University of Michigan
EECS 504: Foundations of Computer Vision
Winter 2020. Instructor: Andrew Owens.

Project proposal information

Posted: Wednesday, February 26, 2020 **Due:** Tuesday, March 10, 2020

Please submit your proposal to Gradescope as a PDF file. Please only submit one proposal per group, and add the other members of your group to your submission.

1 Proposal guidelines

For your final project, you'll study an area of computer vision in significantly more depth. All projects should involve reading research literature, implementing a computer vision system, and evaluating that system.

- You may work in a group of up to 4 people. If you'd like to work in a larger group, please chat with us; our expectations for your project will be higher.
- This project is open-ended: you can either choose from a list of project ideas from the list below, or you can propose a topic of your own. *We highly encourage you to propose your own project.* You'll probably get a lot more out of the experience!
- If you are proposing a new project, you should say: 1) what problem you are addressing, 2) why it is interesting or important, 3) what methods you'll use to solve it, 4) how you'll evaluate your work.

  If you select a project idea from the list, please describe in more detail what you'll do. For example, if you choose an idea from the list that involves reimplementing a paper, please: 1) briefly describe the algorithm in the paper, 2) describe what you will implement, and 3) say how you'll evaluate your model.
- Your project proposal should be short (less than one page is fine).
- Projects that overlap with your research (or your side projects) are generally fine.
- You don't necessarily need to use visual data for your project. You can also apply techniques described in the course to other signals (e.g. audio, LIDAR, medical imagery, etc.).
- If you'd like to significantly change the focus of your project after submission, you may revise your Gradescope submission. For example, if a topic covered later in the course grabs your attention, you may update your proposal to address it instead. We will then review the updated proposal.

- We unfortunately cannot provide GPU resources to groups, beyond what is available on Google Colab. Please keep this in mind when proposing a project. You may simplify your models (e.g. if you are reimplementing a paper, by training them on less data) to help address this.

- We encourage you to come to office hours to discuss your project ideas, so that we can help steer you toward relevant research literature. We may also point you to materials when we grade your proposal.

- The proposal will be worth 5% of your final project grade (and recall that your full final project is worth 30% of your overall class grade).

- Since this is a group project, you unfortunately cannot use late days for your deliverables.

2 Project ideas

We've provided a few project ideas below. Please note that these projects only cover a very small portion of the possible things you can do — most involve reimplementing and extending a paper. We encourage you to propose your own, creative project ideas, and to use these as a starting point! We may also add new project ideas to this list in the coming weeks.

Reconstructing historical scenes. Recently, Luo et al. [1] obtained a dataset of "antique" binocular stereo pairs, recorded using cameras from the 19th century. For this project, follow [1] and reconstruct the 3D structure of these historical scenes using a stereo depth estimation algorithm, and use a view synthesis algorithm to simulate new viewpoints of these scenes. Extend their work in some way (e.g. by trying to improve the depth estimation, inpainting, or view synthesis methods).

Video magnification. Implement a state-of-the-art motion magnification algorithm, such as the method of Wadhwa et al. [2] or a recent deep learning-based method. Try running it on your own videos, too.

Audio-visual source separation. Implement the audio-visual source separation method of Afouras et al. [3]. Optional extensions include replacing the audio network with an image translation network, adding a recurrent network for long-range synthesis, and using a different video representation.

Multi-view 3D reconstruction. Implement a simple structure from motion system, such as Bundler [4] or SFMedu. Apply the model to simple video sequences. Qualitatively demonstrate the effectiveness of your reconstructions (e.g. by using them as part of a view synthesis methods).

Instance segmentation. Implement Mask R-CNN [5]. You may need to simplify the model for performance reasons, such as by using a smaller CNN backbone and training on a subset of the data.

Semantic segmentation. Implement a semantic segmentation method, such as [6], and compare the results to other approaches that use a different network architecture (e.g. compare dilated convolution models to skip connection models).

Self-supervision. Implement a self-supervision method, such as MOCO [7], CMC [8], or audio-visual contrastive learning [9]. Experimentally evaluate how different design decisions affect performance on downstream classification tasks.

Colorizing historical photographs. Implement the interactive colorization method of [10]. Use it to provide color to old black-and-white photos, using a human in the loop to provide your model with helpful hints.

Unpaired image translation. Implement CycleGAN [11]. Extend the model using techniques from GAN literature, and experimentally evaluate how these changes affect performance.

Image captioning. Implement a neural captioning model, such as [12]. Compare the result of your model to a nearest neighbor approach [13].

Motion estimation. Implement an optical flow method, such as a recent neural network based approach [14, 15] or classic optimization-based methods [16, 17].

References

[1] Xuan Luo, Yanmeng Kong, Jason Lawrence, Ricardo Martin-Brualla, and Steve Seitz. Keystonedepth: Visualizing history in 3d. *arXiv preprint arXiv:1908.07732*, 2019.

[2] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013.

[3] Triantafyllos Afouras, Joon Son Chung, and Andrew Zisserman. The conversation: Deep audio-visual speech enhancement. *arXiv preprint arXiv:1804.04121*, 2018.

[4] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM Siggraph 2006 Papers*, pages 835–846. 2006.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[6] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[7] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *arXiv preprint arXiv:1911.05722*, 2019.

[8] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019.

[9] Relja Arandjelovic and Andrew Zisserman. Look, listen and learn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 609–617, 2017.

[10] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S Lin, Tianhe Yu, and Alexei A Efros. Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999*, 2017.

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[12] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[13] Jacob Devlin, Saurabh Gupta, Ross Girshick, Margaret Mitchell, and C Lawrence Zitnick. Exploring nearest neighbor approaches for image captioning. *arXiv preprint arXiv:1505.04467*, 2015.

[14] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8934–8943, 2018.

[15] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2462–2470, 2017.

[16] Deqing Sun, Stefan Roth, and Michael J Black. Secrets of optical flow estimation and their principles. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 2432–2439. IEEE, 2010.

[17] Jia Xu, René Ranftl, and Vladlen Koltun. Accurate optical flow via direct cost volume processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1289–1297, 2017.