

Low-light image enhancement based on U-Net and Haar wavelet pooling

Elissavet Batziou^{1,2}[0000-0003-2645-939X], Konstantinos Ioannidis¹[0000-0001-6767-8762], Ioannis Patras²[0000-0003-3913-4738], Stefanos Vrochidis¹[0000-0002-2505-9178], and Ioannis Kompatsiaris¹[0000-0001-6447-9020]

¹ Information Technologies Institute, Centre for Research and Technology Hellas
6th Km Charilaou-Thermi Road, Thessaloniki, Greece
{batziou.el,kioannid,stefanos,ikom}@iti.gr
<http://www.springer.com/gp/computer-science/lncs>

² School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road London E1 4NS
{e.batziou,ipatras}@qmul.ac.uk

Abstract. The inevitable environmental and technical limitations of image capturing has as a consequence that many images are frequently taken in inadequate and unbalanced lighting conditions. Low-light image enhancement has been very popular for improving the visual quality of image representations, while low-light images often require advanced techniques to improve the perception of information for a human viewer. One of the main objectives in increasing the lighting conditions is to retain patterns, texture, and style with minimal deviations from the considered image. To this direction, we propose a low-light image enhancement method with Haar wavelet-based pooling to preserve texture regions and increase their quality. The presented framework is based on the U-Net architecture to retain spatial information, with a multi-layer feature aggregation (MFA) method. The method obtains the details from the low-level layers in the stylization processing. The encoder is based on dense blocks, while the decoder is the reverse of the encoder, and extracts features that reconstruct the image. Experimental results show that the combination of the U-Net architecture with dense blocks and the wavelet-based pooling mechanism comprises an efficient approach in low-light image enhancement applications. Qualitative and quantitative evaluation demonstrates that the proposed framework reaches state-of-the-art accuracy but with less resources than LeGAN.

Keywords: Image enhancement · low-light images · Haar wavelet pooling · U-Net.

1 Introduction

In the digital era, a plethora of photos are acquired daily, while only a fraction is captured in optimal lighting conditions. Image acquisition in low-lighting conditions is challenging to be improved when low brightness, low contrast, a

narrow gray range, color distortion and severe noise exists, and no other sensor is involved as for example multiple cameras of different specifications. Image enhancement has significantly been improved over the last years with the popularity of neural network architectures that have outperformed traditional image enhancement techniques. The main challenge in low-light image enhancement is to maximize the information involved from the input image on the available patterns and extract the colour information that is hidden behind the low values of luminance. This challenge has been tackled using the traditional approaches of histogram equalization [17] and Retinex theory [21] while neural network approaches either involve Generative Adversarial Networks or Convolutional Neural Networks architectures [14].

Although a wide range of solutions have already been applied in image enhancement, none of them has yet considered photorealistic style transfer. Targeting in improving the quality of under-exposed images, photorealistic style transfer requires as input, one under-exposed image that has been captured in low-lighting conditions, along with an image of normal lighting conditions. The proposed approach is trained using a collection of style images that all have optimal lighting conditions, which are then transferred to the given under-exposed images through an enhancer module.

However, texture, brightness, contrast and color-like information need to be effectively reconstructed in the generated enhanced image, but it has been reported that Haar wavelet-based pooling is superior to max-pooling for generating an image with minimal information loss and noise amplification [36]. Wavelet unpooling recovers the original signal by performing a component-wise transposed-convolution, while max-pooling does not have its exact inverse. Max-pooling that the encoder-decoder structured networks used in the WCT [22] and PhotoWCT [23] are not able to restore the signal and they need to perform a series of post-processing steps for the remaining artifacts. The multi-layer feature aggregation mechanism is adopted as it is presented in [25], but with a Haar wavelet-based transformation to support the Adaptive Instance Normalization for image enhancement in low-lighting conditions.

In this work, we propose an novel framework for image enhancement on low-light conditions. A modified U-Net based architecture is designed that involves dense blocks, which contain dense, convolutional, and wavelet pooling layers to better preserve texture regions in image enhancement. The framework requires as input low-light and normal-light images, where encoding is using low frequency (LL) components and unpooling is performed with high frequency components (LH, HL, HH), which leads to better image enhancement. Our method achieves state-of-the-art results but with less computational resources, both in training and testing phases.

The remainder of the paper is structured as follows. Section 2 provides an overview of state of the art in image enhancement while Section 3 presents the proposed framework analysing and providing details for any of its aspect. Section 4 provides experiments that demonstrate the effectiveness of the proposed approach, and finally Section 5 concludes this work.

2 Related Work

Image enhancement has been initially relied on the statistical method of histogram equalization [4]. Histogram equalization enhances the images through the expansion of their dynamic range either at a global [12] or a local level [17]. In particular, the method [19], which is based on histogram equalization, enhances the image contrast by expanding the pixel intensity distribution range. Other works propose the use of double histogram equalization for image enhancement [2]. The statistical properties of adjacent pixels and large gray-level differences have been utilized by Lee et al. [17] to adjust brightness at local levels. Histogram equalization methods produces visually accepted images with less resources but with many distortion as local pixel correlations are not considered.

Since this type of methods lack physical explanation, several image enhancement methods have been proposed based on Retinex theory. Retinex based methods for image enhancement have estimated illumination and reflectance with more details through the designed weighted variation model [6]. In [10] the authors have estimated a coarse illumination map, by computing the maximum intensity of each RGB pixel channel, and then they refine the coarse illumination map using a structure prior. Another Retinex model has been proposed Li et al. [21] that takes into consideration the noise in the estimation of the illumination map through solving the optimization problem of generating a high-contrast image with proper tonal rendition and luminance in every region. Furthermore, an adaptive multi-scale Retinex method for image enhancement has been proposed in [16] where each single-scale Retinex output image is weighted and it is adaptively estimated, based on both the input content image to generate an enhanced image with a natural impression and appropriate tone reproduction in each area of the image. Additional works have combined the camera response with the traditional Retinex model to obtain an image enhancement framework [27]. Contrary to the methods that change the distribution of image histogram or that rely on potentially inaccurate physical models, Retinex theory has been revisited in the work of [35]. The authors proposed Retinex-Net, that uses a neural network network to decompose the input images into reflectance and illumination, and in an additional layer an encoder–decoder network is used to adjust the illumination. Contrary to the combination of Retinex based methods with neural networks for image enhancement, these methods may generate blurry artifacts that reduce the quality of the generated image.

In the deep learning era, several approaches have been introduced for image enhancement. In [24], the authors apply the LLNet on the low-light image enhancement problem, proposing the stacked sparse denoising of a self-encoder, based on the training of synthetic data to enhance and denoise the low-light noisy input image. In [29] the authors have proposed MSRNet, that is able to learn the end-to-end mapping from low-light images directly to the normal light images with an optimization problem. In [8] the authors introduce a neural network architecture that performs image enhancement with two layers; a data-driven layer that enables slicing into the bilateral grid, and a multiplicative operation for affine transformation to compute a bilateral grid and by predicting local affine

color transformations. In addition, the work of [13] proposes the use of a translation function to learn with a residual convolutional neural network training, aiming to improve both color rendition and image sharpness. The neural network includes a composite perceptual error function that combines content, color and texture losses. A convolutional neural network has also been proposed in [32], where the authors introduce an illumination layer in their end-to-end neural network for under-exposed image enhancement, with the estimation of an image-to-illumination mapping for modeling multiple lighting conditions. The work of [20] has proposed a trainable CNN for weak illumination image enhancement, called LightenNet, which learns to predict the relationship between illuminated image and the corresponding illumination map. A feature spatial pyramid pooling model has been proposed in [30] with a low-light enhancement network, in which the image is decomposed into a reflection and an illumination image and then are fused to obtain the enhanced image. A Global illumination-Aware and Detail-preserving Network (GLADNet) has been proposed in [33]. The architecture of the proposed network is split into two stages. For the global illumination prediction, the image is first downsampled to a fixed size and passes through an encoder-decoder network. The second step is a reconstruction stage, which helps to recover the detail that has been lost in the rescaling procedure. Motivated by the good performance of neural networks, but at a higher computational cost, generative adversarial networks combine two neural networks and have also been involved in image enhancement.

The flourishing of GANs has led to several image enhancement solutions aiming to eliminate paired data for training. In [5] the authors introduce EnhanceGAN, which is an adversarial learning model for image enhancement, where they highlight the effectiveness of a piecewise color enhancement module that is trained with weak supervision and extend the proposed framework to learning a deep filtering-based aesthetic enhancer. Moreover, the unpaired learning method for image enhancement of [3] is based on the framework of two-way generative adversarial networks, with similar structure to CycleGAN, but with significant modifications, as the augmentation of the U-Net with global features, the improvement of Wasserstein GAN with an adaptive weighting scheme, and an individual batch normalization layer for generators in two-way GANs. In addition, Zero-DCE [9] takes a low-light image as input and produces high-order tonal curves as its output, which are then used for pixel-wise adjustment on the input range of to obtain the enhanced image. Furthermore, in [7] the authors present an unsupervised low-light image enhancement method named Le-GAN, which includes an illumination-aware attention module and an identity invariant loss. In a similar alternative, the work of [18] proposes an unsupervised learning approach for single low-light image enhancement using the bright channel prior (BCP), where the definition of loss function is based on the pseudo ground-truth generated using the bright channel prior. Saturation loss and self-attention map for preserving image details and naturalness are combined to obtain the enhanced result. Finally, EnlightenGAN [14] is an unsupervised GAN-based model on the low-light image enhancement, that utilizes a one-way GAN and a global-local

discriminator structure. Although these methods enhance the holistic brightness of the input, they result to the generation of over-exposed regions.

In contrast, the proposed approach adopts a modified U-Net based architecture with dense blocks and wavelet pooling layers. The low frequency (LL) component is involved in encoding while decoding utilizes the high frequency components (LH, HL, HH) of the input image. Consequently, the proposed approach is able to preserve the structure and texture information through this wavelet pooling transformation.

3 Methodology

To achieve photorealism, a model is expected to recover the structural information of a given image, while it enhances the image’s luminance effectively. To address this issue, the proposed U-Net-based network is combined with wavelet transformations and Adaptive Instance Normalization (AdaIN), as it is illustrated in Fig. 1. More specifically, the image recovery is addressed by employing wavelet pooling and unpooling, in parallel preserving the information of the content to the transfer network. The frameworks uses as input low-light images along with normal-light images to generate the enhanced images. Afterwards, dense blocks are used to enhance the quality of feature transferring and skip connections in the transferring process. Intending to a natural stylization effect, the enhanced features are inserted into the image reconstruction process, along with the connected features of different levels from the encoding process.

3.1 Network architecture

Following a U-Net alteration, the proposed network is comprised by the following three sub-components: encoder, enhancer and decoder. The encoder and decoder formulate the symmetric structure of a U-Net, as it has been originally introduced by Ronneberger et al. [28], aiming to preserve the structural integrity of an image. The encoder includes convolutional and pooling layers that also involve a downsampling mechanism, while the decoder has a corresponding upsampling mechanism. The U-Net architecture has been designed with a skip connection between encoder and decoder modules. On top of the existing skip connections, the Photorealistic Style Transfer Network called UMFA [25] includes also multi-layer feature aggregation (MFA) and AdaIN blocks [11], in a module which is called enhancer. The enhancer module and the encoder both include a Haar wavelet-based pooling layer, replacing the traditional max-pooling layer in order to capture smooth surface and texture information. In Figure 1, the red and blue solid lines represent the features of the normal-light and low-light images, respectively, while the Haar wavelet-based pooling is in the the dense blocks of the encoder and the AdaIN block of the enhancer module.

Encoder: The encoder of the proposed framework is illustrated in the green box of Fig. 1 and it has convolutional layers and downsampling dense blocks. Downsampling dense blocks contain a Haar wavelet-based pooling layer, a dense

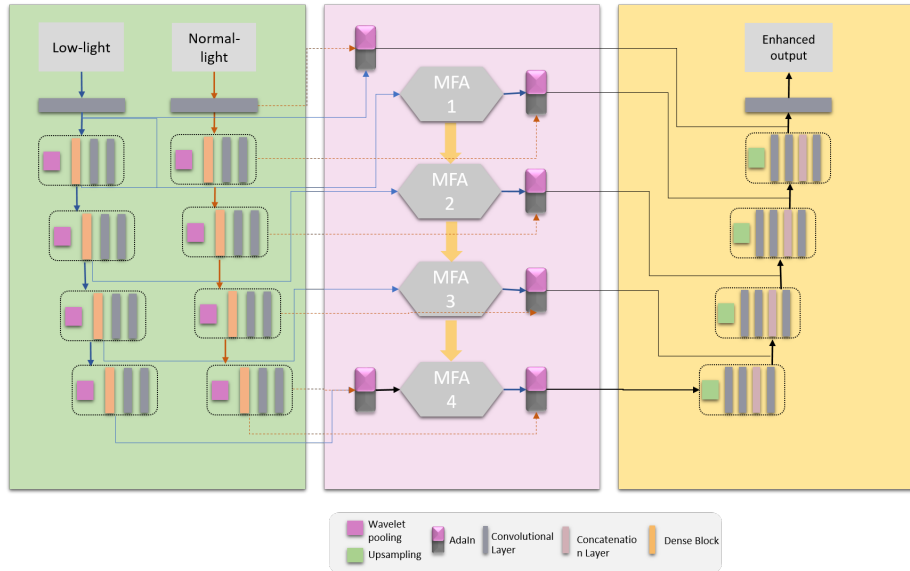


Fig. 1. Proposed image enhancement framework based on U-Net architecture and wavepooling layers.

block and two convolutional layers. A dense block is created by three densely connected convolutional layers. The Haar wavelet transformations of the pooling layer implements downsampling operations to halve the feature size, keeping the dimensions identical to a max pooling layer. The encoder allows to keep multi-scale image information and to process high-resolution images.

Decoder: The decoder, which is illustrated in the yellow box of Fig. 1, mirrors the encoder structure, with four upsampling blocks to formulate a U-Net structure along with the aforementioned encoder. Each block consists of an upsampling layer, three convolutional layers and a concatenation operation that receives the corresponding encoder feature coming from the enhancer module.

Enhancer: In the purple box of Figure 1 the structure of the enhancer module is depicted. This module passes multi-scale features from the encoder to the decoder. Feature aggregation allows the network to incorporate spatial information from various scales and keep the detailed information of the input image. Features are inserted from the previous layer to the next one for all pairs of layers of the encoder, then they are transformed using AdaIN to enhance the input image. The pooling layer is again based on Haar wavelet transformation, briefly described as follows.

Haar wavelet pooling has four kernels, $\{LL^T, LH^T, HL^T, HH^T\}$, where the low (L) and high (H) pass filters are:

$$L^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \quad (1)$$

$$H^T = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \end{bmatrix} \quad (2)$$

The output of each kernel is denoted as LL, LH, HL, and HH, respectively, representing each one of the four channels that are generated from the Haar wavelet-based pooling. The low-pass filter LL contains smooth surface and texture, while the high-pass filters (LH, HL, HH) extract edge-like information of different directions (i.e. vertical, horizontal, and diagonal). Every max-pooling of UMFA method [25] is replaced with the wavelet pooling. Motivated by [36], the max-pooling layers are replaced with wavelet pooling, the low frequency component (LL) is passed to the next encoding layer. The high frequency components (LH, HL, HH) are skipped to the decoder directly. Using the wavelet pooling the original signal can be exactly reconstructed by wavelet unpooling operation. Wavelet unpooling fully recovers the original signal by performing a component-wise transposed-convolution, followed by a summation. Haar wavelet, being one of the simplest wavelet transformations, splits the original signal into channels that capture different components.

3.2 Loss functions

Following the approach of [25], the total loss function L_{total} is computed as a pre-defined weighted sum of enhancement loss, content loss and structured similarity (SSIM) loss functions, respectively:

$$L_{total} = \alpha L_{enh} + \beta L_{cont} + \gamma L_{SSIM} \quad (3)$$

where L_{cont} and L_{SSIM} are computed between the input low-light content image and the enhanced output image. L_{enh} is the enhancement loss between the input normal-light image and the enhanced output image. L_{cont} and L_{enh} are based on the perceptual loss [15]. The parameters α, β, γ represents the weights of the above three losses. For the content loss, the $relu_{2,2}$ is chosen as the feature layer, different from the perceptual loss and the Gram matrices of the features are used for the normal-light image features. The enhancement loss L_{enh} is the sum of all outputs from the four ReLU layers of the VGG-16 network and the SSIM loss is calculated as:

$$L_{SSIM} = 1 - SSIM(C, E) \quad (4)$$

where $SSIM(C, E)$ represents the output of structural similarity (SSIM) [34] measure between the input low-light content image C and enhanced output image E . SSIM is a perception-based model that incorporates important perceptual information, including luminance and contrast terms.

4 Experimental Results

The following section provides insights on the considered datasets, the configurations adopted for the conducted experiments as well as the qualitative and quantitative validation of the corresponding results in comparison with other relevant works.



Fig. 2. Qualitative comparison of the proposed framework with state-of-the-art.

4.1 Datasets

Two benchmark datasets namely LOL [35] and SYN [35] respectively were utilised to accomplish an effective performance comparison on the low-light and enhanced images. LOL dataset has 500 image pairs, where 485 pairs of them are randomly selected for training and 15 pairs for testing. SYN dataset has 1,000 image pairs, and where 950 pairs of them are randomly selected for training and 50 pairs for testing.

4.2 Training process

The training process was initiated by feeding the model with the normal-light and low-light images. For both datasets, the weights for total loss was set as $\alpha = 0.5$, $\beta = \gamma = (1 - \alpha) * 0.5$ following the default values of [25]. Similarly, the number of epochs is set to 150 for the LOL dataset and 100 for the SYN dataset. The batch size is set 4 for both datasets. All input images are 256×256 pixels. Adam optimizer is used for training and the optimal learning rate is 0.0001.



Fig. 3. Qualitative comparison of the proposed framework with state-of-the-art on SYN dataset.

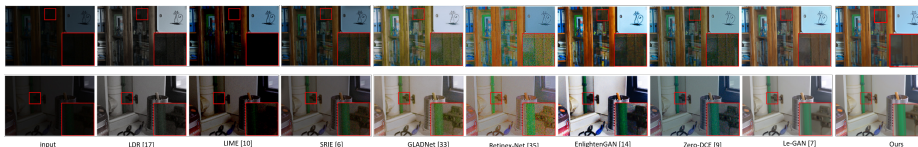


Fig. 4. Qualitative comparison of the proposed framework with state-of-the-art on LOL dataset.

4.3 Results

The baseline methods of the presented comparison are the histogram equalization methods of [1, 17], the Retinex theory based methods of [10, 21, 35, 16, 27, 6], the neural network approaches of [20, 26, 30, 33] and the GANs based approaches of [7, 9, 14] due to their relevance to low-light image enhancement problem and the fact that they cover all different approaches based on histogram equalization, Retinex-based methods, neural networks and GANs. The proposed approach denoted in Table 1 and Fig.2 as "Ours".

Table 1. Quantitative results on LOL and SYN datasets.

2*Methods	LOL		SYN		Times	
	PSNR	SSIM	PSNR	SSIM	training	testing
LIME [10]	15.484	0.634	14.318	0.554	-	-
RobustRetinex [21]	13.876	0.669	16.873	0.714	-	-
JED [26]	13.685	0.657	-	-	-	-
HE [1]	14.800	0.386	-	-	-	-
AMSR [16]	11.737	0.304	-	-	-	-
CRM [27]	12.327	0.472	-	-	-	-
LightenNet [20]	10.301	0.401	-	-	-	-
Zero-DCE [9]	10.770	0.426	15.600	0.796	-	-
FSP [30]	19.347	0.751	-	-	-	-
LDR [17]	15.484	0.634	13.187	0.591	-	-
SRIE [6]	17.440	0.649	14.478	0.639	-	-
GLADNet [33]	20.314	0.739	16.761	0.797	-	-
Retinex-NET [35]	17.780	0.425	16.286	0.779	-	-
EnlightenGAN [14]	18.850	0.736	16.073	0.827	-	-
Le-GAN [7]	22.449	0.886	24.014	0.899	25.6h	8.43ms
Ours	21.266	0.784	19.212	0.716	10.7h	5.27ms

The visual quality of the proposed framework is compared with other methods and the results are shown in Fig 2. It is observed that the GLADNet [33], Retinex-Net [35], EnlightenGAN [14] and Zero-DCE [9] are not able to fully reconstruct the hidden information of the low-light input image on the left. These methods increase the brightness of the image, but the color saturation of the

results is lower compared to the proposed approach. Moreover, they suffer from noise and color bias. In Figs. 3 and 4 we present more qualitative results on SYN and LOL dataset respectively. It is observed that there are some bright regions of the results generated by the compared methods, especially on LOL dataset, which are over-exposed in the enhanced outputs. In contrast, the proposed approach performs well in all datasets with nearly no artifacts and generates quite realistic normal-light images. In addition to the two given datasets, the proposed framework is also tested in a qualitative manner in the no-paired ExDARK [31] dataset. For this purpose as normal-light image is used for the inference mode a random selected cat image from the web and as it is illustrated in Fig.2 the outputs is realistic and enhanced.

Moreover, a quantitative evaluation is provided in Table 1, where the proposed approach is compared with the aforementioned methods for LOL and SYN datasets, as they were reported in [7, 30]. The Peak Signal-to-Noise Ratio (PSNR) scores correspond to the average value of the complete test set of enhanced images in LOL and SYN datasets. We observe that the highest PSNR values for each image test correspond to the output of the proposed framework and the method of Le-GAN [7]. PSNR values show that the generated images from the proposed approach and Le-GAN method, obtain higher proximity to the input normal-light images than the images generated by all other methods. Compared to Le-GAN, the proposed approach is significantly more efficient, without the need of the additional computational resources that are reported in [7] and this is depicted in training and testing times (1). More specifically, the training time of Le-GAN method is 25.6 h, 2.5 times higher the training time of the proposed approach and they use 3 NVIDIA 3090ti GPUs. The training time of our method is 10.7h on the NVIDIA GeForce RTX 2060 SUPER. Moreover, Le-GANs execution time on a 600×400 image is about 8.43ms, 1.5 times up to the execution time of the proposed approach. In Le-GAN, the method requires 3*24GB (standard Memory of 3 NVIDIA 3090ti) contrary to 8GB of ours. As far as SSIM value is concerned, the outputs of the presented framework on LOL dataset is on top-2 higher scores and on SYN dataset the outputs have comparable results with other state-of-the-art methods.

5 Conclusions

In this work, a photorealistic style transfer approach is adapted and modified to low-light image enhancement. The proposed method is based on a U-net architecture using dense blocks and three main modules, namely encoder, enhancer, and decoder. The encoder and enhancer include a Haar wavelet based pooling layer for downsampling that has shown superiority compared to max pooling in transferring brightness, contrast and color-like information from one input image to a generated. Experiments in two benchmark datasets compared the proposed method with recent works in image enhancement, showing that the proposed approach on two benchmark low-light datasets achieves state-of-the-art with at less computational resources, both in training and testing processes.

Future modifications will focus on evaluating further the impact of the wavelet incorporation and their deployment in other levels of the architecture.

Acknowledgment

This work was partially supported by the European Commission under contracts H2020-952133 xR4DRAMA and H2020-958161 ASHVIN.

References

1. Abdullah-Al-Wadud, M., Kabir, M.H., Dewan, M.A.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* **53**(2), 593–600 (2007)
2. Chen, S.D., Ramli, A.R.: Minimum mean brightness error bi-histogram equalization in contrast enhancement. *IEEE transactions on Consumer Electronics* **49**(4), 1310–1319 (2003)
3. 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: *Proceedings of the IEEE Conference on CVPR*. pp. 6306–6314 (2018)
4. Coltuc, D., Bolon, P., Chassery, J.M.: Exact histogram specification. *IEEE Transactions on Image processing* **15**(5), 1143–1152 (2006)
5. Deng, Y., Loy, C.C., Tang, X.: Aesthetic-driven image enhancement by adversarial learning. In: *Proceedings of the 26th ACM international conference on Multimedia*. pp. 870–878 (2018)
6. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2782–2790 (2016)
7. Fu, Y., Hong, Y., Chen, L., You, S.: Le-gan: Unsupervised low-light image enhancement network using attention module and identity invariant loss. *Knowledge-Based Systems* **240**, 108010 (2022)
8. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)* **36**(4), 1–12 (2017)
9. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: *Proceedings of the IEEE/CVF Conference on CVPR*. pp. 1780–1789 (2020)
10. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* **26**(2), 982–993 (2016)
11. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: *Proceedings of the IEEE international conference on computer vision*. pp. 1501–1510 (2017)
12. Ibrahim, H., Kong, N.S.P.: Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* **53**(4), 1752–1758 (2007)
13. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 3277–3285 (2017)

14. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing* **30**, 2340–2349 (2021)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European conference on computer vision*. pp. 694–711. Springer (2016)
16. Lee, C.H., Shih, J.L., Lien, C.C., Han, C.C.: Adaptive multiscale retinex for image contrast enhancement. In: *2013 International Conference on Signal-Image Technology & Internet-Based Systems*. pp. 43–50. IEEE (2013)
17. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing* **22**(12), 5372–5384 (2013)
18. Lee, H., Sohn, K., Min, D.: Unsupervised low-light image enhancement using bright channel prior. *IEEE Signal Processing Letters* **27**, 251–255 (2020)
19. Lee, J., Son, H., Lee, G., Lee, J., Cho, S., Lee, S.: Deep color transfer using histogram analogy. *The Visual Computer* **36**(10), 2129–2143 (2020)
20. Li, C., Guo, J., Porikli, F., Pang, Y.: Lightnet: A convolutional neural network for weakly illuminated image enhancement. *Pattern recognition letters* **104**, 15–22 (2018)
21. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing* **27**(6), 2828–2841 (2018)
22. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. *Advances in neural information processing systems* **30** (2017)
23. Li, Y., Liu, M.Y., Li, X., Yang, M.H., Kautz, J.: A closed-form solution to photo-realistic image stylization. In: *Proceedings of the ECCV*. pp. 453–468 (2018)
24. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* **61**, 650–662 (2017)
25. Rao, D., Wu, X.J., Li, H., Kittler, J., Xu, T.: Umfa: a photorealistic style transfer method based on u-net and multi-layer feature aggregation. *Journal of Electronic Imaging* **30**(5), 053013 (2021)
26. Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J., Yang, M.H.: Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing* **28**(9), 4364–4375 (2019)
27. Ren, Y., Ying, Z., Li, T.H., Li, G.: Lecarm: Low-light image enhancement using the camera response model. *IEEE Transactions on Circuits and Systems for Video Technology* **29**(4), 968–981 (2018)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
29. Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: Msr-net: Low-light image enhancement using deep convolutional network. preprint arXiv:1711.02488 (2017)
30. Song, X., Huang, J., Cao, J., Song, D.: Feature spatial pyramid network for low-light image enhancement. *The Visual Computer* pp. 1–11 (2022)
31. Tao, L., Zhu, C., Xiang, G., Li, Y., Jia, H., Xie, X.: Llenn: A convolutional neural network for low-light image enhancement. In: *2017 IEEE Visual Communications and Image Processing (VCIP)*. pp. 1–4. IEEE (2017)
32. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 6849–6857 (2019)

33. Wang, W., Wei, C., Yang, W., Liu, J.: Gladnet: Low-light enhancement network with global awareness. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). pp. 751–755. IEEE (2018)
34. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
35. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)
36. Yoo, J., Uh, Y., Chun, S., Kang, B., Ha, J.W.: Photorealistic style transfer via wavelet transforms. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9036–9045 (2019)