Check point 1 summary of CSPNet for improving Computer Vision Performance on lower power hardware

Anonymous Authors[1]

Abstract

[Not needed for this checkpoint]

1. Summary of (Wang et al., 2020)

In the world of computer vision there is a growing need for greater accuracy and speed on accessible hardware in tasks for detection and classification. The current solution for gaining accuracy is to make the model deeper or longer, but the computational cost increases exponentially with each addition of layers to the model, and the end result is an increase in accuracy at the cost of speed. This trade-off makes many state of the art models unviable for low power and more available computer hardware like cell phones and laptops. This paper (Wang et al., 2020) attempts to suggest a method to try and reduce the effect of this problem by allowing for a richer gradient combination, while reducing or maintaining the number of parameters. In other words, they want to reduce redundancy in learned feature maps so that there is less chance that 2 feature maps represent the same attribute. In doing so the idea is that you could remove the redundancy, and therefore decrease the number of features that need to be learned, and as such you can decrease the number of parameters in the model. In order to achieve this, they built a new Computer Vision Method called the Cross Stage Partial Network (CSPNet) that can be applied to many existing Computer vision models to improve performance. In order to evaluate the Method, they applied the CSPNet Method to existing state of the art networks like ResNet, DenseNet, ResNeXt and Darknet53, trained the newly constructed model on the ImageNet dataset, and compared the accuracy and speed of the models. The idea is that if the models with the CSPNet modulation perform equivalently or better in-terms of speed and accuracy than their Non-CSP counterparts, then this novel method is successful. CSPNet can be a novel Computer Vision technique that can be applied to current state of the art models to improve speed and accuracy of the models while decreasing the overall computational cost.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the Conference on Computer Vision and Pattern Recognition (CVPR) . Do not distribute.

2. List of Questions

1. How does the CSPNet method improve Gradient variability and feature map redundancy without a feature map comparison?

2. How does the dispersed cardinality of feature maps improve computational speed in comparison to the direct computation of a feature map? (direct computation is using a single convolution filter to compute N depth feature maps. Dispersed cardinality is distributing the N feature map computations to a set of M convolution blocks that are computed separately and will be concatenated to form the final N depth feature map)

3. Unlike other backbones, why does the introduction of an EFM(neck) improve object detection performance but has no effect on the image classification task? In other works, how does the CSPNet maintain spatial information of objects allowing the EFM to perform better, given that the input item has already been down sampled and the spatial information seems to have already been lost?

4. Can this CSPnet method be generalized and applied to AI fields like NLP to improve speech understanding on low power hardware?

Prof. Inouye approved of the paper below over email on September 2, 2020 because it is integral to a research project I am working on.

References

Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. URL https://openaccess.thecvf.com/content_CVPRW_2020/papers/w28/Wang_CSPNet_A_New_Backbone_That_Can_

Enhance_Learning_Capability_of_CVPRW_2020_paper.pdf.