Course Form for PKU Summer School International 2018

Course Title	**Machine Learning in Computer Vision** 计算机视觉中的机器学习
Teacher	Carlo TOMASI
First day of classes	July 2, 2018
Last day of classes	July 15, 2018
Course Credit	2 credits

Course Description

Objective

Machine learning methods have been employed for a long time in computer vision, but their nature and impact on the field has changed dramatically since 2012, with the introduction of deep networks. This course compares samples of pre-2012 and post-2012 machine learning techniques in the context of image recognition, person detection, human body modeling, and optical flow estimation. After an introduction to basic methods of image processing, including the analysis of optical flow, the course surveys decision trees, random forests, and deformable-parts models as a sample of pre-2012 methods, and reviews their use for person detection and human body tracking. Deep convolutional neural networks are then discussed in some detail, including convolutional pose machines and the OpenPose body tracking system. The course closes with a review of supervised and unsupervised deep learning methods for optical flow estimation.

Pre-requisites /Target audience

Senior undergraduate and graduate students with a good grasp of linear algebra, probability, and multivariate calculus. Some light-weight programming may be required for some of the assignments, in a language chosen by the students.

Proceeding of the Course

Ten 3-hour lecture sessions and a final 2-hour exam. Readings are a mix of class notes and papers from the literature. Each session is divided into three one-hour sections with brief breaks in-between.

Assignments (essay or other forms)

Daily reading assignments and short homework assignments with exam-style problems.

Evaluation Details

Attendance 20%; homework 45%; exam 35%

Academic Integrity (If necessary)

CLASS SCHEDULE

(Subject to adjustment)

Session 1：*Overview and Image Processing*	Date: July 2

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Overall structure and contents of the course. Elements of image processing, including convolution, filtering, image differentiation, and image pyramids.

【Questions】

【Readings, Websites or Video Clips】

Class notes

【Assignments for this session (if any)】

Read materials for sessions 1 and 2 (due in session 2)

Session 2：*Optical Flow*	Date: July 3

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)

Two-frame image motion as optical flow: Definitions and variational estimation methods for small displacements. The Shift-Invariant Feature Transform (SIFT). Large Displacement Optical Flow (LDOF).

【Questions】

【Readings, Websites or Video Clips】

- Class notes
- D. Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60(2):91-110, 2004.
- T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Conference on Computer Vision and Pattern*

Recognition, 41-48, 2009.	
【Assignments for this session (if any)】 Homework 1 (due in session 4)	
Session 3：***A Machine Learning Sampler***	Date: July 4
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Basic concepts of machine learning, including decision trees and random forests as examples of "classical" machine learning methods.	
【Questions】	
【Readings, Websites or Video Clips】 Class notes	
【Assignments for this session (if any)】 Read materials for sessions 3 (due in session 4)	
Session 4：***Person Detection and Body Tracking***	Date: July 5
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Methods for finding people in images, and determining their body poses: Histograms of Oriented Gradients (HOG), pedestrian detection, deformable models of body parts.	
【Questions】	
【Readings, Websites or Video Clips】 • Class notes • N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recogntiion, 1:886-893, 2005.* • J. Gall and V. Lempitsky. Class-specific Hough forests for object detection. *IEEE Conference on Computer vision and Pattern Recognition, 143-157, 2009.* • P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision,* 1:55-79, 2005.	
【Assignments for this session (if any)】 Read materials for sessions 4 and 5 (due in session 5)	

Session 5: ***Convolutional Neural Networks***	Date: July 6
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) The basic structure of convolutional neural networks. Stochastic gradient descent as a method for training deep networks.	
【Questions】	
【Readings, Websites or Video Clips】 Class notes	
【Assignments for this session (if any)】 Homework 2 (due in session 6)	
Session 6: ***Training Convolutional Networks. AlexNet.***	Date: July 9
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Back-propagation of gradients through a deep network, and the use of stochastic gradient descent for training. A description of AlexNet, the first broadly successful deep network in computer vision.	
【Questions】	
【Readings, Websites or Video Clips】 • Class notes. • A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems,* 25:1106-1114, 2012.	
【Assignments for this session (if any)】 Read materials for sessions 6 and 7 (due in session 7)	
Session 7: ***The State of the Art in Detection and Body Tracking***	Date: July 10
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Recent deep networks achieve better-than human performance in object detection tasks, and very high performance for body tracking. Two such systems are reviewed, based on the concepts of batch normalization and convolutional pose machines, respectively.	

【Questions】

【Readings, Websites or Video Clips】

- S. Ioffe and C. Szegedy. Batch normalization: accelerting deep network training by reducing internal covariate shift. *International Conference on Machine Learning*, 448-456, 2015.
- S-E. Wei ,V. Ramakrishna, T. Kanade, and Y. Sheikh. Convoutional pose machines. *IEEE Conference on Computer Vision and Pattern Recognition*, 4724-2732, 2017.
- Z. Cao, T. Simon, S-E. Wei, and Y. Sheikh. Real time multi-person 2D pose estimation using part affinity fields. *IEEE Conference on Computer Vision and Pattern Recognition*, 1302-1310, 2017.

【Assignments for this session (if any)】

Read materials for session 8 (due in session 8)

Session 8: ***Supervised Deep Networks for Optical Flow***	Date: July 11

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)

Optical flow is a regression problem, and contracting-expanding networks have proven up to this task. Foundational concepts are reviewed, and a state-of-the-art convolutional flow estimation method is studied.

【Questions】

【Readings, Websites or Video Clips】

A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. FlowNet: Learning optical flow with convolutional networks. *International Conference on Computer Vision*, 2758-2766, 2015.

【Assignments for this session (if any)】

Homework 3 (due in session 10)

Session 9: ***Unsupervised Optical Flow Estimation***	Date: July 12

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)

Training an optical flow estimator requires large amounts of data. Synthetic datasets and data augmentation are two techniques used to address this difficulty. Another one is to learn flow in an unsupervised manner. Tradeoffs and techniques are studied in this session, including spatial transformer networks and an unsupervised flow estimation system.

【Questions】

【Readings, Websites or Video Clips】

- Class notes
- M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks. *Advances in Neural Information Processing* Systems, 2017-2025, 2015.
- Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, and H. Zha. Unsupervised deep learning for optical flow estimation. *AAAI Conference on Artificial Intelligence,* 1495-1501, 2017.

【Assignments for this session (if any)】

Read materials for session 9 (due in session 10)

Session 10：***Review and Concluding Remarks***	Date: July 13

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)

This section reviews what was covered in this course and what was not. The future of machine learning and computer vision is discussed. Questions are answered in preparation for the final exam.

【Questions】

【Readings, Websites or Video Clips】

【Assignments for this session (if any)】

Prepare for the final exam.

Session 11：***Final Exam***	Date: July 15

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)

The final exam takes two hours. It is closed book, closed notes, and covers the main definitions and concepts through questions and simple problems.

【Questions】

【Readings, Websites or Video Clips】
【Assignments for this session (if any)】