Ch. Convolutional Neural Network for Visual Recognition

Date /

§1. Introduction

- Computer Vision
  - Biology / Psychology: Optics, Neuron Science, Cognitive sciences,
  - Computer Science: Graphics, Algorithms, Theory, Systems, Architecture,
  - Mathematics / Physics: Information retrieval, Machine Learning.
  - Engineering: Robotics, Speech, NLP, Image Processing.

- Camera Obscura: 小孔成像. Huble & Wiesel (1959) — electrical signal (视觉皮层) from brain (stimulus ↔ response)

  Block world (1963) (边缘检测); Stages of Visual Representation, David Marr (1970) (视觉是分层的); Pictorial Structure (1973) (图画模型 — 点) & Generalized Cylinder (1979) (圆柱)

  David Lowe (1987) — 边缘检测 (SIFT & object Recognition (1999)); Normalized Cut (1997) (图像分割); Face Detection, Viola & Jones (2001) (AdaBoost)

  PASCAL VOC (2006–2012) (20 object categories); IMAGENET (22k & 15m images) — AlexNet (2012); VGG (2014); Google Net (2014); ResNet (2015); SENet (2017);

- Ingredients for Deep Learning: ① Algorithms ② Data ③ Computation

§2. Image Classification Pipe line

(KNN & Linear Classification)

- Segmantic Gap (激鸿沟): An image is just a grid of numbers between [0, 255]; (<label and pixel values>) (像素矩阵) RGB — 3 channel, (每个 channel 均为一个矩阵) ← 使用 8bit 存储 1 pixel (共 256个)

  Challenges: ① Background Clutter (背影杂乱). ② illumination (照明) ③ Deformation (形变) ④ Occlusion (遮挡) ⑤ intraclass variation (类内代分)

- An image classifier.

```
def classify_image (image):
    # Black Box
    return class_label
```

Machine Learning Data-Driven:

```
def train (images, labels):
    # ML
    return model
```

① Collect a dataset of images and labels — training set.

② Use ML to train a classifier

③ Evaluate the classifier on new images — testing set.

- K-Nearest Neighbour:

① training step: memorize all the training data and label.

② prediction step: predict the label of the most similar training image.

<L_1 Distance> (Manhattan Distance. $d_1(I_1, I_2) = \sum_p |I_1^p - I_2^p|$ ⇐ changing the coordinate frame!

Q. With N examples, how fast are training and prediction?
training → O(1); prediction → O(N)

* instead of copying label from nearest neighbour, take **majority vote** from K closest points.
bigger K ⇒ smooth the boundary.

<L_2 Distance> Euclidean: $d_2 = \sqrt{\sum_p (I_1^p - I_2^p)^2}$ (平方的平方根)

- Hyperparameters: choices about the algorithms than we set rather than learn?
  - └ ① Choose that work best on the data.