Non-local Neural Networks

Liangjie Cao

Aug. 19, 2018

1. Introduction

It opens up a new direction to solve the long-distance dependence in space-time domain in video processing. In this paper, non-local averaging method is used to deal with the relationship between local features and feature points of panorama. This kind of non-local operation can be easily embedded into the existing model, and has achieved good results in the video classification task, and surpassed the Master-CNN of his ICCV best paper in the static image recognition task. And surpass CNN to overcome the shortcomings of CNN network being too concerned about local features. Inspired by the application of NL-Means in image denoising, the task of serialization is to consider all feature points for weighted computation, which overcomes the shortcoming of CNN network that pays too much attention to local features.

2. Method

This is an article by CVPR2018 that introduces non local ideas into video classification. This article is inspired by the traditional non-local mean operation, the core idea is: our non-local operation computes the response at a position as a weighted sum of the features at all positions. It is very local, so in order to achieve the non-local effect (that is, the long range dependency is obtained in the text), it is generally superimposed such a feature extraction layer, so that the receptive field of the high-level network is getting larger and larger. The breadth of information obtained is also increasing. However, this method of continuous superposition will inevitably lead to an increase in the amount of calculation and an increase in the difficulty of optimization, so there is a non local mechanism proposed by the author of this paper.The non-local operation is shown in Equation 2. The meaning of the expression is to use the information of xj near x_i to get y_i.

$$y_i = \frac{1}{C(x)} \sum_{\vee} j f(x_i, y_i) g(x_j) \tag{1}$$

The meaning of x_i and x_j in Equation 2 can be explained by referring to the chart in Figure 1. It is also very easy to understand the non local operation. It is to use the information of the surrounding points when extracting some features. This "around" can be time. Dimensions can also be spatial dimensions. The time dimension is just like the video classification example in this article, which makes better use of timing information.

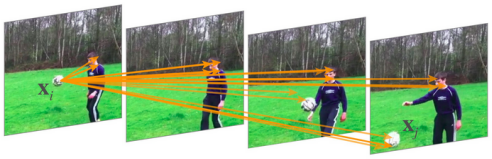

Figure 1. Non-local operation

3. Experiment

Experiments are carried out on video classification, object detection and object instance segmentation tasks that require non-local information association. The ablation study was conducted on the Kinetics dataset to examine the effectiveness of the details of the NL block. The results will not be repeated. [2]

There are different definitions of F (.) in NL blocks, but for better visualization use embedded Gaussian + dot product, the method shown in the formula mentioned above.

The position of NL block is placed in the backbone of the network: put it in the shallow layer, and increase in the upper reaches.The role of NL block deepening: for the shallow backbone network, deepening NL block can improve performance. It is difficult to improve performance for larger and deeper networks, either by adding NL blocks or by deepening the depth of the backbone network.(video task) NL block is better than time alone in time domain or space domain.(video task) compared with C3D [1]: faster and better than C3D [3]. We presented a new class of neural networks which capture long-range dependencies via non-local operations. Our non-local blocks can be combined with any existing architectures. We show the significance of non-local modeling for the tasks of video classification, object detection and segmentation, and pose estimation. On all

tasks, a simple addition of non-local blocks provides solid improvement over baselines. They hope non-local layers will become an essential component of future network architectures.

References

[1] S. Karaman, L. Seidenari, and A. Del Bimbo. Fast saliency based pooling of fisher encoded dense trajectories. In *ECCV*, 2014. 1

[2] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[3] H. Wang and C. Schmid. Action recognition with improved trajectories. In *CVPR*, 2013. 1