levels η, when the histogram equalization (HE), LDR [20], and AGCWD [22] are used for T in (8), respectively.

We construct a guidance image adaptively according to the exposure level η and the edge-aware mask M. We first derive a reversed-S-shaped curve S from a conventional transformation function T by

$$S(l) = \eta T(l) + (1-\eta)T^{-1}(l) \tag{8}$$

where l denotes an input intensity level and T^{-1} is the inverse function of T. Note that a typical transformation function T for CE is concave and has steep and flat slopes near the minimum and maximum intensity levels, respectively. By combining T and T^{-1}, the reversed-S curve S has steep slopes at both the minimum and maximum levels, as shown in Figure 8, so it can enhance both under- and over-exposed regions effectively. Also, we can control the overall output brightness by changing η. Various transformation functions can be used for T. In the default mode, we adopt AGCWD [22] to generate T adaptively according to the intensity distribution of an image.

To enable e-IceNet to control local brightness, we add the edge-aware mask M to the input intensity V,

$$\tilde{V} = V + \lambda M \tag{9}$$

where λ is a parameter controlling the impacts of M, which is fixed to 10. Then, we obtain the guidance image G by transforming each pixel in $\tilde{V}$ via S, given by

$$G(\mathbf{x}) = S(\tilde{V}(\mathbf{x})). \tag{10}$$

4) AUTOMATED INITIALIZATION

It is useful to present a user with an automatically enhanced image and then allow the user to adjust it further. To this end, we provide an initial exposure level η_{init}. Especially, we use polynomial regression to accommodate the preference of a new user with minimal effort. It captures a nonlinear relationship between the entropy h of an input image and the corresponding exposure level. Specifically, an initial exposure level is estimated with a cubic function by

$$\eta_{\text{init}} = c_3h^3 + c_2h^2 + c_1h + c_0. \tag{11}$$

The four coefficients $\{c_i\}_{i=0}^{3}$ are obtained from the observations $\{(h_i, \eta_i)\}_{i=1}^{O}$ using the method of least squares, where η_i is the exposure level selected by the user to enhance the ith image with entropy h_i. We perform this automated initialization when $O > 4$.

FIGURE 9. Qualitative comparison of the proposed e-IceNet with IceNet [11]. The results are obtained at $\eta = 0.65$.

C. LOSS FUNCTIONS

We train e-IceNet by minimizing a weighted sum of two losses,

$$\mathcal{L} = \mathcal{L}_{\mathrm{g}} + w_{\mathrm{d}}\mathcal{L}_{\mathrm{d}}. \tag{12}$$

Here, $\mathcal{L}_{\mathrm{g}}$ is the guidance loss, which is the mean square error between the output image W in (7) and the guidance image G in (10). Also, $\mathcal{L}_{\mathrm{d}}$ is the denoising loss. Images captured in low-light environments tend to contain noise, which may be amplified during CE. Some methods [11], [34] attempt to suppress it by encouraging smooth variations between neighboring pixels, but they may blur edges. Instead, based on the observation that noise is negligible in the intensity map $\tilde{V}$ in (9), we design the denoising loss $\mathcal{L}_{\mathrm{d}}$ as

$$\mathcal{L}_{\mathrm{d}} = \sum_{\mathbf{x}} w_{\mathbf{x}}(\sigma_{\tilde{V},\mathbf{x}} - \sigma_{W,\mathbf{x}})^2 \tag{13}$$

where $\sigma_{\tilde{V},\mathbf{x}}$ and $\sigma_{W,\mathbf{x}}$ are standard deviations over the 15×15 patches around $\mathbf{x}$ in $\tilde{V}$ and W, respectively. Also $w_{\mathbf{x}} = \exp(-\sigma_{\tilde{V},\mathbf{x}})$. Note that we assign a large weight $w_{\mathbf{x}}$ when $\sigma_{\tilde{V},\mathbf{x}}$ is small, because amplified noise is more visible in flat regions.

D. IMPLEMENTATION DETAILS

In MFF, for down-sampling, we horizontally and vertically decimate the spatial resolutions with a sampling rate of 2; for up-sampling, we use bi-linear interpolation with a scale factor of 2. We perform zero padding and adopt ReLU as the activation function in all layers. In (3), (4), and (12), we set $\epsilon = 0.1$, $\delta = 0.95$, and $w_d = 10$, respectively.

We use the same 2,002 training images as [11], randomly selected from the Part1 subset of SICE [44]. They include under-, normal-, and over-exposed images. We crop a 128×128 patch randomly for training. We train e-IceNet using the Adam optimizer with a minibatch size of 32. We start with a learning rate of 10^{-3}. The training is iterated for 50 epochs with an RTX 2080Ti GPU. To emulate user annotations, an exposure level η is randomly selected from $[0.2, 0.8]$, and brightening and darkening scribbles, respectively, are generated 0∼5 times at random positions.

FIGURE 10. Results of e-IceNet during the 2nd user study.

IV. EXPERIMENTS

A. COMPARATIVE ASSESSMENT

We compare the proposed algorithm with a recent interactive algorithm (IceNet [11]) and seven conventional ones (AGCWD [22], LDR [20], SRIE [24], LIME [25], Zero-DCE [34], EnlightenGAN [33], RUAS [9]). We obtain enhanced images of the conventional algorithms using the source codes and parameters provided by the authors.

1) INTERACTIVE CE

We conducted user studies to assess the interactive CE performance of the proposed e-IceNet.

First, using 10 images in DICM [20], we asked 15 participants to provide annotations to e-IceNet and IceNet and vote for better results. 150 votes in total (10 images × 15 participants) were cast. The proposed e-IceNet won significantly more votes: it was preferred in 76% of the tests, while IceNet in only 24%. This is because e-IceNet generates more accurate masks from simple scribbles and provides more natural CE results. In contrast, as shown in Figure 1, IceNet does not provide sufficiently accurate masks. Moreover, in Figure 9, IceNet over-enhances images even at a middle exposure level $\eta = 0.65$.

For more subjective assessment, we collected 50 images by choosing the first 10 indexed images from each of the test sets of NPE [23], LIME [25], MEF [45], DICM [20], and VV [46]. Then, we conducted another user study with the participants. It was designed as follows:

1) A participant provides annotations to e-IceNet, which then yields an enhanced image.
2) The eight enhanced images obtained by e-IceNet and the seven conventional algorithms are presented to the participant in random order.
3) The participant votes for the most pleasing result.

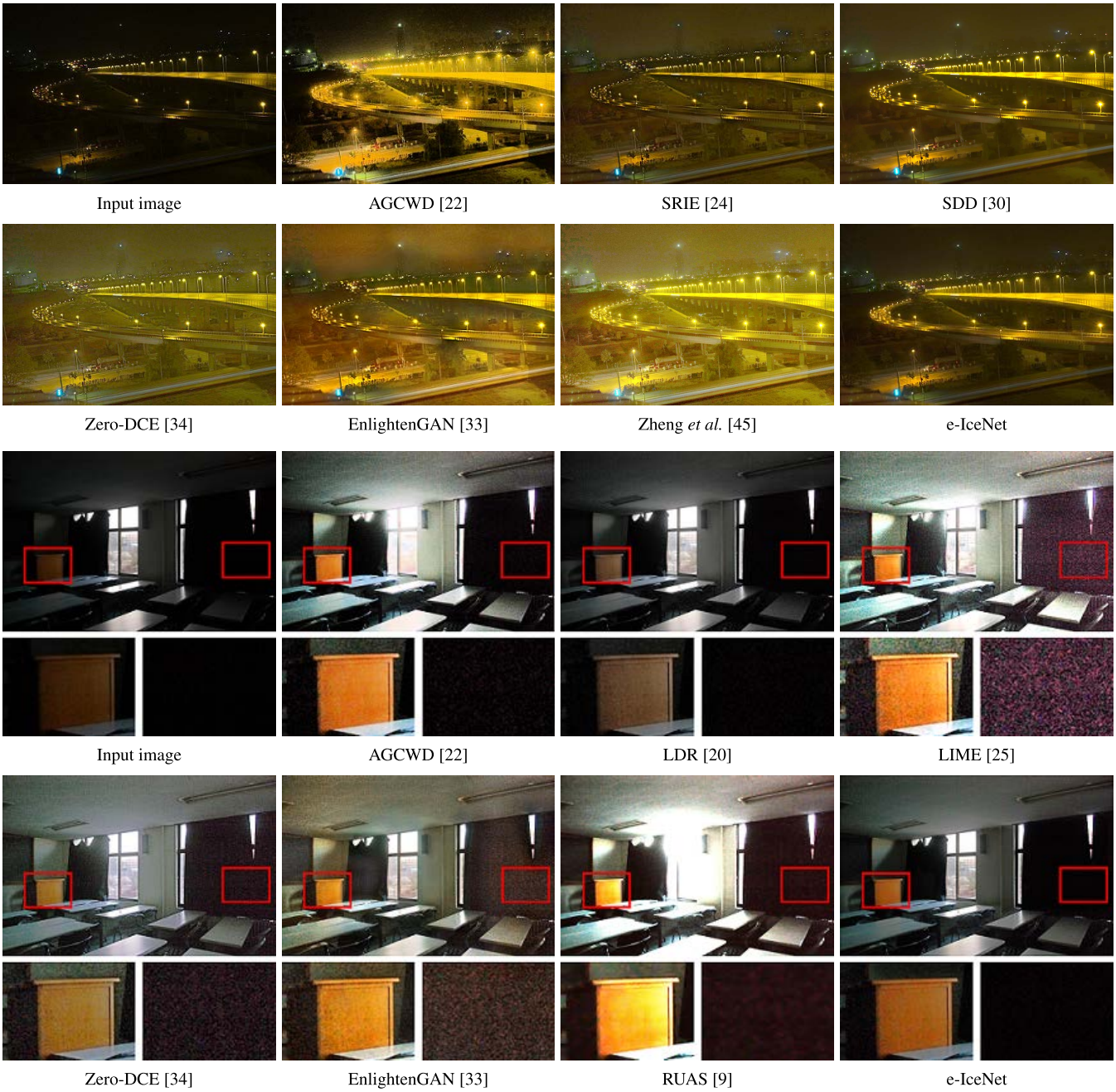


FIGURE 11. Qualitative comparison of the proposed algorithm with the conventional ones. Traditional enhancement methods are in the top row, while CNN-based methods are in the bottom row.

It is recommended to watch the supplemental video for a demo of this 2nd user study. Note that a participant may prefer an automatically enhanced image of a conventional algorithm to the result of e-IceNet. This study was conducted to check whether the participants were sufficiently satisfied with their interactive CE results using e-IceNet.

Figure 10 shows some results of e-IceNet during the 2nd user study. Table 2 summarizes the results. The proposed e-IceNet won the most votes by far. This is because the participants with diverse preferences for contrast were more satisfied with their interactive results than with automatically obtained results. Figure 11 compares qualitative results, in which the result of e-IceNet was obtained by a participant. We see that e-IceNet yields a more natural result by bringing out details without over-enhancement.

2) AUTOMATED INITIALIZATION

Next, we asked the participants to select the preferred image among the automatically enhanced images by LDR [20], SRIE [24], EnlightenGAN [33], and e-IceNet. Note that LDR, SRIE, and EnlightenGAN are the three most preferred algorithms among the conventional methods in Table 2.

TABLE 2. In the 2nd user study, e-IceNet was the most preferred in 53% of the tests, while the second-best SRIE in only 13%.

	AGCWD [22]	LDR [20]	SRIE [24]	LIME [25]
Votes (%)	8.8%	9.9%	13.2%	1.3%
	Zero-DCE [34]	EnlightenGAN [33]	RUAS [9]	Proposed e-IceNet
Votes (%)	3.9%	9.2%	1.1%	**52.7%**

TABLE 3. Subjective assessment of automatically enhanced images.

	LDR [20]	SRIE [24]	EnlightenGAN [33]	Proposed e-IceNet
Votes (%)	10.0%	15.7%	14.0%	**60.3%**

TABLE 4. Full-reference image quality assessment on the Part2 subset of SICE [44]. The best results are boldfaced.

	AGCWD [22]	LDR [20]	LIME [25]	Li *et al.* [26]
PSNR	13.81	13.64	16.17	15.19
SSIM	0.55	0.46	0.57	0.54
	RetinexNet [8]	Zero-DCE [34]	Zheng *et al.* [45]	SDD [30]
PSNR	15.99	16.57	17.59	14.80
SSIM	0.53	0.59	0.63	0.61
	EnlightenGAN [33]	RUAS [9]	IceNet [11]	Proposed e-IceNet
PSNR	16.21	11.84	17.33	**17.71**
SSIM	0.59	0.46	**0.67**	**0.67**

In this test, the proposed e-IceNet automatically generated output images using the exposure levels η_{init} in (11) without requiring annotations. The coefficients for the automatic enhancement were personalized by observing the pairs $\{h_i, \eta_i\}$, which had been selected by each participant during the 2nd user study in Table 2. Then, we collected 20 new test images by sampling five low contrast images from each of NPE, MEF, DICM, and VV. These new images do not overlap with the test images in Table 2. Table 3 summarizes the voting results on these new images: e-IceNet won the most votes again, indicating that it outperforms the conventional methods even without per-image interaction. This confirms that the participants have diverse preferences for contrast and e-IceNet can provide them with satisfactory images automatically through previously collected exposure levels.

To assess automatically enhanced images quantitatively, we use 767 pairs of under- and normal-exposed images in the Part2 subset of SICE [44]. Note that the same test images are used in [34] and [11]. In this test, the coefficients in (11) are determined by the method of least squares, in which 70 pairs of (h_i, η_i) are computed from randomly sampled training images. Table 4 compares the average PSNRs and SSIMs. Being trained on a different paired dataset LoL [8], RUAS [9] performs poorly in this test. The proposed e-IceNet achieves the best PSNR and SSIM scores even without interaction, by generating adaptive results according to the input.

FIGURE 12. Comparison of the proposed mask generation with conventional interactive segmentation techniques [47], [48].

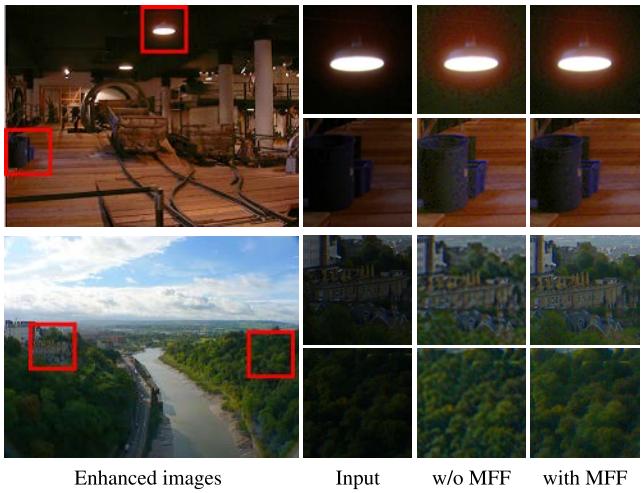


FIGURE 13. Impacts of the proposed MFF module.

B. ABLATION AND ANALYSIS

1) MASK GENERATION

CNN-based interactive segmentation techniques [47], [48], [50], [51], [52] have been developed. However, they train the networks to generate segmentation masks for objects in specific classes in normal-exposed images, yielding unreliable masks for unknown classes in under-exposed images. Figure 12 compares the results of the state-of-the-art interactive segmentation techniques [47], [48] with those of the proposed algorithm. For a fair comparison, we provide the same annotations. The proposed algorithm yields better results on such under-exposed images.

2) MULTI-SCALE FEATURE FUSION

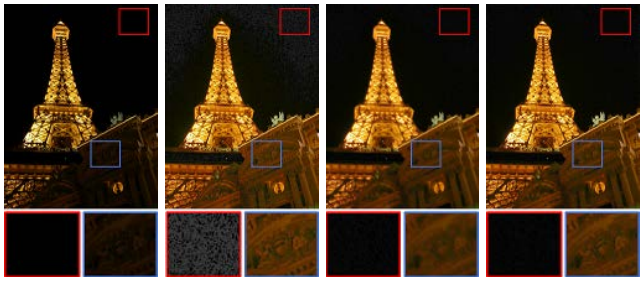
We analyze the impacts of MFF. Figure 13 compares the results of e-IceNet trained without and with MFF. In this test, the output images are generated at $\eta = 0.7$ without scribbles. It is observed that e-IceNet with MFF restores more pleasing images with less noticeable artifacts than that without MFF.

3) TRANSFORMATION FUNCTIONS

We analyze the impacts of transformation functions for guidance images. Figure 14 compares enhanced results

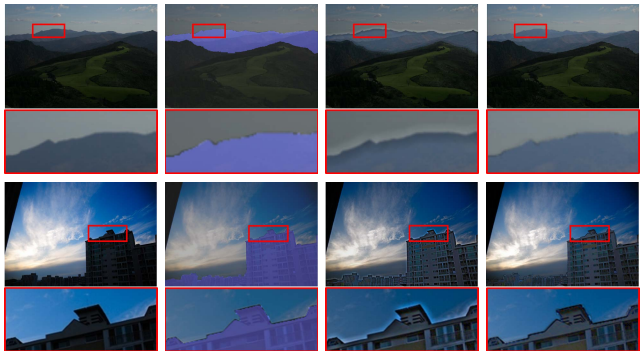
(a) HE (b) LDR [20] (c) AGCWD [22] (d) SPD [53]

FIGURE 14. Comparison of four versions of e-IceNet, trained on guidance images by HE, LDR [20], AGCWD [22], and SPD [53] respectively. Note that differences between (b) and (c) can be seen more clearly in the background (*e.g.* cloud and stone).



Input without $\mathcal{L}_{\mathrm{d}}$ with $\mathcal{L}_{\mathrm{smo}}$ with $\mathcal{L}_{\mathrm{d}}$

FIGURE 15. Comparison of results of e-IceNet trained with different denoising losses. They are obtained at $\eta = 1$ without scribbles.



Input Mask IceNet [11] e-IceNet

FIGURE 16. CE results of IceNet [11] and e-IceNet using the same masks, generated by the proposed MRW simulation.

of e-IceNet trained using transformation functions of HE, LDR [20], AGCWD [22], and SPD [53], some of which are illustrated in Figure 8. Note that SPD is a multi-scale exposure fusion algorithm. Thus, we obtain transformation functions of SPD using multi-scale exposure images. We see that e-IceNet provides different results according to the transformation functions. The original HE tends to over-enhance images, so the corresponding e-IceNet also causes over-stretching. On the other hand, LDR, AGCWD, and SPD can be used to train e-IceNet more effectively. The corresponding versions of e-IceNet enhance the images satisfactorily but in different styles.

4) DENOISING LOSS

Figure 15 shows the results of e-IceNet trained with different denoising losses. We first train e-IceNet without the proposed denoising loss $\mathcal{L}_{\mathrm{d}}$ in Figure 15(b), which incurs severe noise within the red square. Second, we replace $\mathcal{L}_{\mathrm{d}}$ with the smoothness loss $\mathcal{L}_{\mathrm{smo}}$ [11], [34] in Figure 15(c), which yields blurring artifacts within the blue square. Finally, Figure 15(d) shows the result of e-IceNet with $\mathcal{L}_{\mathrm{d}}$. It yields better CE results overall.

5) EDGE-AWARE ENHANCEMENT

Figure 16 shows CE results of IceNet and e-IceNet using the same masks, generated by the proposed MRW simulation. In this test, the results are obtained at low exposure levels to emphasize the differences between the two algorithms. Even when provided with accurate masks, IceNet generates visually annoying rippling artifacts near the mask boundaries. It is because IceNet is trained to encourage smooth variations between adjacent pixels. On the other hand, e-IceNet yields much better results near the boundaries.

6) ROBUSTNESS TO UNRELIABLE GUIDANCE IMAGES

Figure 17 compares enhanced images obtained by AGCWD [22] and e-IceNet at $\eta = 1$ without scribbles. Note that, at $\eta = 1$, $S = T$ in (8), guidance images are

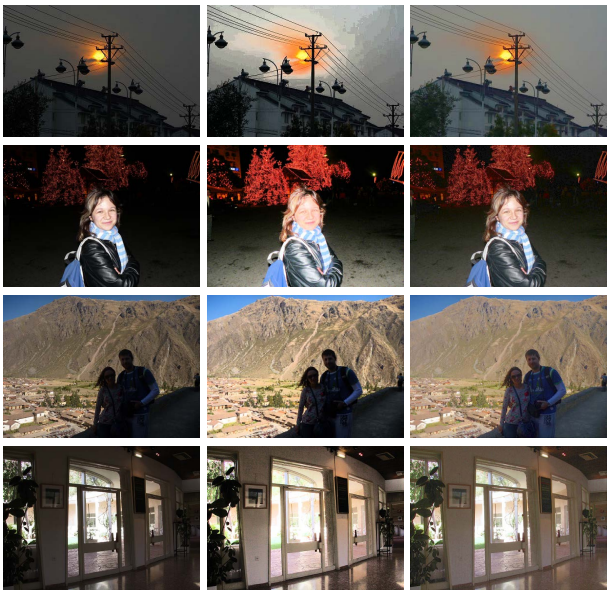


FIGURE 17. Enhanced images obtained by AGCWD [22] and e-IceNet at $\eta = 1.0$ without scribbles.

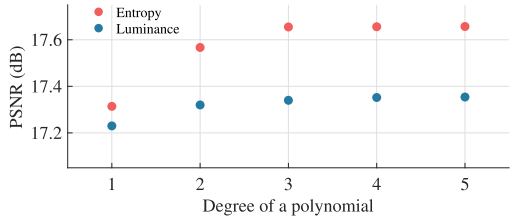


FIGURE 18. PSNR performances on the Part2 subset of SICE [44] according to the degree of a polynomial and the input data.

obtained by AGCWD directly. These results, hence, indicate that e-IceNet yields more reliable results than AGCWD, even though it uses the AGCWD results as the guidance images. Furthermore, e-IceNet can accommodate user scribbles to achieve personalized CE.

7) POLYNOMIAL FUNCTIONS FOR AUTOMATED INITIALIZATION

We use the cubic function in (11) to estimate an initial exposure level. Figure 18 shows that the performance increases up to the degree of three and then saturates, meaning that the cubic function is an effective choice. Also, the blue circles correspond to the PSNRs using the average luminance as input as in [11]. They are poorer than the PSNRs using the entropy. Thus, we adopt the entropy as input.

8) RUNNING TIMES

In this test, we measure the running times of the proposed algorithm with an RTX 2080Ti GPU and a Ryzen 9 3900X CPU. A real-time demo of the proposed algorithm is available in the supplementary video.

Table 5 lists the average running times of the proposed mask generation according to the patch sizes. As the patch size gets larger, the running time increases, but the proposed mask generation is fast enough for practical applications. Thus, we set $K = 32$. Table 6 compares the average running times per image of size 1200×900. Note that the codes in the CPU versions only are available for the conventional algorithms in [20], [22], and [25]. The proposed e-IceNet can be performed in real-time.

TABLE 5. Average running times of the proposed mask generation according to the patch sizes.

Patch size	8×8	16×16	32×32	64×64	96×96
Time (ms)	3	7	15	45	76

TABLE 6. Comparisons of running times in seconds per image.

Method	Time (ms)	Platform
AGCWD [22]	346	Python (CPU)
LDR [20]	49	Matlab (CPU)
LIME [25]	245	Matlab (CPU)
Zero-DCE [34]	2	Pytorch (GPU)
EnlightenGAN [33]	8	Pytorch (GPU)
RUAS [9]	4	Pytorch (GPU)
IceNet [11]	2	Pytorch (GPU)
e-IceNet	3	Pytorch (GPU)

TABLE 7. Average IoU results according to the hyper-parameters.

λ_l	0	1.5	2.5	3.5
	64.48	70.51	**70.65**	67.14
λ_2	0	0.5	1.0	1.5
	66.17	70.61	**70.65**	68.55
λ_3	0	10	10^3	10^4
	70.19	70.26	**70.65**	57.40
σ	3	5	10	15
	61.43	70.19	**70.65**	70.53

9) HYPER-PARAMETERS

Let us discuss the impacts of the hyper-parameters in Figure 3 and Table 1. In this test, we measure the segmentation performance by the intersection over union (IoU) scores with 100 images in the VidSeg dataset [54]. We compute the average IoU scores by changing each parameter in Table 7. In this test, we generate a segmentation mask from one scribble with the initial patch size $K = 64$. The performance increases up to $\lambda_1 = 2.5$, $\lambda_2 = 1.0$, $\lambda_3 = 10^3$, and $\sigma = 10$, respectively, and then decreases.

10) PERSONAL PREFERENCES

Figure 19 shows how two participants enhanced the same input image differently during the user study, which indicates that people have diverse preferences for contrast. Also, by comparing Figure 19(c) with Figure 19(d), we see that the visual quality is affected significantly by user scribbles. The proposed e-IceNet allows the user to specify desired regions

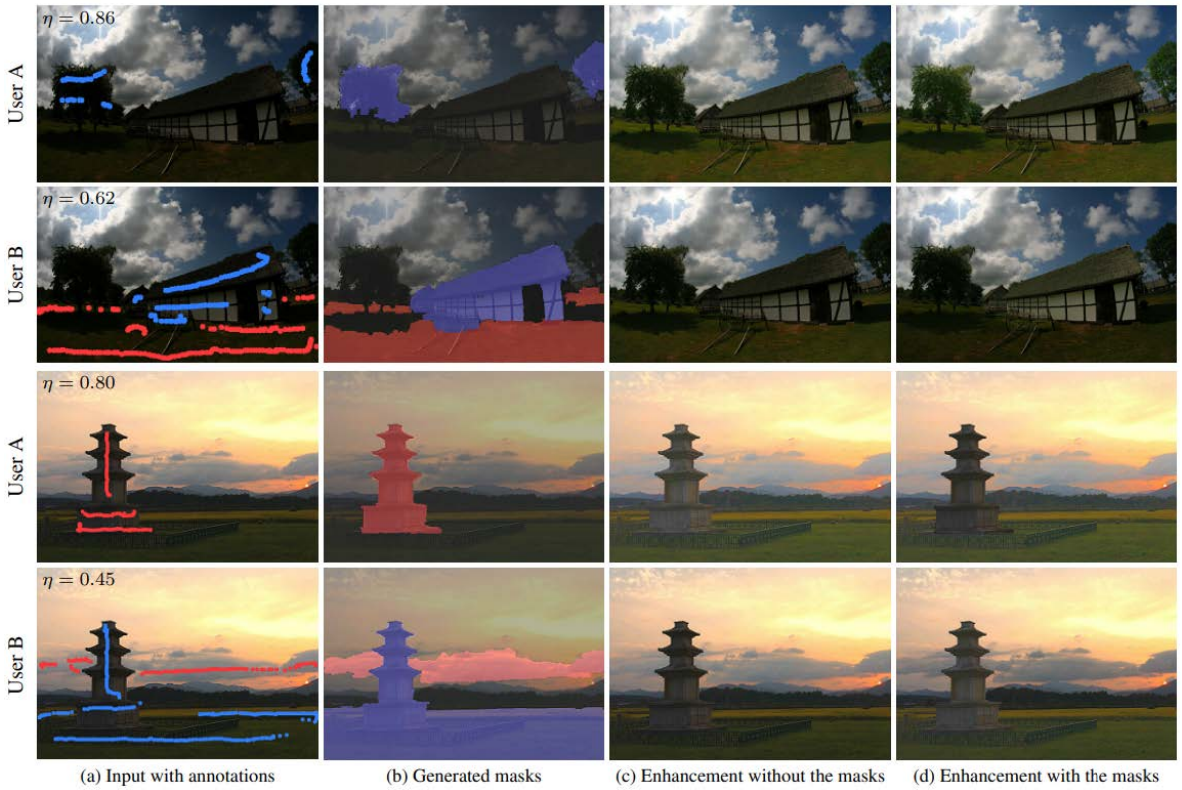


FIGURE 19. **Two participants enhanced the same images differently, which indicates that people have diverse preferences for contrast. Note that the obvious differences between (c) and (d) can be seen around scribbles.**

easily without painstaking annotations, satisfying personal preferences effectively.

V. CONCLUSION

We proposed the edge-aware interactive CE system, composed of the mask generation scheme and e-IceNet. The mask generation scheme enables a user to specify desired regions using only rough scribbles. Then, e-IceNet yields an enhanced image by considering personal preference as well as contextual information. Also, the proposed algorithm can achieve personalized CE without per-image annotation. Extensive experiments demonstrated outstanding CE performance of the proposed algorithm.

REFERENCES

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial Nets," in *Proc. NIPS*, 2014, pp. 1–6.

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[3] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "CycleISP: Real image restoration via improved data synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2696–2705.

[4] K. Ko, Y. J. Koh, and C.-S. Kim, "Blind and compact denoising network based on noise order learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1657–1670, 2022.

[5] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[6] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit. Lett.*, vol. 104, pp. 15–22, Mar. 2018.

[7] 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/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6849–6857.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in *Proc. BMVC*, 2018, pp. 1–12.

[9] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.

[10] H. Kim, S.-M. Choi, C.-S. Kim, and Y. J. Koh, "Representative color transform for image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4459–4468.

[11] K. Ko and C.-S. Kim, "IceNet for interactive contrast enhancement," 2021, *arXiv:2109.05838*.

[12] C. Lee, W.-D. Jang, J.-Y. Sim, and C.-S. Kim, "Multiple random Walkers and their application to image cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3837–3845.

[13] H.-U. Kim, Y. J. Koh, and C.-S. Kim, "PieNet: Personalized image enhancement network," in *Proc. ECCV*, Aug. 2020, pp. 374–390.

[14] J.-H. Kim, W.-D. Jang, J.-Y. Sim, and C.-S. Kim, "Optimized contrast enhancement for real-time image and video dehazing," *J. Vis. Commun. Image Represent.*, vol. 24, no. 3, pp. 410–425, Apr. 2013.

[15] R. W. Liu, Y. Guo, Y. Lu, K. T. Chui, and B. B. Gupta, "Deep network-enabled haze visibility enhancement for visual IoT-driven intelligent transportation systems," *IEEE Trans. Ind. Informat.*, early access, Apr. 27, 2022, doi: 10.1109/TII.2022.3170594.

[16] S. Bae, S. Paris, and F. Durand, "Two-scale tone management for photographic look," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 637–645, 2006.

[17] F. Kou, W. Chen, Z. Li, and C. Wen, "Content adaptive image detail enhancement," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 211–215, Feb. 2015.

[18] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, Sep. 2009.
[19] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.
[20] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.
[21] L. Yuan and J. Sun, "Automatic exposure correction of consumer photographs," in *Proc. ECCV*, Oct. 2012, pp. 771–785.
[22] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
[23] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
[24] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.
[25] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
[26] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust Retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
[27] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.
[28] G. Deng, "A generalized unsharp masking algorithm," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1249–1261, May 2011.
[29] Z. Lu, B. Long, K. Li, and F. Lu, "Effective guided image filtering for contrast enhancement," *IEEE Signal Process Lett.*, vol. 25, no. 10, pp. 1585–1589, Oct. 2018.
[30] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE Trans. Multimedia*, vol. 22, no. 12, pp. 3025–3038, Dec. 2020.
[31] G. Kim, D. Kwon, and J. Kwon, "Low-LightGAN: Low-light enhancement via advanced generative adversarial network with task-driven training," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2811–2815.
[32] T. Ma, M. Guo, Z. Yu, Y. Chen, X. Ren, R. Xi, Y. Li, and X. Zhou, "RetinexGAN: Unsupervised low-light enhancement with two-layer convolutional decomposition networks," *IEEE Access*, vol. 9, pp. 56539–56550, 2021.
[33] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
[34] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. CVPR*, Jun. 2020, pp. 1780–1789.
[35] B. C. Stoel, A. M. Vossepoel, F. P. Ottes, P. L. Hofland, H. M. Kroon, and L. J. S. Kool, "Interactive histogram equalization," *Pattern Recognit. Lett.*, vol. 11, no. 4, pp. 247–254, 1990.
[36] M. Grundland and N. A. Dodgson, "Interactive contrast enhancement by histogram warping," in *Proc. ICCVG* in Computational Imaging and Vision, 2004, pp. 832–838.
[37] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 646–653, 2006.
[38] N. A. Dodgson, M. Grundland, and R. Vohra, "Contrast brushes: Interactive image enhancement by direct manipulation," in *Computational Aesthetics in Graphics, Visualization, and Imaging*. Geneva, Switzerland: Eurographics Association, 2009.
[39] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[40] Z.-G. Wang, Z.-H. Liang, and C.-L. Liu, "A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP," *Displays*, vol. 30, no. 3, pp. 133–139, Jul. 2009.
[41] K.-F. Yang, H. Li, H. Kuang, C.-Y. Li, and Y.-J. Li, "An adaptive method for image dynamic range adjustment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 640–652, Mar. 2019.
[42] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[44] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
[45] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
[46] *VV*. Accessed: Feb. 1, 2022. [Online]. Available: https://sites.google.com/site/vonikakis/datasets
[47] K. Sofiiuk, I. A. Petrov, and A. Konushin, "Reviving iterative training with mask guidance for interactive segmentation," 2021, *arXiv:2102.06583*.
[48] Y. Heo, Y. J. Koh, and C.-S. Kim, "Guided interactive video object segmentation using reliability-based attention maps," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7322–7330.
[49] C. Zheng, Z. Li, Y. Yang, and S. Wu, "Single image brightening via multi-scale exposure fusion with hybrid learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 4, pp. 1425–1435, Apr. 2021.
[50] W.-D. Jang and C.-S. Kim, "Interactive image segmentation via backpropagating refinement scheme," in *Proc. CVPR*, Jun. 2019, pp. 5297–5306.
[51] K. Sofiiuk, I. Petrov, O. Barinova, and A. Konushin, "f-BRS: Rethinking backpropagating refinement for interactive segmentation," in *Proc. CVPR*, Jun. 2020, pp. 8623–8632.
[52] Y. Heo, Y. Jun Koh, and C.-S. Kim, "Interactive video object segmentation using global and local transfer modules," in *Proc. ECCV*, Aug. 2020, pp. 297–313.
[53] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.
[54] W.-D. Jang, C. Lee, and C.-S. Kim, "Primary object segmentation in videos via alternate convex optimization of foreground and background distributions," in *Proc. CVPR*, Jun. 2016, pp. 696–704.

KEUNSOO KO (Student Member, IEEE) received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include image processing and machine learning.

CHANG-SU KIM (Senior Member, IEEE) received the Ph.D. degree (Hons.) in electrical engineering from Seoul National University, in 2000. From 2000 to 2001, he was a Visiting Scholar at the Signal and Image Processing Institute, University of Southern California, Los Angeles. From 2001 to 2003, he coordinated the 3D Data Compression Group in National Research Laboratory for 3D Visual Information Processing in SNU. From 2003 to 2005, he was an Assistant Professor at the Department of Information Engineering, Chinese University of Hong Kong. In September 2005, he joined at the School of Electrical Engineering, Korea University, where he is a Professor. He has published more than 290 technical papers in international journals and conferences. His research interests include image processing, computer vision, and machine learning. He is a member of the Multimedia Systems and Application Technical Committee (MSATC) of the IEEE Circuits and Systems Society. Also, he was an APSIPA Distinguished Lecturer from 2017 to 2018. In 2009, he received the IEEK/IEEE Joint Award for Young IT Engineer of the Year, the Distinguished Dissertation Award from Seoul National University, and the Best Paper Award from *Journal of Visual Communication and Image Representation* (JVCI), in 2014. He was the Editorial Board Member of *JVCI* and an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING. He is a Senior Area Editor of *JVCI* and an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA.