Recovering High Intensity Images from Sequential Low Light Images

Masahiro Hayashi[1], Fumihiko Sakaue[1], Jun Sato[1],
Yoshiteru Koreeda[2], Masakatsu Higashikubo[2] and Hidenori Yamamoto[2]
[1] *Nagoya Institute of Technology, Japan*
[2] *Sumitomo Electric System Solutions Co., Ltd., Japan*

Keywords: Low Light Images, High Intensity Images, Deep Learning, Number Plate Recognition, Sequential Images.

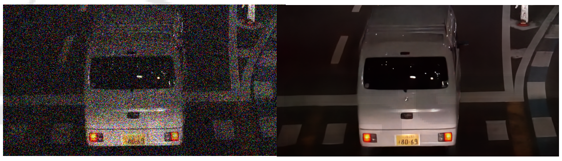
Abstract: In this paper, we propose a method for recovering high intensity images from degraded low intensity images taken in low light. In particular, we show that by using the sequence of low light images, the high intensity image can be generated more accurately. For using the sequence of images, we have to deal with moving objects in the image. We combine multiple networks for generating accurate high intensity images in the presence of moving objects. We also introduce newly defined loss called character recognition loss for obtaining more accurate high intensity images.

1 INTRODUCTION

It is in general difficult to clearly photograph moving objects such as vehicles in low light situations such as at night. When shooting with a long exposure time to obtain a sufficient amount of light, the motion of the object causes a large amount of motion blur. On the other hand, when shooting with a short exposure time to avoid motion blur, large image noise occurs. If we have such motion blur or image noise, we lose high frequency information in the image. As a result, it becomes for example difficult to read the number plate information, which is important for identifying the vehicle from the image.

In order to solve such problems, various methods have been proposed such as imaging techniques using special devices and image processing techniques to recover high quality images from degraded images (Chakrabarti, 2016; Kupyn et al., 2018; Li et al., 2015; Zhang et al., 2017; Remez et al., 2017; Chen et al., 2018). However, the method of using special device is expensive and limited in use. Therefore, in this research, we propose a novel method for recovering high quality images from degraded low light images by using a deep neural network.

Our method uses a sequence of images for recovering high intensity image. The multiple sequential images allow us to obtain more information on the scene and enable us to recover more accurate high intensity image. We also introduce brand new loss called character recognition loss. The new loss enables us to recover high frequency components such as characters in the image, improving the readability of the characters in the recovered high intensity image.



(a) low light image (b) our result

Figure 1: High intensity image recovery. (b) shows a high intensity image recovered from low light image in (a) by using our method. Since the original low light image is too dark, the low light image in (a) is shown with the intensity increased.

Our method can generate highly accurate high intensity images like Fig. 1 (b) from degraded low light images like Fig. 1 (a). The effectiveness of our method is evaluated quantitatively as well as qualitatively. We show that by using our method, the readability of vehicle number plates is drastically improved.

2 RELATED WORK

There are some existing methods for recovering high intensity images from degraded images taken in low light. These existing methods can be divided into two types. The first is an approach that removes mo-

Hayashi, M., Sakaue, F., Sato, J., Koreeda, Y., Higashikubo, M. and Yamamoto, H.
Recovering High Intensity Images from Sequential Low Light Images.
DOI: 10.5220/0010891600003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP*, pages 599-606
ISBN: 978-989-758-555-5; ISSN: 2184-4321

tion blur from images taken with a long exposure time (Shan et al., 2008; Chakrabarti, 2016), and the second is an approach that removes noise from images taken with a short exposure time (Li et al., 2015; Zhang et al., 2017; Remez et al., 2017).

For removing the motion blur, many traditional methods estimate point spread functions (PSF) which represent motion blurs in images and remove the blurs based on the estimated point spread functions (Shan et al., 2008). Recent methods, on the other hand, use deep neural networks for directly removing image blurs without estimating PSFs (Chakrabarti, 2016; Kupyn et al., 2018). However, it is difficult to recover the details of the image by either method.

For denoising low light images, some existing methods estimate image noise first and then enhance low light images with denoising (Li et al., 2015). The deep neural networks are also used for removing noise and enhancing images directly (Zhang et al., 2017; Remez et al., 2017). However, again it is difficult to recover the details of the image as with the deblurring methods. As shown in these existing methods, it is a very difficult problem to recover accurate high intensity images from degraded image taken in low light.

In recent years, a method has been proposed that uses deep learning to recover a clearer high intensity image from a single low light image compared to existing methods (Chen et al., 2018). This method has succeeded in recovering relatively clear image details even in the case of an image with a small amount of noise. However, when there is a lot of noise in the image, the image details cannot be recovered well, and the restoration accuracy is insufficient. This is because the information of the object is largely lost due to image noise, and the information necessary for recovering a clear image is insufficient.

Thus, in this paper, we propose a method that can recover high intensity images with higher accuracy by using multiple low light images. In the proposed method, sequential images of the same object are used as multiple low light images. As a result, information on high frequency components required for accurate recovery can be obtained from multiple images, and as a result, it is expected that more accurate high intensity image can be recovered. However, if we have a moving object in the scene, the position of the object changes in the sequential images. Thus, in this paper, we propose a method for generating a high intensity image while compensating for such a difference in position. We also introduce brand new loss called character recognition loss, which enables us to recover high frequency components and improve the readability of the characters in the recovered high intensity image.

3 PROPOSED METHOD

The network of the proposed method is shown in Fig. 2. As shown in this figure, the proposed method trains three different U-Nets (Ronneberger et al., 2015), that is moving object U-Net, stationary object U-Net and mask image U-Net.

3.1 Alignment of Moving Objects

When dealing with sequential images, point correspondence of moving objects among the sequential images is very important. In case of recovering a high intensity image at time T from T time low light images $\boldsymbol{I}_i$ $(i = 1, \ldots, T)$, the image recovery can be performed more effectively, if the optical flow of the corresponding point is known. Thus, in this research, the optical flow in the sequential images is estimated in advance and used for aligning the moving objects roughly in the images from time 1 to time T. Fig. 3 shows an example of the alignment performed in our method. As shown in this figure, the misalignment of the moving object among the sequential images is almost eliminated by this alignment procedure. By using the sequential images aligned in this way, the improvement of the accuracy in the high intensity image recovery can be expected.

3.2 Generating High Intensity Images from Sequential Low Light Images

In general, image noise occurs randomly for each shot, so even if we shoot in the same scene, we can obtain an image with different noise each time we shoot. Therefore, by using multiple low light images, we can obtain more accurate information about the scene comparing with using only a single low light image.

In this research, we consider a network that inputs the low light images at T times $\boldsymbol{I}_i$ $(i = 1, \ldots, T)$ and outputs a high intensity recovered image $\boldsymbol{I}_R$ at time T, which is the final frame. Therefore, the network can be regarded as the function F shown in the following equation.

$$\boldsymbol{I}_R = F(\boldsymbol{I}_1, \ldots, \boldsymbol{I}_T) \quad (1)$$

In this research, U-Net is used as such a network F. The input of the U-Net is the concat of sequential low light images, and output of the network is the recovered high intensity image. In this way, the network can learn the recovery of accurate high intensity image from multiple noisy low light images.

Suppose the number of vertical pixels in the image is H, the number of horizontal pixels is W, the number

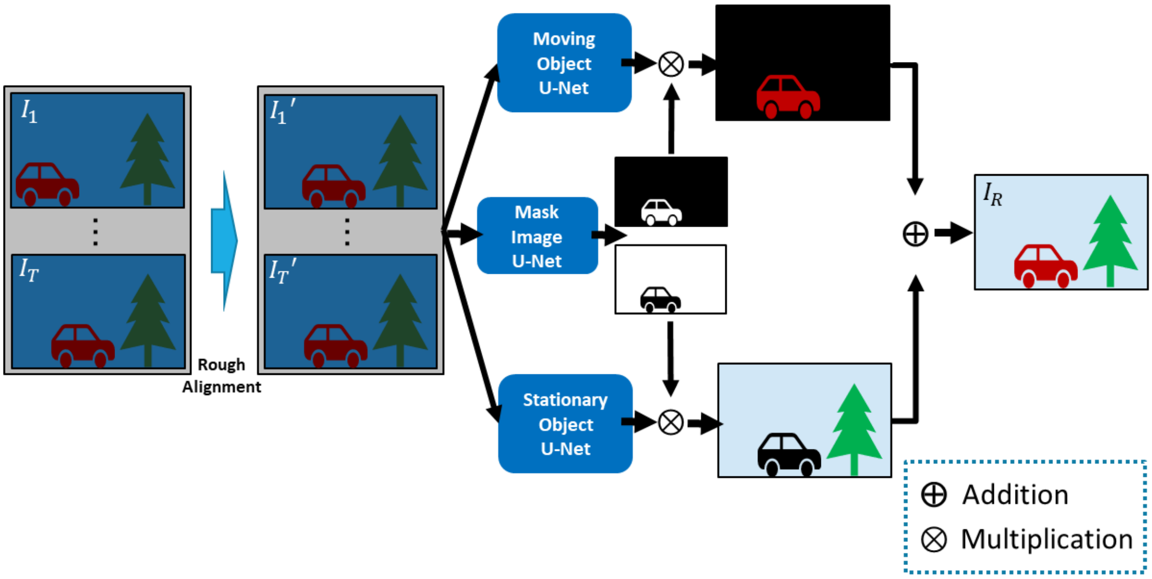


Figure 2: Network of the proposed method. Our network consists of three U-Nets. The moving object U-Net recovers moving objects in the image and the stationary object U-Net recovers stationary objects in the image. The mask image U-Net allows us to separate the image loss of moving objects from the image loss of stationary objects, allowing us to train stationary object U-Net and moving object U-Net separately.

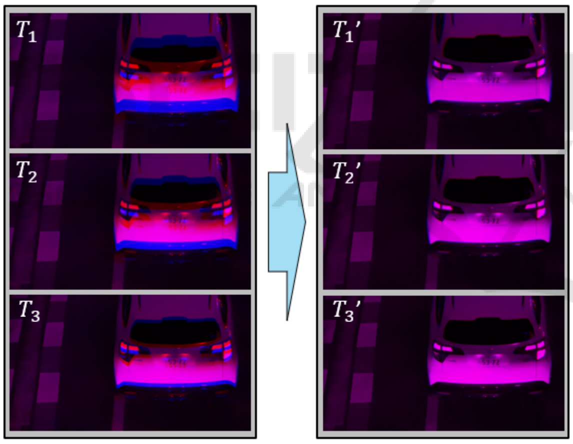


Figure 3: Examples of moving object alignment. The images at time T_1, T_2 and T_3 were aligned with the image at time T_4 by using the optic flow estimation. In order to visualize the misalignment, the image at time T_4 is displayed as R channel, and the images at time T_1, T_2 and T_3 are displayed as B channel.

of channels is C, and the number of times is T. The size of the network input is $H \times W \times CT$, and the size of the network output is $H \times W \times C$. The detail of the network configuration of our U-Net is shown in Fig. 4.

3.3 Multiple U-Nets for Moving and Stationary Objects

Recovery of moving objects is in general more difficult than recovery of stationary objects in sequential images. This is because stationary objects are observed at the same position in sequential images, while moving objects are observed at slightly different positions even if they are aligned by using optical flow. Therefore, in this research, we use two different U-Nets, one is for learning the recovery of moving objects and the other is for learning the recovery of stationary objects, and perform network training according to the characteristics of each to recover the entire image with higher accuracy. we call them moving object U-Net and stationary object U-Net.

For this objective, a mask image for the moving objects and a mask image for the stationary objects are derived by using the third U-Net. We call it mask image U-Net. The network loss of the moving object U-Net and the stationary object U-Net is computed after masking each of the network output image and the ground truth high intensity image with the mask image derived from the mask image U-Net. By learning each of the moving object U-Net and the stationary object U-Net using the loss computed from the mask image, these U-Nets can learn the recovery of the moving object and the recovery of the stationary object respectively.

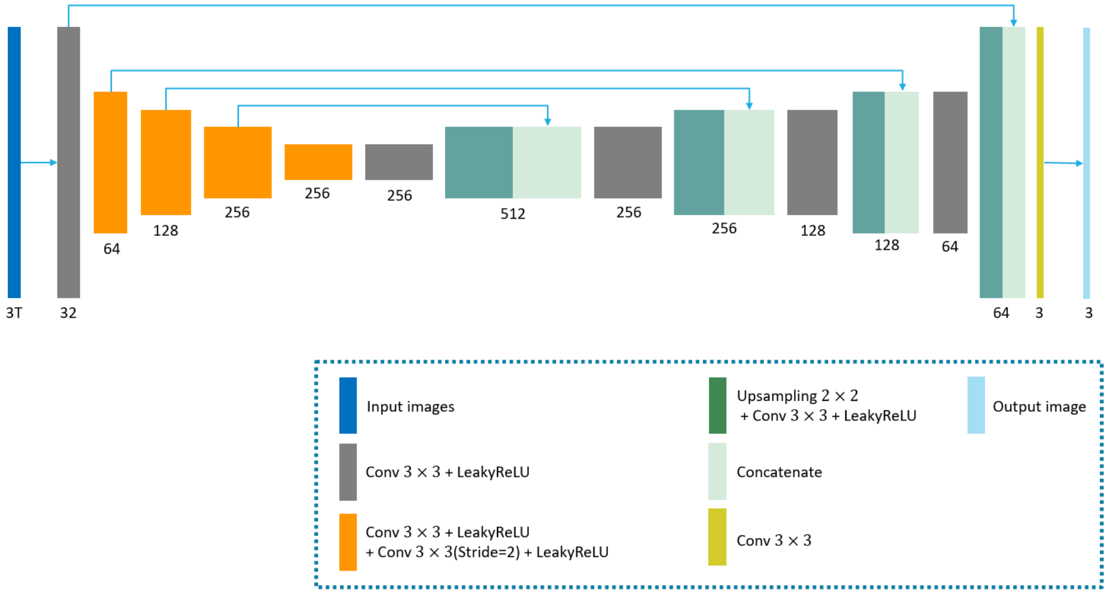


Figure 4: U-Net used in our method.

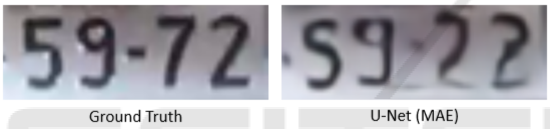


Figure 5: Example where the characters cannot be recovered accurately. The number can be read as ”5922” in the image recovered by using simple L1 loss, but it is actually ”5972” as shown in the ground truth image.

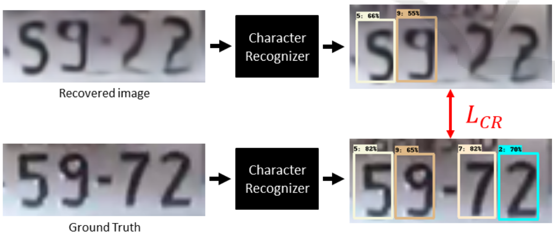


Figure 6: Character recognition loss. The character recognition loss L_{CR} is computed based on the character recognition result when the ground truth image and the recovered image are input to the pre-trained character recognizer.

The loss function for learning the stationary object U-Net is L1 loss between the ground truth image $\boldsymbol{I}_G$ and the output image $\boldsymbol{I}_R$ as follows:

$$L_1 = ||\boldsymbol{I}_G - \boldsymbol{I}_R||_1 \tag{2}$$

The loss function for learning the moving object U-Net uses the L1 loss plus the character recognition loss explained in the next section.

We also train the mask image U-Net that generates a mask image from the input image. The loss function for learning the mask image U-Net is the Binary Cross Entropy L_{BCE} between the ground truth mask image $\boldsymbol{M}_G$ and the generated mask image $\boldsymbol{M}_R$ as follows:

$$L_{BCE} = \boldsymbol{M}_G \log \boldsymbol{M}_R + (1 - \boldsymbol{M}_G) \log (1 - \boldsymbol{M}_R) \tag{3}$$

Since the moving object mask and the stationary object mask have an exclusive relationship with each other, only the moving object mask needs to be estimated by the mask image U-Net.

3.4 Character Recognition Loss

Images recovered by using simple L1 loss shown in Eq. (2) are difficult to recover characters accurately as shown in Fig. 5. In this example, the number can be read as ”5922” in the recovered image, but it is actually ”5972”. Thus, in this research, we introduce character recognition loss in order to improve the accuracy of character restoration, which is important for identifying vehicles in low light images.

The character recognition loss is computed based on the recognition result when the ground truth image and the recovered image are input to the character recognizer, as shown in Fig. 6. In this research, we use Retina Net (Lin et al., 2017) trained by using Street View House Numbers (SVHN) Dataset (Netzer et al., 2011) as a character recognizer. The output of this character recognizer is a character class probability of each extracted character in the input image. For example, if the probability of all classes is less than 0.5, it is considered as the background, and if the probability of ”9” is the highest and is more than 0.5, it is considered as ”9”.

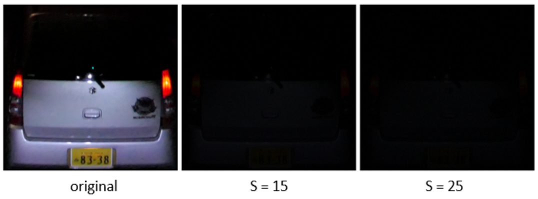


Figure 7: Generated low light images.

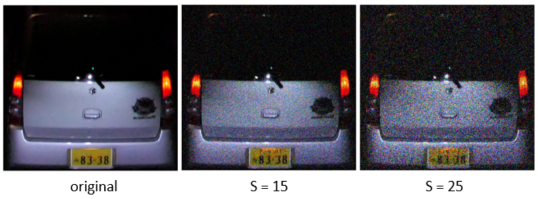


Figure 8: Generated low light images (gain up).

The character recognition loss L_{CR} is computed by taking the L1 loss between the character class probability of the ground truth image P_G and that of the recovered image P_R when the character class probability of the ground truth image P_G is more than 0.5 as follows:

$$L_{CR} = ||P_G^{0.5} - P_R^{0.5}||_1 \tag{4}$$

where, $P^{0.5}$ denotes P when the character class probability of the ground truth image P_G is more than 0.5. In this way, the background area can be ignored and only the character area is considered in the character recognition loss.

The loss function ($Loss$) of the moving object U-Net is defined by using L_1 in Eq. (2) and L_{CR} in Eq. (4) as follows:

$$Loss = (1-\alpha)L_1 + \alpha L_{CR} \tag{5}$$

In our experiment, we used $\alpha = 0.01$.

In order to learn the network of the proposed method, we need pairs of low light sequential image and ground truth high intensity image. However, it is difficult to obtain such image pairs. Therefore, in this research, a low light image with image noise is synthesized from the high intensity image, and a pair of low light image and high intensity image is created and used for learning.

Let us consider a high intensity image $\boldsymbol{I}_0$ of a scene and an image $\boldsymbol{I}_1$ of the same scene taken in low light. Under low light, the S/N ratio of the image decreases, causing large image noise and changing the RGB balance. Therefore, in this research, a low light image with noise is generated by adding noise to the high intensity image as shown below.

First, the high intensity image $\boldsymbol{I}_0$ is divided by S ($S > 1$) to generate a low intensity signal. Then, a random Gaussian noise N of a certain magnitude is added to the low intensity signal. Finally, the low intensity signal is quantized to generate noisy low light image. Thus, the low light image $\boldsymbol{I}_1$ is generated as follows:

$$\boldsymbol{I}_1 = Q\left(\frac{\boldsymbol{I}_0}{S} + N\right) \tag{6}$$

where, Q $(\cdot)$ represents the function that performs quantization.

By increasing the value S, an image with a shorter exposure time and lower light is generated. In this research, we create image pairs under various low light conditions by changing the value S in Eq. (6) from a single high intensity image. An example of the image pair obtained in this way is shown in Fig. 7. Since the generated low light images are too dark in this figure, we show a figure in which the intensity of the low light images is increased in Fig. 8. As shown in Fig. 8, the generated low light image lacks information as when it was taken with a short exposure.

4 EXPERIMENTS

We next show the experimental results obtained from the proposed method. All the networks used in this experiment learn 100 epochs with a learning rate of 0.001.

4.1 High Intensity Image Generation

We first show high intensity images recovered from the proposed method.

For obtaining ground truth high intensity images, the sequential images of 530 moving vehicles were taken using a fixed camera. Of these, 500 vehicles were used as training data and 30 vehicles were used as test data. The corresponding low light images were generated from each high intensity image by the method described in the previous section. In our experiment, low light images with five different values of S in Eq. (6) were generate from a single high intensity image, setting $S =$ 20, 22, 25, 28, 30. For training the network, we used 4 sequential low light images generate with the same S as input of the network and the corresponding high intensity image at time 4 as output of the network. Using the network trained in this way, the high intensity images were recovered from the noisy low light images of the test data.

Fig. 9 shows a set of sequential input low light images and images with simply increased their intensity. As shown in this figure, it is difficult to read the characters on the number plate in the increased intensity

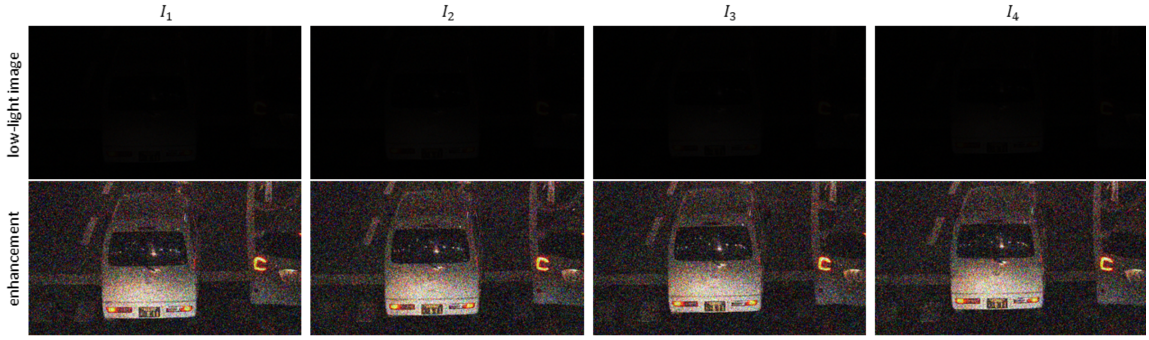
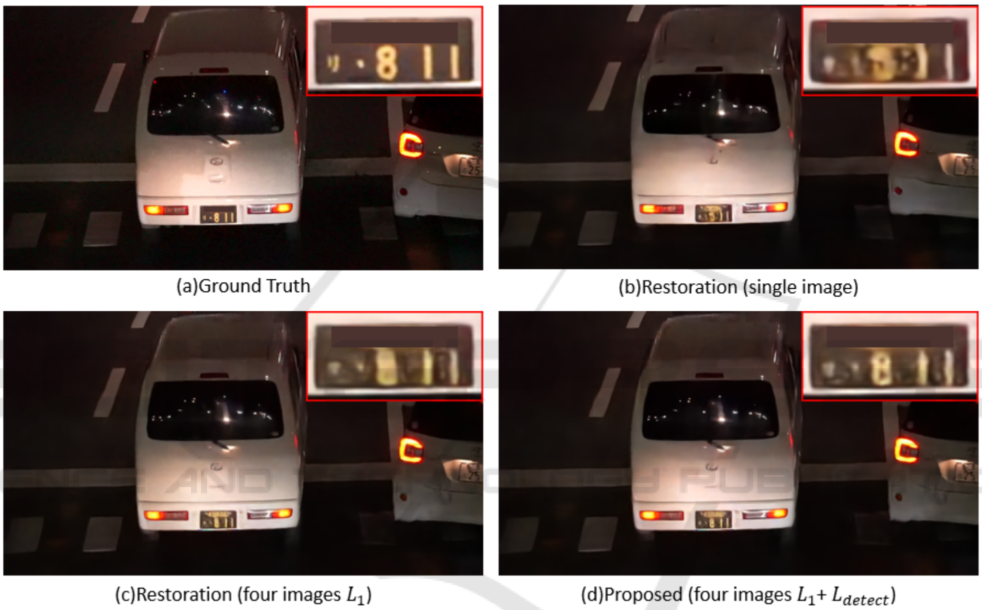


Figure 9: Test input sequential low light images used in our experiments. The first low shows original low light images, and the second low shows images obtained by enhancing their intensity.


(a)Ground Truth

(b)Restoration (single image)

(c)Restoration (four images L_1)

(d)Proposed (four images L_1+ L_{detect})

Figure 10: High intensity images recovered from the proposed method and existing methods. The readability of the number plate shows the effectiveness of the proposed method.

images due to the large image noise. Next, we show in Fig. 10 the high intensity image recovered from the sequential low light images shown in Fig. 9 by using the proposed method. Fig. 10 (a) shows the ground truth high intensity image, and Fig. 10 (b) shows the high intensity image recovered by using a single low light image I_4 in Fig. 9. Fig. 10 (c) shows the high intensity image recovered by using four low light images in Fig. 9 with simple L_1 loss, and Fig. 10 (d) shows the result of the proposed method, that is the high intensity image recovered by using the four low light images with L_1 loss and L_{CR} loss. Note, a part of the number plate is hidden in Fig. 10 for security reasons.

As shown in this figure, the characters ”5972” cannot be read properly in the result of the existing single image based recovery, but it can be read properly in the result of the proposed method shown in Fig. 10 (d). Also, we can see that the characters are recovered more accurately by using the character recognition loss L_{CR} in the proposed method.

Fig. 11 and Fig. 12 shows the results from different test data. Again, we can see that the readability of characters is improved in the image derived from the proposed method as shown in Fig. 12.

4.2 Accuracy Evaluation

We next evaluate the accuracy of the proposed method quantitatively. In this experiment, we focused on the restoration accuracy of the number plate characters of the test data, and evaluated how correctly the characters on the number plate in the recovered image were recognized by the pre-trained character recog-

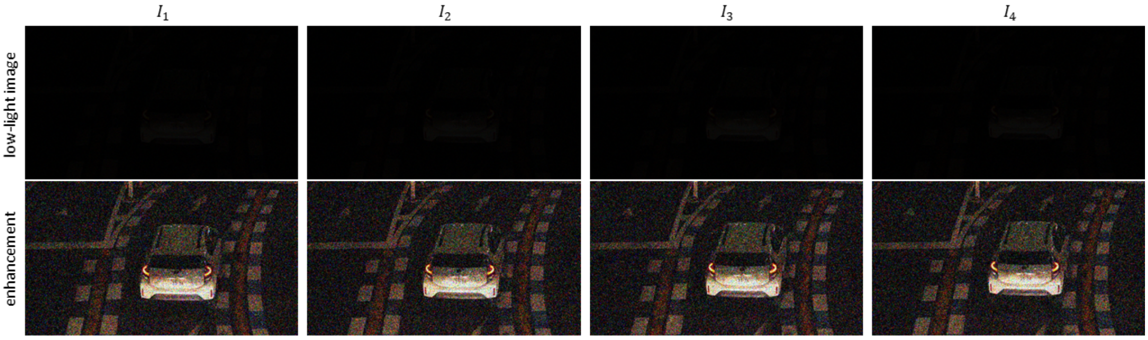


Figure 11: Test input sequential low light images used in our experiments. The first low shows original low light images, and the second low shows images obtained by enhancing their intensity.

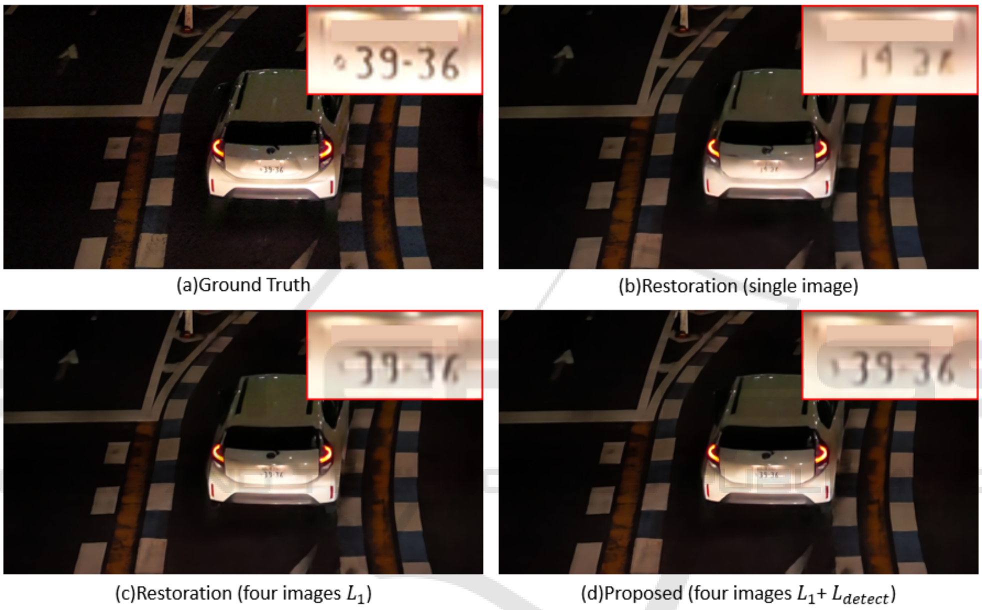

39-36
(a)Ground Truth
(b)Restoration (single image)
(c)Restoration (four images L_1)
(d)Proposed (four images L_1+ L_{detect})

Figure 12: High intensity images recovered from the proposed method and existing methods. The readability of the number plate shows the effectiveness of the proposed method.

nizer. The recovered high intensity images in the test data were input to the pre-trained character recognizer Retina Net (Lin et al., 2017) trained by using Street View House Numbers Dataset (Netzer et al., 2011), and the correct answer rate of the recognizer was evaluated. The correct answer rate is the number of correctly recognized characters divided by the total number of characters in the test data, i.e. 576.

The table 1 shows the correct answer rate of the recognizer derived from the ground truth high intensity images, single image based method with L_1 loss, single image based method with $L1$ and L_{CR} loss, multiple image based method with L_1 loss and the proposed method, that is multiple image based method with $L1$ and L_{CR} loss. From this table, we find that the correct answer rate can be drastically improved by using the sequential multiple images and by using the character recognition loss L_{CR} in the proposed method. We can also find that the correct answer rate of the proposed method is close to that of the ground truth high intensity images.

Table 1: Correct answer rate by trained character classifier. SIM denotes single image based method and MIM denotes multiple image (sequential image) based method. MIM $L_1 + L_{CR}$ is our proposed method.

	correct answer rate
ground truth images	0.896
SIM L_1	0.384
SIM $L_1 + L_{CR}$	0.584
MIM L_1	0.544
MIM $L_1 + L_{CR}$ (proposed)	0.772

5 CONCLUSION

In this paper, we proposed a novel method for recovering high intensity images from degraded images taken in low light. We showed that by using the sequence of low light images, the high intensity image can be generated accurately. For using the sequential images effectively, we used two different U-Nets, one is for recovering stationary objects and the other is for recovering moving objects in the image. The mask image U-Net is also introduced for training the stationary object U-Net and the moving object U-Net efficiently. For obtaining more accurate high intensity images, we used newly defined loss called character recognition loss. The experimental results show that the proposed method can recover highly accurate high intensity images from noisy low light images.

REFERENCES

Chakrabarti, A. (2016). A neural approach to blind motion deblurring. In *European conference on computer vision*, pages 221–235. Springer.

Chen, C., Chen, Q., Xu, J., and Koltun, V. (2018). Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192.

Li, L., Wang, R., Wang, W., and Gao, W. (2015). A low-light image enhancement method for both denoising and contrast enlarging. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 3730–3734. IEEE.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*.

Remez, T., Litany, O., Giryes, R., and Bronstein, A. M. (2017). Deep convolutional denoising of low-light images. *arXiv preprint arXiv:1701.01687*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Shan, Q., Jia, J., and Agarwala, A. (2008). High-quality motion deblurring from a single image. *Acm transactions on graphics (tog)*, 27(3):1–10.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155.