RestoreNet

Quantifying Image Restoration of World War II Photos

Emmanuel Diaz

Abstract

In this project, we look at an approach towards restoring degraded images from the World War II era. These target photos have certain artifacts like “photo creasing” that develop due to natural factors in handling photos and our goal is to devise a method that attempts to restore these images back to their original form. We collected a dataset of World War II images and selected degraded images to analyse features of the degradation process, like Signal-to-Noise Ratio and Structural Similarity Index, in order to further understand reversing this process. Using a Generative Adversarial Network called “Deep Generative Prior”, we show how an inpainting restoration is sensitive to the training data and observe the difficulties of proving a ‘good’ restoration subjectively and quantitatively.

Introduction

The field of digital image restoration has been developed for more than 30 years with its foundations in signal processing techniques. These techniques helped remove noise from ‘degraded’ images which are images that have some form of reduced quality from their original version. However, these image processing techniques have their bounds in how much lost information from degradation may be recovered. The recent advancements of neural networks and collections of huge image datasets have led many researchers to believe that neural networks hold some potential in proper image restoration.

Previous Work

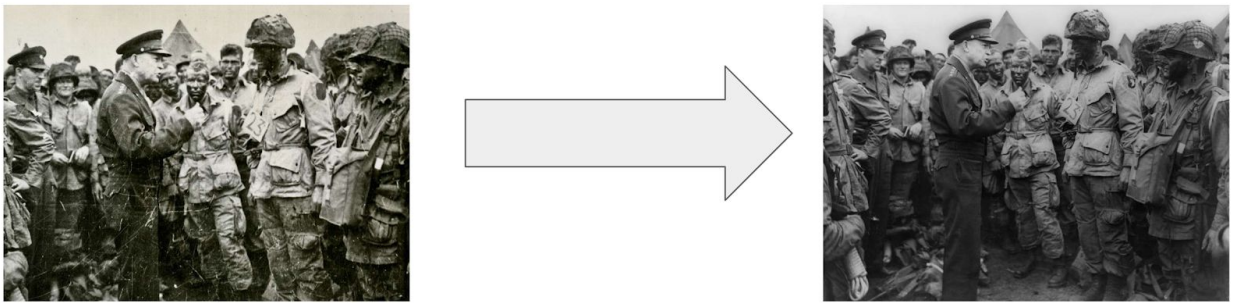
Researchers began looking at different types of neural network architectures to find the best in restoring images. It began with the use of Recurrent Neural Networks (RNNs)[1] which takes the degraded image as input and performs a feedback loop, further restoring the image in each network pass. This method has shown perceptually

good results on small dimensional test images, however RNNs suffered from inherent biases of the initial conditions, causing this method not to generalize well beyond a few examples. The next network challenger up to the task was the Convolutional Neural Network (CNN) Autoencoder[2] which use CNNs as feature extractors and encode the convolutions to a small dimension latent variables then deconvolutional layers learns the upscale process for these variables, giving a ‘cleaned’ image. CNN Autoencoder proved exceptional perceptive results for denoising and inpainting but its complexity and optimization left researchers to keep exploring new methods.

A breakthrough in image restoration came through with the application of Generative Adversarial Networks (GANs)[3] which gave rise to many facets in image processing and computer vision like DeepFakes[4]. GANs are constructed using two networks, the Generator and Discriminator, where the Generator creates constructed samples from the distribution of the training set and the Discriminator classifies these generated samples as being ‘real (could be from the training set)’ or ‘fake’ (not likely in training set). These two networks have contrasting objectives as the Generator wants to create ‘realistic’ samples to fool the Discriminator, and the Discriminator wants to detect as many ‘fake’ samples that the Generator creates. This adversarial battle leads each network to achieve better loss each iteration and once certain epochs are reached, the training is terminated and the Generator is kept to create similar examples from the training set.

Problem

It is widely known that old photos converted from film to digital might have experienced some degradation in the process. World War II photos are no exception and properly restoring these images will benefit us in preserving historical documents. We will investigate if using GANs to approach the ‘image inpainting’ problem, where sections of the image are obstructed and we have to ‘fill in the blanks’, will give us not only a perceptual improvement in quality but also a quantification of how well the restoration occurred.



Data Description

To approach this problem, we start by investigating the population of World War II images. The data collected includes ~3,000 digital images of events from World War II scraped from ww2db.com[5]. The images were separated between three categories: Pre-War, Mid-War, and Late-War.

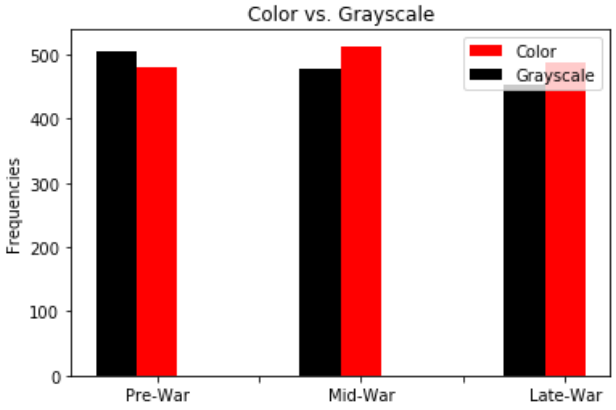
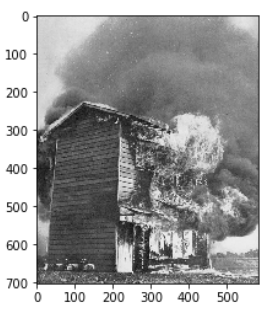


Fig (1)



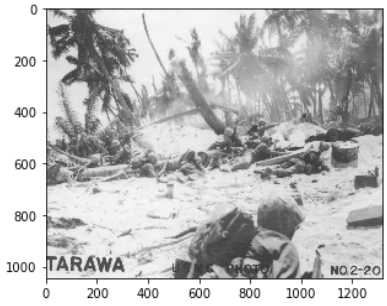


Fig (2)-(4)

Similar amounts of color and grayscale images were collected for each time period in order to have enough data to represent both subsets of the population in further analysis/training. The population we are seeking to refine is the collection of World War II photos and by analysing a sample of these photos, we hope to gain insight behind the degradation process.

In finding similarities between images, I used a blob detector to classify similar images based on number of invariant points and amount of clustering between points. The Difference of Gaussian blob detector was applied to each image and clustered using KMeans.

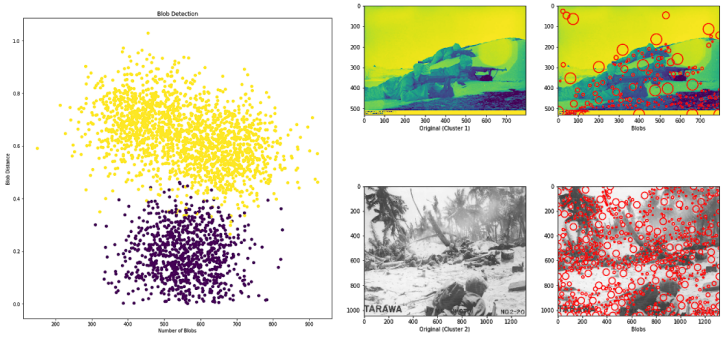


Fig (5)-(6)

Methods

Quantifiable Metric

To start measuring how much 'better' we are getting upon restoration stages, we can quantify this with a measure, namely a combination of Peak Signal-to-Noise Ratio and Structural Similarity Index (See Appendix for definitions).

For iteration t, and chosen iteration ε

Maximize $PSNR + \frac{max(0,t-\varepsilon)*SSIM}{t-\varepsilon}$ over degraded region

Increasing PSNR for early iterations for approximation then adding SSIM for later iterations when PSNR may taper off.

Pre-Processing

Given a target image that is of size $h \times w$, it is loaded into the model where we apply a degradation transformation. To perform this, we use the previously defined binary mask $Mask(x,y)$ and receive the output $Target'(x,y) = Target(x,y) \cdot M(x,y)$. For the block inpainting, we crop a square in the center of the image of sides $min(h,w)$. This block will be what the network will attempt to predict. For the selective mask inpainting, the mask applied is chosen based on the region to be restored (the creases).

Choosing the model

In approaching a GAN architecture, we chose "Deep Generative Prior"[6] which is based on the BigGAN[7] layers/data-loaders as well as "Deep Image Prior"[8], an approach for image restoration learned from parameterizing the network based on the image rather than learning the parameters of the image on the network. *more detail*

Training and parameter selection

A generator is trained for each input image on the cluster closest to the input image in the blob detector. For block and selective inpainting, we have resolution=256x256 with 2000 iterations

Results

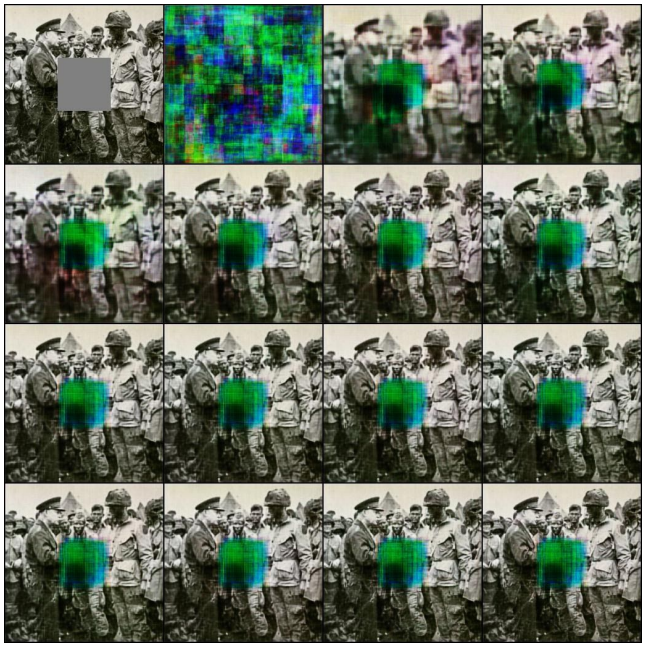


Figure (5): Using block inpainting w/ Random Generator
Metric: +6.9027

Figure (6): Using block inpainting w/ trained weights
Metric: +11.59994

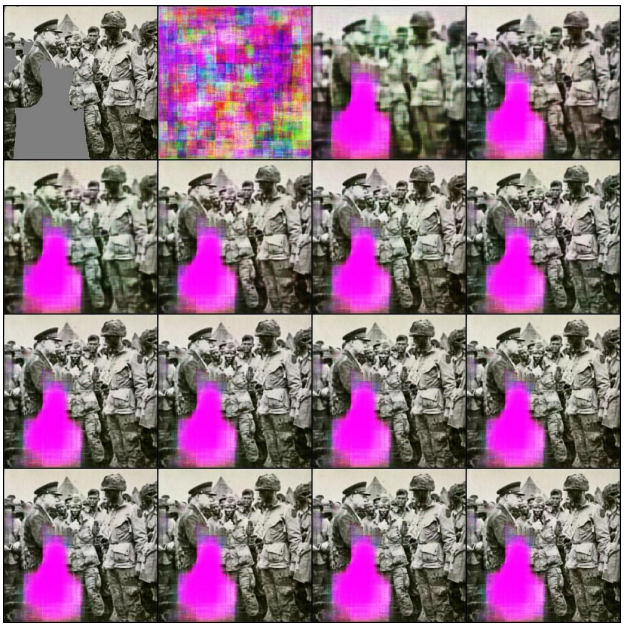


Figure (7): Using selective inpainting mask w/ random weights
Metric: +4.3359

Figure (8): Using selective inpainting mask w/ trained weights
Metric: +12.2342

Discussion

The results seen so far show some promise in using the trained weights over the 'random generator' method. As we see in the first iterations between figure (5) and (6), figure (5) starts with a random image prior and does not converge to a semantically improved restoration as it had no prior knowledge of how to generate the occluded region. In figure (6), we see that the first iteration has aspects of an image learned before and this converges towards a visually better restoration. The metric clearly shows the improvement in restoration in this respect.

Looking at the selective results where masking is applied over known degraded regions, we find the result to be much cleaner. Between figure (7) and (8), it is clear that random generators do not produce any valuable restoration in the inpainting regions. Figure (8) shows more localized improvement on the creases than figure (6).

In the future, we would like to extend this investigation into different restoration objectives like super-resolution, denoising and colorization. We would also like to create an automated selective mask creator that can detect the creases and segment them to make a mask for the selective mask GAN process.

References

[1] *On learning optimized reaction diffusion processes for effective image restoration*, Y Chen, W Yu, T Pock, 2016 https://arxiv.org/pdf/1503.05768v2.pdf

[2] *Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections,* Xiao-Jiao Mao and Chunhua Shen and Yu-Bin Yang, 2016, https://arxiv.org/abs/1606.08921

[3] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014). Generative Adversarial Networks. *ArXiv, abs/1406.2661*.

[4] Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T., & Nahavandi, S. (2019). Deep Learning for Deepfakes Creation and Detection. *ArXiv, abs/1909.11573*.

[5]Chen, C. P. (n.d.). *World War II Database*. Lava Development LLC, Retrieved from https://ww2db.com/photo.php

[6]*Exploiting Deep Generative Prior for Versatile Image Restoration and Manipulation*, Pan, Xingang and Zhan, Xiaohang and Dai, Bo and Lin, Dahua and Loy, Chen Change and Luo, Ping, 2020 https://arxiv.org/pdf/2003.13659.pdf

[7] BigGAN - https://github.com/ajbrock/BigGAN-PyTorch

[8]Ulyanov, D., Vedaldi, A., & Lempitsky, V.S. (2018). *Deep Image Prior*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9446-9454.

Appendix

Peak Signal-to-Noise Ratio:

A measure of reconstruction quality, typically used in lossy compression applications. It is defined as a ratio between the maximum signal power and the amount of noise in the image.

Noise is described as the Mean Squared Error of between the original and an approximation (the reconstructed image). Given an original image I and reconstruction J, we get the MSE as

$$MSE = \frac{1}{3 \cdot m \cdot n} \sum_{x=0}^{m} \sum_{y=0}^{n} \sum_{c=0}^{2} [I(x,y,c) - J(x,y,c)]^2$$

and the maximum signal power as

$$MAX = 2^8 - 1 = 255$$

because there are 8 bits per sample. Thus our PSNR is

$$PSNR = 20 \cdot log_{10}(\frac{MAX}{\sqrt{MSE}})$$

We want our PSNR to be as high as possible, that means we want to decrease our MSE loss as much as possible.

Structural Similarity Index:

- Definition
- Achieving rigid motion invariance for similarity comparison