Proposal for DSA5204

Name (AXXXXXXXX) Shuyuan Shen (A0251155M) Yuhua Wang (A0251546E)
Geer Zhang (A0251176H) Junhao Huang (A0251318L)

1 xxx

xxx [**zhu˙unpaired˙2017**]

2 Introduction to research problem

This paper is aimed at solving the unpaired image-to-image translation problem and applying the technique in many tasks like collection style transfer, season transfer, photo enhancement, etc, where the paired training data does not exist.

Image-to-image translation is one of the main research areas in computer vision, which aims to produce an output image reflecting the target domain's style while keeping unrelated contents of the input source image unchanged. Common image-to-image translation tasks are multi-modal, single-modal and single-image image-to-image translation. Since Generative Adversarial model(GAN) has been proposed, it has been often used in image-to-image translation tasks, and has shown good performance.

When solving these image-to-image tasks, supervised setting with paired input-output training data is the most commonly used. However, in practice, obtaining paired data can be hard, for it may cost much and the desired output is not even well-defined sometimes, like the translation between photographs and pictures in Monet style. Therefore, in this paper, they have learned the relationship between domains, not pairs, i.e. the goal is to learn the mapping G: $X \longrightarrow Y$ given training samples $\{x_i\}_{i=1}^{N} \in X$ and$\{y_j\}_{j=1}^{M} \in Y$ in order to make distribution of images from $G(X)$ indistinguishable from the distribution of Y.

Although GAN is widely used in solving image-to-image problems, in case of unpaired training data, it fails and needs modifications. The mapping G and adversarial loss can just ensure that the distributions of $\hat{Y}$ and Y are the same, but the individual x and y may not be paired up, and the optimization is difficult to make progress due to cases of mode collapse, for example, different inputs may get the same output. Therefore, a "cycle consistent" loss is added to the adversarial loss to prevent the learned mappings from contradicting each other. In the system, there is a translator $G : X \longrightarrow Y$and its inverse$F : Y \longrightarrow X$, they are trained simultaneously with two discriminators D_X and D_Y. The objective is the combination of Adversarial Loss and Cycle-Consistency Loss.

3 Background

The traditional Image-to-image translation can be trace back to texture transfer because image style can be seen as a texture, and if we keep some semantic information while compositing image. Then, we can obtain the result of image-to-image translation. However, at this stage, the transformation was based on pixels, the basic image feature, and didn't include any semantic information. Thus, the performance wasn't ideal.

As it was mentioned above, image-to-image translation can be divided into two parts, image texture extraction and image reconstruction, but, for traditional Image-to-image translation, it can hardly solve the reconstruction part. With the rapid development of deep neural network, researchers found that it can be utilized on image recognition, and CNN (e.g., VGG19) can extract the best feature instead of dividing the image into pixel.

After GAN has been proposed, it obtained great performance on image-to-image translation. What make GAN different from the traditional neural network is the design of the loss function. GAN is composed of generator and discriminator, which used to reconstruct the image and distinguish the image respectively (e.g., CycleGAN, which proposed a new loss, Cycle Consistency Loss).

What we need to learn:

1) The structure of VGG19 and how it was applied in feature extraction.

2) The different between the loss function in traditional neural network and the one in GAN.

3) The structure of GAN.

4) The loss proposed by CycleGAN, why adding this loss can improve the performance.

4 xxx

xxx

5 xxx

xxx

6 xxx

xxx