Extracted the style from the image with Deep Neural Network

Wu Bingzhe
1200010666
The School Of Mathmatical Science

November 13, 2015

1 Problem Description

In the field of visual art,especially painting,humans have mastered the skill to create unique visual experience through composing a complex interplay between the content and style of an image.In other area of computer vision such as object detection and recognition , Convolutional neural networks have recently enjoyed a great success in large-scale image recognition[3] which has become possible due to large public image repositories,such as ImageNet(Deng et al.,2009),and high-performance computing systems,such as GPUs or large-scale distributed clusters[1](Dean et al.2012).In this project ,we use a neural representations to separate and recombine content and style of arbitrary images.This work also offers a algorithmic understanding of how humans create and perceive artistic imagery.

2 The Dataset

In this project,we plan to use the datasets from ImageNet .ImageNet is a dataset of over 15 million labeled high-resolution images belong to roughly 22,000 categories.The images were collected from teh web and labeled by human labelers using Amazon's Mechanical Turk crowd-sourcing tool.Staring in 2010 ,as part of the Pascal Visual Object Challenge,an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has benn held.ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories.In all,there are roughly 1.2 million training images,50,000 validation images,and 150,000 testing images.

3 The Method

3.1 Preprocessing

Before Extracting the features from the image,we plan to rescale the image such that the shorter side was of length 256,and then cropping out the central 256×256 patch from the resulting image.

3.2 Extract the features

We plan to use the convolutional neural network for extracting the features from the images.The model we plan to use is the VGG-model.

4 Software and hardware

4.1 Software

We plan to use Caffe[2](A deeplearning framework) to train our model ,use torch to apply the model to combine the style and the content from two different images.

4.2 Hardware

1. GPU:

   GTX-Geforce TitanX

2. CPU:

   Intel i7 4790k

5 Evaluation Strategy

In our project , there are two different evaluation strategy during different process. First , we use the classify testerror during training the model .During combining the style and content ,we use the loss function

$$L_{total}(\overrightarrow{p}, \overrightarrow{x}, \overrightarrow{a}) = \alpha L_{content}(\overrightarrow{p}, \overrightarrow{x}) + \beta L_{style}(\overrightarrow{x}, \overrightarrow{a}) \tag{1}$$

Where $L_{content}$ and L_{style} is the loss between the image that we get and the original images.More over, we use this :

$$L_{content}(\overrightarrow{p}, \overrightarrow{x}, l) = \frac{1}{2} \sum_{ij} (F_{ij}^{l} - P_{ij}^{l})^2 \tag{2}$$

References

[1] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and Fei-Fei. Large scale visual recognition challenge. In *In http://www.imagenet.org/challenges/LSVRC/2012/index*, 2012.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:2012, 2012.