Application of Deep Learning

Tao Huang

中山大學 SUN YAT-SEN UNIVERSITY 数学学院 笃行工作室

Section 1

Neural Networks

Universal approximation theorem

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a nonconstant, bounded, and continuous function (called the activation function). Let I_m denote the m-dimensional unit hypercube $[0, 1]^m$. The space of real-valued continuous functions on I_m is denoted by $C(I_m)$. Then, given any $\varepsilon > 0$ and any function $f \in C(I_m)$, there exist an integer N, real constants $v_i, b_i \in \mathbb{R}$ and real vectors $w_i \in \mathbb{R}^m$ for $i = 1, \dots, N$, such that we may define:

$$F(x) = \sum_{i=1}^{N} v_i \varphi \left(w_i^T x + b_i \right) \tag{1}$$

as an approximate realization of the function f; that is,

$$|F(x) - f(x)| < \varepsilon \tag{2}$$

for all $x \in I_m$.

Classification and Regression

中山大學 SUN YAT-SEN UNIVERSITY 数學學院 篤行工作室

Classification is about predicting a **label** and regression is about predicting a **quantity**.

Section 2

Computer Vision

Image Classification I

中山大學 SUN YAT-SEN UNIVERSITY 数学学院 笃行工作室

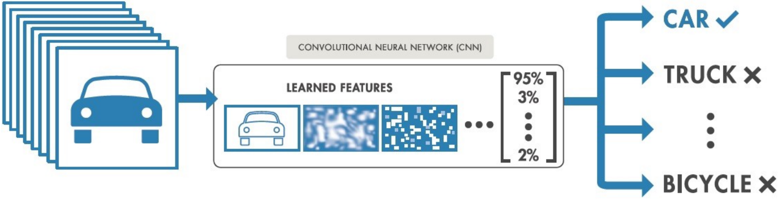


Figure 1

Source : MathWorks `https://goo.gl/zondfq`

Image Classification II

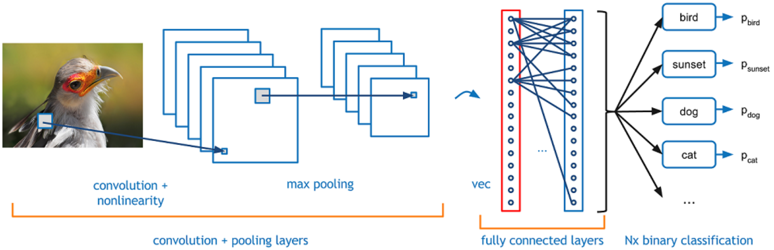


Figure 2: Convolutional neural network

Image Classification III

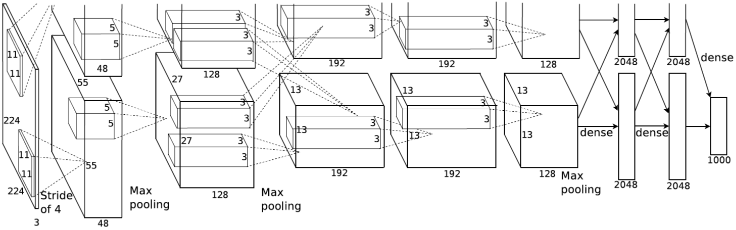


Figure 3: AlexNet

Image Classification IV

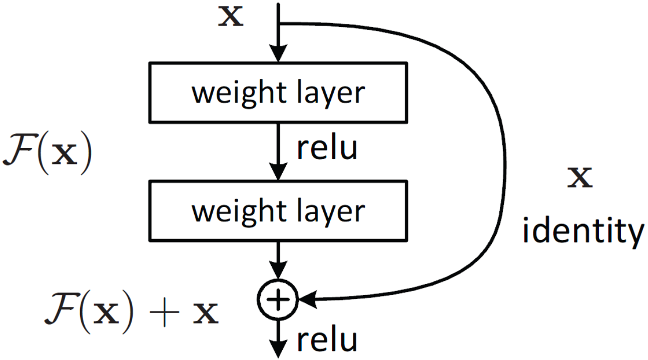


Figure 4: ResNet

Image Classification V

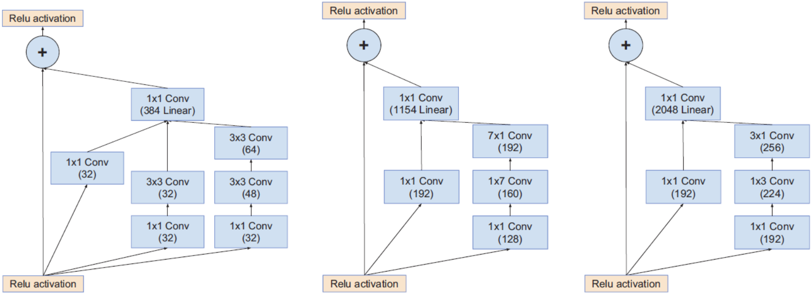


Figure 5: Inception-ResNet-v2

Image Classification VI

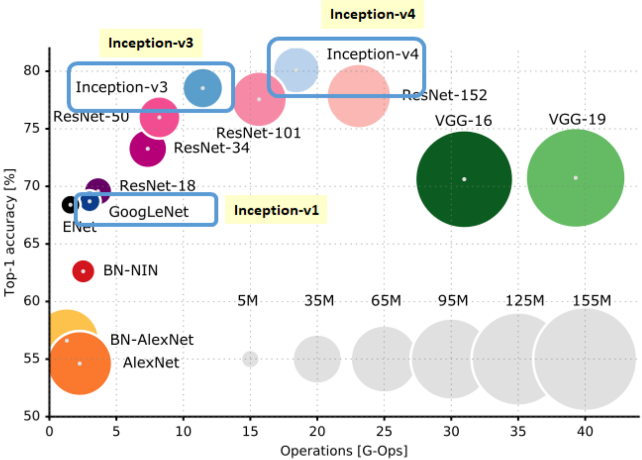


Figure 6: Accuracy against Number of Operations

Object Detection I

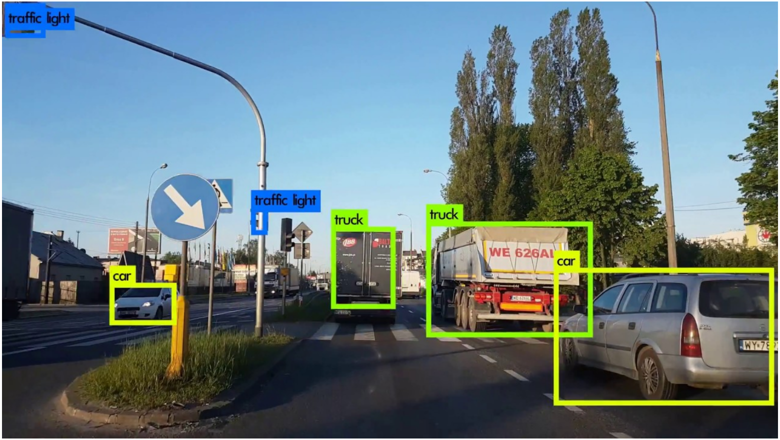


Figure 7

Object Detection II

- Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN)
- Single Shot MultiBox Detector (SSD)
- You Only Look Once (YOLO)

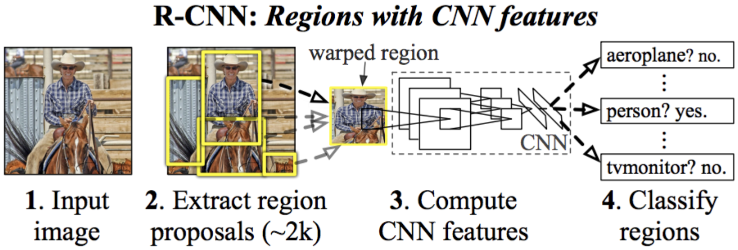


Figure 8: R-CNN

Object Detection III

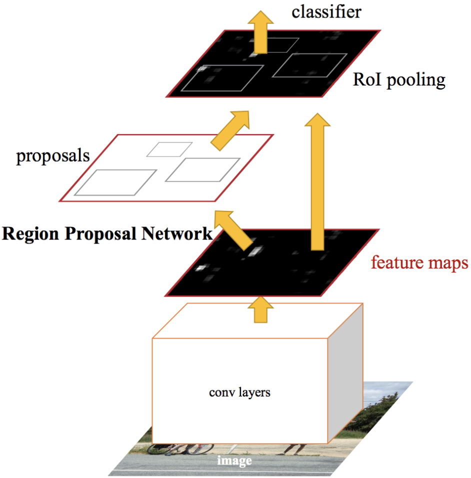


Figure 9: Faster R-CNN

Object Detection IV

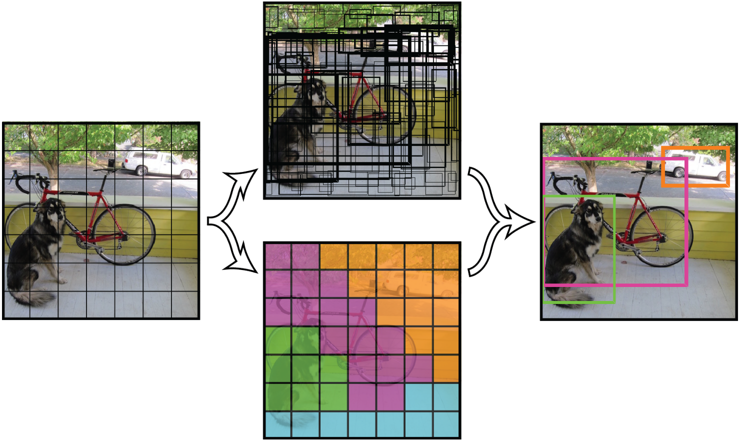


Figure 10: You Only Look Once (YOLO)

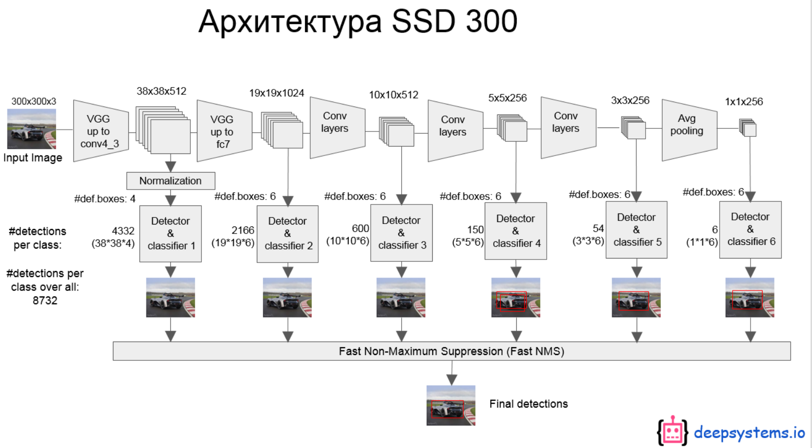


Figure 11: Single Shot MultiBox Detector (SSD)

Semantic Segmentation I

Figure 12

Semantic Segmentation II

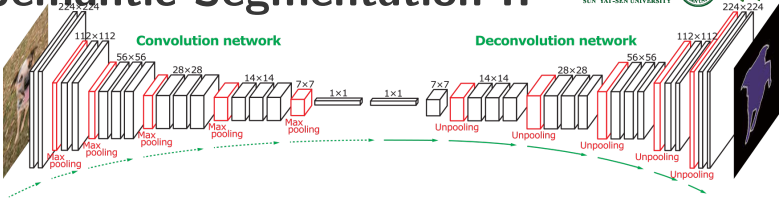


Figure 13: Deconvolution Network

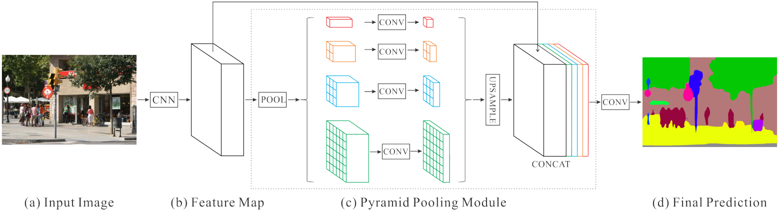


Figure 14: Pyramid Scene Parsing Network

Style Transfer

Figure 15

Image Reconstruction

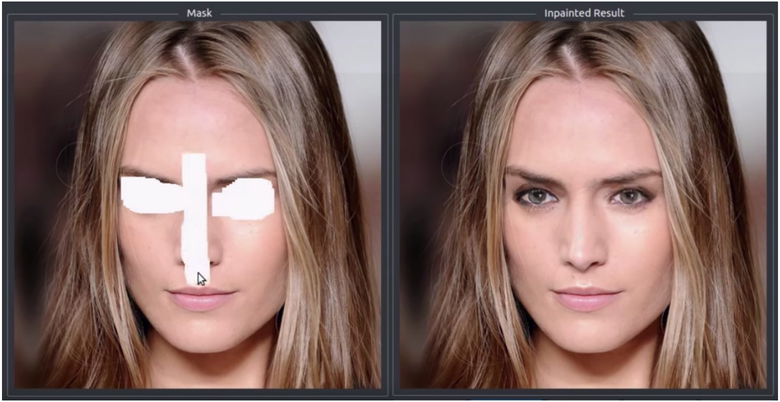


Figure 16

Image Colorization I

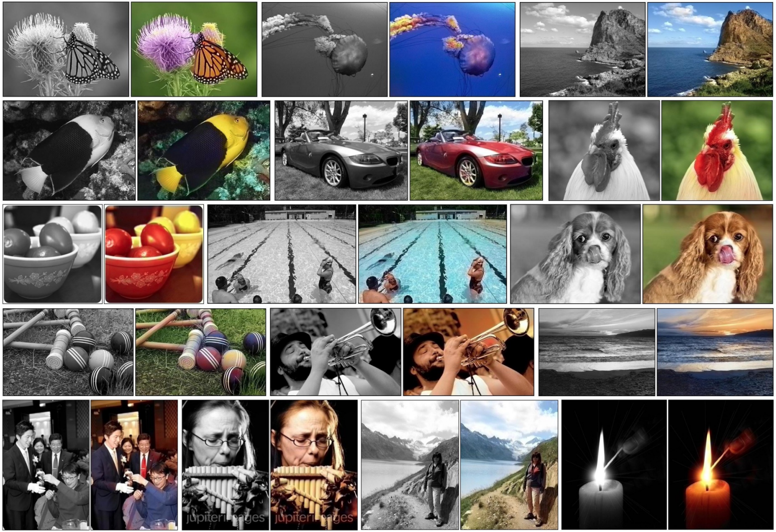


Image Colorization II

Figure 17

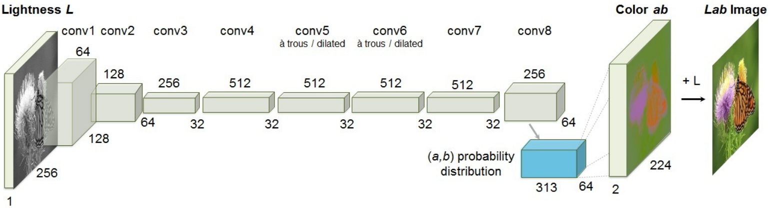


Figure 18

[2]

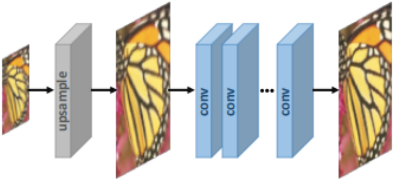
Image Super-Resolution I

bicubic (21.59dB/0.6423)	SRResNet (23.53dB/0.7832)	SRGAN (21.15dB/0.6868)	original

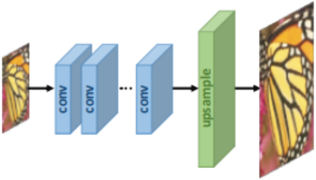
Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

Figure 19

Image Super-Resolution II

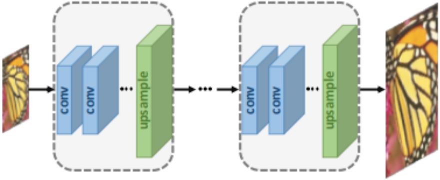


(a) Pre-upsampling SR

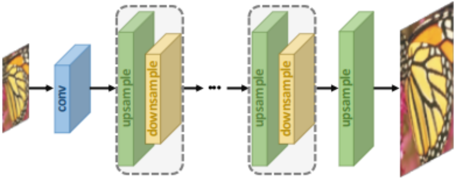


(b) Post-upsampling SR

Image Super-Resolution III



(c) Progressive upsampling SR



(d) Iterative up-and-down Sampling SR

Section 3

Natural Language Processing[1]

Word Embeddings I

中山大學 SUN YAT-SEN UNIVERSITY 数學學院 篤行工作室

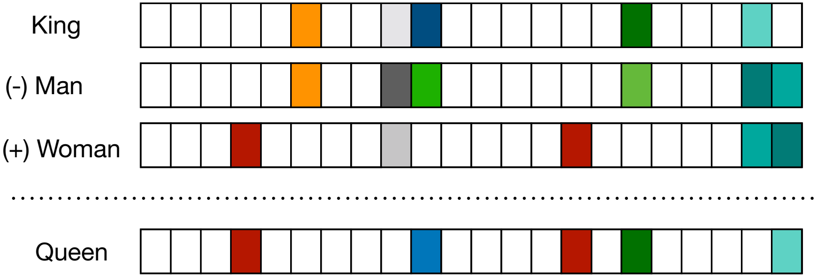


Figure 20: Distributional vectors represented by a **D**-dimensional vector where **D** $<<$ **V**, where **V** is size of Vocabulary.

Word Embeddings II

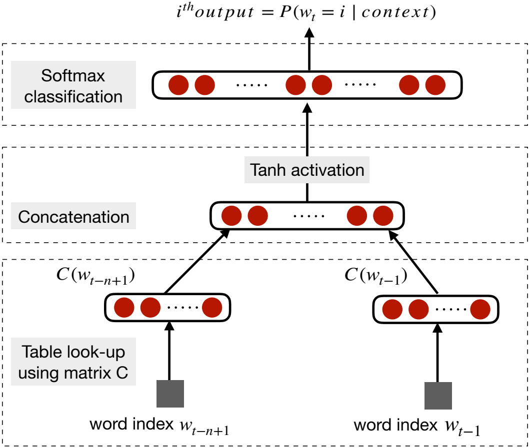
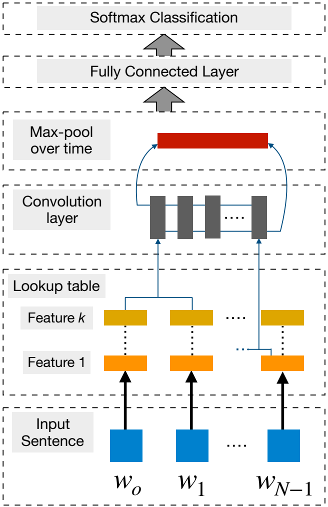


Figure 21: Neural Language Model. $C(i)$ is the i^{th} word embedding.

Convolutional Neural Networks



Recurrent Neural Networks

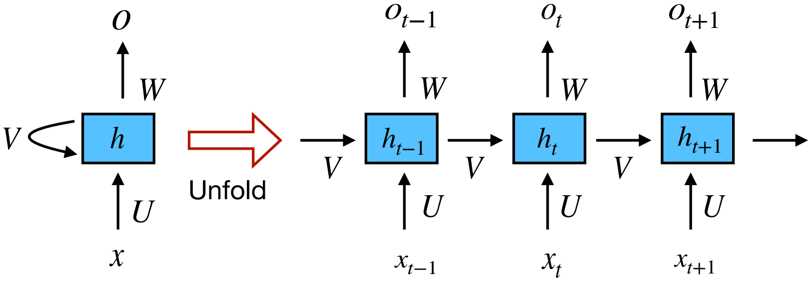


Figure 22: Simple RNN

Long Short-Term Memory

中山大學 SUN YAT-SEN UNIVERSITY 数學學院 篤行工作室

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{h}_{t-1} \\ \boldsymbol{x}_t \end{bmatrix} \tag{3}$$

$$f_t = \sigma(W_f.\mathbf{x} + b_f) \tag{4}$$

$$i_t = \sigma(W_i.\mathbf{x} + b_i) \tag{5}$$

$$o_t = \sigma(W_o.\mathbf{x} + b_o) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c.X + b_c) \tag{7}$$

$$h_t = o_t \odot tanh(c_t) \tag{8}$$

Gated Recurrent Units

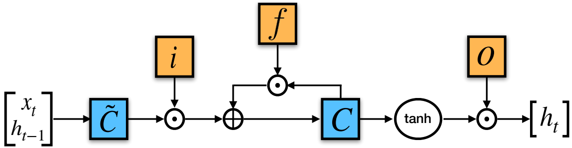
中山大學 SUN YAT-SEN UNIVERSITY 数学学院 笃行工作室

$$\mathbf{z} = \sigma(U_z.\mathbf{x_t} + W_z.\mathbf{h_{t-1}}) \tag{9}$$

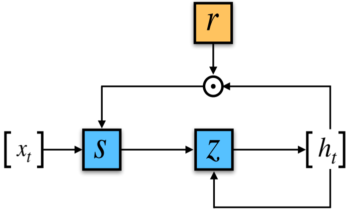
$$\mathbf{r} = \sigma(U_r.\mathbf{x_t} + W_r.\mathbf{h_{t-1}}) \tag{10}$$

$$\mathbf{s_t} = tanh(U_z.\mathbf{x_t} + W_s.(h_{t-1} \odot \mathbf{r})) \tag{11}$$

$$\mathbf{h_t} = (1 - \mathbf{z}) \odot \mathbf{s_t} + \mathbf{z} \odot \mathbf{h_{t-1}} \tag{12}$$



(1) Long Short-Term Memory



(2) Gated Recurrent Unit

Figure 23: Simple RNN

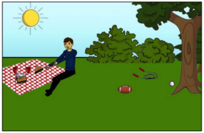
Section 4

other

Visual Question Answering

What color are her eyes?
What is the mustache made of?

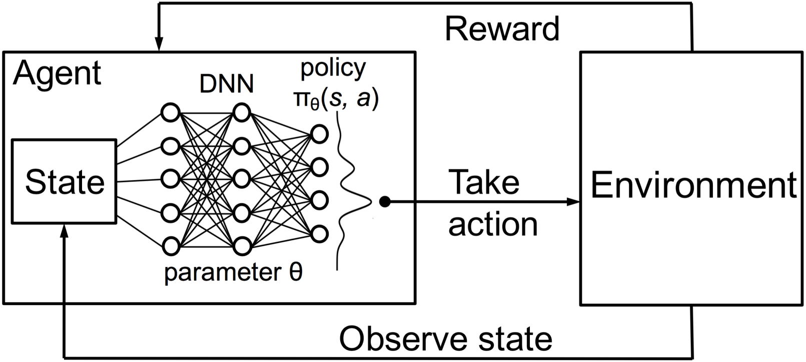
How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

Deep Q-Learning



Questions

Reference I

Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria.
Recent trends in deep learning based natural language processing.
ieee Computational intelligenCe magazine, 13(3):55–75, 2018.

Richard Zhang, Phillip Isola, and Alexei A Efros.
Colorful image colorization.
In *European conference on computer vision*, pages 649–666. Springer, 2016.