Low-light Monocular Vision

Gantogoo Oyunbat, Saranraj Nambusubramaniyan, Priyadharshika Vijayan, Florian Röhrbein

Professorship of Neurorobotics, Faculty of Computer Science
Chemnitz University of Technology, Germany

TECHNISCHE UNIVERSITÄT CHEMNITZ

Abstract

We see a spectacle of different colors on a sunny day. However, during the evening and night hours we need additional light source to see the same sight. Even during the day in a building or a room, if the environment does not allow much light into it and has bad lighting, we would see only small amount of the color detail just as in the evening light. The same is true for cameras, in a well-lit environment cameras perform good but in a dimly-lit room they perform significantly worse as they cannot see the colors anymore.

Brief Introduction to Color Vision

The colors we see do not exist in the form we see them outside of our brain [1, 2, 3, 4, 5]. We humans and several other mammals such as chimpanzees and gorillas have so called trichromatic vision that uses only three main types of light absorbing pigments for color vision [6]. Therefore, the colors we see are non-existent for some animals, while color that fish see are non-existent in other words invisible to humans. Colors are essentially a quality of light [7]. The light has two properties, wave and particle. Which means a light beam has frequency and a wavelength. What we see as colors is light with wavelengths between 350 nm and 700 nm [8]. Our eyes have specialized cells in the retina for detecting these wavelengths [2, 8].

Rods and cones named after their shapes convert light into signals. Rods are especially sensitive to low light, however they cannot detect colors well. Cones on the other hand require more photons, but are good at color detecting [2]. As human vision is trichromatic, the cones correspond to three different colors to which they are most sensitive to, red, green and blue [8]. However, that does not mean that the cones can detect only one color. Our brain computes any other color we see by combining or overlapping the detected wavelengths [2].

Literature Review

There are a few night vision technologies that enable us to see in the dark, low-light environments. Such as:

- **Thermal Imaging** where we see the temperature difference of objects.
- **Image Intensification** in which small amount of photons strike a photocathode emitting electrons which then illuminate an image screen in wavelengths humans can see.
- **Active Illumination** where there is an active Near Infrared (NIR) light source.

For each of the methods a special camera is needed, and the corresponding outputs are blue-red image for thermal cameras, green for intensification cameras, and grayscale for NIR cameras. Each of these methods enable vision in the dark, however, these technologies are usually expensive, and they do not enable the most important aspect of vision, the color vision.

The researchers at University of California Irvine have published a paper last year discussing this problem. They reported that they were able to reconstruct color images from grayscale, NIR images using U-Net like deep learning architecture [9]. The reason for using U-Net for the task of color vision in the dark is the general use of U-Net itself, which is colorization of images in other words segmentation problem. Recent advances in the field of machine learning have shown that U-Net like networks excel in segmentation tasks [10].

The problem with this approach is, however, that the data for NIR and IR images and videos are sparse compared to classic datasets. Our goal is to improve their model and implement it.

How do we get low-light color vision?

In order to recreate RGB images from grayscale or infrared images, we implemented U-Net inspired autoencoder. Browne et al. have also used a U-Net inspired approach to enable vision in poorly illuminated environments [9]. The autoencoder was then tested on grayscale images first to observe whether the input images could be reconstructed. The implemented U-Net model can be seen in Fig 1.

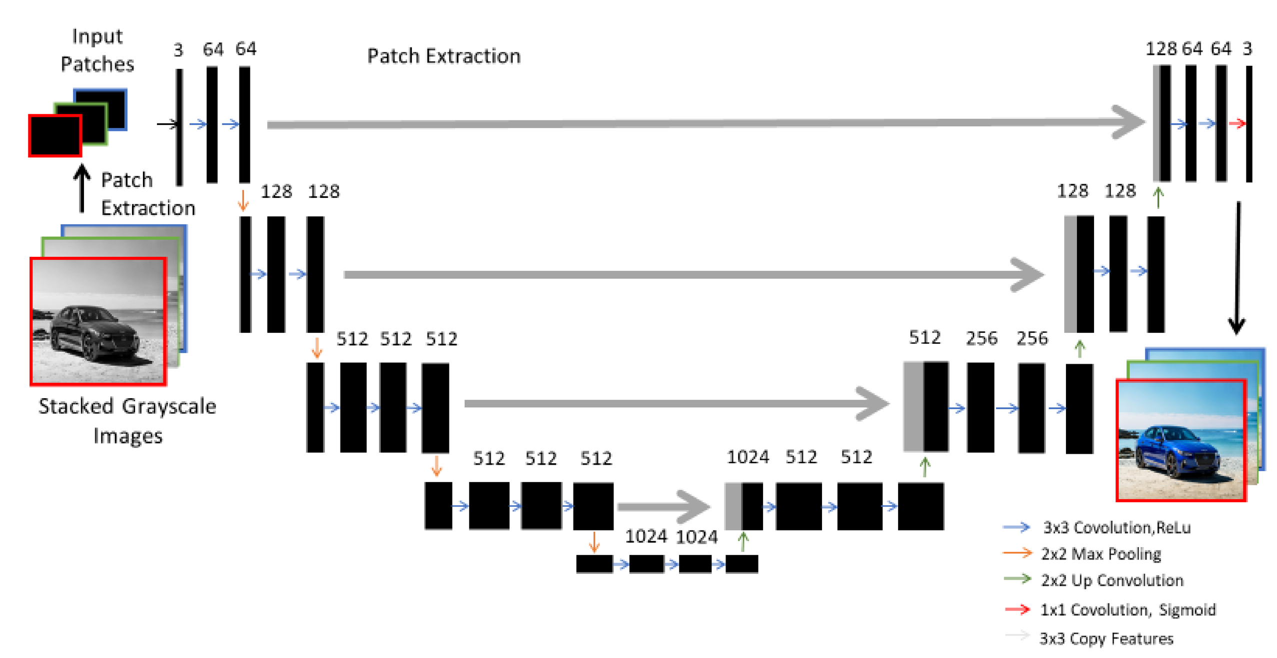

Figure 1. U-Net model.

For image reconstruction as well as colorization of grayscale images, we need for the input and the output 3 channels. The input is then convoluted with a kernel size of 3x3, normalized with batch normalization, and rectified by a ReLU function, each operation used twice for the double convolution. We use padding in the double convolution step to perserve the original image size to facilitate concatenation step later. After the double convolution step we get 64 channels, which goes through max pooling with 2x2 resulting in 128 channels and half of the input image size. The process of double convolution and max pooling then continues for three additional times where we reach the last layer and get 1024 channels. The decoder part of the U-Net starts with up-convolution of the last layer by 2x2 filter and then concatenates it with the feature map of second to last layer from the encoder part. This then is followed by double convolution same as the encoder part, which is then again up-convoluted. The process repeats until it reaches number of channels equal to 64. After one last double convolution, the output is generated with the same size and number of channels as the input.

Training Process

We first trained and tested our U-Net inspired autoenconder on grayscale images to see whether the grayscale images could be reconstructed. For the dataset, we first started with CIFAR-10 dataset. Due to our first prototype of autoencoder having non-flexible concatenation part, we deemed images of small sizes are not fit for our case, however with our current model this error could be solved. Our second choice was CelebA dataset, which was excluded for the moment due to a download error. We then settled on Caltech101 and Caltech256.

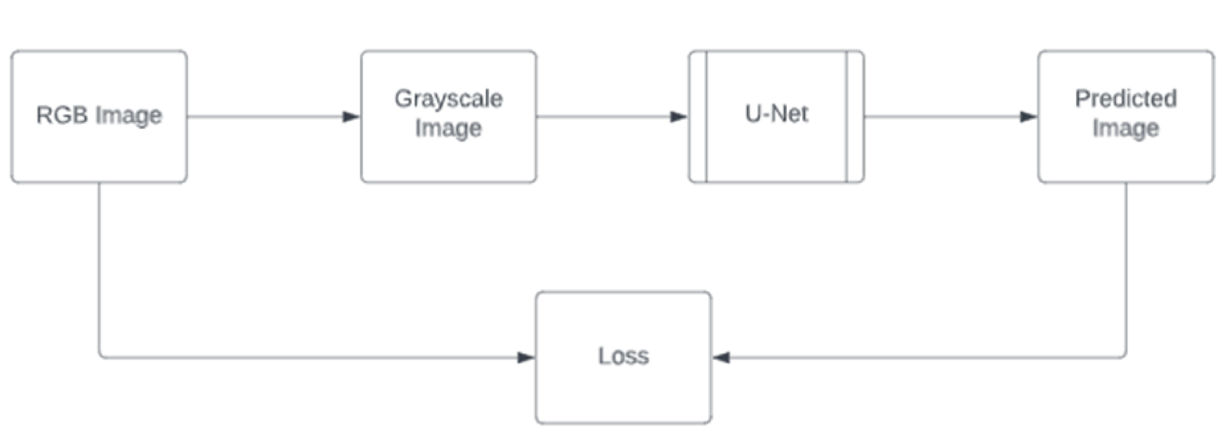

Figure 2. Trainig episode.

Result

The model was trained with batch size of 64, learning rate of 0.0001 and Adam optimization function for 50 epochs with CPU and 100 epochs with GPU.

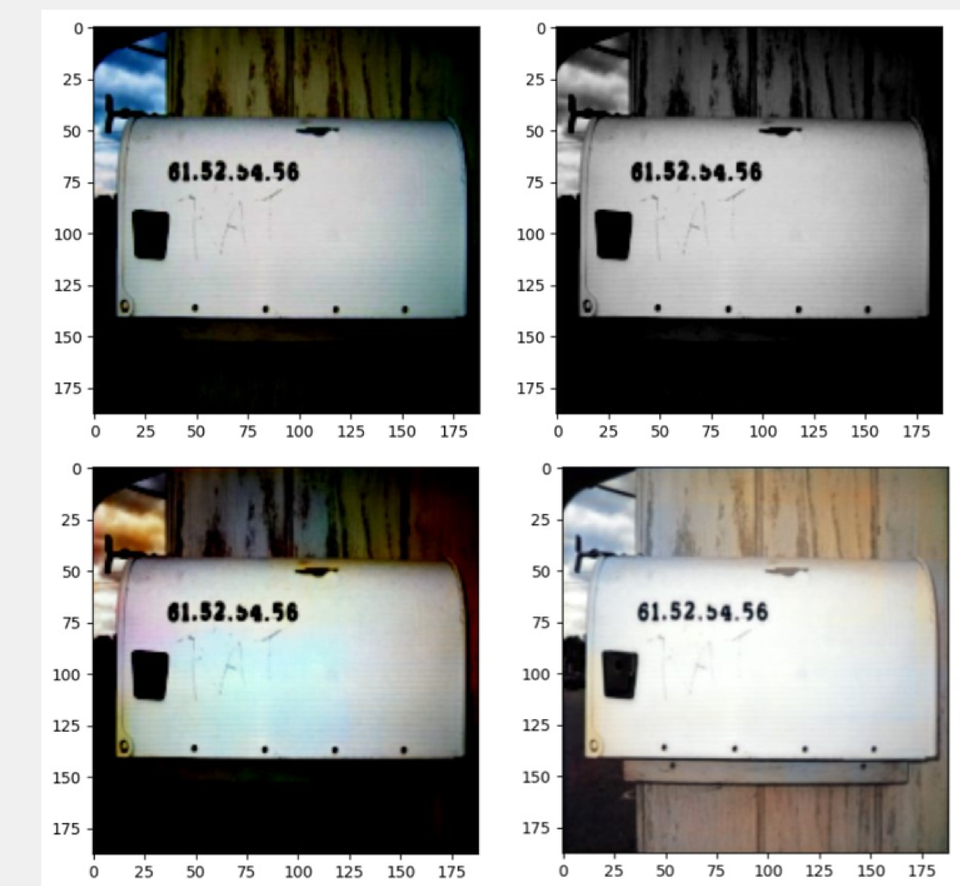

Figure 3. Training results with CPU and GPU.

The original image is shown top left, grayscaled version top right, predicted results with GPU bottom left, and with CPU bottom right.

Conclusion

As we can see from the results, outcomes were a bit different depending on whether we used CPU or GPU for training. When the loss curves were plotted, it could be seen that the curves were still converging. Therefore, we believe that it needs further training time and more data.

References

[1] S. Zeki, *A vision of the brain.* Blackwell scientific publications, 1993.

[2] P. Gouras, "Color vision," 2011.

[3] K. R. Gegenfurtner and D. C. Kiper, "Color vision," *Annual review of neuroscience*, vol. 26, no. 1, pp. 181–206, 2003.

[4] I. Kim, S. W. Hong, S. K. Shevell, and W. M. Shim, "Neural representations of perceptual color experience in the human ventral visual pathway," *Proceedings of the National Academy of Sciences*, vol. 117, no. 23, pp. 13145–13150, 2020.

[5] Y. Sasaki and T. Watanabe, "The primary visual cortex fills in color," *Proceedings of the National Academy of Sciences*, vol. 101, no. 52, pp. 18251–18256, 2004.

[6] G. H. Jacobs, "Photopigments and the dimensionality of animal color vision," *Neuroscience & Biobehavioral Reviews*, vol. 86, pp. 108–130, 2018.

[7] V. Walsh, "How does the cortex construct color?," *Proceedings of the National Academy of Sciences*, vol. 96, no. 24, pp. 13594–13596, 1999.

[8] J. K. Bowmaker and H. Dartnall, "Visual pigments of rods and cones in a human retina.," *The Journal of physiology*, vol. 298, no. 1, pp. 501–511, 1980.

[9] A. W. Browne, E. Deyneka, F. Ceccarelli, J. K. To, S. Chen, J. Tang, A. N. Vu, and P. F. Baldi, "Deep learning to enable color vision in the dark," *PLOS ONE*, vol. 17, pp. 1–15, 04 2022.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[11] K. Seo, "Development of cmos image sensors accelerates to serve a variety of applications," 2021. Accessed: May 16, 2023.

[12] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 4 2018.