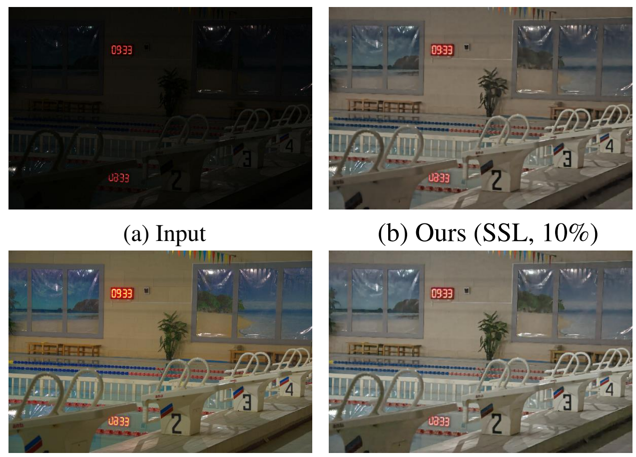
Temporally Averaged Regression for Semi-Supervised Low-Light Image Enhancement

Sunhyeok Lee Donggon Jang Dae-Shik Kim
Korea Advanced Institute of Science and Technology (KAIST)
{sunhyeok.lee, jdg900, daeshik}@kaist.ac.kr

Abstract

Constructing annotated paired datasets for low-light image enhancement is complex and time-consuming, and existing deep learning models often generate noisy outputs or misinterpret shadows. To effectively learn intricate relationships between features in image space with limited labels, we introduce a deep learning model with a backbone structure that incorporates both spatial and layer-wise dependencies. The proposed model features a baseline image-enhancing network with spatial dependencies and an optimized layer attention mechanism to learn feature sparsity and importance. We present a progressive supervised loss function for improvement. Furthermore, we propose a novel Multi-Consistency Regularization (MCR) loss and integrate it within a Multi-Consistency Mean Teacher (MCMT) framework, which enforces agreement on high-level features and incorporates intermediate features for better understanding of the entire image. By combining the MCR loss with the progressive supervised loss, student network parameters can be updated in a single step. Our approach achieves significant performance improvements using fewer labeled data and unlabeled low-light images within our semi-supervised framework. Qualitative evaluations demonstrate the effectiveness of our method in leveraging comprehensive dependencies and unlabeled data for low-light image enhancement.



(a) Input (b) Ours (SSL, 10%)
(c) Ground-truth (d) Ours (SL, 100%)

Figure 1. Visual illustration of the proposed method for low-light image enhancement. (a) is an input low-light image from the LOL dataset [28] and (c) is the corresponding ground-truth image. (d) shows the prediction of our method trained in a fully supervised manner. (b) presents the output of our network trained in a semi-supervised manner using only 10% of the labeled images and the remaining unlabeled low-light images. Our semi-supervised approach with fewer labels outperforms other state-of-the-art comparison methods trained in a fully supervised fashion.

1. Introduction

Low-light conditions significantly degrade the visibility of captured images due to reduced contrast and loss of detail, which can negatively impact the performance of computer vision systems designed for high-quality input images. Researchers have attempted to address the low-light image enhancement problem using methods like Histogram Equalization (HE) [9, 16, 35], which enhance image contrast by expanding the dynamic range, and Retinex-based methods [14] that improve images by decomposing them into reflectance and illumination components. Recent advances in deep learning have led to numerous deep models for low-light image enhancement. However, these models often suffer from limitations such as generating noisy outputs, under/over-enhanced predictions, and the necessity of large amounts of labeled data for supervised learning.

In this paper, we propose a novel end-to-end semi-supervised deep neural network for low-light image enhancement that addresses these limitations. Our method incorporates a Comprehensive Residual Network (CRNet) designed to preserve information-rich features by considering spatial, channel, and inter-layer dependencies. Recognizing the importance of understanding the entire image for image enhancement tasks, our approach exploits both intermediate and high-level features, which allows for a more accurate representation of the image content. We introduce a new progressive enhancement loss function for the training of the proposed structure and a Multi-Consistency Mean-Teacher (MCMT) approach, which extends the Mean-Teacher method [24] by utilizing a Multi-Consistency Regularization (MCR) loss for semi-supervised low-light image enhancement. Our MCMT approach emphasizes multiple consistencies to better leverage unlabeled data for the complex mapping of image enhancement. We encourage the student network to be consistent with the temporally ensembled teacher model by defining distances between their intermediate output results and predictions as the MCR loss function for semi-supervised learning. We update the parameters of the proposed model using the weighted sum of the progressive enhancement loss and MCR loss functions.

Our network demonstrates competitive performance in both synthetic and real paired datasets when trained in a fully supervised manner. Additionally, our proposed method outperforms several state-of-the-art supervised methods when learning in a semi-supervised setting using only 10% of the labels. Qualitative results confirm that our model effectively preserves important features by considering comprehensive dependencies and utilizes unlabeled data for low-light image enhancement. The contributions of our method are summarized as follows:

- We propose a novel network that preserves informative features by considering spatial, channel, and inter-layer dependencies, along with a progressive enhancement loss function to achieve more precise predictions by constraining the intermediate output results.
- We introduce a semi-supervised low-light image enhancement method, the Multi-Consistency Mean-Teacher approach, which effectively utilizes unlabeled data, reducing data acquisition costs for training deep models.
- Our proposed method demonstrates competitive performance when using 100% of the labels. Furthermore, when trained with our semi-supervised method using only 10% of the labels, our approach outperforms several state-of-the-art supervised methods.

2. Related Work

Low-Light Image Enhancement Traditional low-light image enhancement techniques include histogram equalization (HE) methods that enhance image contrast by extending the dynamic range at global or local levels [2, 9, 16, 22, 35], and Retinex-based approaches that decompose images into reflectance and illumination maps, adjusting the illumination maps [6, 10, 11, 14, 17, 21, 26]. However, these methods may struggle to adaptively restore images in various situations. Recent deep-learning-based methods [1, 12, 18, 25, 28, 34] show promising results but can still suffer from artifacts, loss of detail, and color degradation in complex scenes. To address these issues, our proposed architecture takes into account spatial, channel, and inter-layer dependencies.

Semi-Supervised Learning Supervised approaches require a large amount of paired data, resulting in substantial data acquisition costs. By utilizing additional unlabeled data, Semi-Supervised Learning (SSL) can achieve better performance than using only a limited amount of labeled data. Consistency regularization-based methods are popular in SSL, as they assume that predictions from the original input and the perturbed input should be similar. Temporal-ensemble [13] applies augmented input data for consistency regularization and ensembles the results using an exponential moving average of the model's predictions. Inspired by the Mean-Teacher approach [24], which creates a teacher network that is a weighted average of the student network's parameters for consistency targets, we propose a new multi-level consistency loss that leverages temporally emsenbled teacher-generated pseudo-labels and intermediate feature targets to enhance consistency regularization. This approach enables our model to capture both high-level and low-level feature consistency, improving the effectiveness of semi-supervised learning for low-light image enhancement.

Importance Mechanisms Importance mechanisms enable networks to preserve informative features and obtain more accurate results by emphasizing crucial components [7, 8]. Various deep models have incorporated importance mechanisms for tasks such as image generation [32], image classification [7], and image restoration [33]. In the context of super-resolution tasks, HAN [20] introduces the Layer Attention Module (LAM), which considers spatial and channel importance as well as inter-layer correlations to emphasize hierarchical features. Our proposed network incorporates a feature gating mechanism [23] by attaching mask learning convolution layers with a sigmoid function in parallel. This Masked Convolution (MC) module considers both spatial and channel dependencies, effectively integrating importance mechanisms for improved low-light image enhancement performance.

3. Method

Figure 2 presents the proposed Comprehensive Residual Network (CRNet) structure, which accounts for spatial, channel, and inter-layer dependencies. Figure 4 illustrates the overall process of our proposed method, employing

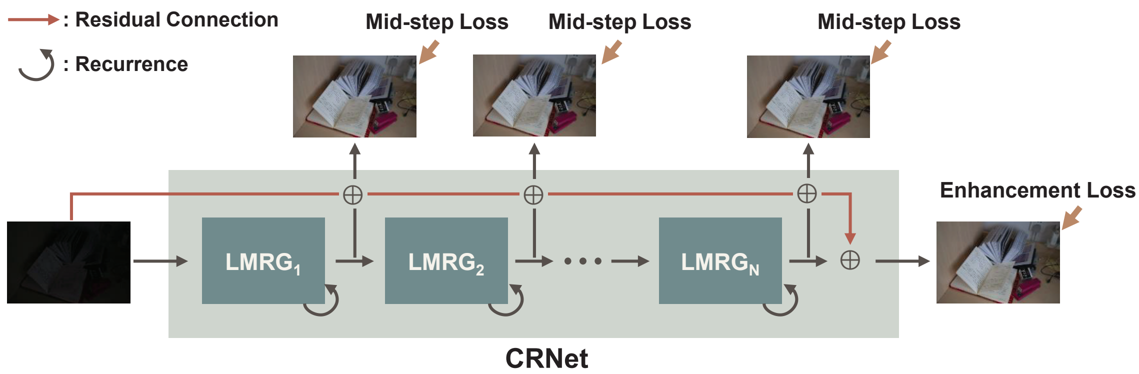


Figure 2. An overview of the CRNet structure. Given an input low-light image, the LMRGs of the CRNet produce intermediate output images. We propose a mid-step loss function to encourage the model to learn a gradual enhancement process. $\oplus$ denotes pixel-wise addition.

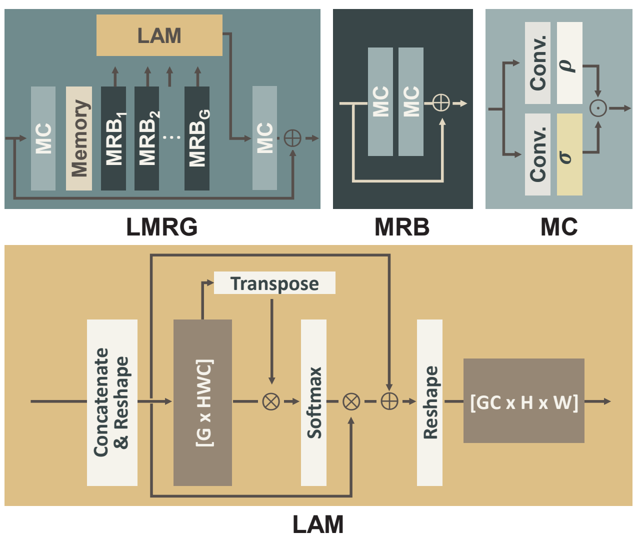


Figure 3. A detailed illustration of the CRNet modules. $\oplus$, $\odot$, and $\otimes$ represent pixel-wise addition, element-wise multiplication, and matrix multiplication, respectively.

multi-consistency regularization for semi-supervised learning (SSL) within our network.

3.1. Progressive Low-Light Enhancement

While existing methods demonstrate promising results, they may generate noisy predictions or misidentify low-light regions as shadows. Furthermore, they often suffer from incomplete details in revealed low-light areas. To address these limitations, we design our network to consider spatial, channel, and layer dependencies, and introduce a novel loss function tailored for low-light image enhancement.

3.1.1 Comprehensive Residual Network

The proposed Comprehensive Residual Network (CRNet) is constructed by stacking masked basic blocks, memory modules, and Layer Attention Modules (LAM) [20]. Figures 2 and 3 depict the detailed structure of our network. The CRNet comprises N Masked Residual Groups with Layer Attentions (LMRG), with each LMRG containing G Masked Residual Blocks (MRB) and an LAM. Each MRB consists of two Masked Convolution Modules (MC) designed to capture spatial and channel dependencies.

$$\begin{aligned} \hat{y} &= \mathbb{E}_x\Big[f_\theta(x)\Big], \\ &= \mathbb{E}_x\Big[g_{\theta,N}(\cdots(g_{\theta,1}(x))\Big]. \end{aligned} \tag{1}$$

In this equation, x represents the low-light input, $f_\theta(\cdot)$ corresponds to the CRNet, and $g_{\theta,n}(\cdot)$ denotes the LMRGs.

Masked Convolution Module Spatial and channel attention mechanisms are widely recognized for preserving informative features essential for image restoration. To enhance feature extraction, we employ a feature gating mechanism [23] that creates soft masks, assigning greater weight to informative features.

Given an input F, the feature extracting convolution ϕ_f with activation ρ generates the input's feature map. Concurrently, the mask learning convolution ϕ_m with sigmoid σ creates a soft mask identifying informative features. We then acquire an improved feature map that considers spatial and channel dependencies through element-wise multiplication of the extracted feature map and the generated weight map, as illustrated in Eq. (2).

$$MC_{\theta,b}(F_i) = \rho\{\phi_{\theta,f}(F_i)\} \odot \sigma\{\phi_{\theta,m}(F_i)\}. \tag{2}$$

The Masked Residual Block (MRB) is constructed by combining two sequential MCs and a skip connection from the input to the output of the final layer, as demonstrated in Eq. (3). We utilize the MRB as a foundational block for building our residual and recurrent network.

$$MRB_{\theta,g}(F_i) = F_i + M_{\theta,2}(M_{\theta,1}(F_i)) \tag{3}$$

Next, we designate a single MC as the LMRG's head part. We then add a convolutional memory module, G MRBs, and a tail MC. The LMRG features a long skip connection from the input to the output of the tail layer (Figure 3) and facilitates R recurrent predictions. Subsequently, we construct the baseline network (CRNet) with N distinct LMRGs, as depicted in Eq. (1) and Figure 2.

Layer Attention Module Although the masked convolution modules within our CRNet capture the spatial and channel-wise dependencies of the features, the mask learning processes operate independently across layers, potentially overlooking inter-layer correlations. To address these inter-layer dependencies, we incorporate the Layer Attention Module (LAM) [20] into each masked residual group of the proposed network, forming the LMRG. Each LMRG generates advanced feature maps that account for hierarchical features.

To calculate inter-layer attention scores, we concatenate the intermediate feature maps of the LMRG, resulting in a dimension of $(GC \times H \times W)$. The LAM reshapes the integrated feature map $(G \times CHW)$ and multiplies it by its transpose, subsequently applying the softmax function to obtain the attention map $(G \times G)$. This attention map reflects the correlation between different layers. We derive improved features from the matrix multiplication of the integrated feature map and the attention map. By adding the residual connection from the input and the predicted attention map with a scale factor τ, and reshaping the output with a dimension of $(GC \times H \times W)$, the LAM generates the final prediction.

$$LAM(\mathbf{F}_i) = \mathbf{F}_i + \tau \sum_{j=1}^{G} w_{j,k} \cdot F_{i,j}, \tag{4}$$

where $\mathbf{F}_i$ denotes the concatenated feature map, $F_{i,j}$ is the j-th feature of the $\mathbf{F}_i$. The initial value of τ is 0, and the network learns the value adaptively. $w_{j,k}$ denotes the inter-layer weight of the j-th and k-th layers.

3.1.2 Progressive Enhancement Loss Function

We define the structural difference (negative structural similarity (SSIM) [27]) between the final output of the CRNet and the ground truth as the enhancement loss. We also define the average *L1* distance between the intermediate stage output of each LMRG and the ground-truth image as the mid-step loss. We propose the progressive enhancement loss L_{PE} by adding the weighted mid-step loss L_{ms} to the enhancement loss L_E.

$$L_{PE} = L_E + \alpha \cdot L_{ms}, \tag{5}$$

where α is the weight for the mid-step loss,

$$\begin{aligned} L_E &= -SSIM(\hat{y}, y), \\ L_{ms} &= \frac{1}{(N-1)} \sum_{n=1}^{N-1} \Big[|\mathbb{E}_x[g_{\theta,n}(x)] - y|_1 \Big]. \end{aligned} \tag{6}$$

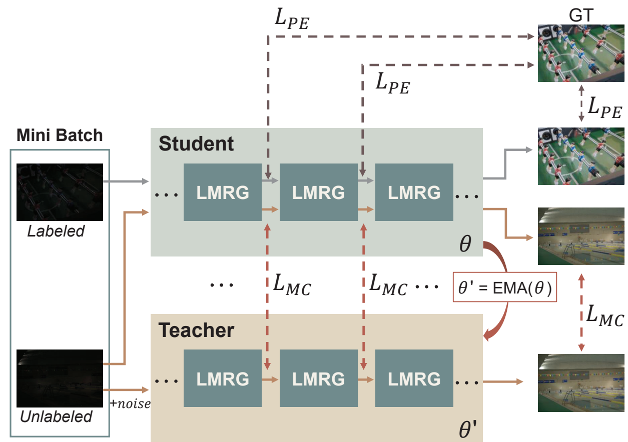


Figure 4. An overview of the MCMT. Given an input batch of labeled and unlabeled data, our student network processes both sets of data, while the teacher network processes the data with added Gaussian noise. We compute L_{PE} using labeled data and the student network output. We also calculate L_{MC} using both networks' output results. We update the parameter θ using L_{SSL}, the weighted sum of L_{PE} and L_{MC}.

3.2. Multi-Consistency Mean-Teacher

We propose the Multi-Consistency Mean-Teacher (MCMT) method, based on [24], to train our model while reducing data acquisition costs. To the best of our knowledge, this is the first end-to-end semi-supervised method for low-light image enhancement.

Weighted Averaged Consistency Target The mean-teacher method [24] employs two separate models with identical structures, referred to as the student (with weights θ) and teacher networks (with weights θ'). The consistency loss L_C is defined as the distance between the student's and teacher's predictions.

$$L_C = \mathbb{E}_{x,x'}\Big[|f_\theta(x) - f_{\theta'}(x')|_2^2\Big]. \quad (7)$$

The student model's parameter θ_t is updated using the consistency loss, while the teacher's parameter θ'_t is defined as the exponential moving average (EMA) of θ_t at training step t.

$$\theta_t = \lambda\theta'_{t-1} + (1-\lambda)\theta_t. \quad (8)$$

Multi-Consistency Regularization Loss Inspired by the performance improvement observed in supervised learning when applying progressive enhancement loss (Table 2), we propose a new multi-consistency regularization loss that maintains consistency between intermediate outputs. We add the weighted mid-consistency loss L_{mc} to L_C, resulting in the multi-consistency loss function L_{MC}.

$$\begin{aligned} L_{MC} &= \mathbb{E}_{x,x'}\Big[|f_\theta(x) - f_{\theta'}(x')|_2^2\Big], \\ &+ \beta \cdot \mathbb{E}_{x,x'}\Big[\tfrac{1}{N-1}\sum_{n=1}^{N-1}|f_{\theta,n}(x) - f_{\theta',n}(x')|_2^2\Big]. \end{aligned} \quad (9)$$

In this equation, $f_{\theta,n}(x)$ represents the output of the n-th LMRG of f_θ, and β is the weight for L_{mc}. Our proposed method guides the student model to maintain more constrained consistency while leveraging information from the unlabeled data.

3.3. The Objective Function

Our CRNet learns from both the paired dataset in a supervised manner and unlabeled data in a semi-supervised manner. For fully supervised training, we only use the progressive enhancement loss L_{PE}. To incorporate unlabeled data into the deep model's training, we employ the total loss L_{SSL} for end-to-end semi-supervised learning, which is calculated as the weighted sum of L_{PE} and L_{MC}.

$$L_{SSL} = L_{PE} + \gamma \cdot L_{MC}. \quad (10)$$

The weight γ for the multi-consistency loss function is empirically set to 1.

4. Experiments

In this section, we first evaluate our proposed structure against other state-of-the-art methods on both synthetic and real paired datasets in a fully supervised setting. Next, we reduce the number of labeled samples in the real dataset and train our CRNet using the proposed loss for semi-supervised learning, comparing its performance with previous semi-supervised and fully supervised methods. We reproduce other state-of-the-art methods using their original codes and settings for comparison purposes.

4.1. Settings

Datasets Our model is compared with other state-of-the-art methods on synthetic and realistic paired datasets [28]. The authors of [28] collect 1000 raw images from RAISE [3] and generate a synthetic dataset by adjusting the histogram of the Y channel. We divide the 1000 image pairs of the synthetic dataset into 900 training and 100 testing pairs. The real dataset [28] consists of 485 image pairs for training and 15 images for testing. For the semi-supervised learning experiment, we randomly select a portion of pairs and use them as labeled pairs. We also evaluate our network and other comparison methods on unlabeled real-world low-light images [6, 15].

Implementation Details Our CRNet consists of 4 LMRGs, with each LMRG having 2 recurrences. We set the number of MRBs to 5. Convolution layers are applied with a kernel size of 3, a stride of 1, and padding of 1. The input, intermediate, and output convolution channels are 6, 32, and 3, respectively. For training, we randomly crop 30 patches of 100 × 100 pixels from each input image. The coefficients α, β, and γ for the loss function are all set to 1. We train the model for 100 epochs using the Adam optimizer with default parameters. The EMA coefficient, λ, is set to 0.99. The learning rate is initially set to 0.0005 and halved at epochs 20, 40, 60, and 90. We train our model on NVIDIA Titan Xp, RTX, and V GPUs.

Table 1. The quantitative comparison results with other state-of-the-art methods on the synthetic and real datasets [28]. The proposed method achieves the new state-of-the-art performance on both the synthetic and the real datasets. **Bold** and <u>underline</u> indicate the best and the second-best scores, respectively.

	Synthetic [28]		Real [28]	
Methods	PSNR	SSIM	PSNR	SSIM
CLAHE [35]	12.58	0.5604	9.46	0.3854
BPDHE [9]	12.50	0.5771	12.10	0.3559
Dong [4]	17.02	0.7539	17.38	0.5895
DHECE [19]	18.14	0.8157	17.97	0.5187
MF [5]	17.75	0.7916	18.03	0.6292
EFF [30]	17.93	0.8096	14.91	0.6866
CRM [31]	19.83	0.8733	18.08	0.7318
LIME [6]	17.67	0.7935	18.10	0.6007
JED [21]	17.05	0.7507	14.17	0.7127
RRM [17]	17.31	0.7471	14.24	0.7150
RetinexNet [28]	18.50	0.8274	17.73	0.7742
KinD [34]	22.34	0.9203	21.56	0.8870
<u>DRBN</u> [29]	<u>23.61</u>	<u>0.9478</u>	<u>22.59</u>	<u>0.8961</u>
CRNet	**24.85**	**0.9613**	**24.01**	**0.9281**

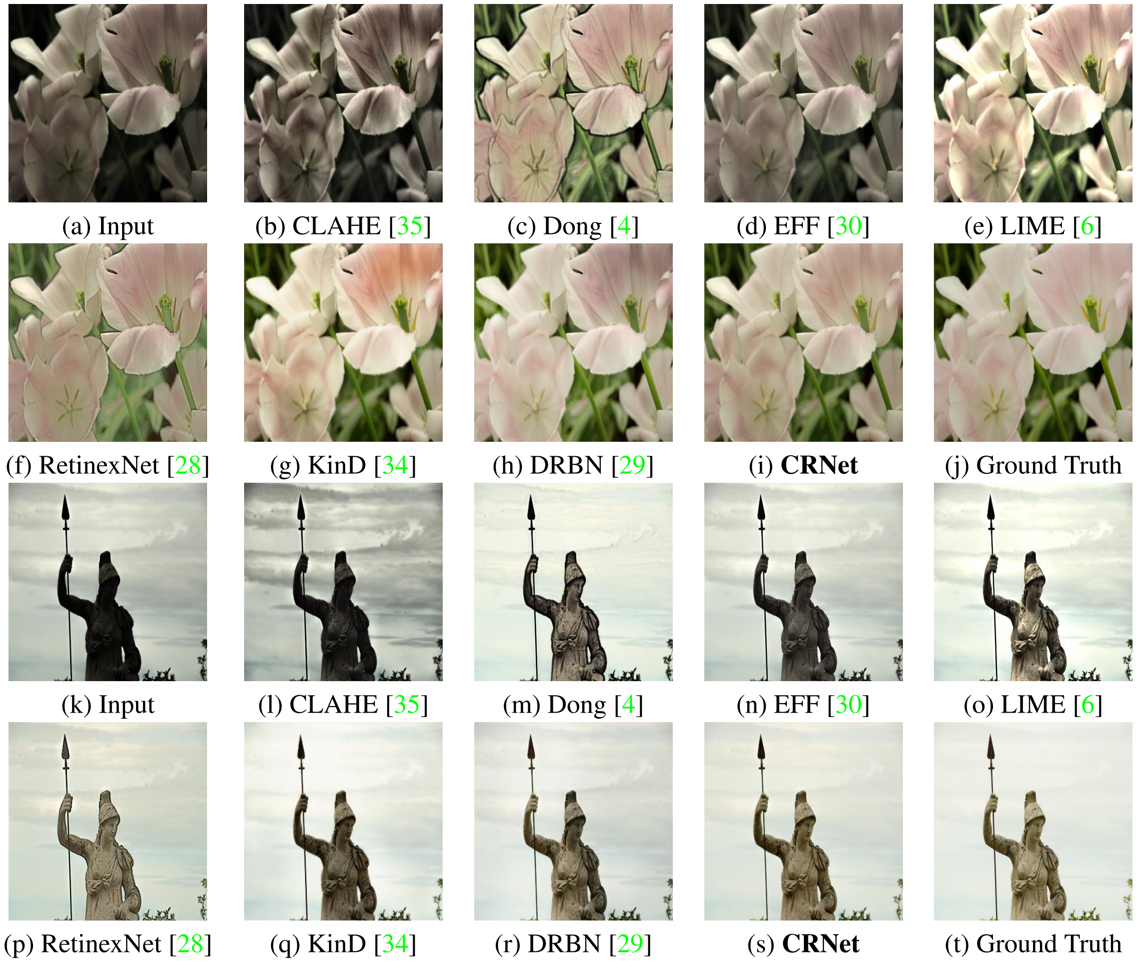


Figure 5. Qualitative evaluation results on the synthetic data [28] in a fully supervised manner.

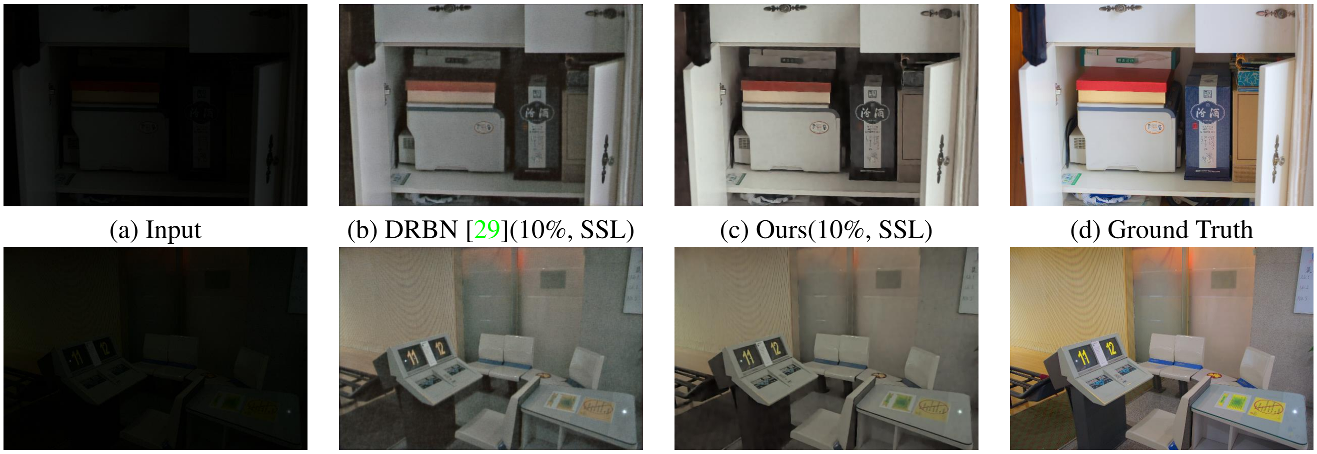
4.2. Comparison with Other Supervised Methods

Quantitative Evaluation on Paired Datasets We assess the performance of the CRNet and other state-of-the-art methods in a supervised setting. Table 1 demonstrates that our method achieves superior results on both synthetic and real datasets. CRNet surpasses other methods by at least +1.24 dB on the synthetic dataset and +1.42 dB on the real dataset in terms of *PSNR*. Furthermore, our proposed method attains the highest results in terms of *SSIM*, with scores of 0.0135 on the synthetic dataset and 0.0320 on the real dataset, respectively.

Qualitative Evaluation on the Paired Dataset Figure 5 showcases the qualitative comparison between our method and other methods on the synthetic dataset. Previous methods tend to underexpose images and struggle to capture the color distribution of the input images accurately. In Figure 5 (h), the method fails to brighten the petals as effectively as our result in (i). In Figure 5 (l-r), other methods manage to brighten low-frequency areas, such as the sky, but they perform poorly in predicting high-frequency detail areas compared to the CRNet result in (s). These qualitative results demonstrate that our model effectively restores natural illumination close to the ground-truth images while preserving high-frequency details, such as the edges of petals and the statue.

4.3. Comparison with Semi-Supervised Methods

Quantitative Evaluation of Semi-Supervised Methods Figure 7 displays the evaluation results of our method and another semi-supervised comparison method. Our SSL method with 10% labeled data (right, red) achieves superior performance compared to the state-of-the-art comparison method [29] trained with 100% labeled data (left, blue). When comparing our method with the supervised model using only 10% of labels (both, gray), it becomes evident that the proposed method benefits significantly from the unla-



(a) Input (b) DRBN [29](10%, SSL) (c) Ours(10%, SSL) (d) Ground Truth

(e) Input (f) DRBN [29](10%, SSL) (g) Ours(10%, SSL) (h) Ground Truth

Figure 6. Qualitative evaluation results on the LOL [28] in a semi-supervised manner using only 10% of the labeled data. Our semi-supervised method trained with 10% of labels successfully suppresses noise and artifacts compared to the previous method [29].

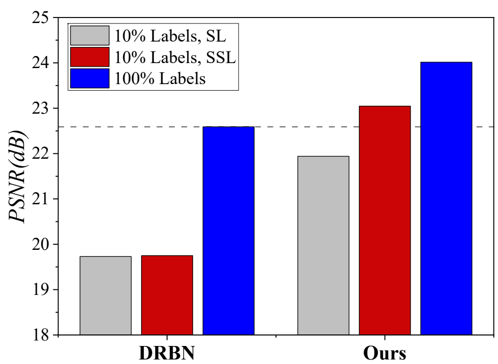


Figure 7. Comparison of semi-supervised low-light image enhancement methods. Our semi-supervised approach using 10% of labels (right, red) achieves significant performance gains from the unlabeled data and outperforms the fully supervised previous method (left, blue).

beled data, more so than the previous semi-supervised approach.

Qualitative Evaluation of Semi-Supervised Methods Figure 6 presents the comparison results between the previous semi-supervised model and our method. We use only 10% of the paired data and the remaining unlabeled low-light images for training. The qualitative results demonstrate that our semi-supervised method effectively reduces noise and artifacts, improving the perceptual quality of the recovered output images. In contrast, the comparison method generates noisy and under-enhanced output results.

Further Evaluation on Unlabeled Real-World Images We compare our method with state-of-the-art methods on unlabeled real-world datasets [6, 15]. Figure 8 illustrates the comparison results on these datasets.

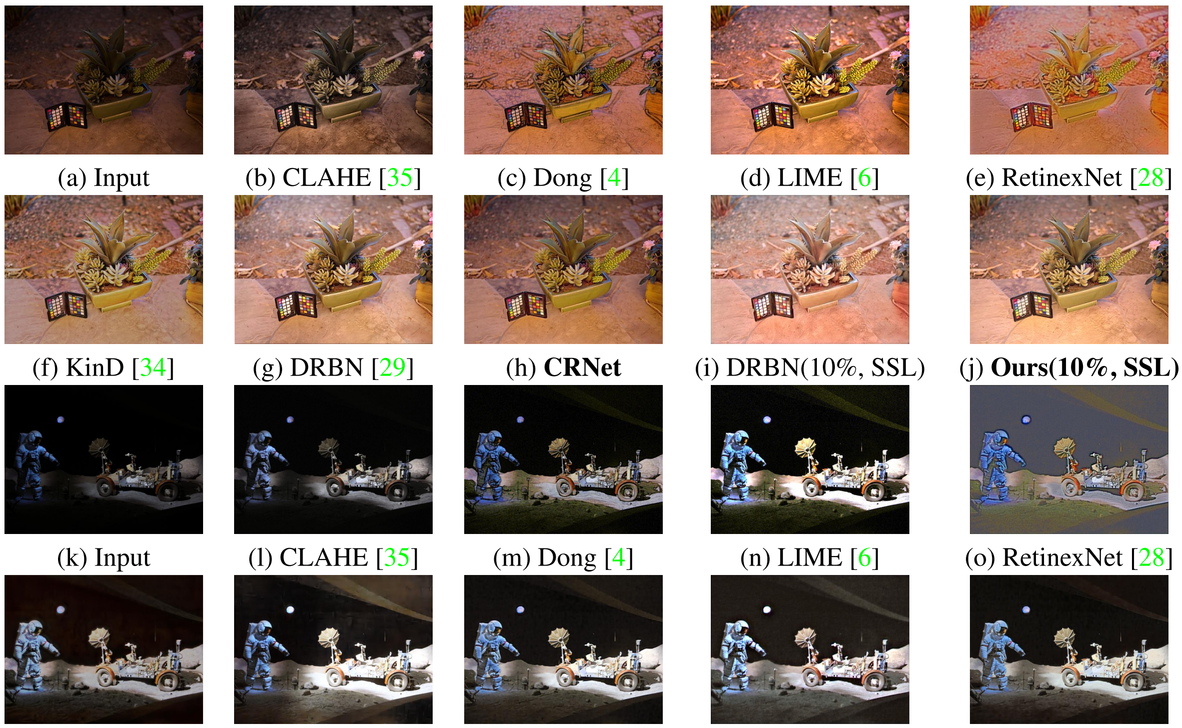
Table 2. Ablation study of the proposed architecture, the mid-step loss, and the multi-consistency loss on the real-world dataset [28].

Method	MC	LA	L_{ms}	L_C	L_{mc}	PSNR	SSIM
RN	-	-	-	-	-	20.83	0.8904
MRN	+	-	-	-	-	22.17	0.9287
MRN+	+	-	+	-	-	22.38	0.9321
CRNet-	+	+	-	-	-	23.67	**0.9326**
CRNet	+	+	+	-	-	**24.01**	0.9281
CRNet(10%)	+	+	+	-	-	21.94	0.9086
Ours-(10%,SSL)	+	+	+	+	-	22.50	**0.9370**
Ours(10%,SSL)	+	+	+	+	+	**23.05**	0.9354

Our model successfully enhances the input low-light images while preserving their content and details, as well as suppressing noise and artifacts. Our approach, as shown in Figure 8 (j) and (t), demonstrates superior light-enhancement performance using only 10% of labeled data, whereas other methods tend to produce suboptimal results with artifacts or noise. Our method retains the shadow in Figure 8 (h) and (j), while other models in Figure 8 (c-g) and (i) mistakenly treat the shadow as low-light regions. These results suggest that our method is effective in recovering complex, unseen low-light images in real-world situations.

4.4. Ablation Studies

Table 2 presents the results of ablation studies to analyze the contributions of the proposed network structure and SSL framework. Each component of the proposed method contributes to its advanced performance. RN refers to the network created by removing the LA and MC from the proposed model. MRN is the structure that adds MC to RN, and MRN+ denotes the method of applying L_{ms} to MRN.



(a) Input (b) CLAHE [35] (c) Dong [4] (d) LIME [6] (e) RetinexNet [28]

(f) KinD [34] (g) DRBN [29] (h) **CRNet** (i) DRBN(10%, SSL) (j) **Ours(10%, SSL)**

(k) Input (l) CLAHE [35] (m) Dong [4] (n) LIME [6] (o) RetinexNet [28]

(p) KinD [34] (q) DRBN [29] (r) **CRNet** (i) DRBN(10%, SSL) (s) **Ours(10%, SSL)**

Figure 8. Qualitative evaluation results on the unlabeled real-world data (upper rows: [6], lower rows: [15]). Our CRNet effectively enhances the real-world low-light images while suppressing noise and artifacts.

CRNet- is the method of removing L_{ms} from our CRNet. Ours-(10%, SSL) is trained without L_{mc}. Each component of the network contributes to our method by preserving informative features and successfully performing low-light image enhancement through progressive recovery. Additionally, the proposed loss functions effectively enhance the performance gains of our approach.

5. Conclusion

Deep models show potential in low-light image enhancement but can produce noisy outputs with lost details. Additionally, acquiring labeled data for supervised learning is costly. To tackle these challenges, we proposed a novel network structure and an end-to-end semi-supervised framework that leverages unlabeled real-world data. Consequently, our method achieved state-of-the-art performance on both synthetic and real paired datasets. Notably, our semi-supervised models using only 10% of labels outperformed existing supervised methods. Our approach effectively enhances low-light images, highlighting the potential of semi-supervised learning in this field.

6. Acknowledgement

This work was supported by the SHINSEGAE-KAIST AI R&D Center funded by the SHINSEGAE I&C Inc. (Project No. G01220220).

References

[1] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 2

[2] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 2

[3] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. Raise: A raw images dataset for digital image forensics. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 219–224, 2015. 5

[4] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu. Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2011. 5, 6, 8

[5] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 5

[6] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 2, 5, 6, 7, 8

[7] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 2

[8] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. 2

[9] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 1, 2, 5

[10] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 2

[11] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997. 2

[12] Dokyeong Kwon, Guisik Kim, and Junseok Kwon. Dale: Dark region-aware low-light image enhancement. *arXiv preprint arXiv:2008.12493*, 2020. 2

[13] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016. 2

[14] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 1, 2

[15] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *2012 19th IEEE International Conference on Image Processing*, pages 965–968. IEEE, 2012. 5, 7, 8

[16] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 1, 2

[17] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 2, 5

[18] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[19] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 445–449. IEEE, 2013. 5

[20] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *European Conference on Computer Vision*, pages 191–207. Springer, 2020. 2, 3, 4

[21] Xutong Ren, Mading Li, Wen-Huang Cheng, and Jiaying Liu. Joint enhancement and denoising method via sequential decomposition. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018. 2, 5

[22] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004. 2

[23] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015. 2, 3

[24] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2, 4

[25] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019. 2

[26] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 2

[27] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4

[28] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2, 5, 6, 7, 8

[29] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3063–3072, 2020. 5, 6, 7, 8

[30] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017. 5, 6

[31] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3015–3022, 2017. 5

[32] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International Conference on Machine Learning*, pages 7354–7363. PMLR, 2019. 2

[33] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 2

[34] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019. 2, 5, 6, 8

[35] Karel Zuiderveld. Contrast limited adaptive histogram equalization. *Graphics gems*, pages 474–485, 1994. 1, 2, 5, 6, 8