ICANN19

28th International Conference on Artificial Neural Networks

Research on Image-to-Image Translation with Capsule Network

Jian Ye, Qing Chang* ,and Xiaotian Jia (changqing@ecust.edu.cn)
School of Information Science and Engineering,
East China University of Science and Technology, Shanghai 200030, China

Abstract

Recently, Capsule Network (CapsNet) with great potential has been proposed. It is able to preserve more input's information, especially location information by it's unique capsule structure and dynamic routing algorithm.

We propose Capsule conditional Generative Adversarial Network (CapscGAN) for performing image-to-image translation tasks.

CapscGAN utilizes CapsNet to encode image into one capsule called PixelCapsule and combines it with Markovian discriminator (PatchGAN) as discriminator. Multiple datasets' results demonstrate that our model has higher translation quality than convolutional image translation framework.

Introduction

Traditionally, different kinds of tasks are solved by corresponding specific models or methods whereas this makes it inefficient to complete multiple different tasks. Pix2Pix which uses conditional Generative Adversarial Networks to generate different styles of images tackled this problem.

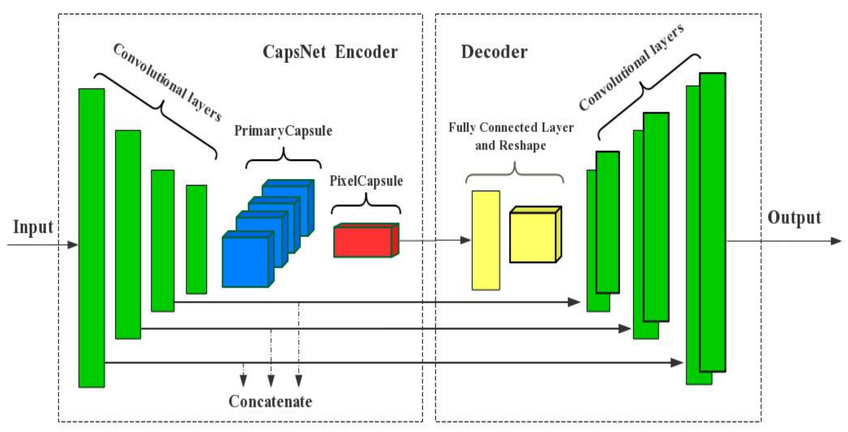
Convolution Neural Network has been the most popular framework in computer vision tasks, due to its remarkable flexibility and performance. However, it does not take into account the spatial hierarchy and rotation invariance between features. Our work is based on Pix2Pix framework. We enhance translation ability by using capsule rather than scalar to represent a specific entity. We use CapsuleGAN's strategy in our model's discriminator. In order to achieve satisfactory translation quality, we have made a series of changes in CapsNet.

Application of CapsNet in Generator

Our model's generator is composed of CapsNet encoder and deep convolutional decoder with "U-net".

For generator, because the task and input size are different from original CapsNet , there are three main differences:

- Multiple convolution layers with small kernel size instead of convolution layers with large kernel in CapsNet are used.
- Capsule's number and dimension in PrimaryCapsule and DigitCapsule layers are modified. Especially, in DigitCapsule layer, the output has only one capsule that contains input latent information.
- We choose Leaky ReLU as the activation function in routing part. This change maintains the entity's inherent information while routing.



Effect of PixelCapsule's Dimension

For our model, capsule's dimension in PixelCapsule layer determines capacity of encoding information. We do a series of experiments on this.

The model with 256 dimensions is the most suitable one.

Method	Pixel acc.	Class acc.	Class IoU.
CapscGAN-256	**0.79**	**0.25**	**0.20**
CapscGAN-384	0.62	0.23	0.19
CapscGAN-512	0.66	0.22	0.17
CapscGAN-1024	0.56	0.18	0.14

Analysis of the Activation Function

For out task,activation function should only filter the image information or extract advanced features through the dynamic routing algorithm, and do not change their distribution ratio.We use Squashing, Normalization and Leaky ReLU as activation function for experiments, respectively. Leaky ReLU has the best effect in our experiments.

Note that it is not necessarily the most suitable. We believe that other functions which do not change state of information distribution are likely to achieve better results.

Input Squashing Normalization Leaky ReLU Ground-Truth

Role of CapsNet in Discriminator

Translation ability has been promoted to some extent on account of CapsNet's use in generator. We have explored the role of CapsNet in discriminator.

PatchGAN fail to adapt the change in generator's translation capability and could only translate simple figure outline. But CapsNet used in discriminator and generator makes the two more compatible.

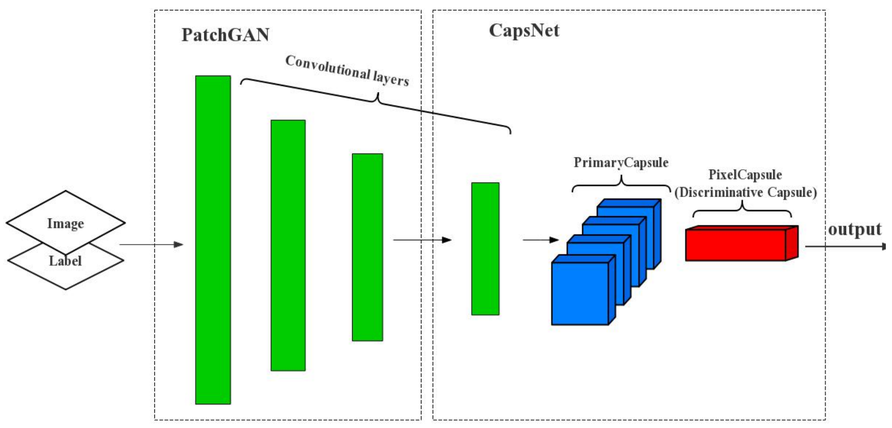
lnput PatchGAN PacthGAN with CapsNet Ground Truth

CapsNet in Discriminator

We add a CapsNet after PatchGAN and use combination of the two as our model's discriminator. The discriminator focuses on local image patches first, and then take them as a whole to CapsNet.

In our model, PatchGAN is used to get high-level information about each patch of image and classify it. Then CapsNet identifies input's authenticity by this information.

CapsNet in discriminator is modified in the same way as the one in generator

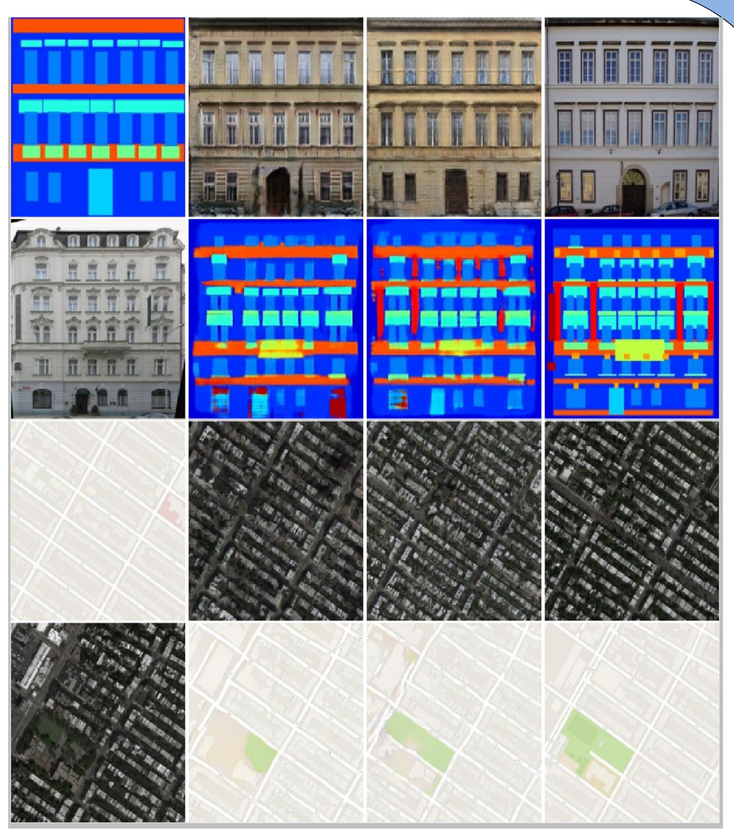


Result and Conclusion

Our model is compared with Pix2Pix in multiple datasets. We keep the setting and parameters of our model same with Pix2Pix except CapsNet and fully connected layer.

On multiple metrics,our method has higher values and translation results are closer to ground truth in terms of color parts and overall structure.

A series of experiments on multiple datasets show that the images produced by our model are more precise and CapscGAN has higher translation ability than convolutional image translation framework.



Input Pix2Pix CapscGAN-256 Ground-Truth

ManuscriptID:101