# Weekly Report

Dr. KHODJA AbdErraouf
abderraouf_khodja@zjnu.edu.cn

05 Dec. 2021

## 1 Context

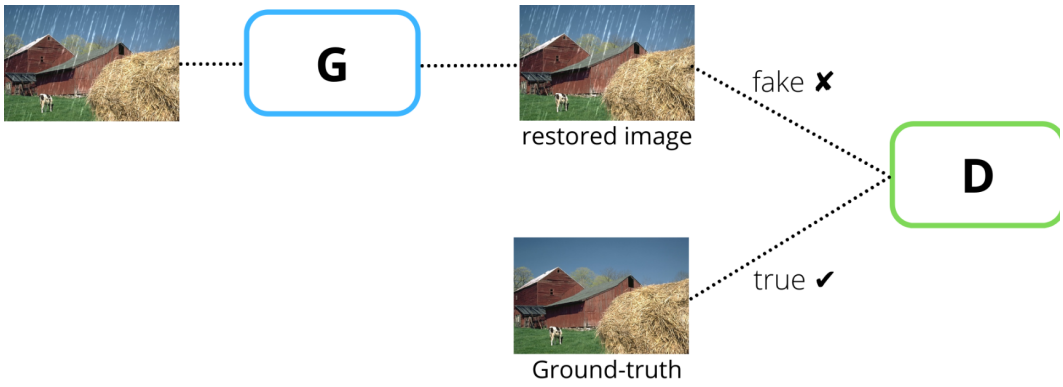


Figure 1: Supervised approach for image restoration through a GAN framework.

In the context of self supervised images restoration I have recently explored Generative Adversarial Network (GAN) to train image restoration models in an unsupervised fashion. This aligns with the philosophy of image-to-image translation which learns the mapping of from an input image to an output images [6]. In this context the goal is to learn the mapping from degraded images $Y$ to clean outputs $X$.

In a supervised fashion, the task would be for the generator $D$ to take as input degraded images $Y$ and generate $D(Y)$ a restored (fake) image. Then, a discriminator $D$ take on the role learn the restored (fake) images from the clean ground-truth $X$ as shown in Figure 1.

It is also still possible to achieve image restoration without having access the paired corrupted-clean images. The GAN generative models showed impressive performance when it comes learn the distribution of a given set of images. We can train our model learn to differentiate between corrupted and clean images without necessarily having access to paired data Figure 2. Thus, the discriminator would learn to differentiate corrupted natural images from clean ones. And hopefully signaling the generator to generate restored images and achieving self supervised image restorations.

Therefore, we can reformulate the GAN restoration problem as training a generator to restore clean images by only looking at unpaired corrupted-clean images as depicted in Figure 3.

### ViT and Robustness

Vision Transformers (ViT) have gain a lot of attention in the computer vision community since the seminal work of Dosovitskiy *et al.*[3]. Further works [1] have extensively explored the robustness of ViT on out-of-distribution benchmarks [5]. As quoted on their paper "*The dominion of CNNs on visual recognition has been challenged by the recent findings that Transformers appear to be much more robust than CNNs. For example, Shao et al. [7] observe that the usage of convolutions may introduce a negative effect on model's adversarial robustness, while migrating to Transformer-like architectures (e.g., the Conv-Transformer hybrid model or the pure Transformer) can help secure models' adversarial robustness. Similarly, Bhojanapalli et al. [2] report that, if pre-trained*


Figure 2: Paired and paired images for training.

*on sufficiently large datasets, Transformers exhibit considerably stronger robustness than CNNs on a spectrum of out-of-distribution tests (e.g., common image corruptions [5], texture-shape cue conflicting stimuli [4]).*".

## Assumption

*ViT might be a key feature for a self supervised GAN image restoration model. The built-in robustness of ViT to various degradation motivate us to explore its effectiveness combined with a GAN framework.*

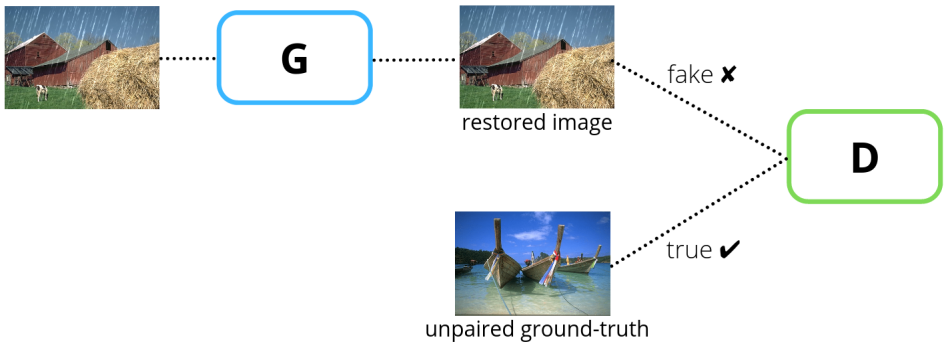


Figure 3: Unsupervised approach for image restoration through a GAN framework.

## 2 Goal

As part of our proposal, we are motivated to develop a ViT-GAN image restoration model for unpaired corrupted datasets (unsupervised learning).

## 3 Challenges

### Self supervision with GAN

Training restoration generator $G$ with adversarial cost alone may introduce visual artifacts in certain regions of the generated restored output, but the clean image discriminator $D$ can still end up classifying it as real data rather than generated data, which harmfs the restoration performance.

In our experiment we could successfully train a self supervised GAN model to restore noisy unpaired noisy-clean MNIST images. This task becomes much more difficult when using color natural images. Although, the noise is being removed from the generated images, generative artifacts are introduced in the output of the generator. Moreover, the colors between the input and the generated output are inconsistent. Suggesting that a style transfer has occurred between the unpaired training sets.

### ViT GANs

ViT are relatively new in the computer vision field. The challenge is that GAN training becomes highly unstable when coupled with ViTs, and that adversarial training is frequently hindered by high-variance gradients (or spiking gradients) [1] in the later stage of discriminator training. Furthermore, conventional regularization methods such as gradient penalty, spectral normalization cannot resolve the instability issue even though they are proved to be effective for CNN-based GAN models. As unstable training is uncommon in the CNN-based GANs training with appropriate regularization, this presents a unique challenge to the design of ViT-based GANs.

# References


[1] Yutong Bai, Jieru Mei, Alan Yuille, and Cihang Xie. Are transformers more robust than cnns? In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[2] Srinadh Bhojanapalli, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. Understanding robustness of transformers for image classification. *arXiv preprint arXiv:2103.14586*, 2021.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[4] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018.

[5] Dan Hendrycks and Thomas G Dietterich. Benchmarking neural network robustness to common corruptions and surface variations. *arXiv preprint arXiv:1807.01697*, 2018.

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[7] Rulin Shao, Zhouxing Shi, Jinfeng Yi, Pin-Yu Chen, and Cho-Jui Hsieh. On the adversarial robustness of visual transformers. *arXiv preprint arXiv:2103.15670*, 2021.