World Journal of Pharmaceutical Science and Research
www.wjpsronline.com
Research Article
ISSN: 2583-6579
SJIF Impact Factor: 3.454
Year - 2024
Volume: 3; Issue: 2
Page: 248-256

ENHANCING UNDEREXPOSED IMAGES USING U-NET ANDDEEP LEARNING TECHNIQUE

Vinod Kumar Dhiman*

Vice President, DCRUST Sonepat Haryana India.

Article Received: 20 February 2024 || Article Revised: 11 March 2024 || Article Accepted: 02 April 2024

Corresponding Author: Vinod Kumar Dhiman
Vice President, DCRUST Sonepat Haryana India.
DOI: 10.5281/zenodo.10968370

ABSTRACT

This research addresses the challenge of enhancing underexposed images in the context of computer vision and photography. Under- exposed images, characterized by low levels of captured photons, pose significant limitations for various applications, including autonomous vehicles, medical imaging, and photography. I propose a novel approach that leverages the U-Net architecture, a convolutional neural network known for its effectiveness in image-processing tasks. Our model is designed to recover underexposed images while preserving sharpness, color accuracy, and cleanliness. I also explore integrating generative adversarial networks (GANs) and attention mechanisms to improve image quality further. Through extensive experimentation and evaluation 1, I demonstrate the potential of our method in significantly enhancing underexposed images, making them suitable for critical applications such as object detection, medical diagnosis, and photography. Future research directions include addressing model generalization across sensors and optimizing real-time processing capabilities.

KEYWORDS: Enhancing underexposed images, generative adversarial networks, generalization.

1. INTRODUCTION

1.1 What research problem am I trying to solve?

Social media and computer vision applications demand high-quality images, but current CMOS sensors used in cameras and smartphones struggle to perform well in challenging lighting conditions due to physical limitations. Low levels of photons captured by CMOS pixels result in dark, under-exposed images that are difficult for computer analysis and human perception. Traditional physical methods such as enlarging pixels, aperture, or increasing ISO can improve the Signal-to-Noise ratio, but at the cost of resolution, depth of field, or noise in the image. Therefore, our research aims to develop a better computational approach using a probabilistic neural network model to recover under-exposed images while maintaining their sharpness, color accuracy, and cleanliness.

1.2 Why is this problem important?

The potential of solving this problem can extend from mere convenience to saving lives. For example, I can use cameras for object detection and obstacle avoidance instead of lidars in computer vision and autonomous vehicles in low light conditions. Similarly, a similar method can be used in search and rescue operations. In radiation imagery, such as X-ray images used in the medical field, underexposed pixels can lose crucial details, potentially hiding a small tumor or other irregularities that medical professionals would otherwise detect. On the other hand, in photography, underexposed images can be recreated with higherbrightness, which can merely be convenient for people.

2. Related Works

The literature has thoroughly examined the computational processing of low-light images. I will give a brief overview of the methods currently available.

Convolutional Neural Networks: Researchers have explored the use of deep neural networks for extreme low-light imaging, but the approach has limitations and shortcomings. One issue is that the convolutional network[1] is individually tuned for each camera sensor, requiring cross-sensor generalization to be effective across different cameras. The network's hyper-parameters, such as amplification factors, also need to be manually tuned, indicating the need for an auto ISO feature to improve efficiency. Additionally, the absence of HDR tone mapping and dynamic objects in the dataset limits the network's ability to enhance such images. Artifacts in the final image could also potentially be reduced. Finally, the long processing time of 0.38-0.66 seconds per image could be problematic for real-time applications. These limitations must be addressed to improve the practicality and effectiveness of deep neural networks for image enhancement.

Low-light image enhancement: Various methods are available for enhancing images, including histogram equalization and gamma correction. Histogram equalization aims to balance the entire image's histogram, while gamma correction amplifies the brightness of dark areas and compresses bright pixels. More advanced techniques involve comprehensive analysis and processing, like the inverse dark channel prior, wavelet transform[2], Retinex model[3], and estimation of the illumination map.[5] However, these methods assume that the images adequately depict the scene content.

Nevertheless, my current focus centers on extreme low-light imaging, characterized by elevated noise and color distortion levels surpassing the capabilities of existing enhancement pipelines.

3. METHODOLOGY

I propose a generative network model designed to analyze and sample from the distribution of under-exposed images and generate well-exposed images for comparison against the discriminate network using CNNs with ground truth as input. I will delve deeper into the structure of the network and steps later.

4. METHOD

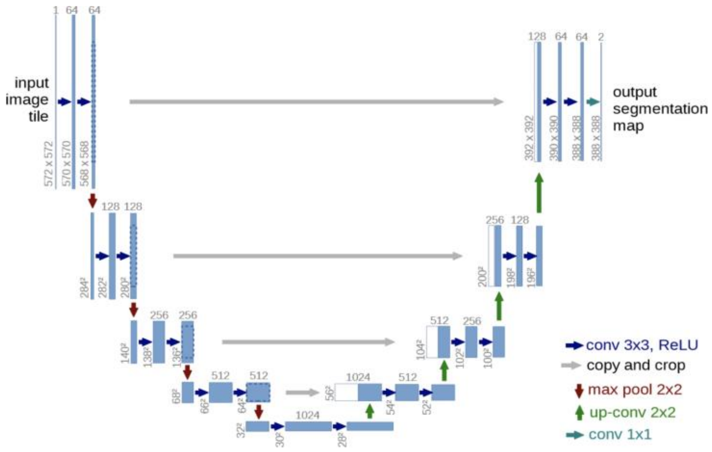

Fig. 1: Architecture of the U-Net Model for Semantic Segmentation.

4.1 Training

The training process for the proposed image enhancement method consists of several key steps, including data pre-processing, model training, optimization, and learning rate scheduling. This section presents a comprehensive overview of the training procedure.

Data Preprocessing: Before training the model, the training dataset undergoes essential pre-processing steps. Random cropping is employed to extract image patches from the input data, ensuring diversity in the training samples. The process involves randomly selecting starting positions within the images and cropping them to a specified size. Additionally, specific pre-processing techniques, such as normalization and data format conversions, are applied to align the input images with the target ground truth images. These pre-processing steps facilitate effective learning and improve the model's generalization of unseen data.

Model Training and Optimization: The training step involves optimizing the model's parameters based on the available training data. During training, the model is presented with preprocessed images. The network output is then supervised under the ground truth and compared against them. This comparison calculates a loss metric quantifying the discrepancy between the model's prediction and the desired outputs. Optimization involves minimizing this loss by adjusting the model's parameters through backpropagation and gradient-based optimization techniques. Specifically, the gradients are computed with respect to the model's parameter using the chain rule, and an optimization algorithm is employed to update the parameters in the direction that reduces the loss. This iterative optimization process incrementally refines the model's capabilities and enables it to approximate the desired image enhancement function better.

Loss Function: I used L1 loss to calculate the loss for each iteration in the network model output. The choice of L1 loss function is motivated by its ability to preserve image details and promote spatially consistent enhancements. Unlike other loss functions, such as the L2 loss (mean squared error), the L1 loss is less sensitive to outliers and encourages sparser errors. This property benefits image enhancement, where preserving fine details and avoiding over-smoothing are desirable.

5. ALGORITHM

Algorithm 1 presents the training procedure for the proposed image enhancement method using the U-Net architecture 1. The algorithm outlines the steps in training the model and optimizing its parameters.

Input: Training dataset: images taken in the dark (input) and corresponding ground truth long exposure images (target)

Input: Hyperparameters: learning rate, step size, gamma.

Output: Trained U-Net model

Preprocessing: Randomly crop input and corresponding ground truth images to extract image patches of a specified size. Apply normalization and data format conversions to align the input images with the target ground truth images.

Initialize Model: Create an instance of the U-Net model. Set the initial model parameters.

Initialize Optimizer: Select an optimization algorithm, such as Adam. Initialize the optimizer with the model parameters and a specified learning rate.

Initialize Learning Rate Scheduler: Configure the learning rate scheduler with the selected scheduler (e.g., StepLR). Set the step size and gamma values according to the predefined parameters (step_size =10, gamma = 0.8).

Training Loop: repeat

Model Training: Set the model in training mode. Clear the gradients of the optimizer. Retrieve a batch of cropped input images and corresponding ground truth images. Transfer the data to the device (e.g., GPU). Perform additional preprocessing steps, such as packing the raw images (if applicable) and adjusting exposure ratios. Pass the preprocessed input images through the U-Net model to obtain the output predictions. Calculate the loss between the predictions and the ground truth images using a suitable loss function (e.g., L1 loss). Backpropagate the gradients through the model to compute the gradients with respect to the model parameters. Update the model parameters by taking an optimization step using the optimizer.

Learning Rate Scheduling: Adjust the learning rate according to the predefined schedule. Update the learning rate based on the stepsize and gamma values. Modify the optimizer's learning rate using the updated value.

Until *convergence or a predefined number of iterations*;

Output: Return the trained U-Net model.

Algorithm 1: Training the U-Net for Image Enhancement

The training algorithm follows a standard procedure for training deep neural networks. It involves preprocessing the data, initializing the model, optimizer, and learning rate scheduler, and then iteratively updating the model parameters while adjusting the learning rate. By optimizing the model with respect to the loss function, the algorithm aims to train the U-Net to accurately enhance dark images compared to their corresponding ground truth long exposure images.

6. THEORY

The proposed method for direct single-image processing of fast, low-light images involves advanced deep learning techniques, specifically, a U-Net architecture. The U-Net is a convolutional neural network effective in image segmentation[6] and other image processing tasks.

In this approach, the U-net is used to improve the image quality of underex-posed images while recovering the ground truth data. The U-net is trained on a dataset of paired under-exposed and well-exposed images, where the low-light image is the input, and the highlight image is the ground truth. The U-Net learns to map the low-light image to the high-light image, recovering the lost information due to the low signal-to-noise ratio.

The U-Net is trained with a generative model, such as a variational autoen- coder (VAE), to enhance the overall image quality further. The VAE provides artistic enhancements to underexposed images, such as adjusting the bright- ness, contrast, and color saturation. The VAE is trained on a large dataset of high-quality images and then used to enhance the underexposed images.

To optimize the overall image quality, the U-Net and VAE are trained in parallel, allowing the VAE to improve the overall appearance of the image while the U-Net improves the accuracy of the image. The loss function used to train the model combines the objectives of the U-Net and VAE, encouraging the model to produce high-quality images close to the ground truth.

The model is trained on a dataset of paired short-exposure and long-exposure images, allowing the model to learn from multiple examples of the same ground truth data. Finally, the U-Net and VAE output are combined and evaluated using a discriminator 7 network to ensure that the final image is of high quality.

7. Additional Experiments

7.1 UNet with Generative Adversarial Network

I experiment on improving the UNet structure proposed in SID paper[1] by attaching a discriminator network and providing adversarial loss to the generator, UNet. Since I am using UNet as the generator, I chose UNet as our discriminator to lower the difference between the generator and the discriminator. The UNet discriminator architecture and losses are adapted from another paper.[4]

7.2 U-UNet – Double UNet with skip-z

I want to improve on the washed-out color of the output from our baseline model 2 3, therefore, I experiment on whether chaining another model on top of the pre-trained baseline model will help with improving the color. I called this model U-UNet because I not only feed the output from the previous UNet, but I also use skip connections between the upsampling layer of the previous UNet and a downsampling layer of the next UNet model 5, similar to the skip connection from the downsampling layer to the upsampling layer.

7.3 Deeper UNet

I also experimented with deep U-net architectures, aiming to leverage the in- creased capacity of the model to capture more complex features and improve overall performance. However, I observed worse performance compared to baseline U-Net mode, with the models struggling to converge and the loss unable to prop- agate effectively through the deeper layers. This limitation suggests that simply increasing the U-Net depth may not improve performance. Further investigations are necessary to identify the underlying causes of the issue and explore potential strategies to overcome it.

7.4 UNet with Attention

Another approach I explored to enhance our image restoration efforts was to utilize U-net models with an attention mechanism. By incorporating the attention mechanism into the U-Net architecture, I aimed to improve the model's ability to focus on relevant image features. This attention mechanism allows the network to dynamically allocate its resources to important regions, effectively enhancing the model's ability to capture fine details and improve overall image quality. Through our experimentation, I observed that U-Net with attention demonstrated promising results, showcasing improved restoration performance compared to the baseline U-Net model. However, despite the visual improvements, I observed that the numerical results were slightly worse than the baseline U-Net model.

8. RESULTS

The following is a summary of our results

8.1 Test Results

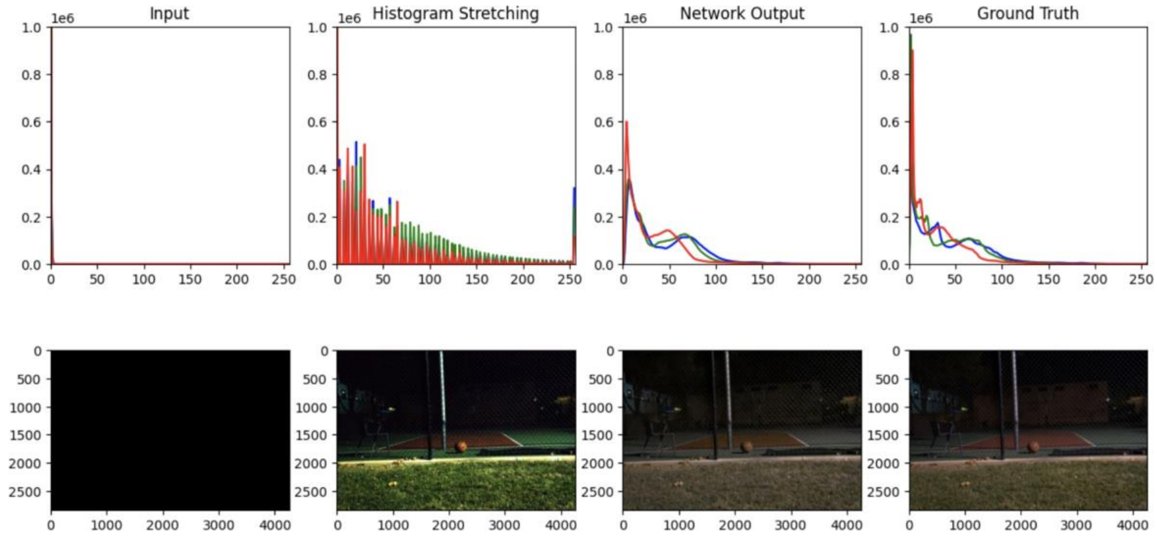

Fig. 2: Sample output from Baseline Model.

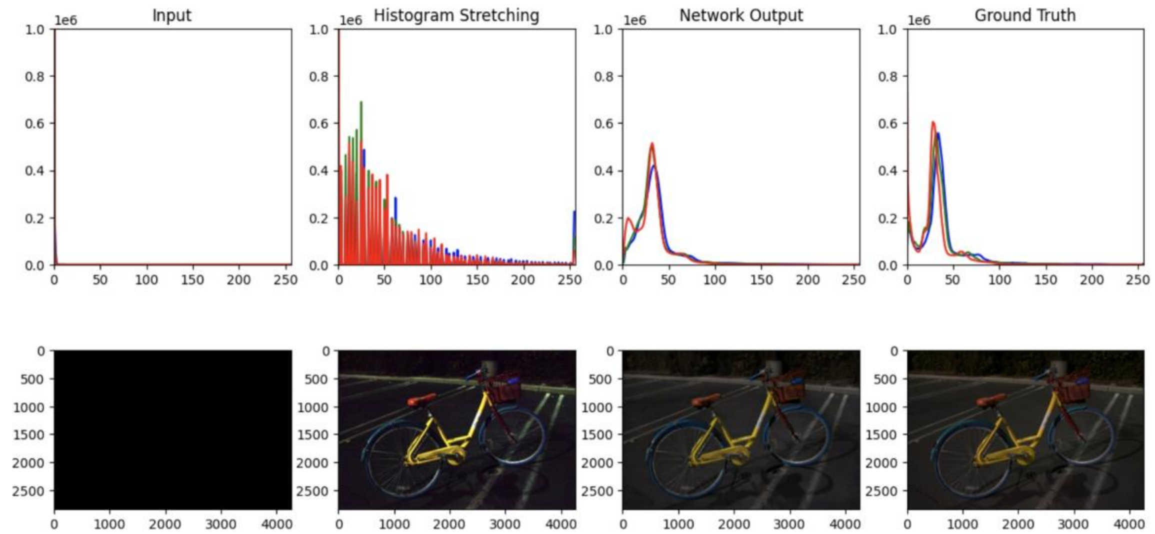

Fig. 3: Sample output from Baseline Model.

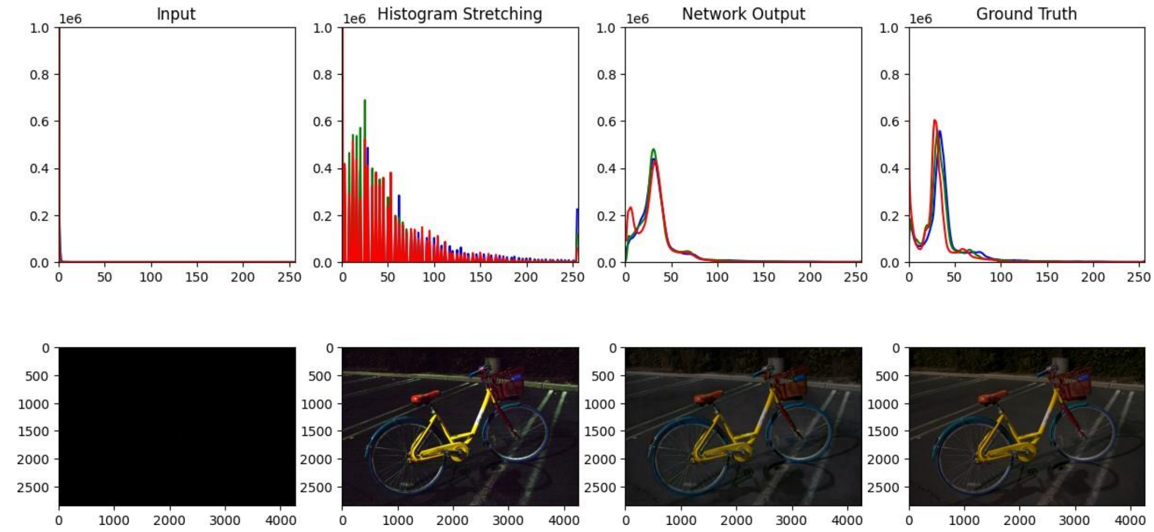

Fig. 4: Sample output from UNet GAN Model.

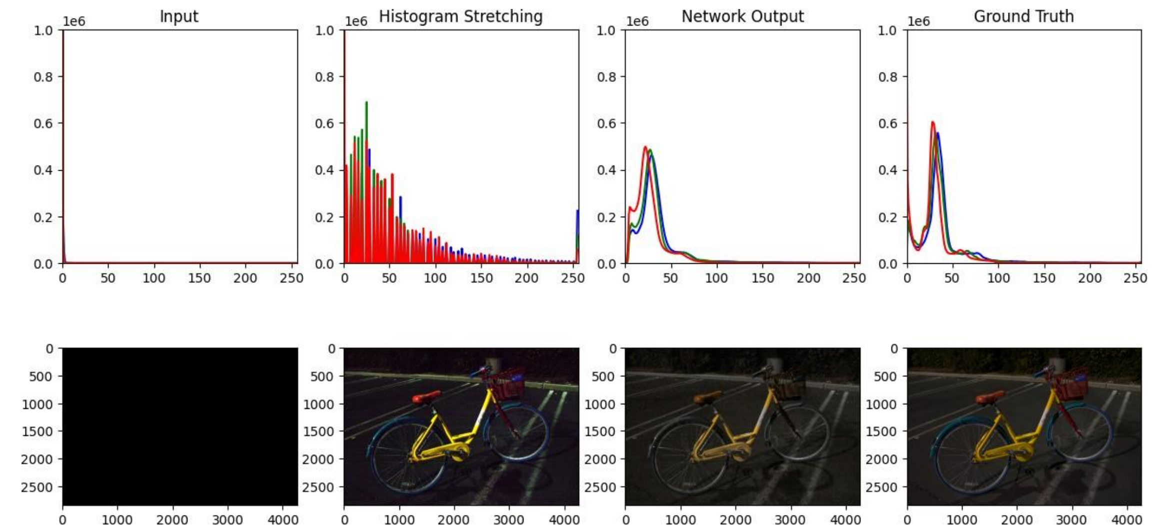

Fig. 5: Sample output from UNet Attention Model.

8.2 Evaluation

I evaluate the performance of models that produce acceptable results using PSNR, Peak Signal-to-Noise Ratio, and SSIM, Structural Similarity Index Measure:

Table 1: Evaluation metrics on test set.

Model	PSNR	SSIM
Baseline	27.251	0.756
UNet with Attention	25.447	0.727
UNet with GAN	27.178	0.747

9. DISCUSSION

9.1 Improving Results

Our image enhancement method’s current network architecture has advantages and limitations. The architecture can handle 14-bit images, ensuring sufficient dynamic range for capturing subtle details in dark scenes. However, for practical applications such as obstacle avoidance on roads, where 8-bit images are commonly used, a minimum of 10 bits may limit its suitability. Another limitation is the lack of dynamic input size support, which restricts its flexibility in handling images of varying resolutions. On the positive side, adopting the U-Net architecture brings advantages such as skip connections, enabling the fusion of low-level and high- level features for better contextual understanding. However, it is worth noting that U-net may suffer from issues like excessive memory consumption and limited receptive field size. Therefore, future improvements to our image enhancement results could involve exploring alternative network architectures that address the limitations mentioned while leveraging the benefits of skip connections and adapting to different input sizes, ultimately enhancing the model's applicability and performance in various real-world scenarios.

9.2 Other Architectures

Our research question poses some restrictions on the types of network architecture 6 I can choose. When I choose network architecture for our experiments, I need to consider whether the network can learn the desired features from inputs and the network capabilities of handling massive inputs, in our case, around 12M pixels. This restriction prevents us from using architectures with any linear layer, as the flattening process will result in variate vector lengths. Thanks to the non-linear bottleneck, I stick with UNet because it can handle variable input sizes without changing the structure. From our experiment result with GAN 4 loss and deeper UNet, this architecture might have reached its capability for our task. I am still actively finding different architectures with similar properties to push the upper bounds further.

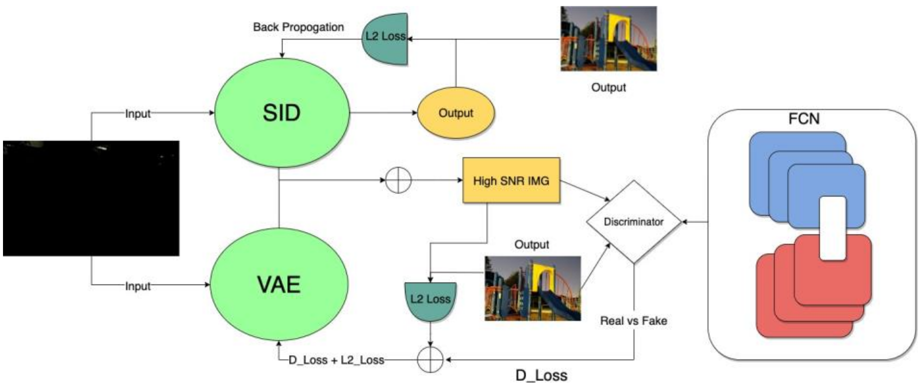

Fig. 6: Various alternative network architectures were explored and experimented with; however, despite our efforts, I could not achieve significantly improved results. Including the VAE-GAN architecture shown above.

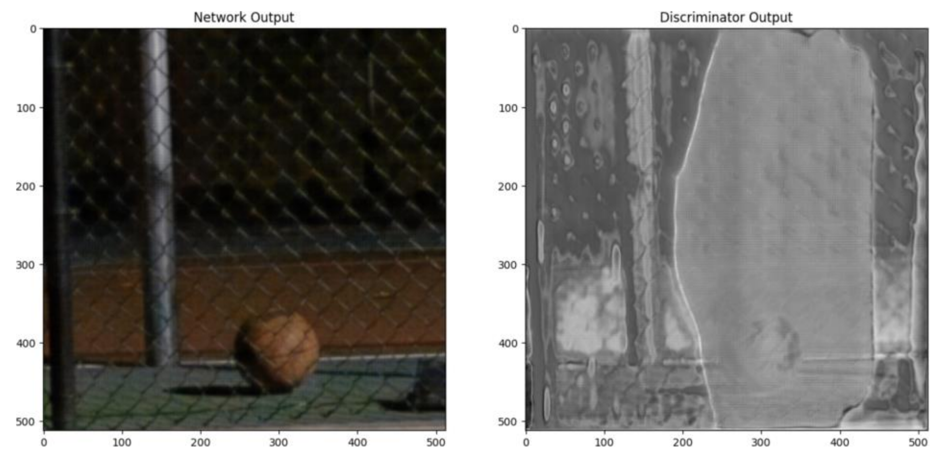

Fig. 7: Discriminator network output obtained from the Unet-gan architecture.

10. CONTRIBUTION

11. CONCLUSION

In our research, I tackled the problem of underexposed images by leveraging the power of the U-net architecture in conjunction with advanced deep learning tech- niques. The U-Net, a convolutional neural network, is known for its effectiveness in image segmentation tasks. I adapted this architecture to address underexposure by designing a modified U-Net model for artistic enhancement and detail recovery. The network was trained on a dataset comprising pairs of underexposed images and their corresponding long-exposure ground truth images.

Moving forward, future work should focus on further refining the proposed model and addressing its limitations. The model's generalization ability should be improved across camera sensors and lighting conditions. It is also essential to optimize the training process to reduce the computational time required for real-time applications. Furthermore, the developed solution can be extended to handle other challenging image conditions, such as overexposure and low-light video processing. By advancing the field of image recovery from underexposed photographs, I can unlock new possibilities for improving visual perception and analysis in a wide range of domains.

REFERENCES

1. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. CoRR abs/1805.01934 (2018), http://arxiv.org/abs/1805.01934
2. Loza, A., Bull, D., Hill, P., Achim, A.: Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients. Digital Signal Processing: a Review Journal, Dec 2013; 23(6): 1856–1866. https://doi.org/10.1016/j.dsp. 2013.06.002
3. Park, S., Yu, S., Moon, B., Ko, S., Paik, J.: Low-light image enhancement using variational optimization-based retinex model. IEEE Transactions on Consumer Electronics, 2017; 63: 178–184. https://doi.org/10.1109/TCE.2017.014847
4. Schönfeld, E., Schiele, B., Khoreva, A.: A u-net based discriminator for generative adversarial networks (2021).
5. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation (06 2019). https://doi.org/10.1109/CVPR.2019.00701
6. Yin, X.X., Sun, L., Fu, Y., Lu, R., Zhang, Y.: U-net-based medical image segmentation(apr 2022)