A Real-Time Semantic Segmentation Model Using Iteratively Shared Features In Multiple Sub-Encoders Supplemental Materials

Tanmay Singha[1], Duc-Son Pham, Aneesh Krishna

School of EECMS, Curtin University, Bentley, 6102, Western Australia, Australia

Keywords: semantic segmentation, deep convolution neural networks, multi-encoder, decoder, feature scaling, feature aggregation, feature reuse, resource-constrained applications, mobile devices

1. Introduction

Due to limited space in the manuscript and as per the reviewers' instructions, we needed to present some technical implementation details as supplemental materials to keep the main paper within the page limit. The implementation details are important for the reader to faithfully reproduce the results in our work. This document is kept on the Github repository for the paper.

1.1. Implementation Details

We train our model using a computer equipped with three NVIDIA Titan RTX GPUs. We set the batch size equal 4 for high-resolution (Cityscapes, Indoor objects) and 8 for low-resolution (KITTI and CamVid) input images. For parallel and distributed processing, the study uses CUDA 10.2 and `horovod` 0.19. Deep learning packages such as `tensorflow` 2.1.0 and `keras` 2.3.1 are used to implement the model. For experiments measuring inference speed, we use TensorRT 6.0.1 and follow the common practice to convert a keras model to a TRT-based deployment model and use the TensorRT engine to run inference. We also measure the size of the deployment model as an indication of GPU memory requirement.

Motivated by [1, 2, 3], we select the polynomial decay strategy for setting the learning rate (LR) with the power being 0.9. Such a strategy requires the *initial* learning rate LR_i, which is an upper bound, and the *end* learning rate LR_e, which is a lower bound, and the actual learning rate is scheduled to decrease polynomially from LR_i to LR_e as training progresses as follows

$$LR = (LR_i - LR_e)(1 - epoch/steps)^{power} + LR_e.$$

To find the lower and upper bounds of LR, we use a common rate finder method [4]. Essentially, this algorithm changes the learning rate exponentially from very small to very large at each batch

[1]Corresponding Author. Email `tanmay.singha@postgrad.curtin.edu.au`

update within the first few epochs. A plot of the loss vs learning rate within this first few epochs is carefully examined. Such a plot is illustrated in Fig. 1. When LR is less than 10^{-4}, the loss curve is flat meaning that the network does not learn effectively. When LR is more than 10^{-1}, the loss explodes meaning that the network learns incorrectly. The steepest region from 10^{-4} to 10^{-1} is where the network learns most effectively and hence determines the upper and lower bounds. To minimize the model loss, the distributed synchronous stochastic gradient decent (SGD)

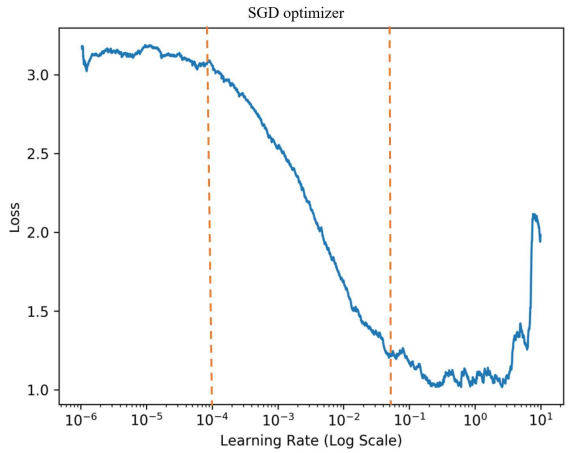

Figure 1: Loss vs learning rate curve for the first few epochs used in the rate finder algorithm.

optimizer is used. Due to the distributed `horovod` framework, gradient descent may face challenges to find the minimal at the early stage of of the training process. To overcome this problem, the study makes use of the gradual warm-up strategy [5]. It basically ramps up the LR in the first few epochs to the maximum value and then, LR starts converging as training progresses. As standard in segmentation, this study employs various data augmentation techniques, such as cropping, resizing, clipping by value, horizontal and vertical flipping, adjusting brightness, contrast, and saturation of the input to address the limited sample size problem. To handle model over-fitting issue, a ℓ_2 regularization and a dropout layer with a dropout rate of 0.35 are deployed.

References

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proc. ECCV, 2018, pp. 801–818.

[2] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proc. CVPR, 2017, pp. 2881–2890.

[3] R. P. Poudel, S. Liwicki, R. Cipolla, Fast-SCNN: fast semantic segmentation network, arXiv preprint arXiv:1902.04502.

[4] L. N. Smith, Cyclical learning rates for training neural networks, in: Proc. WACV, IEEE, 2017, pp. 464–472.

[5] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, K. He, Accurate, large minibatch SGD: Training ImageNet in 1 hour, arXiv preprint arXiv:1706.02677.