

PixelBrush: Art Generation from text with GANs

Jiale Zhi jz2@stanford.edu

Introduction

Artworks especially paintings are an indispensable part of a lot of people's life. They are drawn by skilled artists. But good artists are rare, and good paintings are usually expensive, not everyone has the opportunities to enjoy artworks. Even for people who can afford good paintings, most people don't have access to personalized painting due to high price and low number of supply.

To make artworks more accessible to everyone, we'd like to utilize recent development of computer vision so that we can train computer algorithms to generate artistic paintings that look like painted by artists.

Our Goal

- Train a computer algorithm to generate artistic painting
- Personalize these paintings based on text descriptions

Applications

- Personalized screen savers
- Home decorations
- Art education for students

Data

	Aero	Bird	Boat	Chair	Cow	Table	Dog	Horse	Sheep	Train	Total
Train	74	319	862	493	255	485	483	656	270	130	3463
Val	13	72	222	140	52	130	113	127	76	35	865
Test	113	414	1059	569	318	586	549	710	405	164	4301
Total	200	805	2143	1202	625	1201	1145	1493	751	329	8629

Figure 1: Oxford paintings dataset class statistics

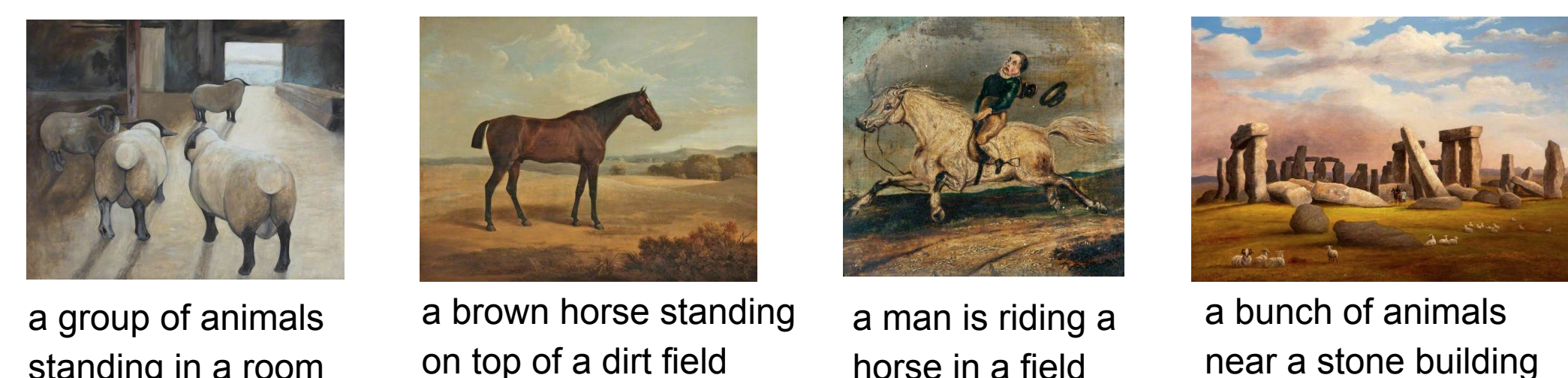
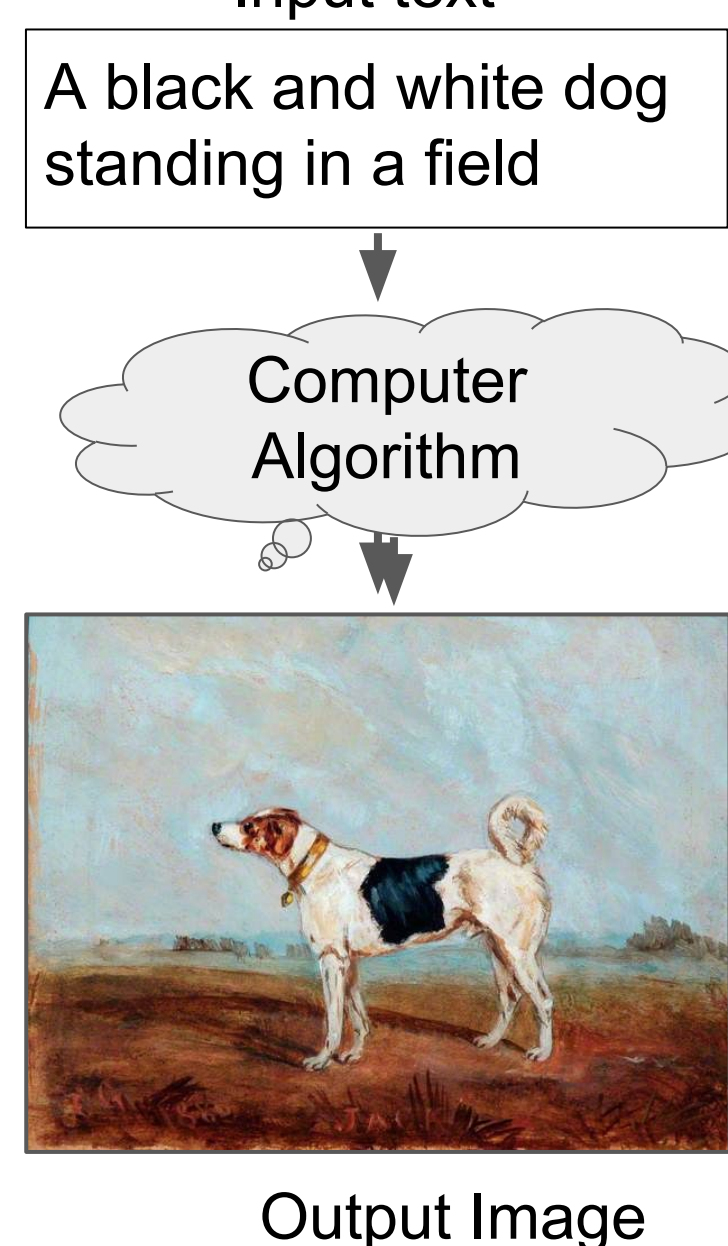
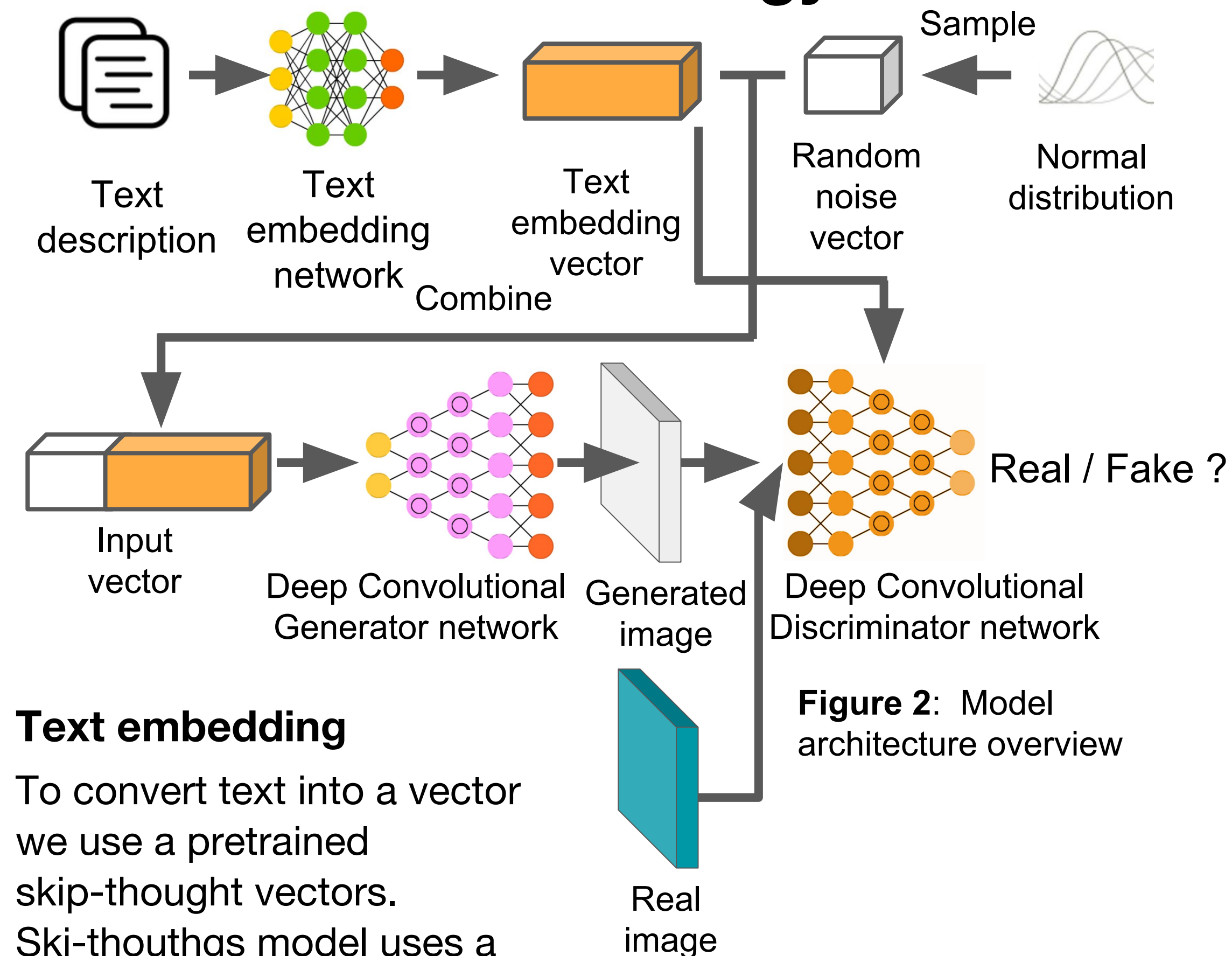


Image captions are generated from Andrej Karpathy's NeuralTalk 2 with human editing for wrong and non-accurate captions.

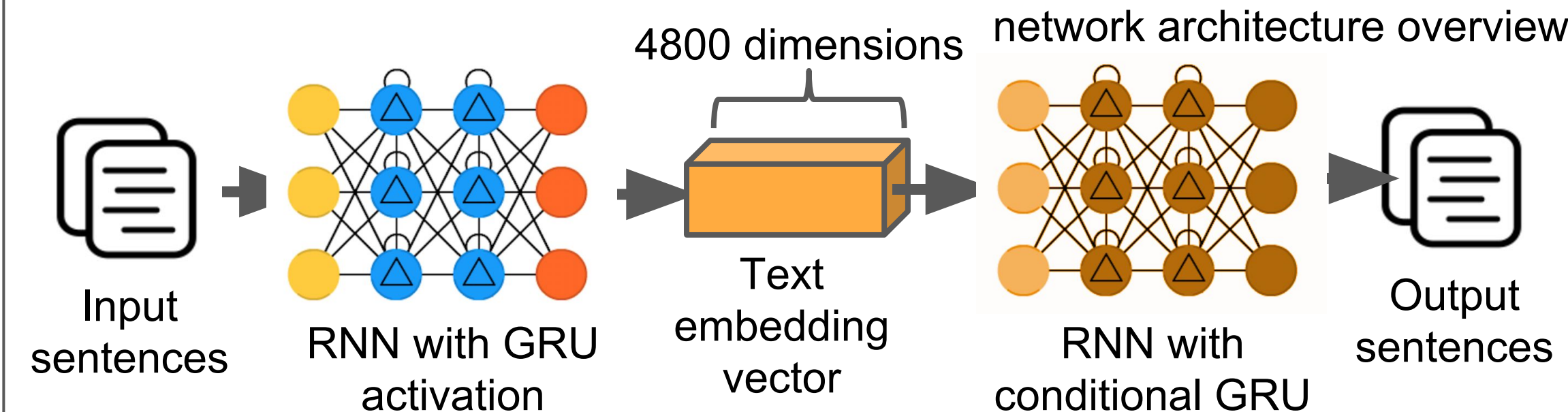


Methodology



Text embedding

To convert text into a vector we use a pretrained skip-thought vectors. Skip-thoughts model uses an encoder-decoder architecture.



Deep convolutional GAN architecture

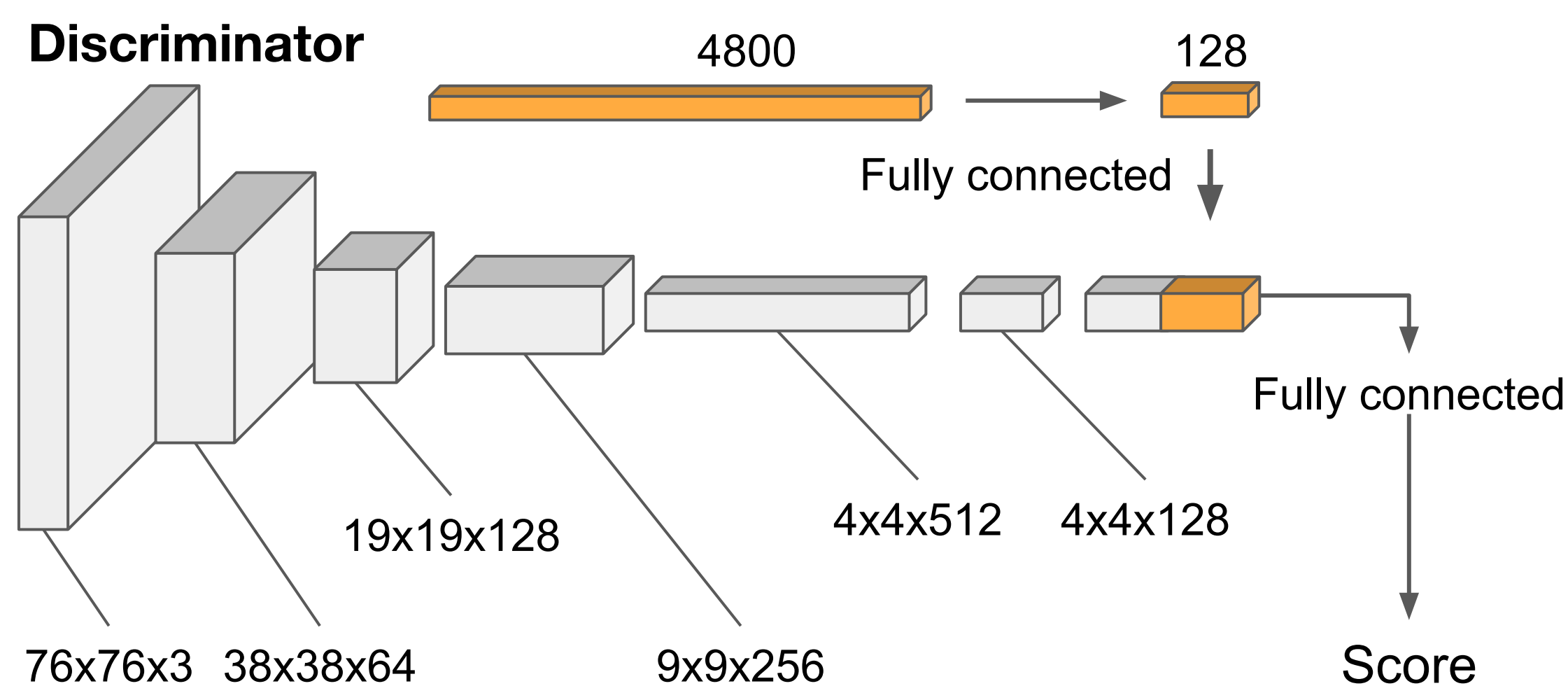


Figure 4: Discriminator network architecture overview

Results & Analysis

Baseline



Figure 5: Art generated from DCGAN without any conditions.

We use DCGAN as our baseline. We train DCGAN on the same dataset that we use for our work. We can see the result of DCGAN is sharp and color distribution looks like training data. But the images generated from DCGAN doesn't show any real objects.

Our work

a clock tower in the middle of a city
a brown horse standing on top of a dirt field
a brown horse standing in a field of grass
a group of people riding horses on a dirt road



Figure 6: Images generated from text descriptions on test set, first column are real images.

In our work, images are generated according to text descriptions, we can see generated image roughly shows what input text describes. Note that text descriptions are generated from image caption network and have errors.

Future Work

We plan to train our conditional deep convolutional GAN on a larger dataset so that discriminator can learn from more examples and generator will generate more clear images. During training, we also found the result images are blurred, the can be because training images are not sharp, but the exact reason needs to be investigated.

We also plan to utilize newly published Wasserstein GAN to improve our training stability and generated image quality.