# Cartoon Autopainter

Anil Kumar Gupta and Kritagya Dabi

Computer Science and Engineering Department, Indian Institute of Technology Kanpur

## Objectives

The Objectives of this project are:

- To build a conditional GAN [1] based model for translating sketches of cartoons into colorful and appeasing images. The architecture of the model is based the pix2pix model [2].
- A comparison of outputs was performed with different losses like GAN loss, pixel loss, feature loss and total variance loss.
- Analysis part of the project involved comparison of outputs of the Auto-painter in the cases when it was trained with missing features vs when it was trained without them.

## Introduction

Humans have the ability to comprehend hand-drawn sketches and extrapolate these to have realistic and colorful visual imagery of these sketches. This ability is obtained through experience, by actually seeing the colorful images. Similar to the human ability a machine learning model can be trained to perform the same task given sufficient data to do the learning. Hence a machine learning model (more precisely a neural network) can be constructed to paint a gray-scale image or hand-drawn sketch of a cartoon to pleasing colorful object.

The image generation process from sketches can be regarded as an image synthesis problem. In the past, many non-parametric models were proposed by matching the sketch to a database of existing image fragments. Realistic image generation from black and white images, sketches or outlines using deep neural networks has become a hot topic recently in visual recognition. Colorful realistic images can be generated at pixel or patch level by training from a large data-set.

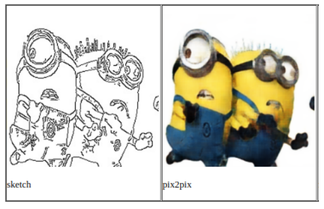


Figure 1: Pix2Pix model output

## Work Done

- **Literature Survey :** We went through the following papers related to the project. Image-to-image translation with conditional adversarial networks [2] has the Pix2Pix model which uses conditional GAN along with the U-net architecture. Auto-painter [3] improves the pix2pix model by taking more losses i.e. feature loss and total variance loss into the picture.
- **Pix2Pix Model Implementation :** We have implemented the code for Pix2Pix from scratch using the code of Pix2Pix implemented in pyTorch by Jun-Yan Zhu as reference.
- **Data-sets :** We have also manually constructed a data-set of images of Minions. For training we need images in pairs of sketch and corresponding cartoon, for this we have written the code for edge detection. Using this edge detection code we generate sketch from the colorful cartoon images.
- **Adding Loss Functions :**
  1. **Feature Loss :** This was the L1 Loss of output of the fourth layer of VGG16. Taking the original code for 16 layer VGG as reference, we coded the VGG till 4 layers and modified the load state function of VGG to load weights of model pre-trained on ImageNet to load these weights till the $4^{th}$ layer only.
  2. **Total Variance Loss :** The variance at point i,j is $\sqrt{(image1[i,j] - image2[i,j])^2 + (image1[i,j] - image3[i,j])^2}$. The total variance is the sum of variance at all the pixels.

## Analysis

We performed the analysis of missing feature prediction of Autopainter and pix2pix. For this we trained the code of Autopainter two times, first when the train dataset had some images with eyes removed and the other in which the train dataset had no such image. We tested both models on a dataset which had some images from which the eyes were removed. As expected, the output when the model was trained with such images was better than when trained without such images. When trained with images with missing eyes, the model was mostly able to reconstruct the eyes in many cases. We did a similar analysis on pix2pix model too. Autopainter performs slightly better than pix2pix for predicting the missing eyes.

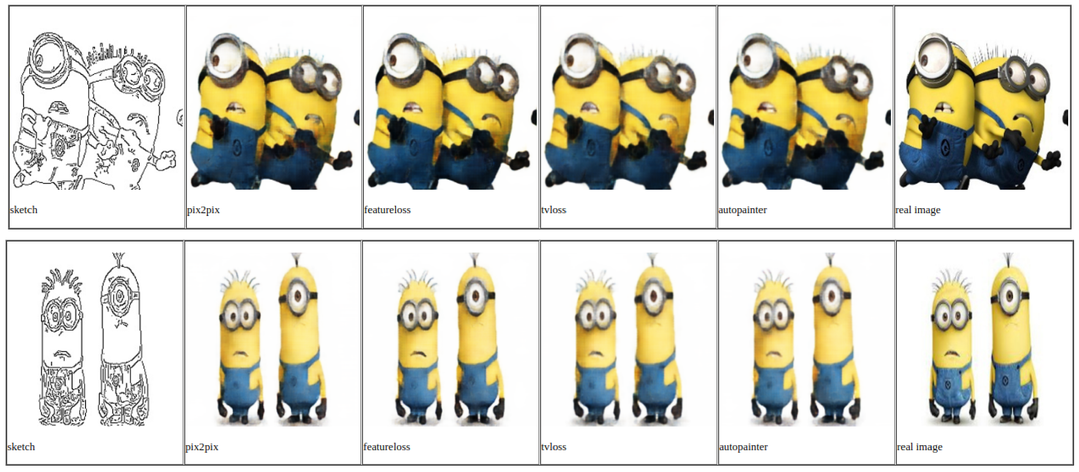


Figure 2: Comparison of performance of all the models

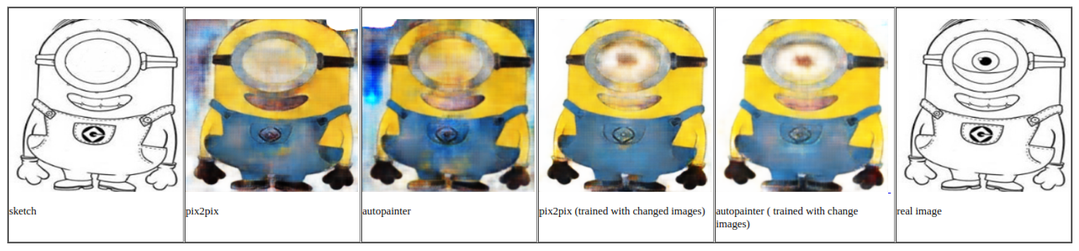


Figure 3: Result comparison on input image with missing features(eyes).

## Inferences

- The output produced by the Autopainter and TVLoss models are relatively smooth, that is there is no sudden change of color at any point in the image and color gradients are uniform as compared to pix2pix and feature-loss model.
- The total variance loss added tends to reduce the change of colors, thus leading to loss of feature information, though it leads to more even images, it reduces the sharpness in the image.
- Feature loss is used to preserve small details/ features of input image, but sometimes it causes unintended white spots in final images.
- When testing on minion images with background, auto-painter gives better results.
- Overall Autopainter surely produces more realistic looking images in the cases when there are no random white patches because of presence of even tone and absence of abrupt color changes.

## References

[1] Osindero S. Mirza, M. Conditional generative adversarial nets. *CoRR abs/1411.1784*, 2014.

[2] Tinghui Zhou Alexei A. Efros Phillip Isola, Jun-Yan Zhu. Image-to-image translation with conditional adversarial networks. *Berkeley AI Research (BAIR) Laboratory*, 21 Nov 2016.

[3] Zhenbo Luo Yifan Liu, Zengchang Qin and Hua Wang. Auto-painter: Cartoon image generation from sketch by using conditional generative adversarial networks. *Intelligent Computing and Machine Learning Lab, School of ASEE Beihang University, Beijing 100191, China*, 7 May 2017.