Network In Network

Min Lin[1,2], Qiang Chen[2], Shuicheng Yan[2]
[1]Graduate School for Integrative Sciences and Engineering
[2]Department of Electronic & Computer Engineering
National University of Singapore, Singapore
{linmin,chenqiang,eleyans}@nus.edu.sg

Deep learning journal Club

2018-04-19

Conventional CNN

- Linear convolution layers
- fully connected layers for classification

NIN

- MLP convolution layers
- Global average pooling

LeNet-5

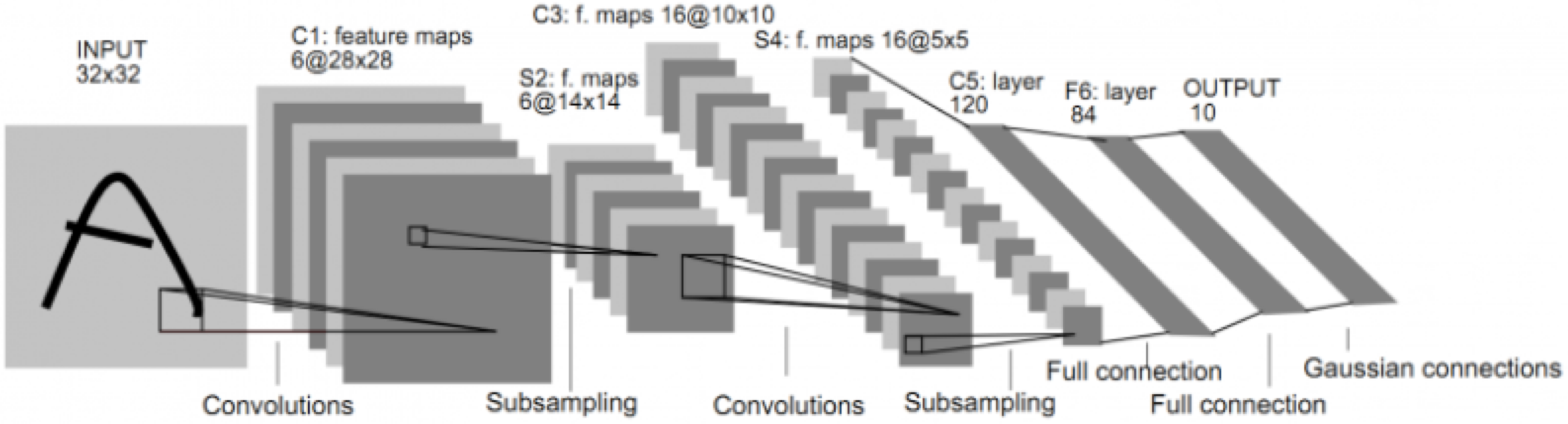


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Network In Network

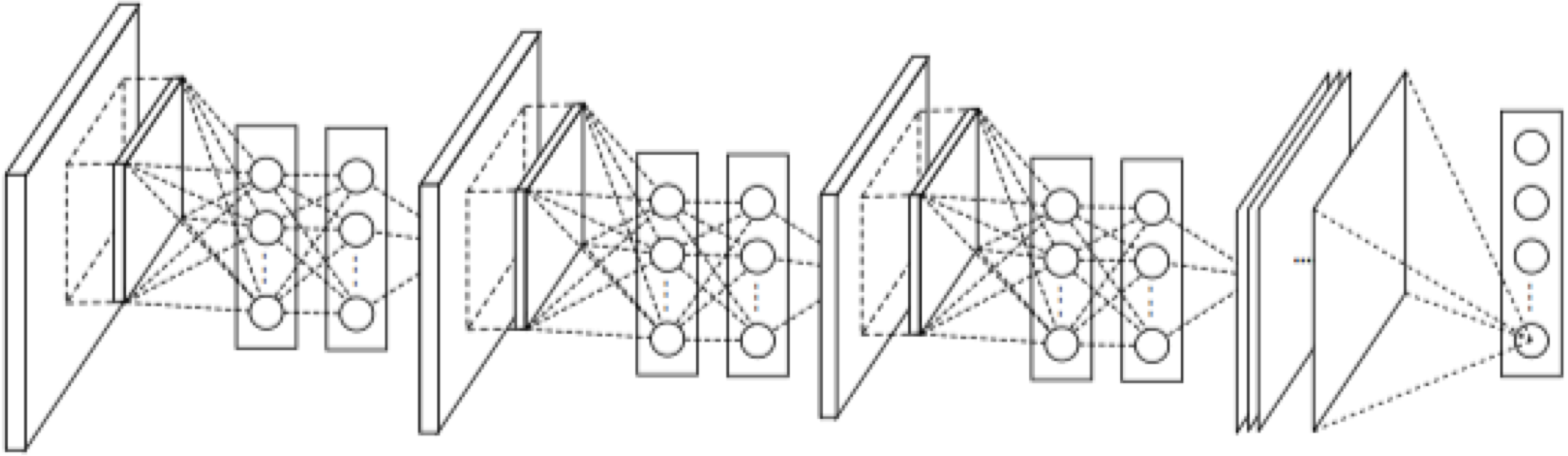
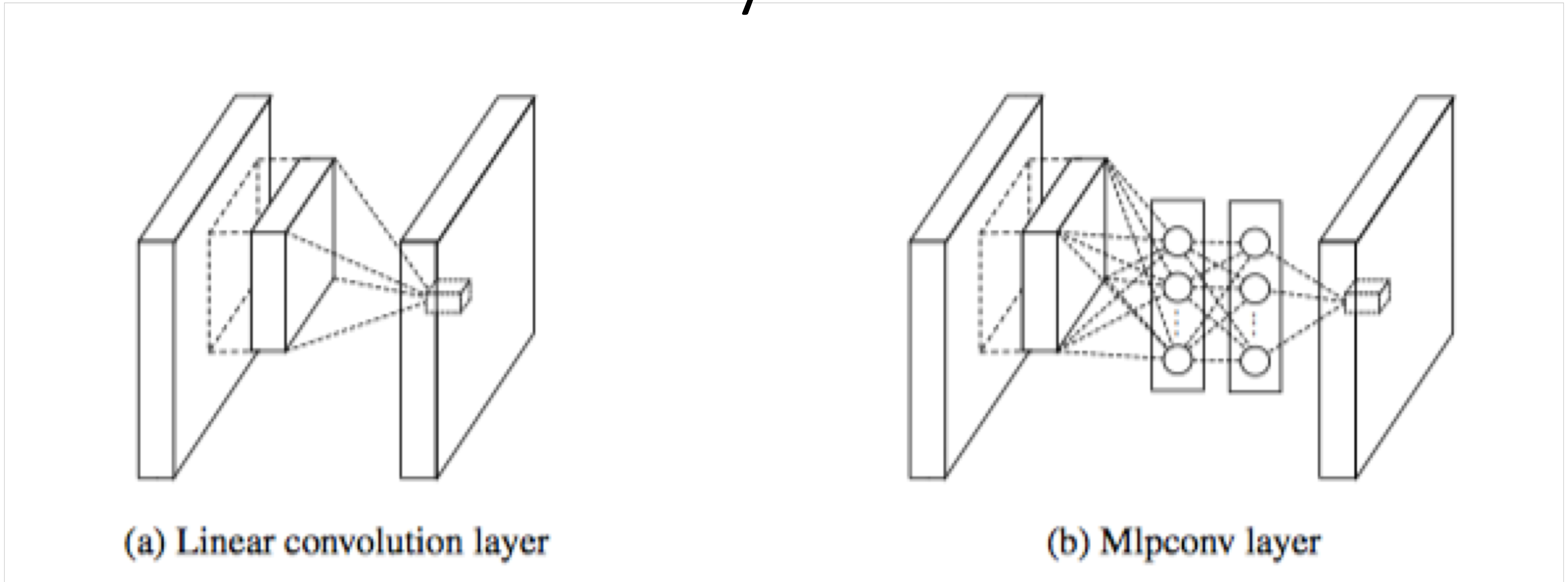


Figure 2: The overall structure of Network In Network. In this paper the NINs include the stacking of three mlpconv layers and one global average pooling layer.

- The convolution filter in CNN is a generalized linear model (GLM) for the underlying data patch, and we argue that the level of abstraction is low with GLM
- Conventional CNN implicitly makes the assumption that the latent concepts are linearly separable

- Replacing the GLM with a more potent nonlinear function approximator can enhance the abstraction ability of the local model
- In NIN, the GLM is replaced with a ”micro network” structure which is a general nonlinear function approximator

1 MLP Convolution Layers



The mlpconv maps the input local patch to the output feature vector with a multilayer perceptron (MLP) consisting of multiple fully connected layers with nonlinear activation functions

$$
\begin{aligned}
f_{i,j,k_1}^1 &= \max({w_{k_1}^1}^T x_{i,j} + b_{k_1}, 0). \\
&\vdots \\
f_{i,j,k_n}^n &= \max({w_{k_n}^n}^T f_{i,j}^{n-1} + b_{k_n}, 0).
\end{aligned} \tag{2}
$$

Here n is the number of layers in the multilayer perceptron. Rectified linear unit is used as the activation function in the multilayer perceptron.

- multilayer perceptron is compatible with the structure of convolutional neural networks, which is trained using back-propagation
- multilayer perceptron can be a deep model itself, which is consistent with the spirit of feature re-use

2 Global Average Pooling

- generate one feature map for each corresponding category of the classification task in the last mlpconv layer
- the resulting vector is fed directly into the softmax layer
- advantages
  - more native to the convolution structure
  - overfitting is avoided at this layer
  - more robust to spatial translations of the input

Result

Table 1: Test set error rates for CIFAR-10 of various methods.

Method	Test Error
Stochastic Pooling [11]	15.13%
CNN + Spearmint [14]	14.98%
Conv. maxout + Dropout [8]	11.68%
NIN + Dropout	**10.41%**
CNN + Spearmint + Data Augmentation [14]	9.50%
Conv. maxout + Dropout + Data Augmentation [8]	9.38%
DropConnect + 12 networks + Data Augmentation [15]	9.32%
NIN + Dropout + Data Augmentation	**8.81%**

Table 2: Test set error rates for CIFAR-100 of various methods.

Method	Test Error
Learned Pooling [16]	43.71%
Stochastic Pooling [11]	42.51%
Conv. maxout + Dropout [8]	38.57%
Tree based priors [17]	36.85%
NIN + Dropout	**35.68%**

Result

Table 3: Test set error rates for SVHN of various methods.

Method	Test Error
Stochastic Pooling [11]	2.80%
Rectifier + Dropout [18]	2.78%
Rectifier + Dropout + Synthetic Translation [18]	2.68%
Conv. maxout + Dropout [8]	2.47%
NIN + Dropout	2.35%
Multi-digit Number Recognition [19]	2.16%
DropConnect [15]	**1.94%**

Table 4: Test set error rates for MNIST of various methods.

Method	Test Error
2-Layer CNN + 2-Layer NN [11]	0.53%
Stochastic Pooling [11]	0.47%
NIN + Dropout	0.47%
Conv. maxout + Dropout [8]	**0.45%**

Table 5: Global average pooling compared to fully connected layer.

Method	Testing Error
mlpconv + Fully Connected	11.59%
mlpconv + Fully Connected + Dropout	10.88%
mlpconv + Global Average Pooling	10.41%