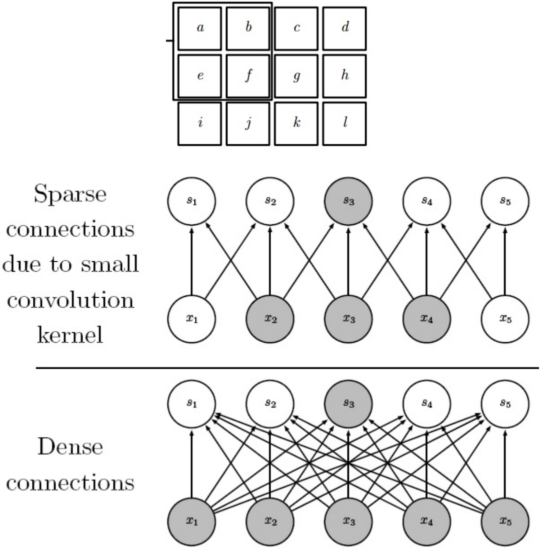
Convolutional Neural Net

COMP4211

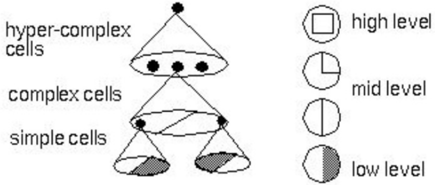
THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Sparse Connectivity

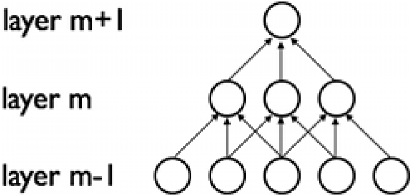


- receptive field

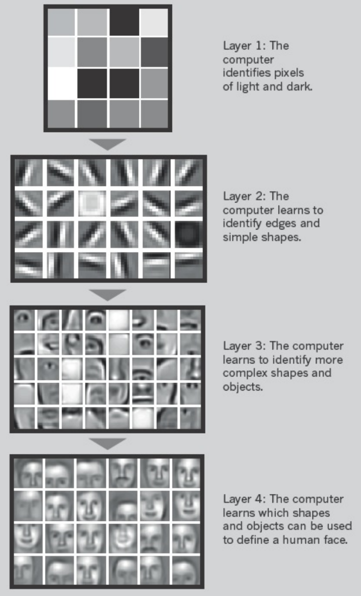
Feature Hierarchy



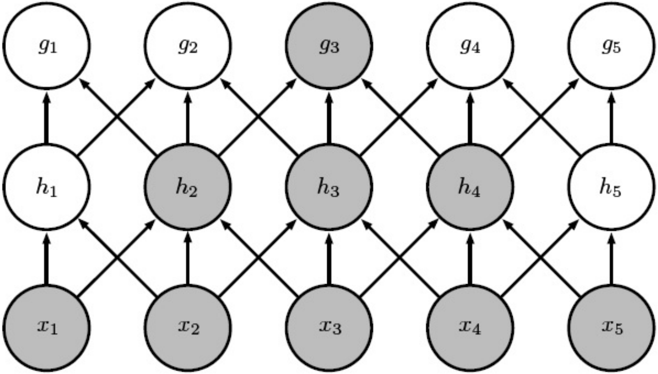
- hidden units are connected to a local subset of units in the previous layer



Example: Face Recognition



Growing Receptive Field



Shared Weights

- each local receptive field is replicated across the entire image

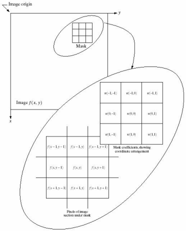
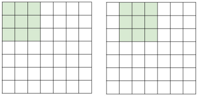
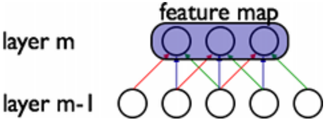


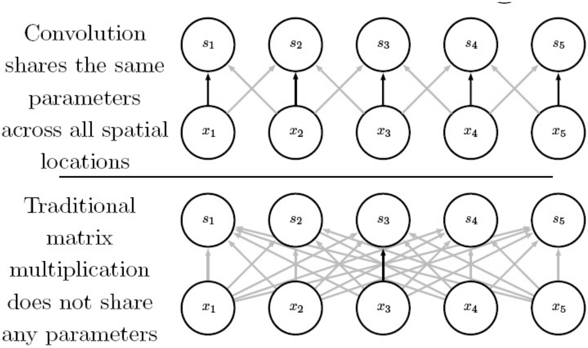
FIGURE 3.32 The mechanics of spatial filtering. The magnified drawing shows a 3 × 3 mask and the image section directly under it; the image section is shown displaced out from under the mask for ease of readability.



- weights of the same color are shared (constrained to be identical)



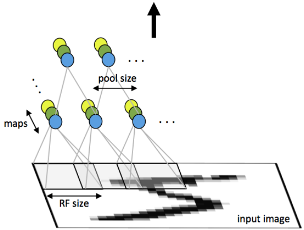
Parameter Sharing



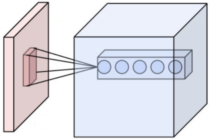
- allows for features to be detected regardless of their position in the image
  - robustness to shifts of the input

Convolutional Layer

- multiple feature maps look at the same region of the input



- stack the activation maps for all filters along the depth dimension



- 1×1 convolution
  - perform convolution without looking at neighboring pixels
  - dimension reduction

Efficiency of Convolution

- parameter sharing greatly reduces the number of free parameters to learn

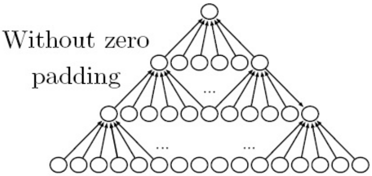
Input size: 320 by 280
Kernel size: 2 by 1
Output size: 319 by 280

	Convolution	Dense matrix	Sparse matrix
Stored floats	2	319*280*320*280 $>$ 8e9	2*319*280 = 178,640

Nonlinearity

- Convolution is a linear operation
- need nonlinearity
  - otherwise 2 convolution layers would be no more powerful than 1
- common to apply a rectified linear unit (ReLU): $y = \max(z, 0)$

Zero-Padding



- representation shrink at each layer
- limits the number of layers

Zero-padding

- adding zeros to each layer
- allows the use of an arbitrarily deep convolutional network

Pooling Layer

motivation

once a feature has been detected, only its approximate position relative to other features is relevant

Example

the input image contains

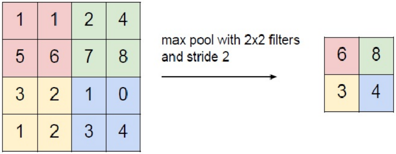
1. the endpoint of a roughly horizontal segment in the upper left area
2. a corner in the upper right area
3. the endpoint of a roughly vertical segment in the lower portion

the input image is a seven

- positions are likely to vary for different instances of the character
- spatial invariance

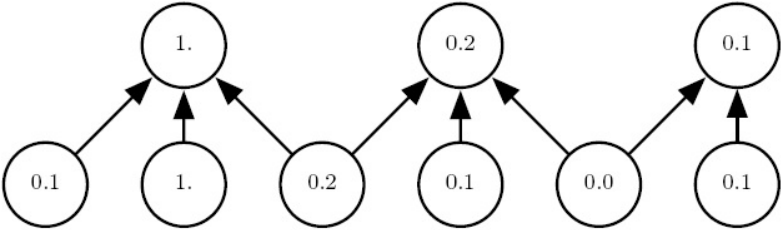
Pooling

- max-pooling
  - for each such sub-region (e.g., over a 2×2 area in the previous layer), outputs the maximum value



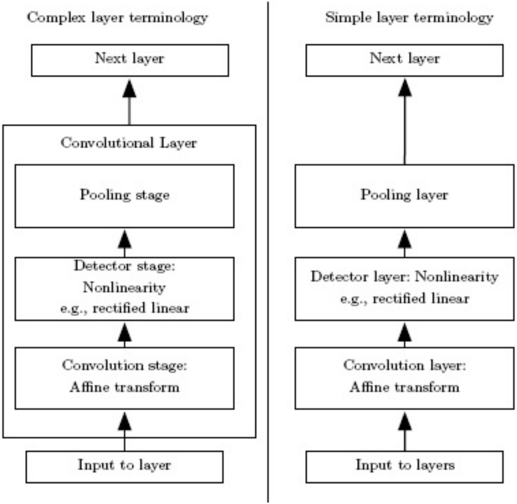
- can also have average pooling

Pooling with Downsampling

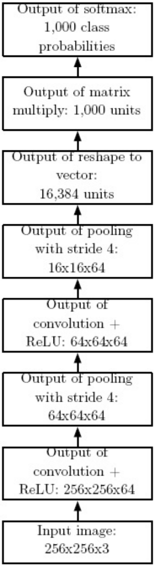


- stride of two
- reduces the representation size by a factor of two
- reduces the computational and statistical burden on the next layer

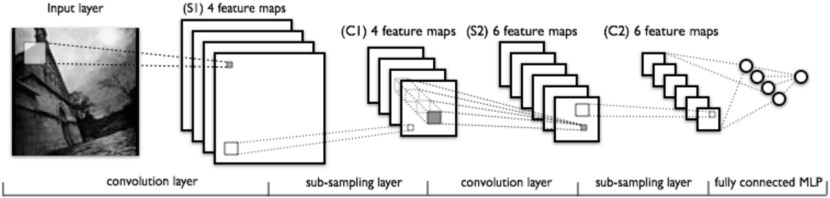
Convolutional Network Components



Example Classification Architecture



Example



- lower-layers: alternating convolution and max-pooling layers
- fully-connected (traditional MLP)
- classification error