Deep Learning Approaches for Robust Flare Artifact Correction in Imaging Systems

David Tomarov[1,2], Wenrui Wu[1], Anuj Chavan[1]
University of Toronto, Canada[1], Tel Aviv University, Israel[2]

Motivation

- Flare artifacts are a persistent challenge arising when strong light sources interact with a camera's lens system, manifesting into diverse, seemingly random patterns, often reducing image quality and usability.
- Convolutional neural networks (CNNs) techniques have shown promise in handling complex imaging and segmentation, which may struggle with modeling real-life in flare artifacts when trained on synthetically-generated flares.
- Vision Transformers (ViTs) and their variants, such as transformer unet (TransUnet), surpass CNNs in various image processing tasks by effectively capturing global context for real-world images.

Project Goals

- Investigate whether Vision Transformers, outperform CNNs for flare artifact removal.
- Bridge the gap between simulated-flare modified images and real-world captures performance, ensuring broad applicability.

Related Work

- CNNs are a promising tool for flare and glare removal, however fail to capture global contexts.
- Wu et al. [1] proposed a CNN-based UNet model trained on semi-synthesized flare-corrupted images, demonstrating effectiveness in reducing flare artifacts. Although it faced significant challenges with overfitting, limiting its ability to generalize to images taken with different lens systems.
- Huang et al. [2] introduced UNet3+, which takes advantage of full-scale skip connections and deep supervisions. The improvements in full-scale skip connections and deep supervisions make it a great application for our problem of flare removal.
- Dosovitskiy et al. [3] shows transformers have emerged as a powerful alternative in vision tasks due to their ability to capture long-range dependencies and global context through self-attention mechanisms, suggesting transformers could be well-suited for flare removal which we explore through this project.

References

[1] Wu, He, Xue, Garg, Chen, Veeraraghavan, Barron, How to Train Neural Networks for Flare Removal, ICCV, 2021.
[2] Huang, Lin, Tong, Hu, Zhang, Iwamoto, Han, Chen, Wu. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation, ICASSP, 2020.
[3] Alexey, Dosovitskiy. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv: 2010.11929 (2020).

New Technique

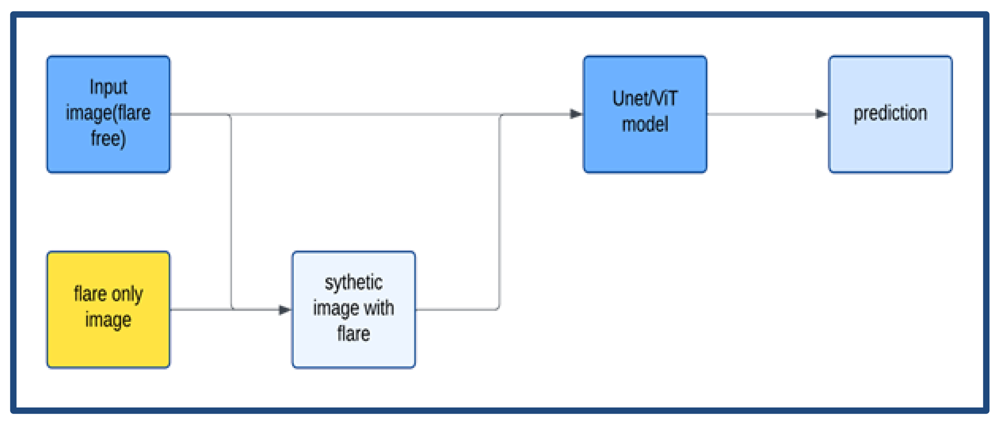


Figure 1: The pipeline of training

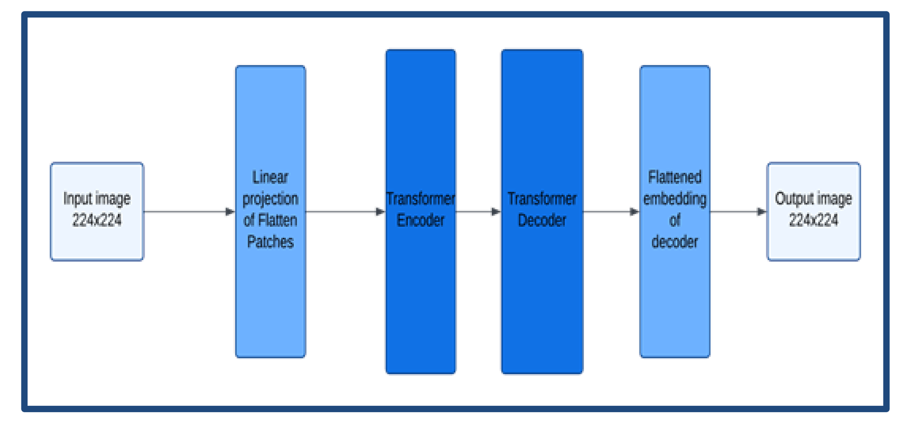


Figure 2: ViT model

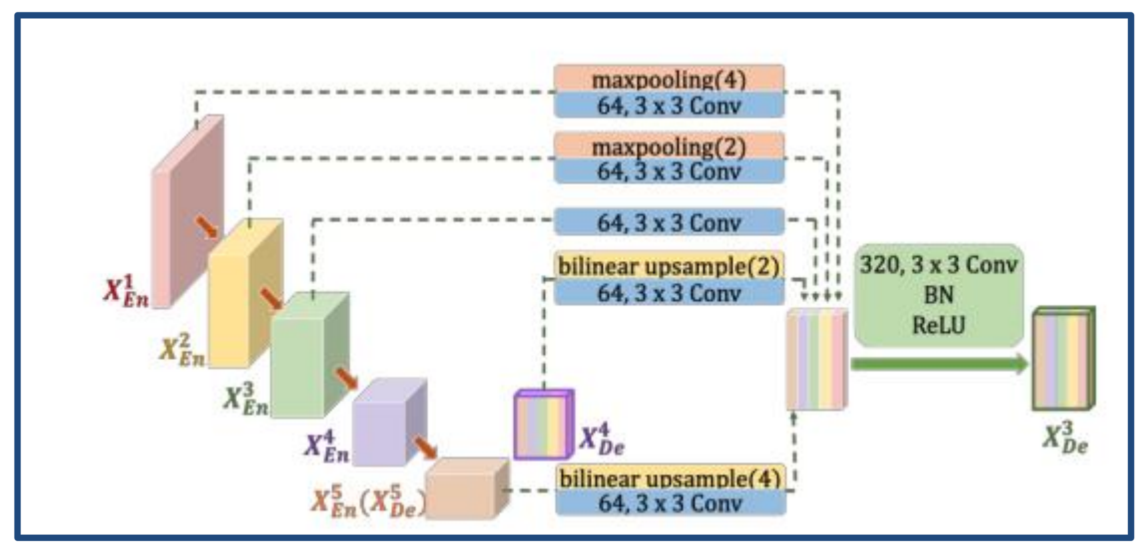
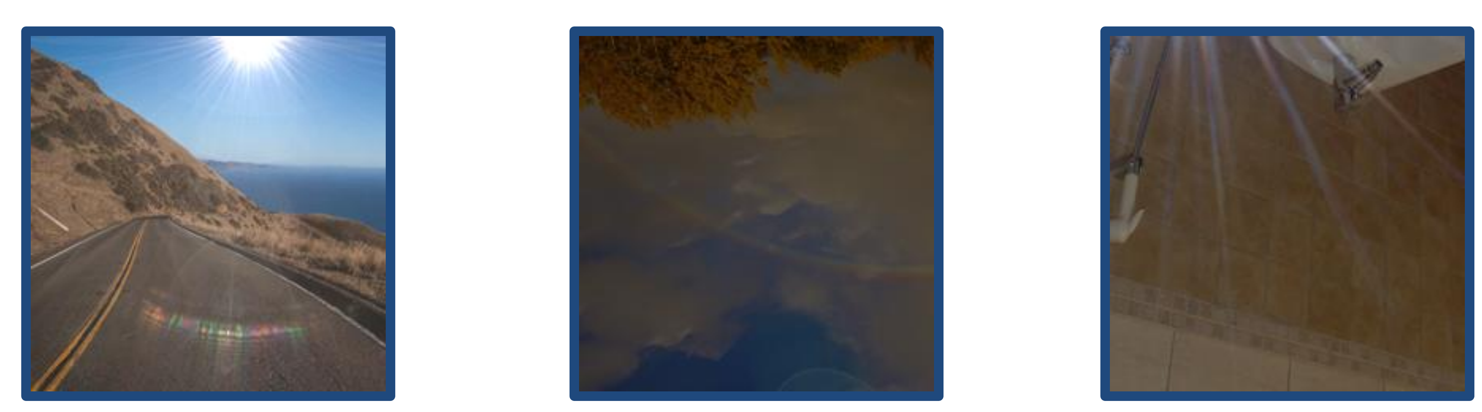


Figure3: Unet3+

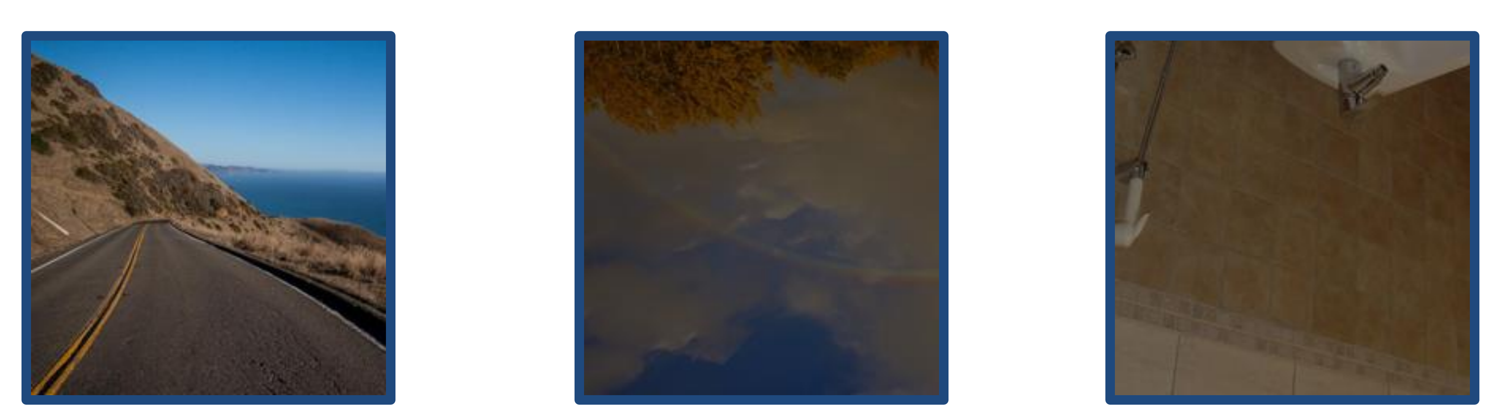
- In the original paper, they trained with image size 512x512. Due to our computational resource limitation, we decided to downsample the data into 224x224 and train on that.
- Training a vision transformer required a hugh dataset, but our dataset is relatively small. So we adapt a pretrained ViT on ImageNet-21k. The pretrained model is used for a classification task so to fit it in our task, we removed the CLS-token layer and link it with a decoder.

Experimental Results

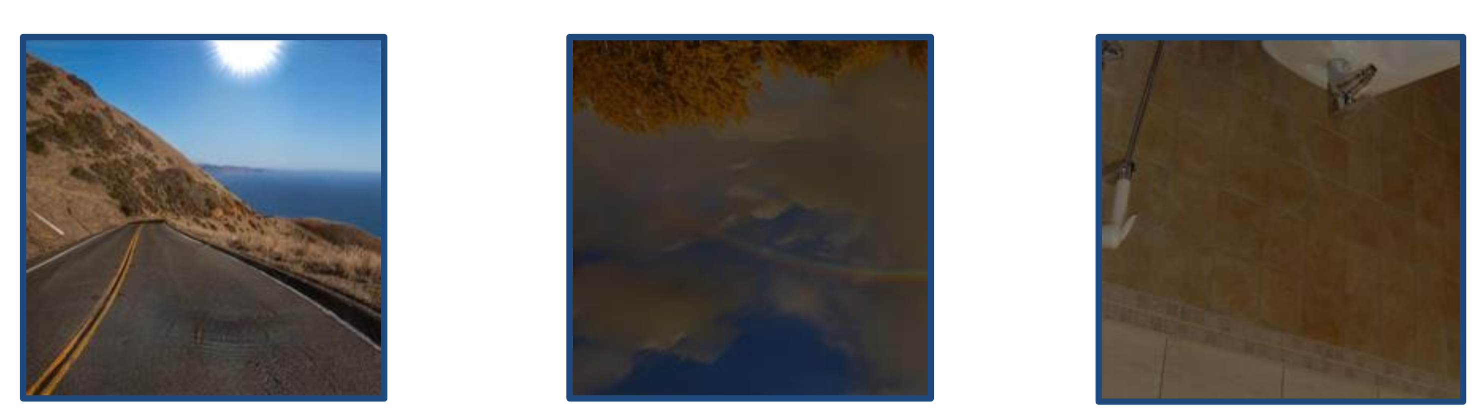
Input image with flare



Ground truth

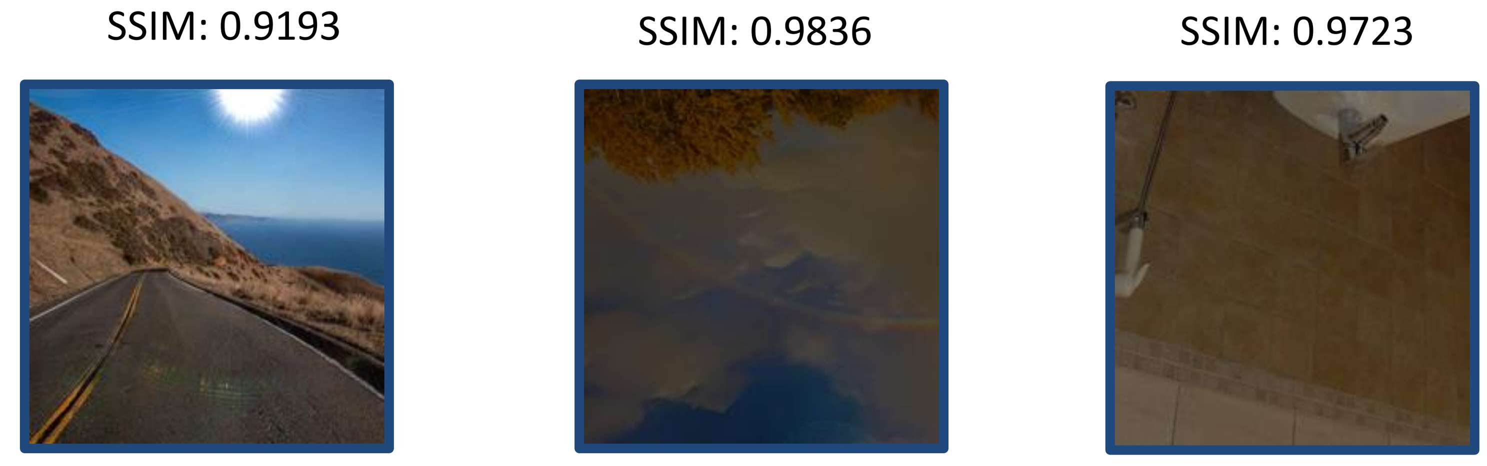


Prediction with Unet3+



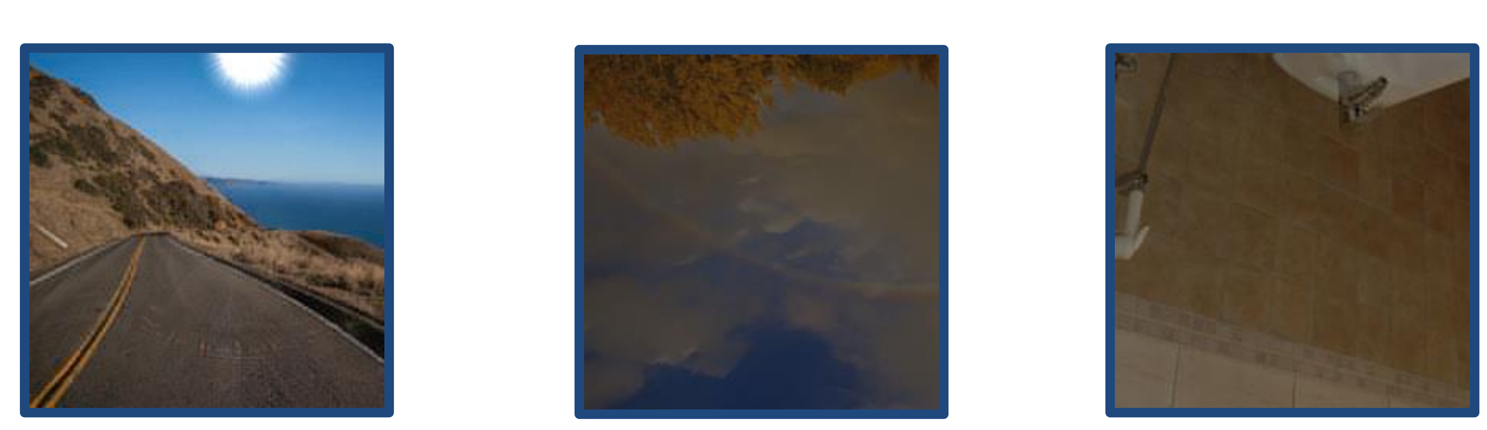
PSNR: 31.38	PSNR: 33.91	PSNR: 29.84
SSIM: 0.9193	SSIM: 0.9836	SSIM: 0.9723

Prediction with Unet



PSNR: 27.57	PSNR: 31.41	PSNR: 34.00
SSIM: 0.8770	SSIM: 0.9445	SSIM: 0.9916

Prediction with ViT



PSNR: 31.63	PSNR: 37.23	PSNR: 34.23
SSIM: 0.8583	SSIM: 0.9629	SSIM: 0.9688