GoogLeNet

Paper review:
“Going Deeper with Convolutions”
Christian S. et al

2015-6-30

uoguelph-mlrg
He Ma

Outline

- Related works
- Main contribution
- Some Ideas
- Implementation
- Strength
- Limitation
- Future work

Related works

- The Inception model is inspired by a neuroscience model of the primate visual cortex [Serre et al.,2007].
- Apply Network-in-Network [Lin et al.,2013] for reducing dimension.
- Apply Regions with Convolutional Neural Network [Girshick et al.,2014] in detection.
- Compare with AlexNet [Krizhevsky et al.,2012] in terms of parameter amount and accuracy.

1. Whose author claim that human visual processing is hierarchical, aiming to build invariance to position and scale first and then to viewpoint and other transformations.
2. This is heavily used in the inception model to constrain input width before doing the actual convolution to avoid model size growing too large
3. This is used as the basic structure of their model for the object detection task.
4. To show the advantage of the model.

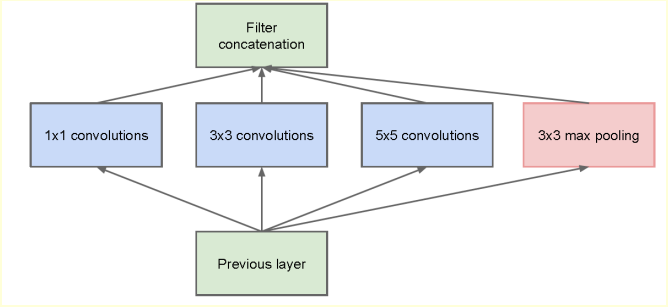
Main contribution

Uniformly increasing network size has two drawbacks:

- More prone to overfitting and a lot computation wasted.
- We need to cluster neurons with high correlated outputs.
- A way of approximating expected sparse matrix is by dense building blocks.

Main contribution

- Learning an Inception is a way to automatically constructing non-uniform network.
- Allowing abstracting features from different scales simultaneously

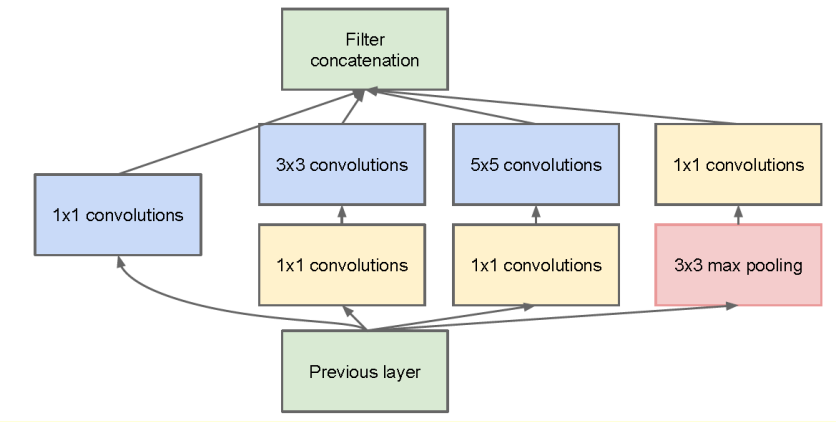


Limitation:

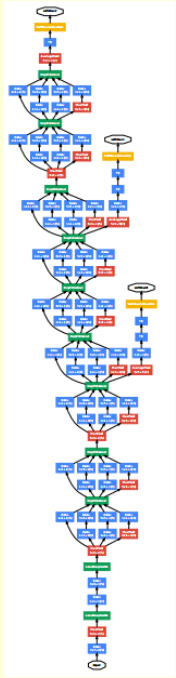
Limitation: Output size is increasing layer by layer. Too computationally expensive if feed the previous layer with large number of filters directly to those convolution layer. We need to reduce the incoming number of filters.
So we can not stack too many of this inception layer in a network.

Main contribution

- Judiciously applying dimension reduction and projection



This is a way to control the incoming filter number and, as a result, to control computation inside a inception model



Some Ideas

- Auxiliary classifiers
- Training policy:
  asynchronous sgd
  fixed learning rate schedule
  poly works better

Auxiliary classifier: encourage discrimination in the lower stages in the classifier, increase the gradient signal that gets propagated back, and provide additional regularization

Since “as features of higher abstraction are captured by higher layers, their spatial concentration is expected to decrease”
Author suggests the ratio of 3x3 and 5x5 outputs among those concatenated output filters should increase as moving to higher layers

Some Ideas

Number of models	Number of Crops	Cost	Top-5 error	compared to base
1	1	1	10.07%	base
1	10	10	9.15%	-0.92%
1	144	144	7.89%	-2.18%
7	1	7	8.09%	-1.98%
7	10	70	7.62%	-2.45%
7	144	1008	6.67%	-3.45%

- Ensemble in testing
- Aggressive cropping in training and testing

Ensemble: 7 independent models

The ensemble also gives better detection result.

Implementation

- DistBelief [Jeffrey D. et al. ,2012] distributed machine learning system
- Modest amount of model and data-parallelism.
- using few high-end GPUs within a week

- 64s per 5120 images using caffe cudnn on K40c

Didn't mention how long is each epoch or every 5120 images. How much GPU memory needed in training.

100-120s on theano using cudnn

Strength

- Comparing with AlexNet
- Fewer parameters in the model
  51M vs. 232M
- Low memory footprint ?

- Higher accuracy.
- Single model 10.6% vs. 19%
- Ensemble 6.7%

10.6%: aggressive cropping used in training.
19%: random cropping
16%: data augmentation

Limitation

- Using Inception model only at higher layers. Infrastructure inefficiency.
- Backpropagate gradients was a concern
- Computational budget.

AlexNet cuDNN Titan X	GoogLeNet cuDNN K40c
231ms = 70ms+161ms	1685ms = 562ms+1123ms

Computation will be high if inception work with larger images than 28x28.

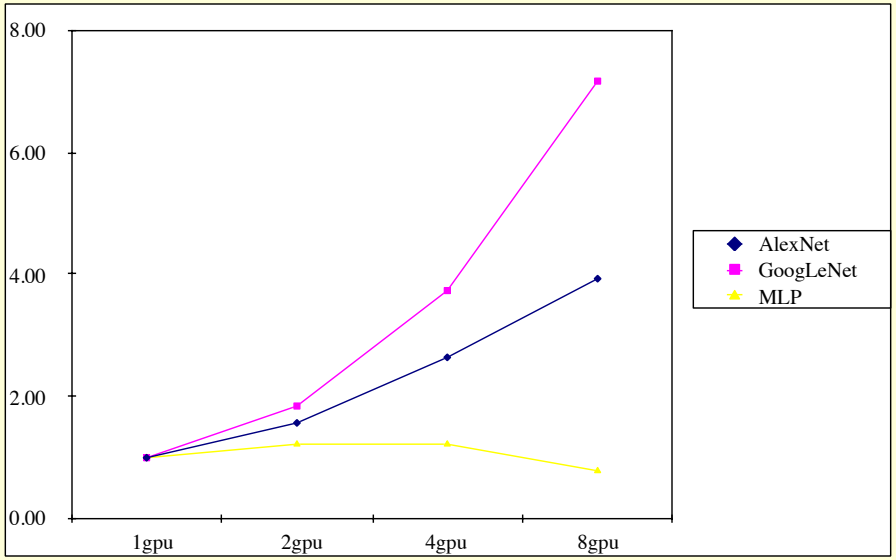
I think the infrastructure inefficiency also means in the inception layer the computation of those internal layers may not be well parallelized, 1x1conv is faster than the rest. Or even their results are computed in series then that will be way slower.

BP: because the intermediate layers need to improve their discriminating ability.

Future work

- The author suggests future work towards creating sparser and more refined structures in automated ways similar to using inception.

MPI speedup


8.00
6.00
4.00
2.00
0.00
1gpu
2gpu
4gpu
8gpu
AlexNet
GoogLeNet
MLP