Suggested projects for ECE-GY 6123 Image and Video Processing (Spring 2020)

Updated 3/1/2020

Please email the contact person to set up an appointment to learn more about the project. If you want to choose a project outside this list, you should discuss your idea with the instructor before deciding.

- 360 Degree Video Saliency Estimation and/or View Prediction (contact: Zhipeng, zf606@nyu.edu)

Background needed: deep learning

References:

Chenge Li, Weixi Zhang, Yong Liu, Yao Wang, “Very Long Term Field of View Prediction for 360-degree Video Streaming“, invited paper in IEEE Multimedia Information Processing and Retrieval (MIPR), 2019.

Pan, Junting, et al. "Shallow and deep convolutional networks for saliency prediction." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

Wu, Chenglei, et al. "A Dataset for Exploring User Behaviors in VR Spherical Video Streaming." *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 2017.

Lo, Wen-Chih, et al. "360 Video Viewing Dataset in Head-Mounted Virtual Reality." *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 2017.

Bao, Yanan, et al. "Viewing 360 degree videos: Motion prediction and bandwidth optimization." *Network Protocols (ICNP), 2016 IEEE 24th International Conference on*. IEEE, 2016.

Bao, Yanan, et al. "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos." *Big Data (Big Data), 2016 IEEE International Conference on*. IEEE, 2016.

https://code.facebook.com/posts/118926451990297/enhancing-high-resolution-360-streaming-with-view-prediction/

- Brain lesion segmentation from MRI images (Contact: Zhipeng Fan, zf606@nyu.edu)

Background needed: deep learning

References:

Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study

https://www.ncbi.nlm.nih.gov/pubmed/31190607

Learning joint lesion and tissue segmentation from task-specific hetero-modal datasets

https://openreview.net/forum?id=HJeZW_QxxN

Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI https://www.sciencedirect.com/science/article/pii/S2213158217302486

Segmentation using a Data Management and Processing Infrastructure
https://www.nature.com/articles/s41598-018-31911-7

- Denoising using deep networks with unknown and possibly spatially varying noise levels ods. (Contact: Amirhossein Khalilian akg404@nyu.edu).

Background needed: deep learning

- Video compression using deep learning based frame prediction and/or interpolation

(Contact: Haojie Liu, hl3933@nyu.edu)
Background needed: deep learning and image compression

Refs.:
Video Compression through Image Interpolation,
https://arxiv.org/abs/1804.06919

Learned Video Compression, https://arxiv.org/abs/1811.06981

Background: image compression and deep learning

- Layered image coding using multiresolution neural networks (Contact: Yao Wang, yaowang@nyu.edu Haojie Liu, hl3933@nyu.edu)

Background: image compression and deep learning.

Refs:
Jia C, Liu Z, Wang Y, Ma S, Gao W. Layered image compression using scalable auto-encoder. In2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) 2019 Mar 28 (pp. 431-436). IEEE.
H. Ma, D. Liu, R. Xiong and F. Wu, "iWave: CNN-Based Wavelet-Like Transform for Image Compression," in IEEE Transactions on Multimedia.
Huang, C., Liu, H., Chen, T., Shen, Q., & Ma, Z. (2019, December). Extreme Image Coding via Multiscale Autoencoders with Generative Adversarial Optimization. In *2019 IEEE Visual Communications and Image Processing (VCIP)* (pp. 1-4). IEEE.

- Learning multi-resolution representations (contact: Yao Wang yaowang@nyu.edu)

- Video interpolation: predicting intermediate frames from surrounding frames using deep learning. (contact: Ran Wang, rw1691@nyu.edu)

Xu X, Siyao L, Sun W, et al. Quadratic video interpolation[C]//Advances in Neural Information Processing Systems. 2019: 1645-1654.

Liu Z, Yeh R A, Tang X, et al. Video frame synthesis using deep voxel flow[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 4463-4471.

Niklaus S, Mai L, Liu F. Video frame interpolation via adaptive separable convolution[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 261-270.

Bao W, Lai W S, Ma C, et al. Depth-aware video frame interpolation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3703-3712.

Jiang H, Sun D, Jampani V, et al. Super slomo: High quality estimation of multiple intermediate frames for video interpolation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 9000-9008.

- Video prediction: predicting future frames from prior frames using deep learning method (contact: Ran Wang, rw1691@nyu.edu)

[PredRNN] Y. B.Wang, Z. F. Gao, M. S. Long, J. M.Wang, and P. S. Yu, “Predrnn++: Towards a resolution of the deep-intime dilemma in spatiotemporal predictive learning.,” in ICML. 2018, vol. 80 of JMLRWorkshop and Conference Proceedings, pp. 5110–5119, JMLR.org.

[TBnet] J. W. Xu, B. B. Ni, Z. F. Li, S. Cheng, and X. K. Yang, “Structure preserving video prediction,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

Older papers:

[DVF] Z. W. Liu, R. A. Yeh, X. O. Tang, Y. M. Liu, and A. Agarwala, “Video frame synthesis using deep voxel flow.,” in ICCV. 2017, pp. 4473–4481, IEEE Computer Society.

[DualGAN] X. D. Liang, L. Lee, W. Dai, and E. P. Xing, “Dual motion gan for future-flow embedded video prediction.,” in ICCV . 2017, pp. 1762–1770, IEEE Computer Society.

[EpicFlow] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, “Epicflow: Edge-preserving interpolation
of correspondences for optical flow.,” in CVPR . 2015, pp. 1164–1172, IEEE Computer Society.

[NextFlow] N. Sedaghat, “Next-flow: Hybrid multi-tasking with next-frame prediction to boost optical-flow estimation in the wild.,” CoRR, vol. abs/1612.03777, 2016.

[CNDA] C. Finn, I. J. Goodfellow, and S. Levine, “Unsupervised learning for physical interaction through video prediction.,” in NIPS , Daniel D. Lee, Masashi Sugiyama, Ulrike V. Luxburg, Isabelle Guyon, and Roman Garnett, Eds., 2016, pp. 64–72.

[MCnet] R. Villegas, J. M. Yang, S. Hong, X. Y. Lin, and H. Lee, “Decomposing motion and content for natural video sequence prediction.,” CoRR, vol. abs/1706.08033, 2017.

M. Mathieu, C. Couprie, and Y. LeCun, “Deep multiscale video prediction beyond mean square error.,” CoRR, vol. abs/1511.05440, 2015.

- Segmentation of ultrasound images of mouse embryos. (contact: Ziming Qiu zq415@nyu.edu)

Background: deep learning

Ziming Qiu, Jack Langerman, Nitin Nair, Orlando Aristizabal, Jonathan Mamou, Daniel H. Turnbull, Jeffrey Ketterling, Yao Wang, “Deep BV: A Fully Automated System for Brain Ventricle Localization and Segmentation in 3D Ultrasound Images of Embryonic Mice“, IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2018, arXiv preview link.

- Tracking the movements of plants, as part of our PlantTracer Project (http://planttracer.com/) (Contact: Yixiang Mao, yixiang.mao@nyu.edu)
    - Tracking plant apex using classical tracking approaches (e.g. KLT tracker, multi-view block matching)
    - Using deep learning approach (you will help to annotate ground truth data as well as training and testing the network)

- Moving Foreground detection in video using Robust Principal Component Analysis or other sparse-representation-based optimization methods. (Contact: Amirhossein Khalilian akg404@nyu.edu).

Background needed: Linear Algebra. Convex optimization background is desired but not necessary.

Ref:

[1] Candès, Emmanuel J., et al. "Robust principal component analysis?." Journal of the ACM (JACM) 58.3 (2011): 11.

[2] X. Cao, L. Yang and X. Guo, "Total Variation Regularized RPCA for Irregularly Moving Object Detection Under Dynamic Background," in IEEE Transactions on Cybernetics, vol. 46, no. 4, pp. 1014-1027, April 2016.

[3] Lin, Zhouchen, et al. "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix." Coordinated Science Laboratory Report no. UILU-ENG-09-2214, DC-246 (2009).

See the competition http://www.changedetection.net/ for data samples.

- Image/video deblurring or super-resolution with or without knowing the blurring kernels (Contact: Amirhossein Khalilian akg404@nyu.edu)

Background needed: Linear Algebra and DSP background is required. Convex optimization background is desired but not necessary.

Ref:

[1] S. Farsiu, M. D. Robinson, M. Elad and P. Milanfar, "Fast and robust multiframe super resolution," in IEEE Transactions on Image Processing, vol. 13, no. 10, pp. 1327-1344, Oct. 2004.

[2] Zhao, Ningning, et al. "Fast Single Image Super-Resolution Using a New Analytical Solution for l2-l2 Problems." IEEE Transactions on Image Processing 25.8 (2016): 3683-3697.

[3] Chan, S. H., Khoshabeh, R., Gibson, K. B., Gill, P. E., & Nguyen, T. Q. (2011). An augmented Lagrangian method for total variation video restoration. *IEEE Transactions on Image Processing*, *20*(11), 3097-3111.

- Image/video cutout: the goal of interactive image / video segmentation is to develop a user guided segmentation tool. The User provides hints to the region of interest and the segmentation tool will provide the result. The results can be further improved when the user changes hint points. An effective implementation is possible using graph-cut algorithm. (Contact: Amirhossein Khalilian akg404@nyu.edu)

Background needed: Good understanding of the image segmentation lecture in class. Graph signal processing is desired but not necessary.

Ref:

[1] Li, Y., Sun, J., Tang, C. K., & Shum, H. Y. (2004). Lazy snapping. *ACM Transactions on Graphics (ToG)*, *23*(3), 303-308.

[2] Wang, J., Bhat, P., Colburn, R. A., Agrawala, M., & Cohen, M. F. (2005). Interactive video cutout. *ACM Transactions on Graphics (ToG)*, *24*(3), 585-594. (Also see https://www.juew.org/projects/VideoCutout/VideoCutout.htm for more sample results.)

[3] Vicente, S., Kolmogorov, V., & Rother, C. (2008, June). Graph cut based image segmentation with connectivity priors. In *2008 IEEE conference on computer vision and pattern recognition* (pp. 1-8). IEEE.

[4] Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *International journal of computer vision*, *70*(2), 109-131.

- Denoising using deep networks with unknown and possibly spatially varying noise levels (Contact: Amirhossein Khalilian akg404@nyu.edu)

Background needed: Understanding of classical denoising techniques and implementing deep neural networks is desired.

Ref:

[1] Mohan, S., Kadkhodaie, Z., Simoncelli, E. P., & Fernandez-Granda, C. (2019). Robust and interpretable blind image denoising via bias-free convolutional neural networks. *arXiv preprint arXiv:1906.05478*.

[2] Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, *27*(9), 4608-4622.

[3] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2018). Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

[4] Liu, D., Wen, B., Jiao, J., Liu, X., Wang, Z., & Huang, T. S. (2020). Connecting image denoising and high-level vision tasks via deep learning. *IEEE Transactions on Image Processing*, *29*, 3695-3706.

- 360 degree or panoramic video or image stitching (contact: Yixiang Mao ym1496@nyu.edu)

Background needed: Image registration and warping

Ref:

Brown, Matthew, and David G. Lowe. "Automatic panoramic image stitching using invariant features." *International journal of computer vision* 74.1 (2007): 59-73.
Szeliski, Richard. "Image alignment and stitching: A tutorial." *Foundations and Trends in Computer Graphics and Vision* 2.1 (2006): 1-104.
Shum, Heung-Yeung, and Richard Szeliski. "Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment." *International Journal of Computer Vision* 36.2 (2000): 101-130.

- Medical Image registration (cross-modality, or within the same modality) (Contact: Amirhossein Khalilian akg404@nyu.edu)

- Few shot image segmentation using deep learning. (Contact: Ziming Qiu, zq415@nyu.edu)

Wang, Kaixin, et al. "Panet: Few-shot image semantic segmentation with prototype alignment." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

Liu, Jinlu, and Yongqiang Qin. "Prototype Refinement Network for Few-Shot Segmentation." *arXiv preprint arXiv:2002.03579*(2020).

- **3D Human Face Reconstruction w/o 3DMM face model from RGB image [Zhipeng Fan zf606@nyu.edu]**

In this project, you are required to implement a deep neural net based model to reconstruct the human faces from single RGB images. A statistical model (3DMM Face Model: https://ibug.doc.ic.ac.uk/media/uploads/documents/0002.pdf) will be employed to regularize the reconstructed face. A base approach will be to use the deep net to learn the parameters of the face model while an advanced version will be to employ an additional model to reconstruct details on the human face that is beyond the representation space of the 3DMM. Either the base model or the advanced model will be sufficient for the course project.

*** This project will use a little computer graphic skills, but necessary functions will be provided.

Base approach will be like the coarse part of:

Guo, Y., Cai, J., Jiang, B., & Zheng, J. (2018). Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. *IEEE transactions on pattern analysis and machine intelligence*, *41*(6), 1294-1307.

Advanced approach will be like:

Tran, L., & Liu, X. (2018). Nonlinear 3d face morphable model. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7346-7355).

Tran, L., Liu, F., & Liu, X. (2019). Towards high-fidelity nonlinear 3D face morphable model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1126-1135).

- **Improve neural network generalization for image classification with adversarial information bottleneck.** [Ran Wang, rw1691@nyu.edu]

The target of this project is to investigate improving the generalization power of deep neural networks. Previous works explain the network generalization from an information theory point of view [1]. During training, the network forgets redundant information of input whilst remembers information related to the task. Such information refinement is named as information bottleneck. In [3], such information bottleneck is explicitly used as a loss to improve network generalization ability. However the method only achieves an approximation of the actual information bottleneck.

In this project, we are going to borrow idea of the information bottleneck and develop a new approach for generalize the network for image classification task.

Requirement: Deep learning.

[1] Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle[C]//2015 IEEE Information Theory Workshop (ITW). IEEE, 2015: 1-5.

[2] Wu T, Fischer I, Chuang I L, et al. Learnability for the information bottleneck[J]. Entropy, 2019, 21(10): 924.

[3] Alemi A A, Fischer I, Dillon J V, et al. Deep variational information bottleneck. arXiv 2016[J]. arXiv preprint arXiv:1612.00410.

[4] Alemi A A, Fischer I, Dillon J V. Uncertainty in the variational information bottleneck[J]. arXiv preprint arXiv:1807.00906, 2018.

- Other possible projects in medical image analysis: you can pick one of the challenges in the link below and work on it. Please discuss with the instructor once you have several choices after you have read the background material. I will gauge whether you have sufficient background and time to do it or suggest a subset to work on.
https://grand-challenge.org/challenges/

- Other possible projects in computer vision: you can pick one of the challenges in the links below and work on it. Please discuss with the instructor once you have several choices after you have read the background material. I will gauge whether you have sufficient background and time to do it or suggest a subset to work on.
http://www.robustvision.net/

http://www.icme2019.org/conf_challenges

- Optical flow estimation using deep learning:

[PWCnet] Deqin Sun, Xiaodong Yang, Ming-Yu Liu, Jan Kautz, PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. CVPR 2018. https://arxiv.org/abs/1709.02371

[Flownet2] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, “Flownet 2.0: Evolution of optical flow estimation with deep networks.,” in CVPR . 2017, pp. 1647–1655, IEEE Computer Society
https://lmb.informatik.uni-freiburg.de/Publications/2017/IMKDB17/

[FlowFields] C. Bailer, B. Taetz, and D. Stricker. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In IEEE Int. Conference on Computer Vision (ICCV), 2015.

Older papers:
[DeepFlow] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. In IEEE Int. Conference on Computer Vision (ICCV), 2013.

[EpicFlow] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

- Depth estimation from single image:

Ref: Zhengqi Li and Noah Snavely, “Megadepth: Learning single-view depth prediction from internet photos,” in Computer Vision and Pattern Recognition (CVPR), 2018.

- Image inpainting: filling holes in an image

G. L. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, “Image inpainting for irregular holes using partial convolutions.,” in ECCV (11) . 2018, vol. 11215 of Lecture Notes in Computer Science , pp. 89–105, Springer.