Text Replacement and Generation in Images Using GANs

Names: Andrew Sharp, Christian Gabor, Joseph Son
SUNet IDs: awsharp, gaborc, joeson

Introduction

Currently, it is not easy to replace text in images while maintaining the same style and font. Replacing logos and text appearing in real world images requires manual labor.

We aimed to create a GAN that can automatically replace text in an image with any similar length input text string. We have developed a model capable of replacing a single character. In the future, this model could be extended to work on longer strings as well.

We see several possible uses for such a model, such as for generating synthetic data for optical text recognition to increase the size of existing datasets or for quick replacement of text in graphic design, such as for logos, animations, or video games.

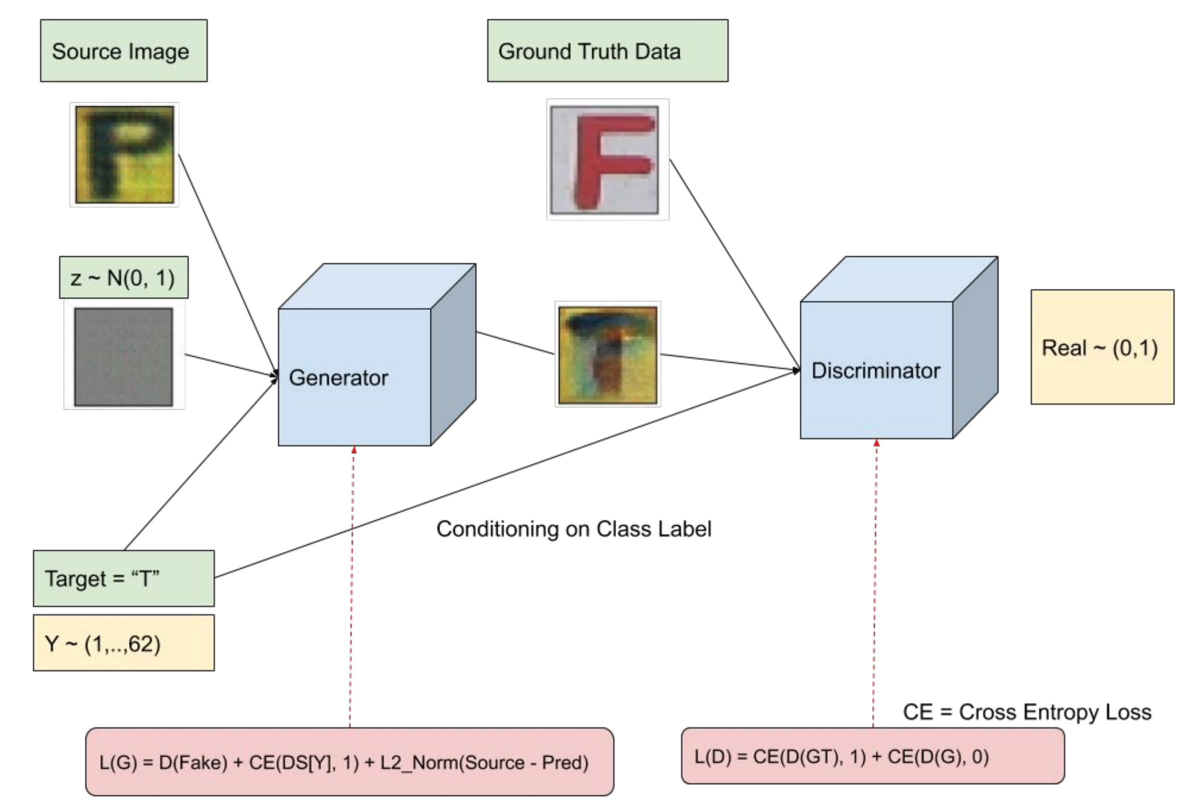
Data

Dataset
The GAN is trained on the Chars74K dataset which has images of over 7,000 characters from natural images. Each image is composed of RGB channels of varying resolutions.

Preprocessing
We normalize the images to have intensity values in the range [−1,1] and resize them to size 28×28.

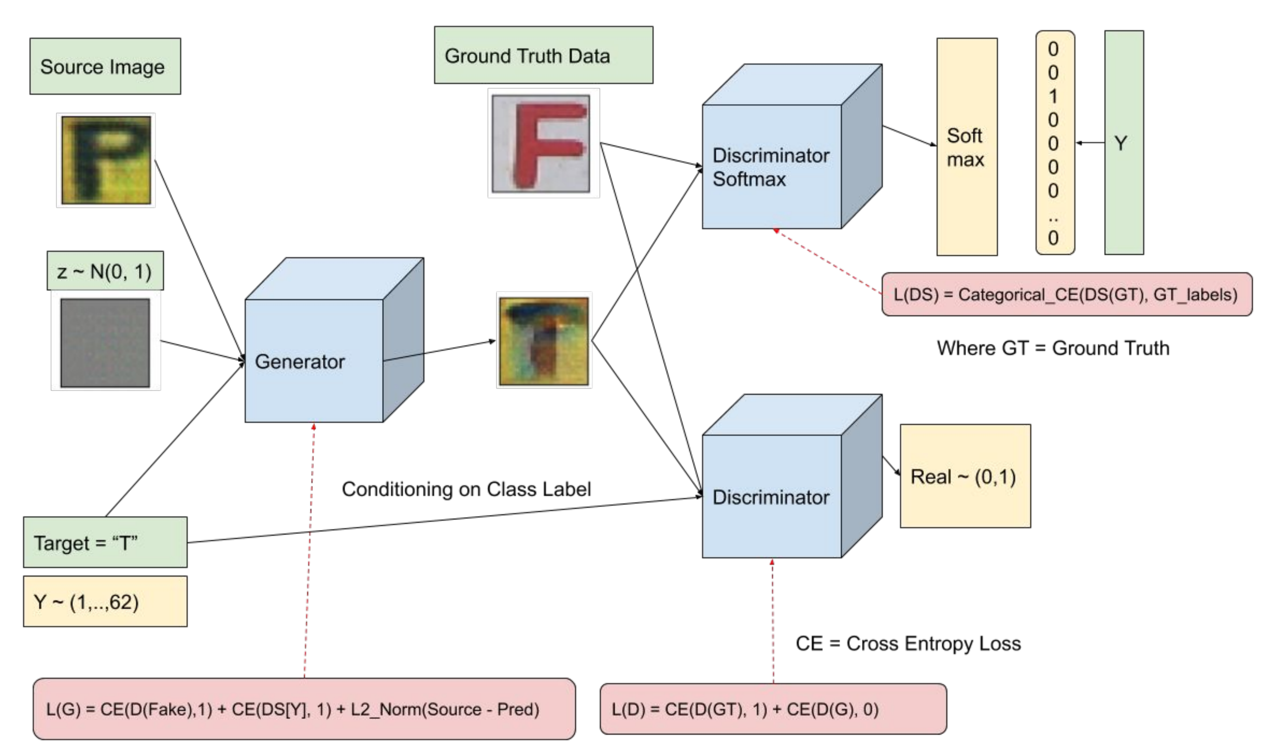
Conditional GAN Models

Our current model uses a target character class which is input into both the discriminator and the generator of a DCGAN architecture and turned into a class embedding. This embedding allows us to use the target class label to select the desired character output. The generator then attempts to produce an image which the discriminator will believe comes from the target class.



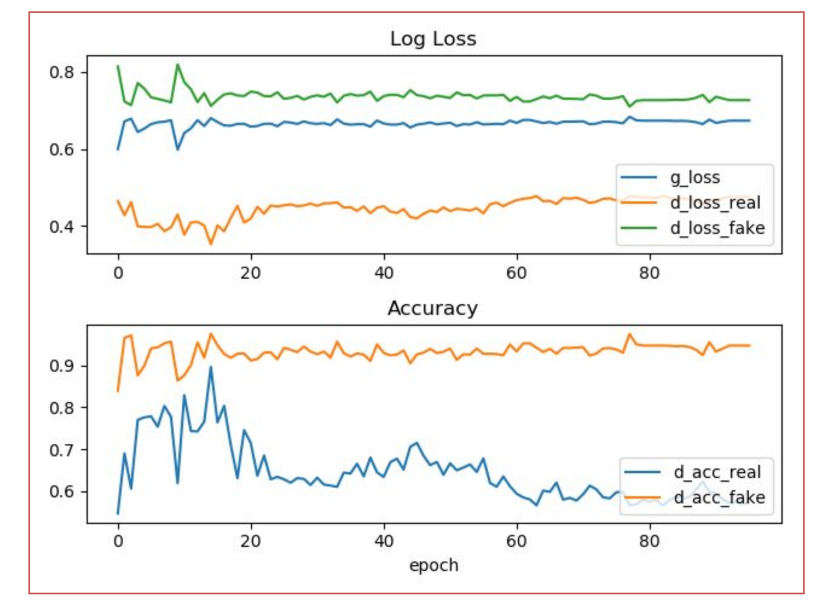
GAN with Softmax Discriminator

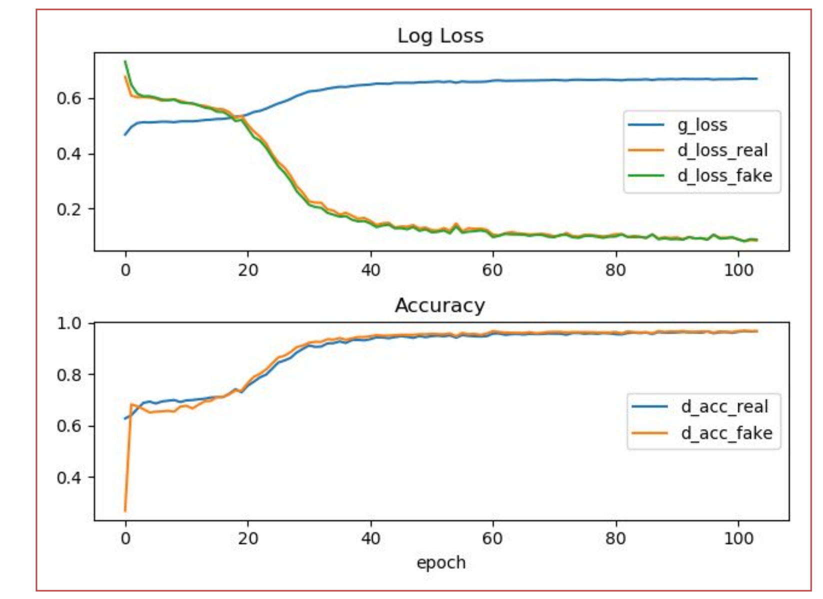
We add a second discriminator to the generator loss function that trains on ground truth images and labels. This additional loss penalizes the generator when it makes an image that does not look like it belongs to the target class.

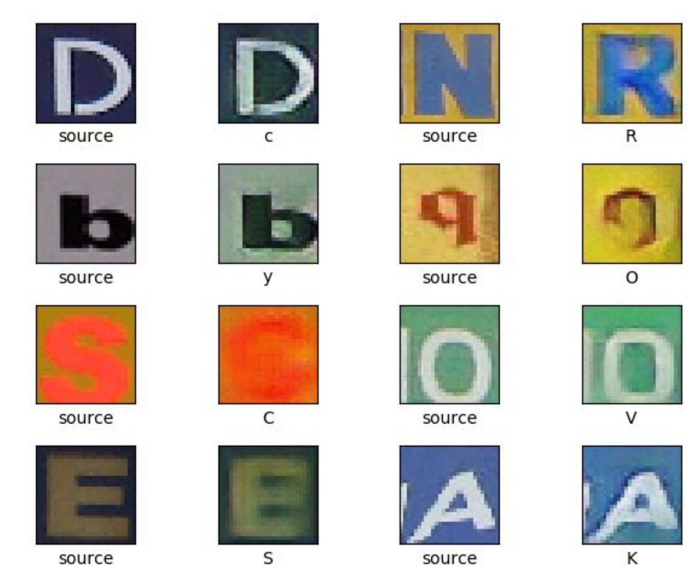


Results

The generator is more stable without the softmax loss, which ultimately degrades picture quality. Left: without softmax, Right: with softmax.







Conclusions

Adding the softmax discriminator resulted in more generator images that looked like the target class. Roughly 1% of images from the original model changed to the proper class image by human examination, while the softmax model converted 17%. The training hyperparameters need to be tuned better to achieve stable performance and a better balance between maintaining style and updating because it often diverges during training.