Non-local Neural Networks

Liangjie Cao

Aug. 17, 2018

1. Background

The article is mainly inspired by NL-Means in image denoising applications. The task of serialization is to consider all feature points for weighted calculation, which overcomes the shortcomings of CNN network focusing on local features.

Image denoising is a very basic and very necessary study. Denoising is often performed before more advanced image processing and is the basis of image processing. The noise in the image is often approximated by Gaussian noise. An effective way to remove Gaussian noise is image averaging. The result of averaging N identical images will reduce the variance of Gaussian noise to one-Nth. Now the better denoising algorithm is based on this. An idea to algorithm design.

The full name of NL-Means is: Non-Local Means [1], which is a non-local average. It was proposed by Baudes in 2005. The algorithm uses the redundant information that is common in natural images to denoise. Different from commonly used bilinear filtering, median filtering, etc., using image local information to filter, it uses the entire image to denoise, find similar regions in the image block, and then seek for these regions. On average, it is better to remove the Gaussian noise present in the image.

The usual CNN network simulates the cognitive process of humans. Local connections are used between adjacent layers of the network to obtain the local characteristics of the image. It is generally believed that people's perception of the outside world is from local to global, and the spatial connection of images. It is also localized that the pixel connections are tighter, while the farther pixel correlation is weaker. Therefore, it is not necessary for each neuron to perceive the global image. It only needs to perceive the local part, and then integrate the local information at a higher level to obtain global information. The idea of network part connectivity is also inspired by the visual system structure in biology, the bottom layer captures contour information, the middle layer's combined contour information, and the high-level combined global information. Finally, different global information is finally synthesized, but due to sampling and Information is transmitted layer by layer and loses a lot of information, so traditional cnn has limitations in global information capture.

2. Non-local Neural Networks

Non-local operations have the following advantages: (a) In contrast to the repetitive behavior of cyclic operations, non-local operations directly capture long-range dependencies by calculating the interaction between any two locations, without the need for two-position positions. Distance constraint. (b) As we have shown in our experiments, non-local operations are highly efficient and achieve the best results with only a few layers. (c) Finally, our non-local operations maintain the size of the input variables and are easily combined with other operations (such as convolution operations).

They will demonstrate the effectiveness of non-local operations in video classification applications. In video, long-range interactions occur between long-distance pixels in space or time. A non-local [2] block is our basic unit and can capture this spatiotemporal dependency directly through feedforward. In some non-local blocks, our network structure is called a non-local neural network and has a more accurate video classification effect than a 2D or 3D convolution network (including its variants). In addition, non-local neural networks have lower computational overhead than 3D convolutional networks. We conducted detailed studies on the Kinetics and Charades datasets (optical flow, multiscale testing, respectively). Our approach yields better results on all datasets than the latest methods.

In order to prove the versatility of non-local operations, we further carried out experiments on target detection/segmentation and pose estimation on the COCO dataset. Based on MaskR-CNNbaseline, our non-local blocks require only a small extra computational overhead to improve accuracy in three tasks. Experiments in video and images have shown that non-local operations can be a common part of designing deep neural networks.

Figure 1. Example of the behavior of a non-local module on res3

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1

[2] X. Zeng, W. Ouyang, J. Yan, H. Li, T. Xiao, K. Wang, Y. Liu, Y. Zhou, B. Yang, Z. Wang, et al. Crafting gbd-net for object detection. *TPAMI*, 2017. 1