- CSCI-GA.3033-020
  (19202) Integrating Machine Learning to Computer Vision  T 7:10-9:00PM CIWW 202

Course Description:  The course will investigate how Unsupervised Learning, Recurrent Networks, and Supervised Learning, are becoming the foundations of Computer Vision.  We will study it through the vision topics of: Object Classification, Image Segmentation, Face Recognition, Stereo Vision, Motion, Action Recognition, and  Image Annotation.

Course Format: There will be a review of papers for each topic, at least two papers per class. The professor will make one paper presentation and a student (or group of two students) will make another paper presentation. All students will be expected to read the papers before class. There will be a (final) project to be presented at the end of the year (students can make groups of two), and in the middle of the course, the students will describe their project proposals. Suggestions of projects will be given, but students are free to propose their own project. Students will be evaluated by paper presentation, class participation, and final project.

**1. From Wavelets and Image processing to Scattered Wavelet Networks (unsupervised learning).** One class.

A. Wavelet Material. Professor Geiger.

B. Invariant Scattering Convolution Networks.  Joan Bruna and Stephane Mallat. TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 35, NO. 8, AUGUST 2013.

https://www.di.ens.fr/~mallat/papiers/Bruna-Mallat-Pami-Scat.pdf

relevance: first unsupervised network to demonstrate mathematically and empirically the importance of multilayer wavelet processing with local non-linearity.

**2. Object Classification and Recognition** (two classes)

2. A.  ImageNet Classification with Deep Convolutional Neural Networks . Krizhevsky, Sutskever, and Hinton. Nips 2012.

https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Relevance was the original deep CNNs in  computer vision. Bring some concepts, such as drop out, that are not currently being used today, but are interesting and may have future impact.

2.B. Very deep convolutional networks for large-scale image recognition. Karen Simonyan and Andrew Zisserman . ICLR2015. https://arxiv.org/pdf/1409.1556v6.pdf

Relevance: shows that convolutional neural networks have to have a deep network of layers in order for this hierarchical representation of visual data to work

3.A. Going Deeper with Convolutions. Christian Szegedy et. al. CVPR 2015.

http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf

Relevance: Excellent machine, works very well, and it is available and trained within Tensor Flow on image net. Great benchmark to compare.

3.B. Deep Residual Learning for Image Recognition. Kaiming He et. al. Dec 2015. https://arxiv.org/pdf/1512.03385v1.pdf

Relevance: The ResNet model is the best and simplest CNN architecture today.

**4. Image Segmentation** (one class)

4.A. Graph Cut algorithm (Unsupervised Technique)

Graph Cuts—Combinatorial Optimization in Vision (Chapter 2). Hiroshi Ishikawa. In *Image Processing and Analysis with Graphs: Theory and Practice*, Edited by Olivier Lézoray and Leo Grady, CRC Press, July 2012.

Relevance: best unsupervised technique and developed at NYU.

4. B. Fully Convolutional Networks for Semantic Segmentation Jonathan Long, Evan Shelhamer and Trevor Darrell UC Berkeley. 2014.

https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Relevance: combining the concepts of unsupervised technique (3.A.) with supervised

Technique

**5**. **Face Detection, Boosting and Convolution Neural Networks**

(1 class)

5.A. Rapid Object Detection using a Boosted Cascade of Simple Features. Paul Viola Michael Jone. CVPR2001. https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf

Relevance: original seminal paper on learning face detection using boosting.

5.B. Synergistic Face Detection and Pose Estimation with Energy-Based Models. NIPS2004. http://yann.lecun.com/exdb/publis/pdf/osadchy-04.pdf

Relevance: expand the problem to also consider pose estimation.

**6. Stereo Vision: From local features to gradient descent Gaussian curvature (unsupervised)** (one class)

6.A. Professor Geiger lecture on Wavelet Local Features.

Relevance: Basis to construct Stereo.

6.B. Spectrally regularized surfaces. Yuanhao Gong. ETH-Zürich (2015). http://dx.doi.org/10.3929/ethz-a-01043829

Relevance: unsupervised technique based on gradient descent flow that minimize Gaussian Curvature.

**7**. **Motion (integration of unsupervised and supervised methods)** (one class)

7.A. Lucas-Kanade 20 Years On: A Unifying Framework SIMON BAKER AND IAIN MATTHEWS. International Journal of Computer Vision 56(3), 221–255, 2004

http://www.ncorr.com/download/publications/bakerunify.pdf

Relevance: best and simplest optical flow method: window matching and affine transformation to them.

7.B. Spatial Transformer Networks. Max Jaderberg et. al. 2016. https://arxiv.org/pdf/1506.02025.pdf

relevance: it suggests that techniques of motion, affine invariance (Lucas Kanade), can be seen as unsupervised learning and can be used to reduce data sets for supervised learning.

Similarly, suggests we can put learning on top of Lucas Kanade.

**8. Class to discuss proposals for all Projects, that will be presented at the end of the Course. Each group will have fifteen minutes to present their proposal.**

**9. Reinforcement Learning** (one classes)

9.A. Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3-4):229–256, 1992.

Relevance: foundation.

9.B. Richard S Sutton, David A McAllester, Satinder P Singh, Yishay Mansour, et al. Policy gradient methods for reinforcement learning with function approximation. In NIPS, volume 99, pages 1057–1063, 1999.

Relevance: also foundation.

9.C. Haitham Bou Ammar, Rasul Tutunov, and Eric Eaton. Safe policy search for lifelong reinforcement learning with sublinear regret. The Journal of Machine Learning Research (JMLR), 2015.

Relevance: new directions.

**10. Driverless Cars : Reinforcement learning and Deep Neural Networks** (one class)

10.A Keith Naughton. Human drivers are bumping into driverless cars and exposing a key flaw. http://www.autonews.com/article/20151218/OEM11/151219874/human-drivers-arebumping-into-driverless-cars-and-exposing-a-key-flaw , 2015.

Relevance: interesting points.

10.B Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving Shai Shalev-Shwartz, Shaked Shammah, Amnon Shashua. https://arxiv.org/abs/1610.03295

Supplementary video files. https://www.dropbox.com/s/136nbndtdyehtgi/doubleMerge.m4v?dl=0

Relevance: best research group in the topic.

10.C. NVIDIA research paper End-to-End Learning for Self-Driving Car - Mariusz Bojarski et. al.

http://images.nvidia.com/content/tegra/automotive/images/2016/solutions/pdf/end-to-end-dl-using-px.pdf

Relevance: attempt to do purely supervised.

**11. Adversarial Networks and Wavelets** (**unsupervised**) (one classes)

11.A. Generative Adversarial Networks, Goodfelllow et. al.

https://arxiv.org/pdf/1406.2661v1.pdf

Relevance: the foundation paper.

11.B. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. Emily L. Denton et. al. Nips 2015.

Relevance: integration with wavelets (first class).

11.C. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Alec Radford et. al. 2016.

https://arxiv.org/abs/1511.06434

Relevance: integration with deep convolution networks (second class).

**12. Action Recognition** (one class)

12.A Action Recognition using Visual Attention. Shikhar Sharma, Ryan Kiros, Ruslan Salakhutdinov, Feb 2016.

https://arxiv.org/abs/1511.04119 (more info at http://shikharsharma.com/projects/action-recognition-attention/)

Relevance: best suggestion of approach so far.

12.B. Human Action Recognition by Learning Bases of Action Attributes and Parts Bangpeng Yao et. al.

https://people.csail.mit.edu/khosla/papers/iccv2011_yao.pdf

Relevance: interesting suggestion.

12.C. https://www.robots.ox.ac.uk/~vgg/publications/2016/Bilen16a/bilen16a.pdf

Dynamic Image Networks for Action Recognition. Hakan Bilen e.t al. 2016

Relevance: a recent new idea. Worth to consider.

13. **Variational Auto Encoders** (one class)

13.A. Tutorial on Variational Autoencoders(VAEs), Carl Doersch**.** August, 2016.

https://arxiv.org/abs/1606.05908

Relevance: sampling from distributions inside deep networks, and can be trained with stochastic gradient descent. VAEs have already shown promise in generating many kinds of complicated data, including handwritten digits, faces, house numbers, CIFAR images, physical models of scenes, segmentation, and predicting the future from static images.

13.B. "Variational Convolutional Networks for Human-Centric Annotations," *13th Asian Conference on Computer Vision, 2016.*

Tsung-Wei Ke, Che-Wei Lin, Tyng-Luh Liu and Davi Geiger, Relevance: Use of VAEs to annotate automatically images.

Relevance: Integration to vision. Professor is an author.

14. Presentation of the Papers by students. (one class + finals)