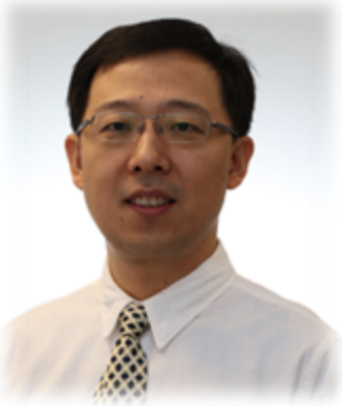
AI6121 Computer Vision

Overview

Asst Prof Lu Shijian (Shijian.Lu@ntu.edu.sg)

Instructor (https://personal.ntu.edu.sg/shijian.lu/)



Asst Prof LU Shijian, PhD
School of Computer Science and Engineering
Nanyang Technological University,
Singapore, 628798

Email: Shijian.Lu@ntu.edu.sg
Phone: (65) 6790 4928
Office: N4-02C-108

Research interests:
- Image and video analytics
- Visual intelligence
- Cognitive vision
- Machine learning

Research topics:
- Scene text detection and recognition
- Object detection and scene classification in satellite imagery
- Perception for unmanned vehicles
- Human action detection and human re-identification
- Facial expression classification and generation

Topics and Schedule

Week	Topic	Week	Topic
1	• T0: Overview • T1: Introduction		Recess
2	• T2: Image Filtering (Spatial)	8	• T8: 3D Vision (Imaging Geometry)
3	• T3: Image Filtering (Spectral)	9	• T9: 3D Vision (Stereo Vision)
4	• T4: Image Features (Points)	10	• T10: Learning in Computer Vision
5	• T5: Image Features (Edges)	11	• T11: Neural Networks
6	• T6. Image Classification	12	• T12: Image Synthesis*
7	• T7. Visual Attention (e-learning)	13	• T13: Image Segmentation*

The topics highlighted by * are optional, covering them or not depends on the progress.

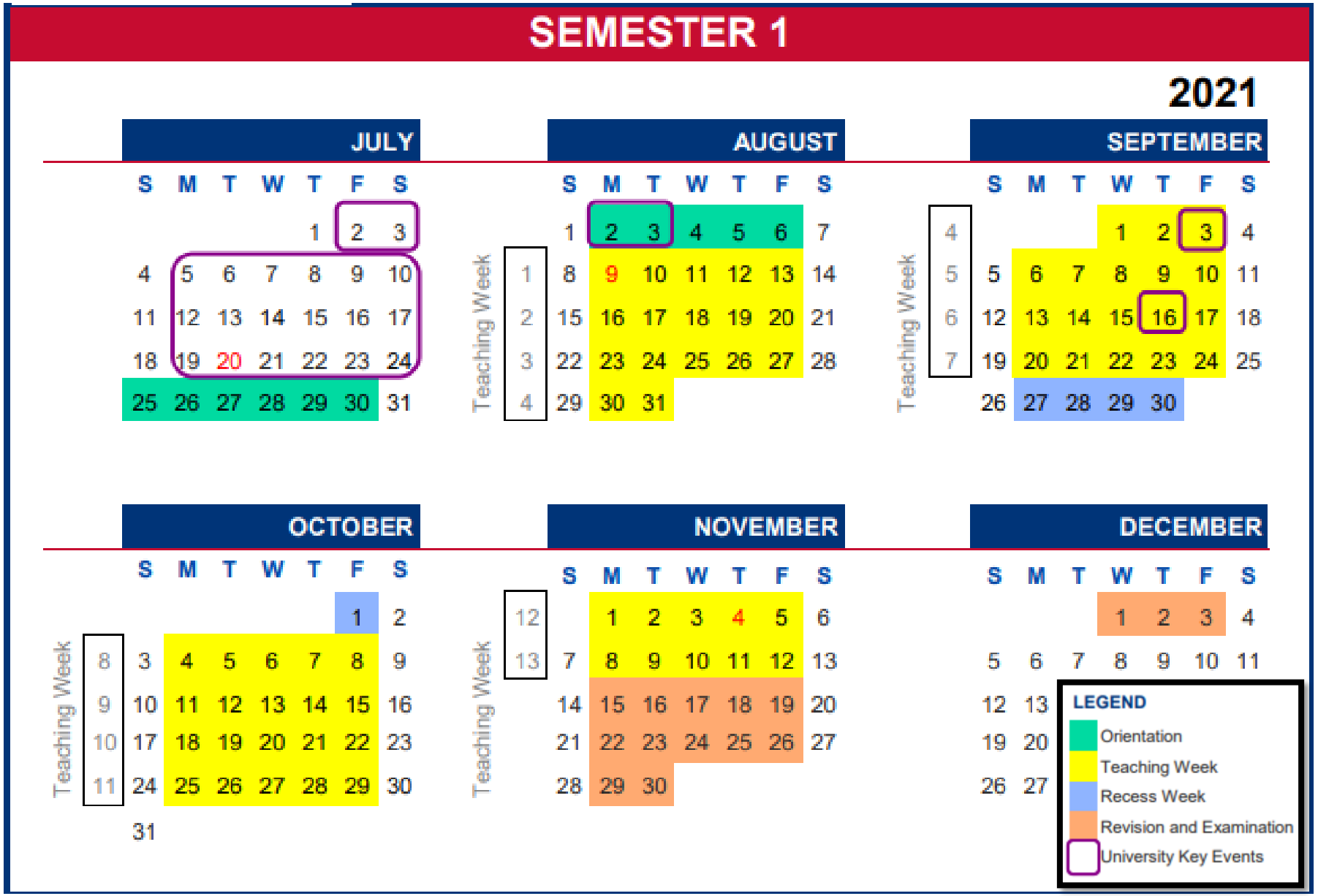
Course Overview

- Why open this course
- Focus on computer vision basics
- Wide coverage of both traditional and latest computer vision techniques
- Link up with the latest computer vision development whenever possible
- Link up with real-world computer vision applications whenever possible

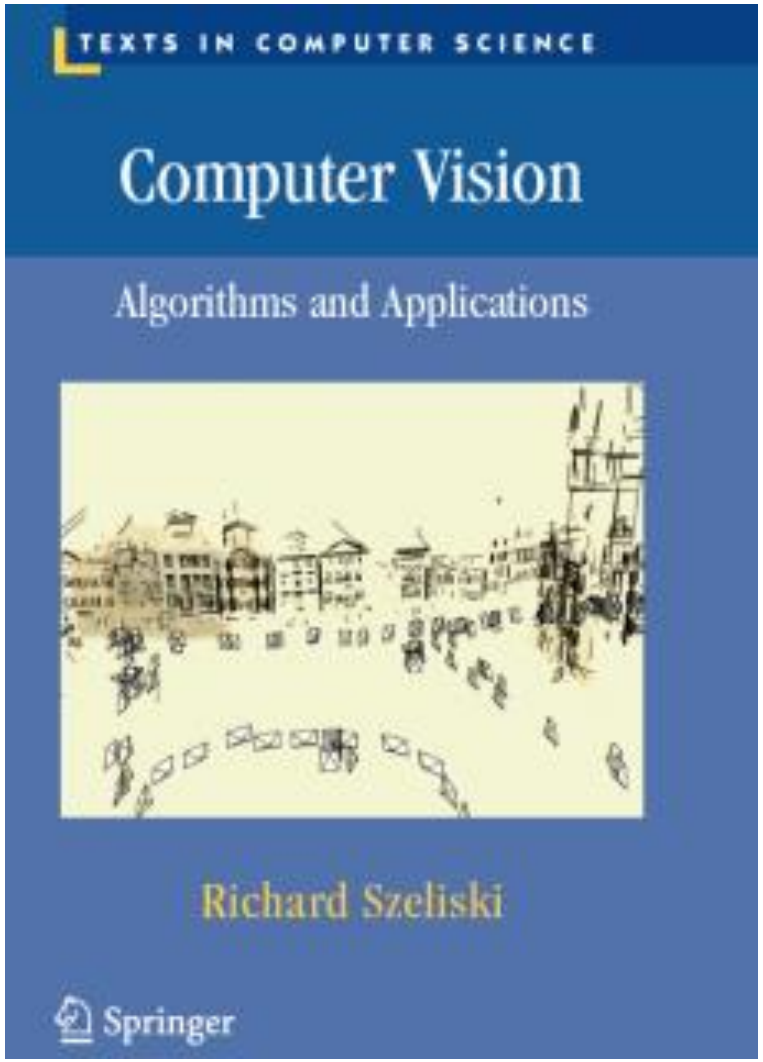
Course Structure

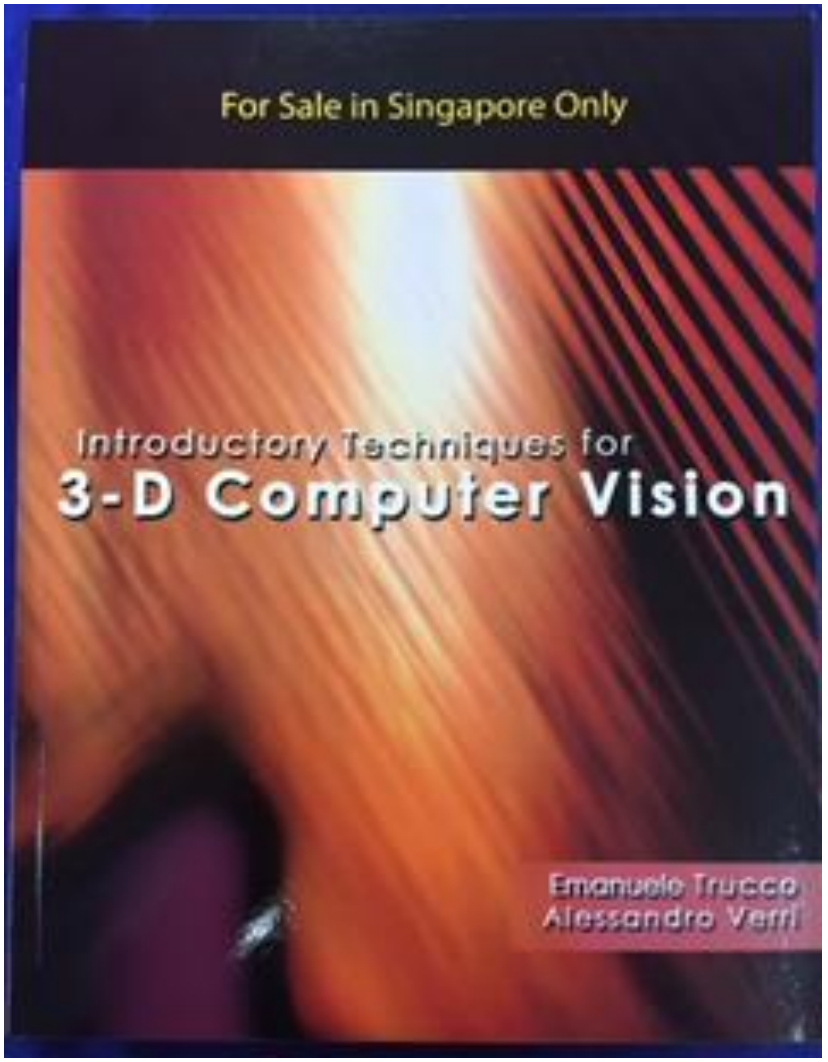
- 12 weeks lectures
- 1 week self-learning
- The assessment consists of
  - One paper reading and literature review (Oct 1st)
  - Two assignments (Sept 15th and Nov 1st)
  - One group project – max with 3 members (Nov 20th)
  - Submit your reports via NTULearn
- Grading
  - Paper reading (20) + Assignment (20*2) + Project (40)

Schedule

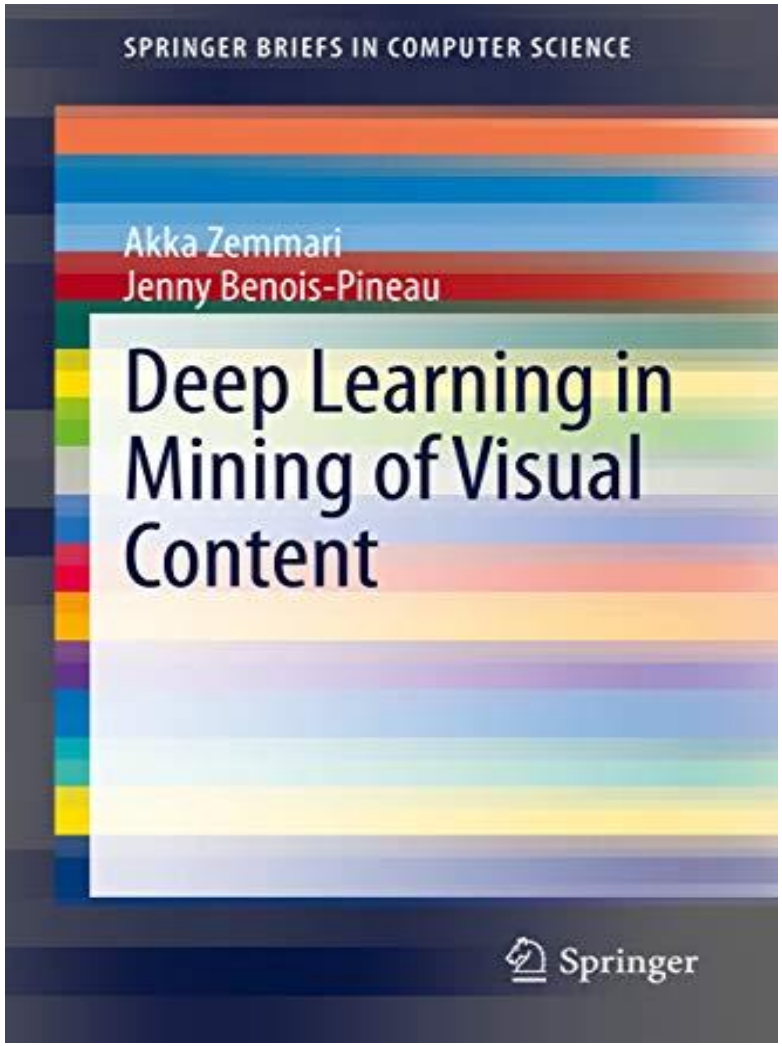

SEMESTER 1
2021
JULY
S M T W T F S
1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31
AUGUST
Teaching Week
S M T W T F S
1 2 3 4 5 6 7
1 8 9 10 11 12 13 14
2 15 16 17 18 19 20 21
3 22 23 24 25 26 27 28
4 29 30 31
SEPTEMBER
Teaching Week
S M T W T F S
4 1 2 3 4
5 5 6 7 8 9 10 11
6 12 13 14 15 16 17 18
7 19 20 21 22 23 24 25
26 27 28 29 30
OCTOBER
Teaching Week
S M T W T F S
1 2
8 3 4 5 6 7 8 9
9 10 11 12 13 14 15 16
10 17 18 19 20 21 22 23
11 24 25 26 27 28 29 30
31
NOVEMBER
Teaching Week
S M T W T F S
12 1 2 3 4 5 6
13 7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30
DECEMBER
S M T W T F S
1 2 3 4
5 6 7 8 9 10 11
12 13
19 20
26 27
LEGEND
Orientation
Teaching Week
Recess Week
Revision and Examination
University Key Events

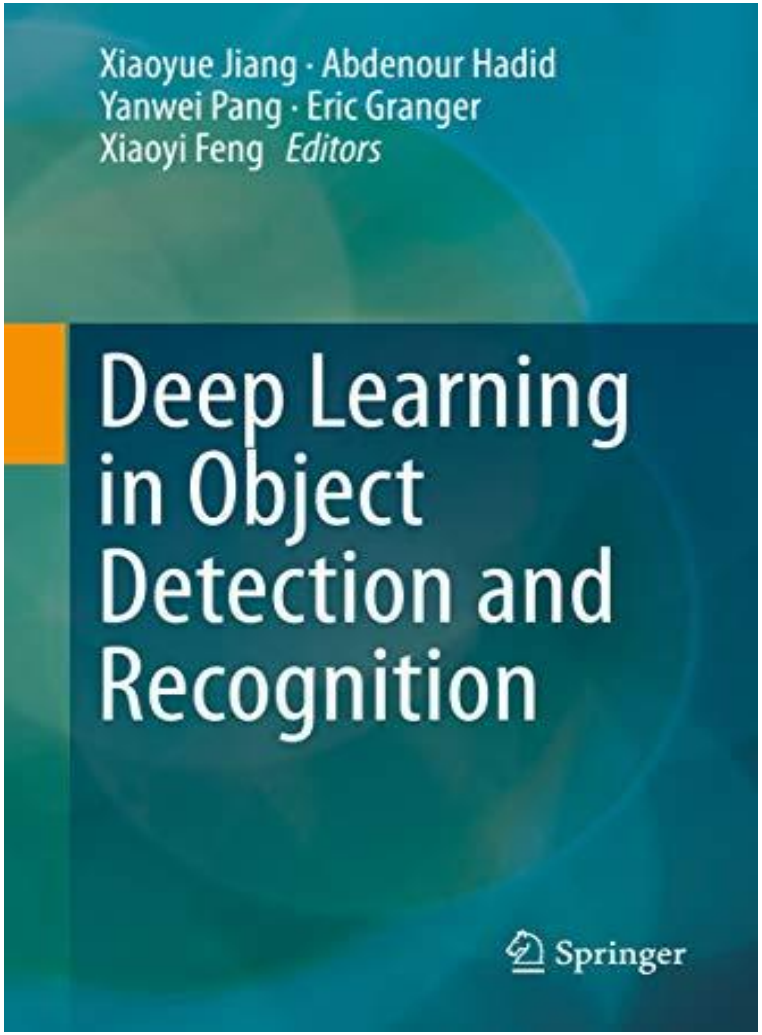
Recommended Text Books





Recommended Text Books





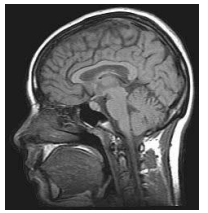
CE6902 Computer Vision

Introduction

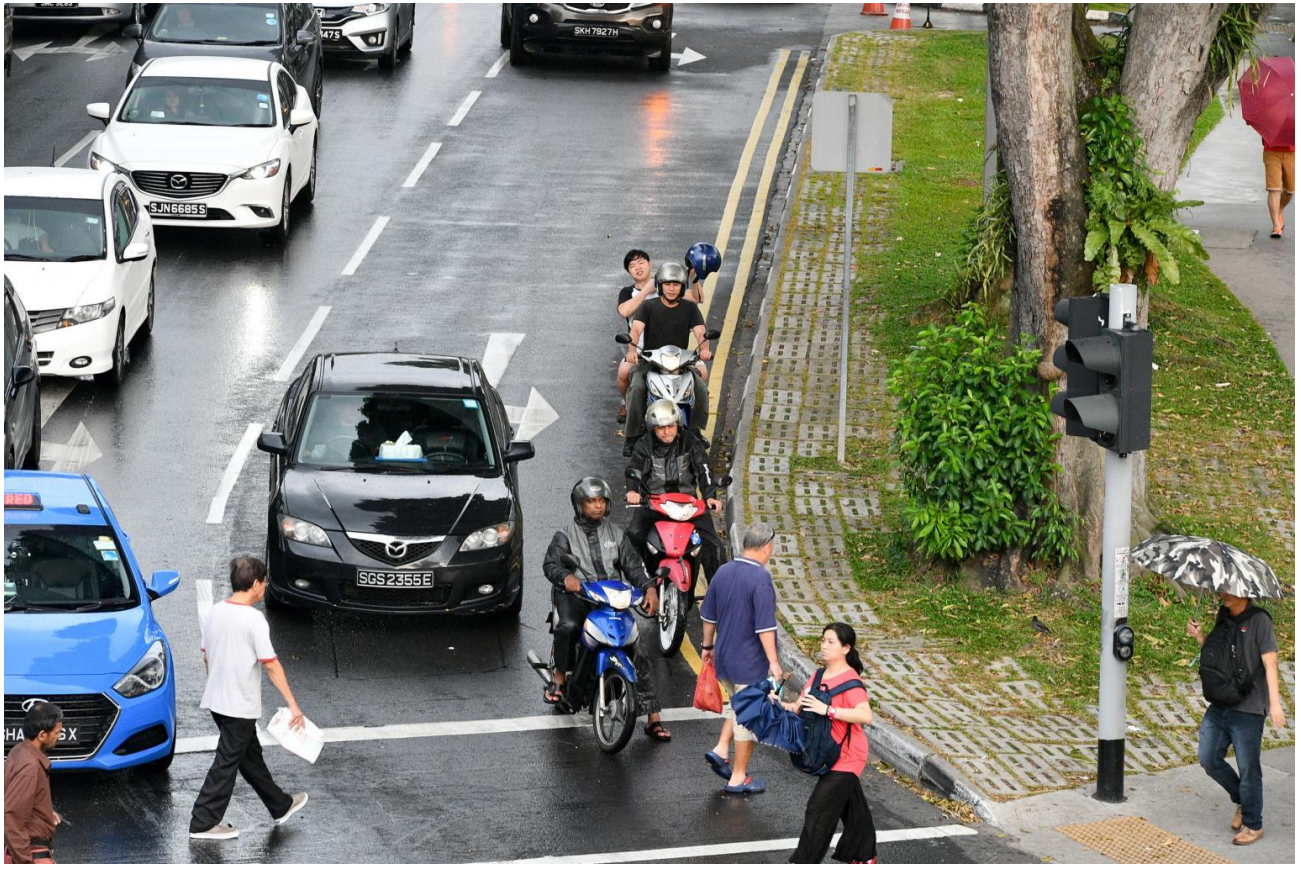
Asst Prof Lu Shijian (Shijian.Lu@ntu.edu.sg)

What is Computer Vision?





Computer vision is the science and technology that enable machines to see. The goal is to write computer programs that can interpret visual data of different modalities such as videos, depth images, multiple-view images, multi-dimension images, point cloud, etc.



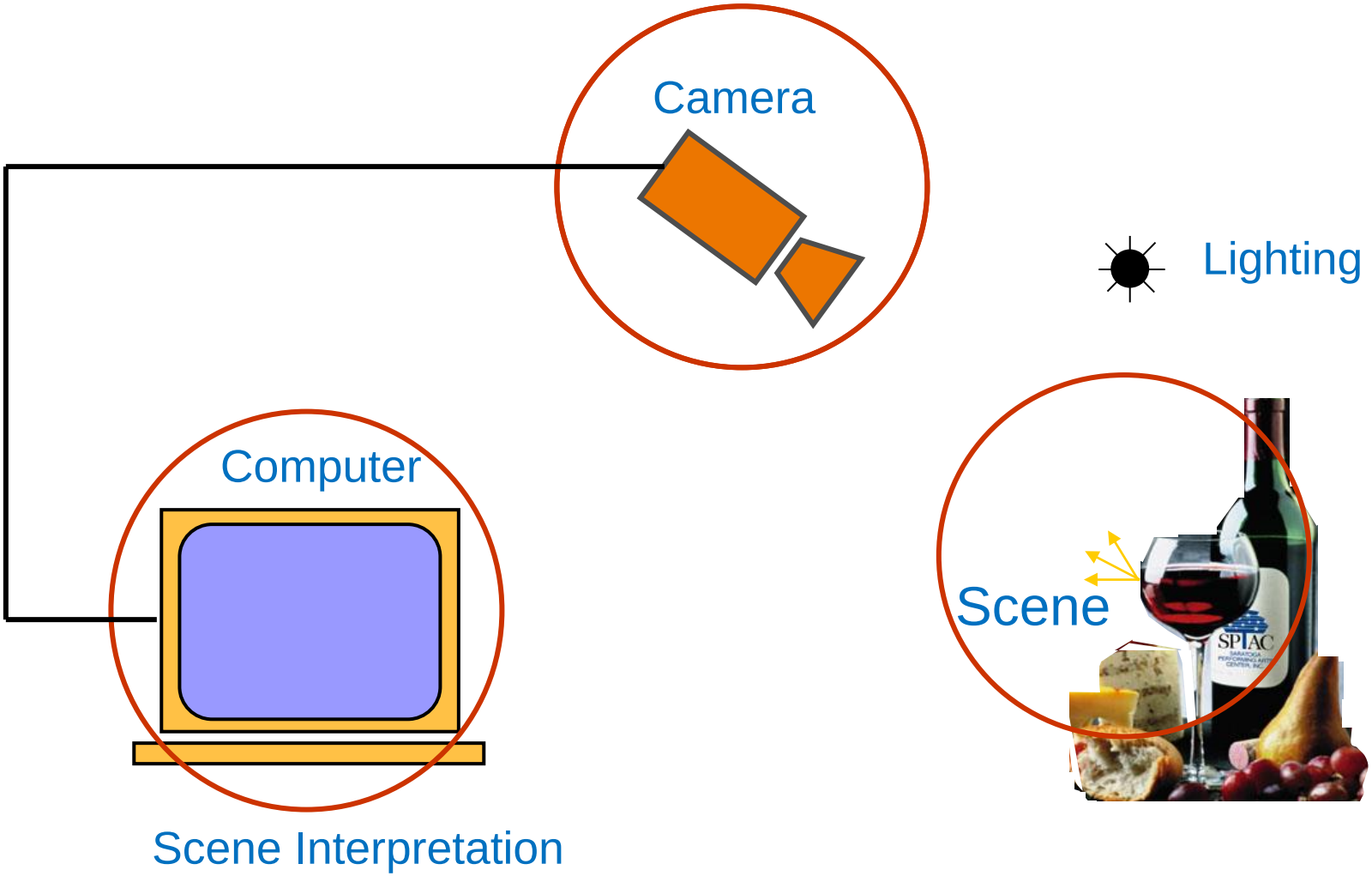
What are in the scene?

Where are the objects in the scene?

How are the objects interact with each other?

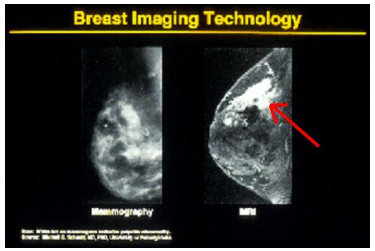
…

Composition of Computer Vision Systems



Why Study Computer Vision?

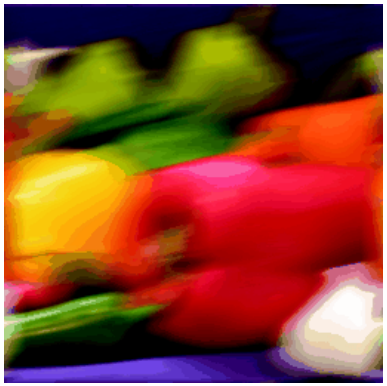
- Images and videos are everywhere
- Fast-growing collection of useful applications
  - Computer-aided disease diagnosis
  - Automated surveillance (who's doing what)
  - Autonomous driving
- Various deep and attractive scientific mysteries
  - How visual data are stored and retrieved?
  - How does object recognition work?
- Greater understanding of human vision



Spatial Image Filtering

- The images we are facing everyday often suffer from various types of noises, artefacts, distortion, etc.
- Spatial image filtering aims to remove undesired information from image pixels, and/or transform them to certain desired forms.
- Spatial image filtering can be a standalone task for producing images with certain desired special features, or a preprocessing process for facilitating certain ensuing tasks.






Spectral Image Filtering

- Instead of filtering images in spatial domain, spectral image filtering transforms images into spectral space, filters the spectral representation, and converts it back to spatial space.
- Spectral image filtering allows to access certain image structures and geometry characteristics which are encoded within certain specific frequency bands.
- Spectral transform decomposes images into multiple frequency components which helps in many advanced computer vision tasks.

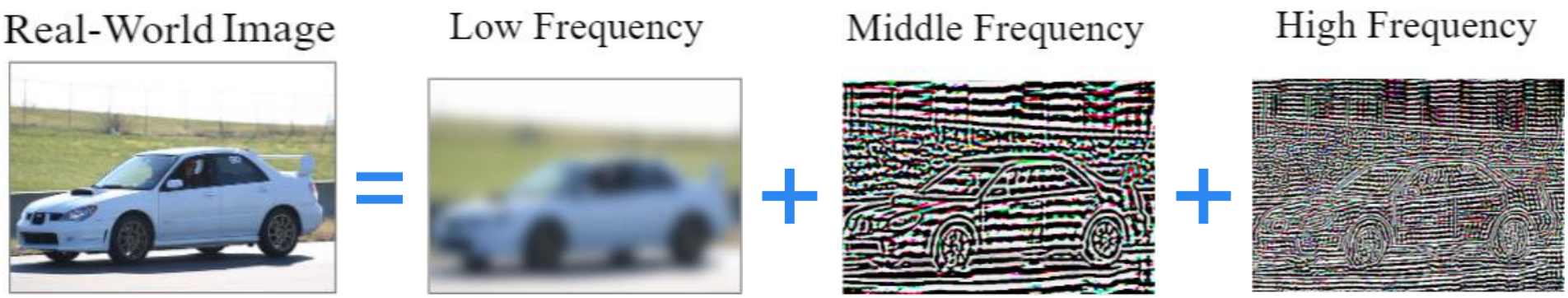


Image Features - Points

- The visual world is largely defined by appearance and structures, the latter largely defined by points, straight lines and curves.
- How to represent a visual scene for various visual recognition tasks (e.g. image captioning) under various photometric and geometric transformation – a bag of points like many words in a document.
- How to detect points reliably under various transformations?

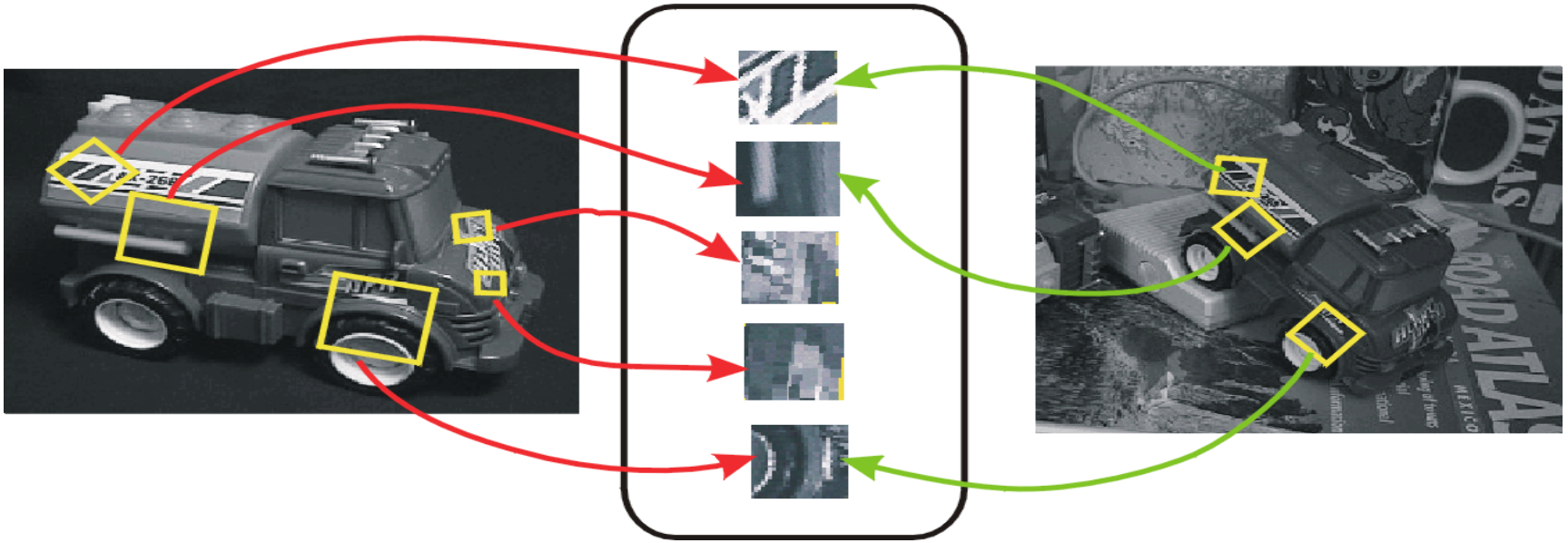


Image Features – Edges/Lines

- Line is another types of features that have been widely used in computer vision tasks.
- Line features can often be located based on image gradient and image edges.
- Lines, edges, and noises often have similar features and properties, suppressing one tends to affect another.

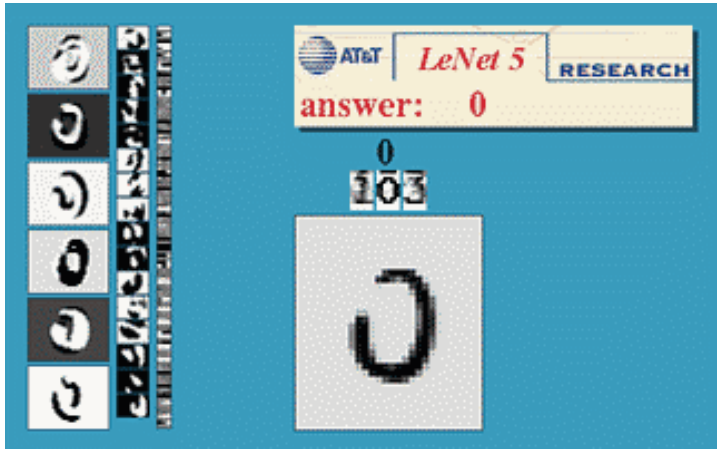
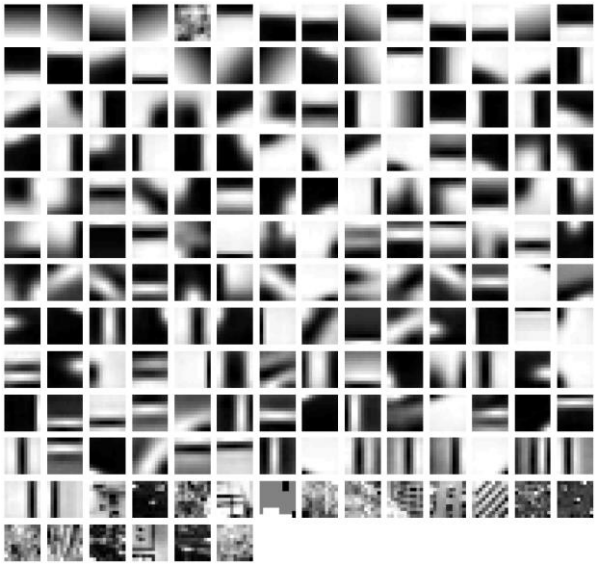
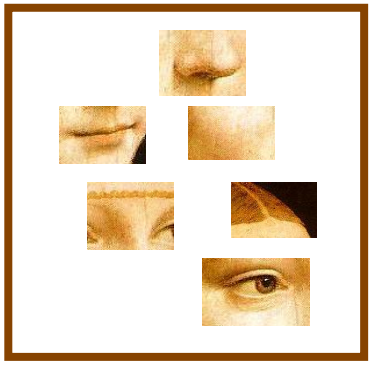
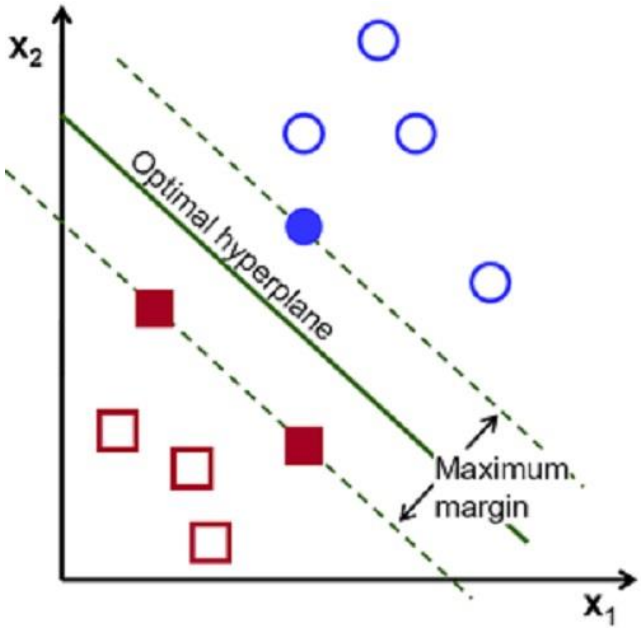


Image Recognition

- Each training image can be converted to a feature vector based on the learnt dictionary.
- A new image can then be recognized by a classifier that is trained by using the feature vector of training images.







Visual Attention

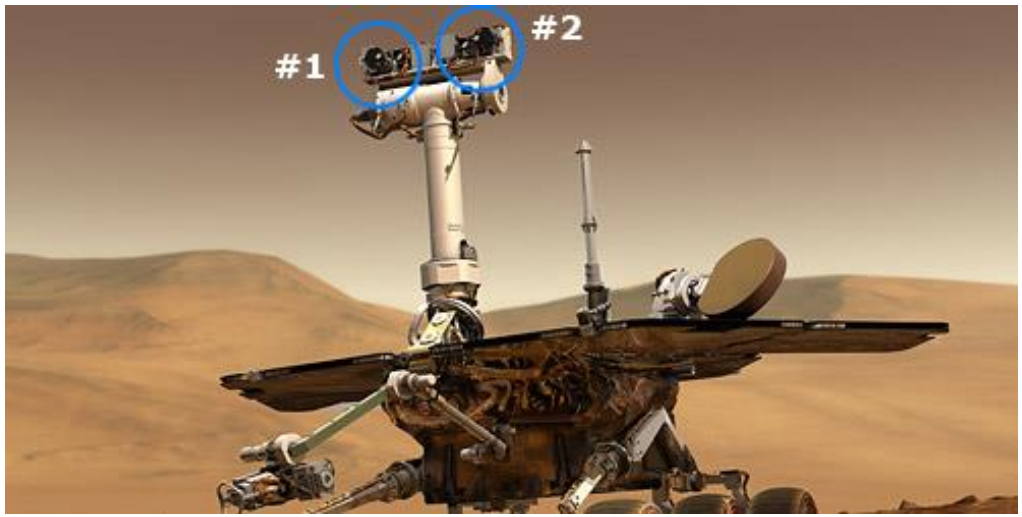
- The human vision system has no sufficient capacity to process all sensed visual information in parallel. It employs visual attention for sequential processing.
- This applies to machine as well, where visual attention can help focus on more critical information.
- Visual attention can be spatial-wise, temporal-wise, channel-wise, etc., depending on applications.
- Visual attention can be either bottom-up or top-down.

Imaging Geometry

- The imaging processing projects 3D visual world into 2D images which has witnessed many successful applications.
- The inverse process that restores 3D information from 2D images is similarly important to many computer vision tasks.
- Can 3D-2D imaging process be reversed for 3D reconstruction?

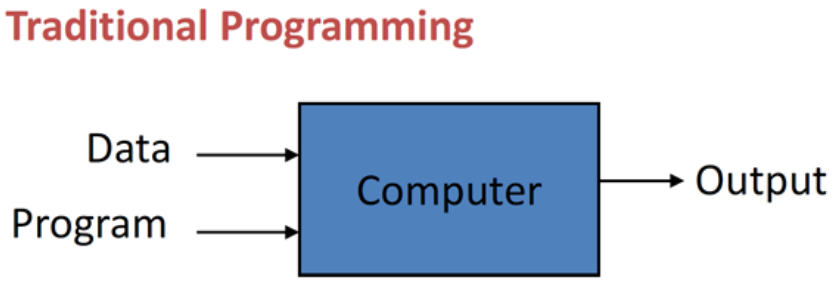
Stereo Vision

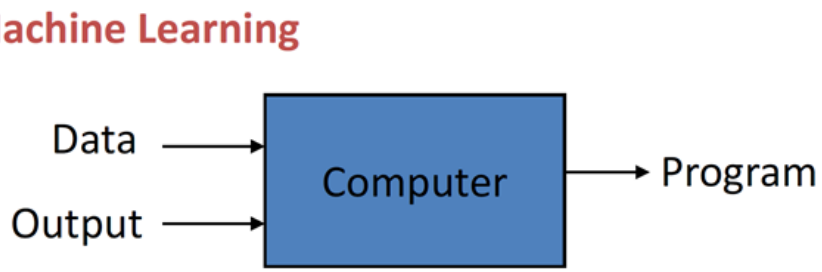
- A single image from a single camera often does not have insufficient information for accurate 3D information estimation.
- Stereo cameras simulate the human eye mechanism which exploit the disparity from two cameras for more accurate depth and 3D information estimation.
- Stereo cameras have been widely used in many computer vision applications such as mobile eyes in autonomous driving.



Machine Learning

- Traditional programming usually ‘programs’ parameters and algorithms manually according to expert experience for certain target output. Machine learning instead learns programs with training data and the desired outputs.
- Machine learning refers to algorithms that can improve their performance measure *P* using training data. It involves a task *T* and experience *E*. A well-defined learning task is given by <P, T, E>.
- Machine learning can be supervised, semi-supervised, unsupervised, self-supervised, contrastive, etc.





Neural Networks

- Neural network is a learning method that have many applications in text mining, speech recognition, image recognition, etc.
- Neural networks model the brain and nervous system, which are highly parallel and can simulate very complex functions and very complex behaviours.
- Convolutional neural networks have achieved tremendous success in various computer vision tasks.

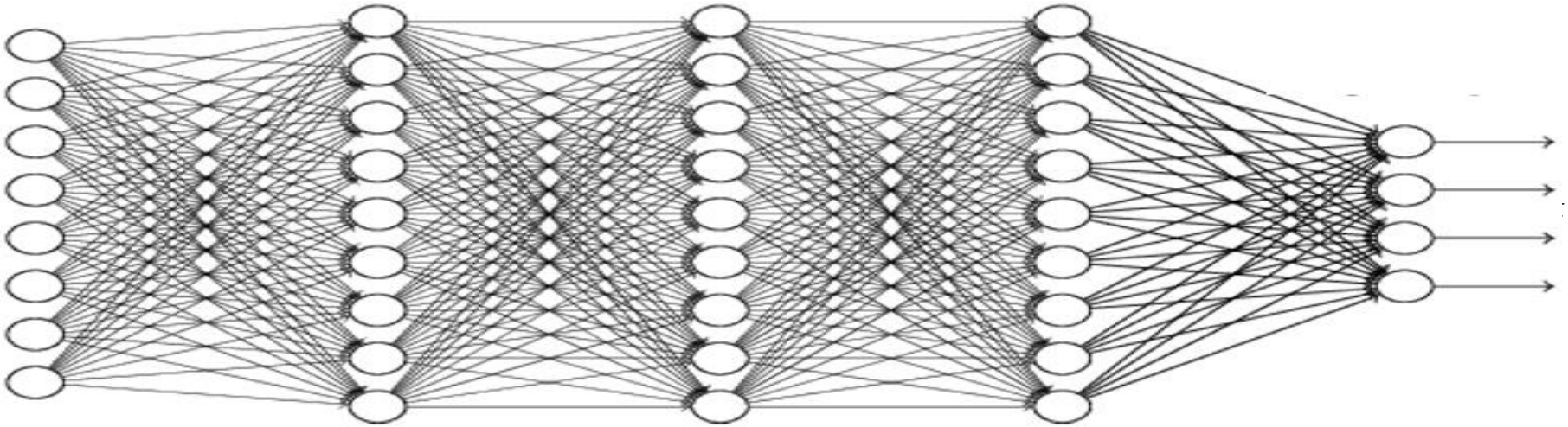
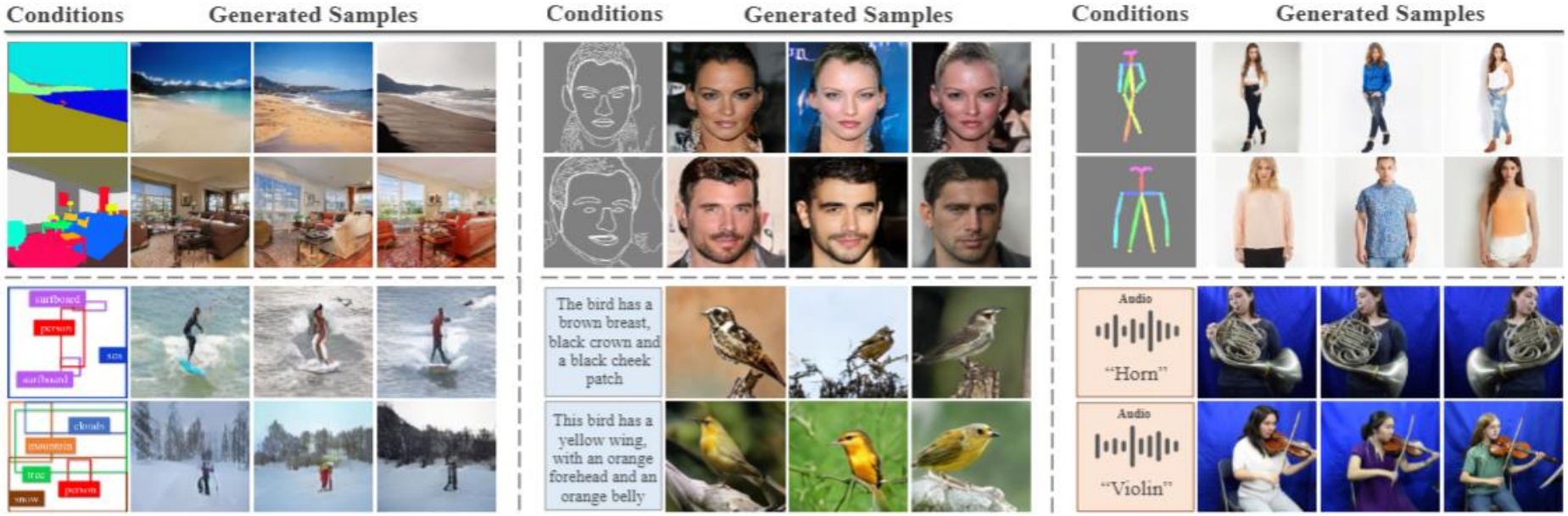


Image Synthesis

Big data and the corresponding annotations have become the major bottleneck for efficient development and deployment of CNNs.

Image synthesis generates annotated data automatically which is one typical approach for mitigating the data collection and annotation challenge in CNN design and training.

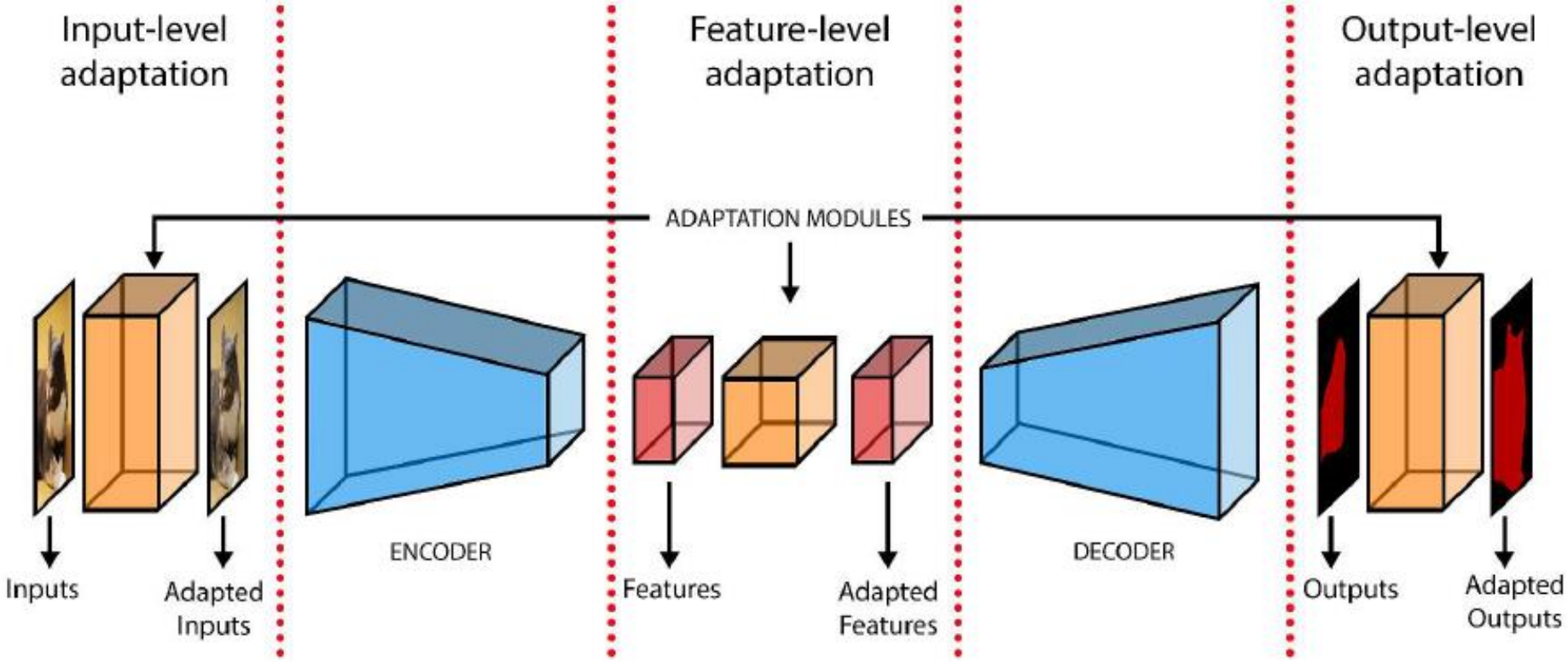
Beyond image generation from noises, images can be synthesized through crossing domain translation, editing, and composition.



Semantic Segmentation

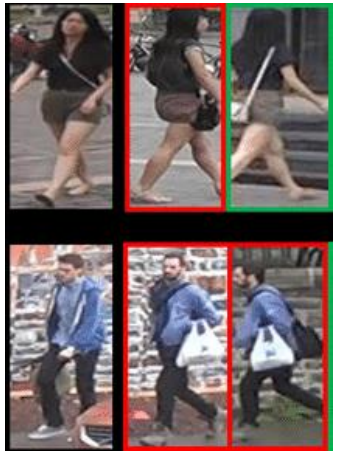
Semantic segmentation predicts a semantic class for each image pixel densely which requires huge efforts in image annotation.

Unsupervised domain adaptation learns a well-performed model in unlabelled target domains by adapting from annotated data of source domains. It aims to learn domain-invariant features.



Computer Vision Applications

Computer vision techniques have been successfully applied to different applications such as surveillance, autonomous driving, etc., though various challenges still need to be conquered…



Direct Reading and Literature Review (20 marks)

With the advance of deep learning, we have witnessed rapid progress in a wide spectrum of computer vision techniques in *image segmentation* (e.g. semantic/instance/panoptic segmentation), *image classification*, *object detection* (e.g. one/two-stage detection, anchor-free detection, transformer), *image generation* (e.g. image synthesis, image composition, image translation), etc.

To address various challenges in computer vision tasks, we also witnessed the fast development of machine learning techniques in supervised learning, semi-supervised learning, few-shot learning, self-supervised and unsupervised learning, transfer learning, etc.

This direct reading aims to equip you with capabilities of reading scientific papers in computer vision. You are expected to select one paper of the above listed topics, and produce a paper reading report.

Direct Reading and Literature Review (20 marks)

Your report should cover: what is the work about; what are gaps of existing work, and what are motivations of the performed research; how does the proposed technique address the gaps; what evaluations were designed to validate the designs; what are constraints of the proposed technique; what are possible future work, etc.

The report will be evaluated based on its contents and presentation. You need to work out the report independently and the report should reflect your understanding of the selected paper. For presentation, the report should have good clarity, logical flow, elegance, etc.

You need to submit your report in PDF format. No standard report templates, and no requirements of specific report length.

You need to submit the report through NTULearn before the deadline on Oct 1st 2021. There will be penalty for late submission.

Summary

1. Syllabus and schedule
2. Assessment
3. Brief introduction to computer vision (CV)
4. Brief introduction of the covered CV topics