Revision:

Emerging Topics in Machine Learning & Computer Vision

What's deep learning

- Improving object classification: Better classifiers? Better features?
- Conventional computer vision feature representation (hand crafted) vs. deep features
- Multi-level hierarchical feature representation: Hard for hand-crafting
- Learning hierarchical features (structure) holistically
- Supervised learning vs. unsupervised learning
- Deep learning vs. shallow learning
- Image classification vs. object recognition (detection + classification)
- The importance of labelled datasets

What’s CNN & AlexNet

- Why convolutional neural networks (CNN)
- What is CNN and how does it work
  - Logistic regression & activation function
- Design of AlexNet – what are the key features & rational
- AlexNet filter design considerations
- What is max pooling and what is it for
- What is feed-forward in CNN and what is it for
- What is back-propagation in CNN and what for
- What is dropout in CNN and what for

Deep learning for super-res

- What is Super-Resolution (SR)
- How does it work – what are the steps of computation
- What do the convolutional layers in a deep SR model represent
- How does deep learning SR compare with conventional methods such as sparse-coding
- What advantages does deep learning have over conventional models

Attribute & transfer learning

- What are visual attributes
- Attributes vs. conventional visual features
- How to compute attributes, assumptions & limitations
- Why domain adaptation & transfer learning
- Design a CNN transfer learning model
    - Approaches to deep learning for domain adaptation
    - Differences to a CNN
    - Key considerations

R-CNN for object detection

- What's R-CNN: Image Classification vs. Object Recognition
    - Key considerations, R-CNN vs. AlexNet
    - R-CNN and CNN feature layers and fine-tuning CNN layers
    - CNN feature layer visualisation: What does it tell us about the layers
    - Size of original auxiliary data size vs. fine-tune data size
    - What's region proposals vs. sliding window
- Selective search
    - Segmentation vs. object localisation vs. candidate selection
    - Hierarchical image representation
    - Perceptual grouping & four basic similarity measures / metrics

CNN & action recognition

- Action “context” and how to compute it
- Action recognition vs. multi-instance learning
- Why multi-instance learning
  - Objective function
  - MIL vs. conventional supervised learning (e.g. SVM)
- Weakly supervised learning
- R-CNN as context modelling
  - R-CNN vs. Fast R-CNN vs. Faster R-CNN
  - What’s R*CNN and key considerations
  - R*CNN vs. R-CNN

Deeper models

- Why going deeper
- What is rational for NIN, VGG and GoogLeNet
- More training data vs. computational resources
- Challenges in training deeper models
  - GoogLeNet multiple softmax output layers in training
- Key considerations in designing GoogLeNet
  - ReLU vs. Sigmoid
  - Hebbian principle
  - Narrow filters – 1x1 vs. 3x3 vs. 5x5
  - Inception filter design – MLP key ideas
  - Network parameter size
- Network in Network key ideas
- VGG design considerations and why

Deep learning from web data

- Why deep learning works
  1. Scalability: Millions – 10^2 millions of parameters (vs. conventional models of 10^3 – 10^4 parameters)
  2. Large parameter size -> Encoding big data: 10-100 millions of data samples, from parameterised model to nearest neighbour data sample encoding

- Exploring big data with cheap labels

  1. Harnessing big data with noisy (weak) labels: Lots of parameters -> big cleanly labelled data (location & categorisation), “easy/clean” (Google) vs. “hard/noisy” data (Flickr-Yahoo)
  2. Overfitting vs. Generalisation: Large parameter size with small data vs. small parameter size with large data
  3. Exploring R-CNN with confusion matrix and graph relation layer