



Lecture Notes on  
Pattern Recognition

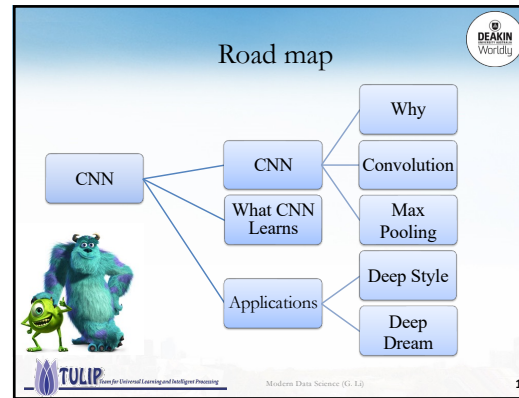
Module 09(C): Convolution Neural Network

Gang Li  
School of Information Technology  
Deakin University, VIC 3125, Australia


Modern Data Science (G. Li)

0



1

Convolution Neural Network



- Why CNN?
- Convolution
- Max Pooling

TULIP

Modern Data Science (G. Li)

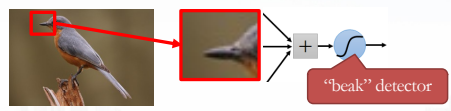
2

Why CNN for Image (1)

- Some patterns are much smaller than the whole image

A neuron does not have to see the whole image to discover the pattern.

Connecting to small region with less parameters



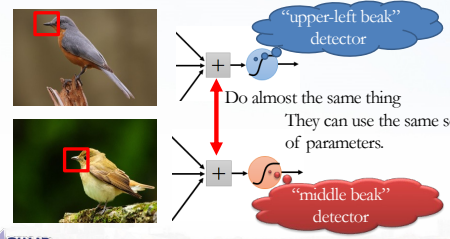
TULIP

Modern Data Science (G. Li)

3

Why CNN for Image (2)

- The same patterns appear in different regions.



TULIP


Modern Data Science (G. Li)

4

Why CNN for Image (3)

- Subsampling the pixels will not change the object

bird



bird

subsampling

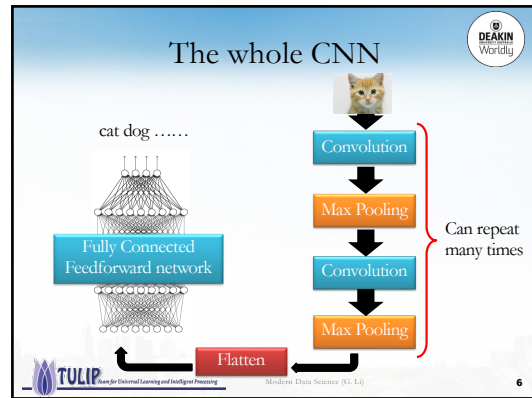
We can subsample the pixels to make image smaller

→ Less parameters for the network to process the image

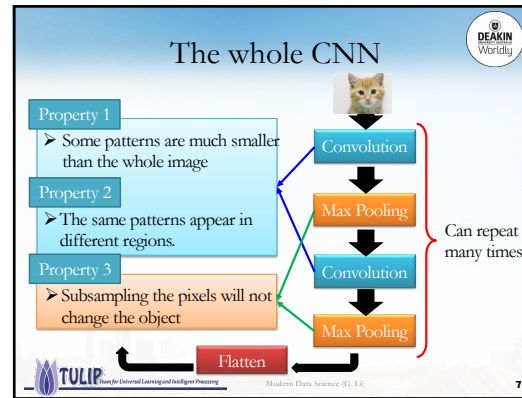
TULIP

Modern Data Science (G. Li)

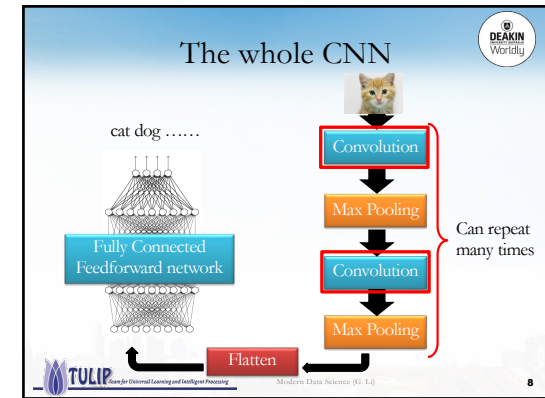
5



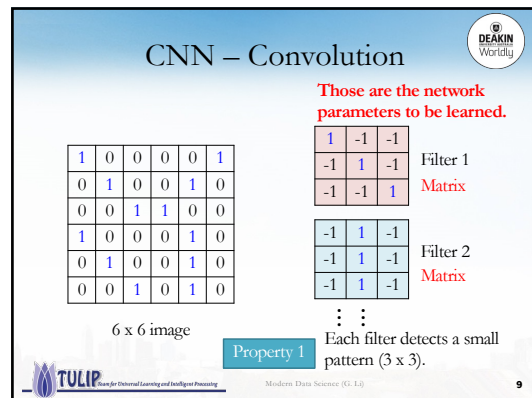
6



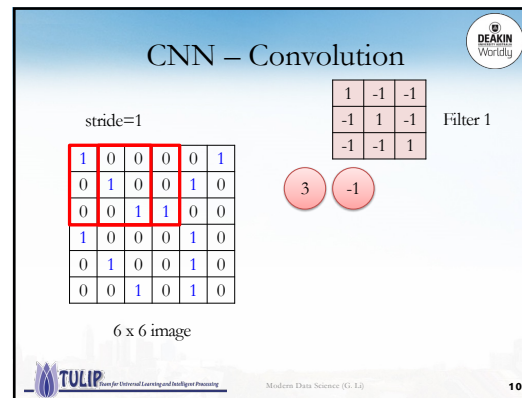
7



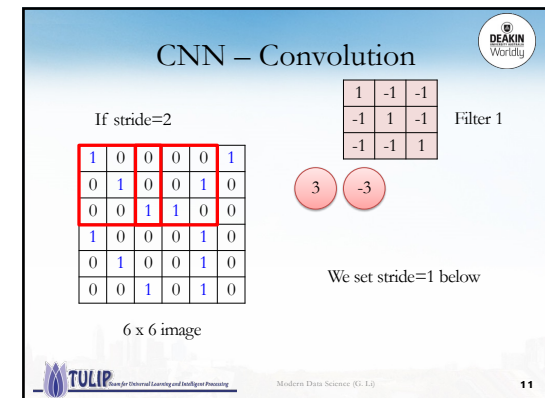
8



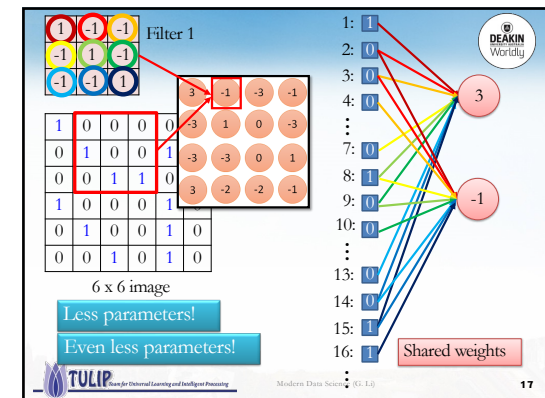
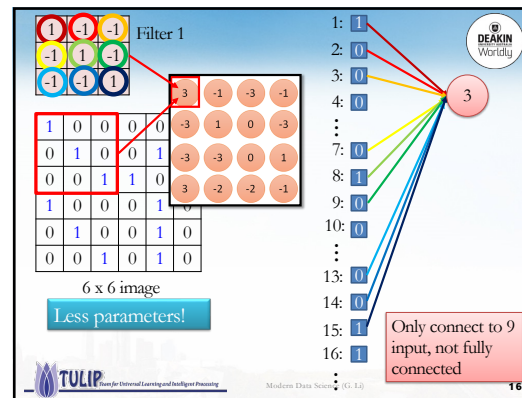
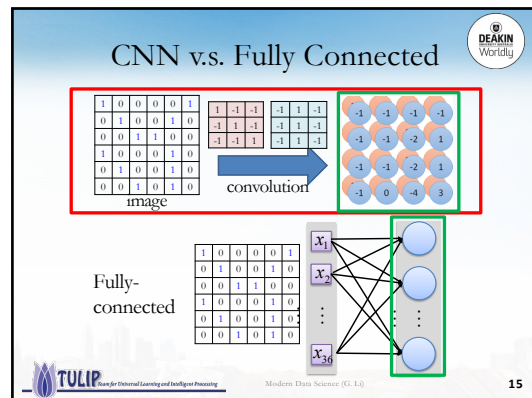
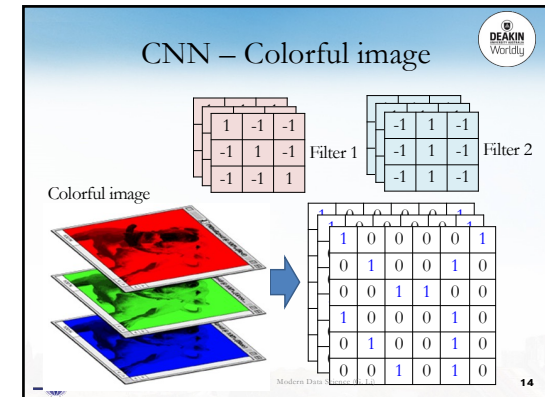
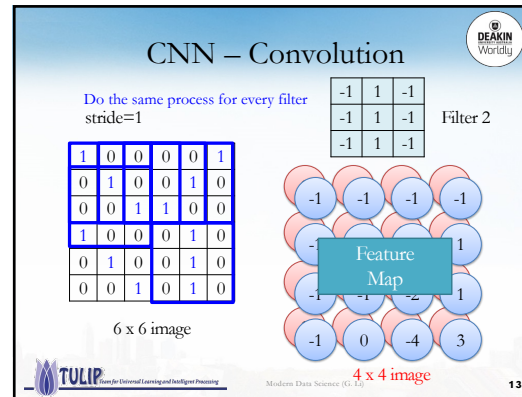
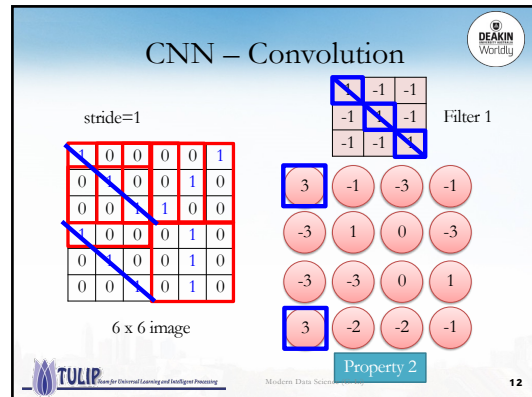
9

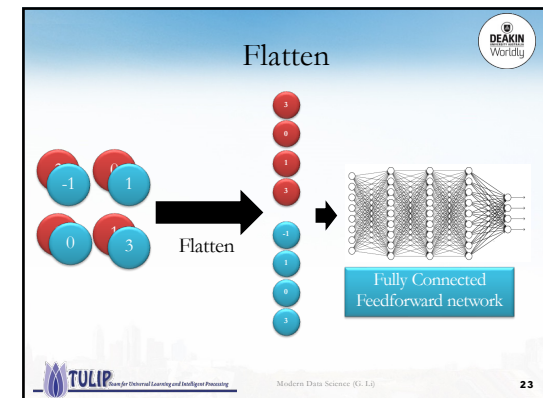
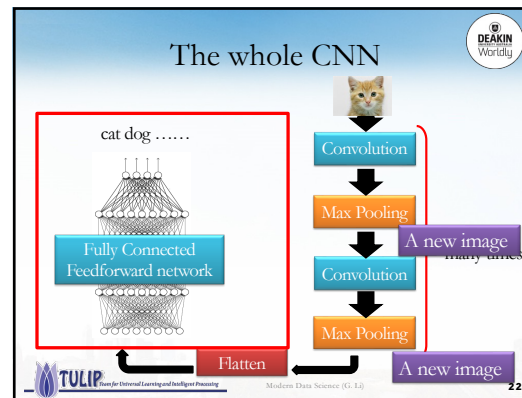
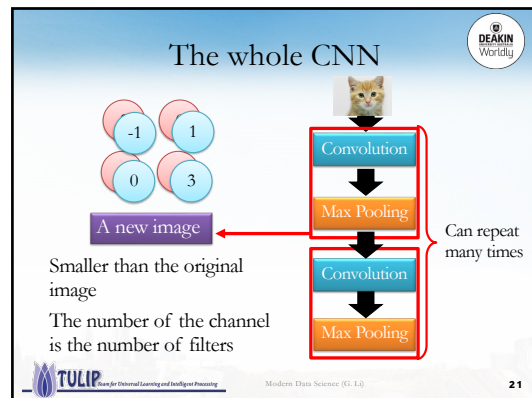
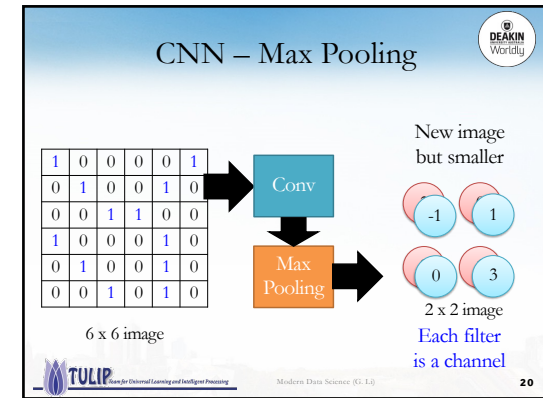
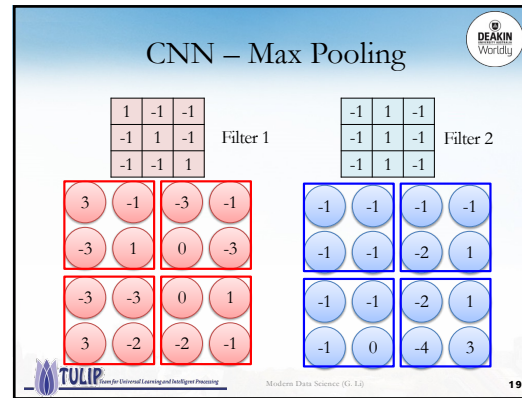
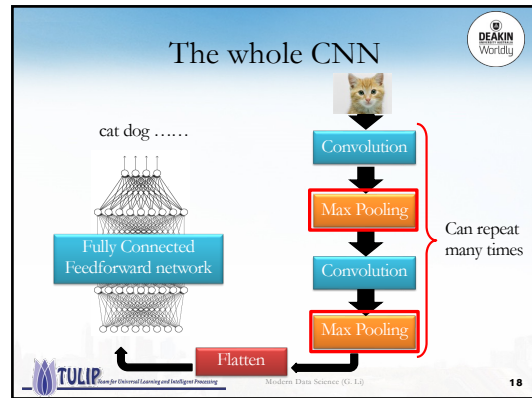


10



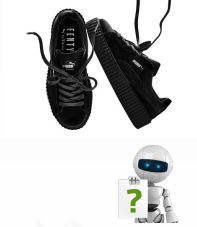
11





## What CNN Learns?

- What CNN Learns?

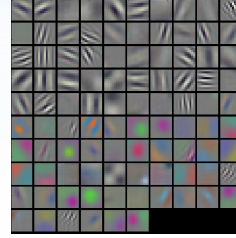


TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 24

24

## First Convolution Layer

- Typical-looking filters on the trained first layer



11 x 11  
(AlexNet)

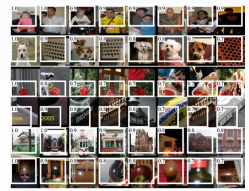
<http://cs231n.github.io/understanding-cnn/>

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 25

25

## How about higher layers?

- Which images make a specific neuron activate



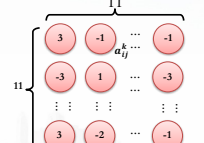
- Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR, 2014

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 26

26

## What does CNN learn?

The output of the k-th filter is a 11 x 11 matrix.  
Degree of the activation of the k-th filter:  
 $x^* = \arg \max_x a^k$  (gradient ascent)

$$a^k = \sum_{i=1}^{11} \sum_{j=1}^{11} a_{ij}^k$$


11

$\frac{\partial a^k}{\partial x_{ij}}$

25 3x3 filters

x input

Convolution

Max Pooling

50 3x3 filters

Convolution

Max Pooling

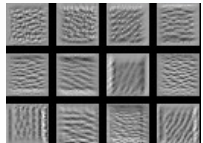
50 x 11 x 11

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 27

27

## What does CNN learn?

The output of the k-th filter is a 11 x 11 matrix.  
Degree of the activation of the k-th filter:  
 $x^* = \arg \max_x a^k$  (gradient ascent)

$$a^k = \sum_{i=1}^{11} \sum_{j=1}^{11} a_{ij}^k$$


11

$\frac{\partial a^k}{\partial x_{ij}}$

25 3x3 filters

x input

Convolution

Max Pooling

50 3x3 filters

Convolution

Max Pooling

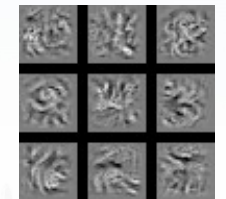
50 x 11 x 11

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 28

28

## What does CNN learn?

Find an image maximizing the output of neuron:  
 $x^* = \arg \max_x a^j$



Each figure corresponds to a neuron

input

Convolution

Max Pooling

Convolution

Max Pooling

Flatten

$a_j$

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 29

29



### What does CNN learn?

Over all pixel values

$$x^* = \arg \max_x y^i$$

$$x^* = \arg \max_x \left( y^i - \sum_{i,j} |x_{ij}| \right)$$

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

30

30

### What does CNN learn?

dumbbell cup dalmatian  
bell pepper lemon husky

- Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR, 2014

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

31

31

### What does CNN learn?

$\left| \frac{\partial y_k}{\partial x_{ij}} \right|$   $y_k$ : the predicted class of the model

Pixel  $x_{ij}$

- Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR, 2014

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

32

32

### What does CNN learn?

True Label: Pomeranian True Label: Car Wheel True Label: Afghan Hound

- Reference: Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In Computer Vision-ECCV 2014 (pp. 818-833)

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

33

33

### Applications

- Deep Dream
- Deep Style

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

34

34

### Deep Dream Generator

<https://deeplearninggenerator.com>

guillermo ponce

22 hours ago

TULIP: Tools for Universal Learning and Intelligent Processing

Modern Data Science (G. L.)

35

35

## Deep Style

- Given a photo, make its style like famous paintings

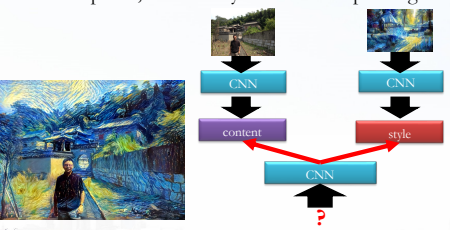


TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 36

36

## Deep Style

- Given a photo, make its style like famous paintings



TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 37

37

## Deep Style

- Given a photo, make its style like famous paintings

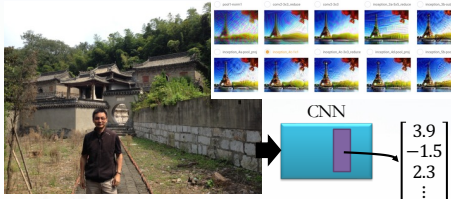


TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 38

38

## Deep Dream

- Given a photo, machine adds what it sees .....




TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 39

39

## Deep Dream

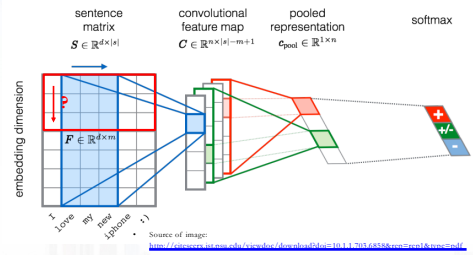
- Given a photo, machine adds what it sees .....



TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 40

40

## More Application: Text



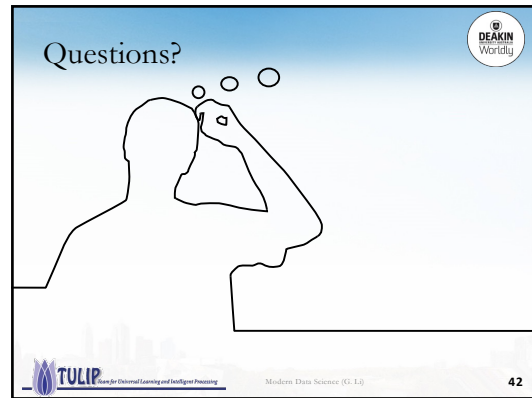
sentence matrix  $S \in \mathbb{R}^{d \times |x|}$   
 convolutional feature map  $C \in \mathbb{R}^{m \times |x| - m + 1}$   
 pooled representation  $C_{\text{pool}} \in \mathbb{R}^{1 \times n}$   
 softmax

embedding dimension  
 $F \in \mathbb{R}^{d \times m}$   
 Love my pet Lipstick

Source of image:  
<https://paperswithcode.com/sota/download?doi=10.1.1.703.6858&rep=1&view=code>

TULIP Team for Universal Learning and Intelligent Processing Modern Data Science (G. Li) 41

41



42