Sparsity Invariant CNNs

Group 66: Taoyue Wang, Shipeng Liu, Zhewen Gao, Mingkang Tang

Introduction

Traditional CNNs lead to suboptimal results on sparse datas where only a low proportion of input pixels carry valid information. The SparseConvNet[1] solve the problem by introducing a sparse convolutional layer that takes additionally a mask indicating the validity of each pixel as an input. The mask makes it possible for the layer to average only valid pixels covered by the kernel in the forward pass, and is maintained across layers with max-pooling.

$$f_{u,v}(\boldsymbol{x}, \boldsymbol{o}) = \frac{\sum_{i,j=-k}^{k} o_{u+i,v+j}\, x_{u+i,v+j}\, w_{i,j}}{\sum_{i,j=-k}^{k} o_{u+i,v+j} + \epsilon} + b$$

$$f^{o}_{u,v}(\boldsymbol{o}) = \max_{i,j=-k,..,k} o_{u+i,v+j}$$

Datasets

The experiments are conducted on:

- **MNIST**: digits classification dataset of 28 x 28 grayscale images of handwritten digits 0 - 9
- **Caltech-101**: contains pictures of objects belonging to 101 categories, each with 40 to 800 images sized roughly 300 x 200 pixels

Experiments

We modify existing codes[2] on the KITTI dataset to test the performance on MNIST and Caltech-101. Input images with different sparsity levels are generated. We adjust the structure of the original network to be compatible with new datasets.

For MNIST, we truncate several convolutional layers and add linear layers. We train 20 epochs with learning rate 0.005. The assessment metric is the classification accuracy for test set.

Sparsity	50%	60%	70%	80%	90%
Sparse ConvNet	94.59%	94.39%	92.77%	86.09%	76.23%
ConvNet	95.07%	94.17%	90.43%	85.32%	69.27%

For Caltech-101, we add batch normalization and pooling layers for the sparse convolution block, and dropout for linear layers.

Sparsity	5%	10%	20%	30%	40%	50%	60%	70%
Sparse ConvNet	62.80%	61.98%	61.29%	62.68%	60.65%	60.94%	59.32%	59.26
ConvNet	43.93%	41.90%	40.34%	39.18%	38.83%	38.07%	38.60%	37.84%

Conclusions

For MNIST, based on the intrinsic properties of the dataset, high sparsity levels were mainly tested. The model with sparse convolution has better performance and generalization ability than model with ordinary convolution.

For Caltech-101, similar to MNIST, a sparse convolution based model than ordinary convolution based model is more robust to changes of sparsity level and results higher accuracy.

[1] Jonas Uhrig et.al., "Sparsity Invariant CNNs." arXiv:1708.06500 (2017)
[2] https://github.com/chenxiaoyu523/Sparsity-Invariant-CNNs-pytorch