1 Deep Image: Scaling up Image Recognition

The latest attempt in image classification with an error 5.98% in ImageNet data set is reported by Ren Wu et al.[1] of Baidu research.They developed an end to end deep learning system named Deep Image. It uses a highly optimized parallel algorithm to implement large deep neural network with augmented input data. The network is trained using stochastic gradient decent algorithms (SGD)[ref] on a custom built high performance system comprised of 36 server nodes, each with 2 six-core Intel Xeon E5-2620 processors and 4 NVIDIA Tesla K40m GPUs . System uses an InfiniBand network for interconnections. Parallelism strategies used in their network are model-data parallelism and data parallelism. This methods have been proposed by Alex Krizhevsky [2] and Omry Yadan et al.[3] for training convolutional neural networks with SGD on a multiple GPU systems. But it is not easy extend the same strategies to multiple GPU cluster because of the communication overhead. So the Baidu Team focused on minimizing network data transfers and overlapping the computation. They uses butterfly synchronization and lazy update strategies to achieve data parallelism in gradient computation. Their results shows model-data parallelism is better when number of GPUs is less than 16. Implementation of Data parallelism in large number of GPU cluster is better because of the constant communication requirements.

The authors have explored different data augmentation techniques to increase the number of labeled images in the training set. This includes color casting, Vignetting , Lens distortion , Rotation , Flipping and Cropping. Instead of using the same resolution on all images, they have trained separate models at different scales, combined results by averaging softmax class posteriors. Data set used in this experiment was subset of ImageNet data set , used in the competition ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)[4]. This data set includes 1.2 million images which contains 1,000 categories.

Major contribution of this work is the demonstration of tremendous computational power to achieve high accuracy in image classification. It also shows , augmented multi-scale images can be combined to achieve less error rate in convolutional network in the context of the image classification .

2 VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan and Andrew Zisserman [5] evaluated the effect of network depth in image classification using very small convolution filters. Their deep network architecture comprise of fixed size input layers , a stack of convolution layers , three Fully-Connected (FC) layers and 5 max-pooling layers for spatial spooling over a 2 2 pixel window with stride 2. Hidden layers are modeled using Rectified Linear Units(ReLU)[6]. On the hardware side , it uses a multi-GPU system with NVIDIA Titan Black GPUs. Netwok is trained using multinomial logistic regression based on back-propagation with momentum of 0.9 and batch size 256.

In this work authors formed a conclusion that greater depth with small convolution filters and preinitialization of certain layers will cause the learning process to converge in less number of epochs. This model of convolution network does not differ from the classical architecture proposed by LeCun et al.[7]. But the authors reported a significant improvement in the performance using an increased depth. This implementation results in an error rate of 6.8% in ILSVRC 2014 of ImageNet.

3 REFERENCES

References

[1] Ren Wu, Shengen Yan, Yi Shan, Qingqing Dang, and Gang Sun, "Deep Image: Scaling up Image Recognition," *arXiv Prepr. arXiv1501.02876*, 2015.

[2] Alex Krizhevsky, "One weird trick for parallelizing convolutional neural networks," pp. 1–7, 2014.

[3] O Yadan and Keith Adams, "Multi-GPU Training of ConvNets," *arXiv*, pp. 10–13, 2013.

[4] Alex Berg and J Deng, "Imagenet large scale visual recognition challenge 2010," *Challenge*, 2010.

[5] Karen Simonyan and Andrew Zisserman, "V d c n l -s i r," pp. 1–13, 2015.

[6] Vinod Nair and Geoffrey E Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, , no. 3, pp. 807–814, 2010.

[7] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, pp. 2278–2323, 1998.