Computational Photography CS 413
Project Proposals 2020

February 2020

1 Introduction

Below is a list of project proposals for the CS 413 course, Spring semester 2020. Clarifications about each proposal can be obtained from the corresponding supervisor TA. The **deliverables** explain what is expected from you to submit by the end of the semester, aside from presentations/reports.

2 Projects

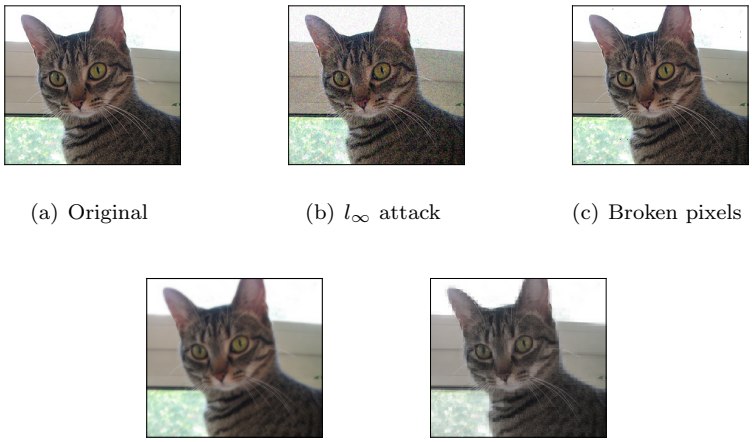
2.1 Attack v.s. Defense on Corrupted Images (2 Teams)

Synopsis: Recent studies have shown that image classifiers based on deep learning models are vulnerable to adversarial attacks [16]. Small but imperceptible perturbations can cause the state-of-the-art models give wrong predictions with very high confidence. The research on robust image classification is divided into two parts: from the attack side, we search for the input perturbations that do not change the semantic meanings of the image but fool the classifier; from the defense side, we need to construct the models resistant to these adversarial attacks. Formally, we let $g(\theta, x)$ be the loss function of the model parameter θ and input x. For a training set D, we are interested in the following problem:

$$\min_{\theta} \mathbb{E}_{x \sim D} \max_{x' \in \mathcal{S}_\epsilon(x)} g(\theta, x') \tag{1}$$

The attack focuses on the inner maximization problem while the defense focuses on the outer minimization problem. In addition, $\mathcal{S}_\epsilon(x)$ is the set of all allowed perturbed images of x, whose size is parameterized by ϵ. In many existing works [8], the set $\mathcal{S}_\epsilon(x)$ is defined by a l_p norm based ball in the neighborhood of x: $\mathcal{S}_\epsilon(x) = \{x' | \|x - x'\|_p \leq \epsilon\}$. Here we study other reasonable settings of $\mathcal{S}_\epsilon(x)$:

1. Broken pixels: [14] $\mathcal{S}_\epsilon(x) = \{x' | \|x - x'\|_0 \leq \epsilon\}$ means we can perturb at most ϵ pixels. We do not constrain the magnitude of perturbation to each perturbed pixel.



(a) Original (b) l_∞ attack (c) Broken pixels

(d) Blurred image (e) Mosaic image

Figure 1: The original image and different kinds of perturbed images.

2. Blurred images: $\mathcal{S}_\epsilon(x) = \{x \circledast W | \|W - I\|_1 \leq \epsilon\}$ means the blurred images by a convolutional kernel $W \in \mathbb{R}^{2k+1 \times 2k+1}$. Here I represent the identity kernel i.e. $I_{k,k} = 1$ and other entries of I are 0.

3. Mosaic images: In each $k \times k$ image patch, we calculate the average value $x^{(patch)}$. For each pixel x, the mosaical pixel $x' = \epsilon x^{(patch)} + (1 - \epsilon)x$.

Figure 1 shows some examples of perturbed images under different adversarial budgets. We encourage students to explore other reasonable image perturbations. We use CIFAR10 and CIFAR100 as the main dataset of our experiments.

This project has two teams. The first half of the project is the same: finding adversarial examples in the sets defined above. Since we initially have no idea about the reasonable values of ϵ in each case, we then ask each team to find the smallest value of ϵ such that they can fool the prediction accuracy below a threshold e.g. 20%. The team finding a smaller value of ϵ wins bonus points.

The second half of the project is defense, we need to construct a robust model resistant against such perturbations. In this part, one team is allowed to denoise the perturbed images but not allowed to change the training method of the model. That is to say, the model must be trained in the normal way, but denoising technique such as [17] can be applied to 'remove' the adversarial perturbations. On the contrary, other team is allowed to change of model training algorithm, such as adversarial training [12], but not allowed to post-processing the input data. We will compare the performance of both teams and cross-test their methods i.e. test the performance of team A's method under team B's attack. The teams will lose points if their cross-test performance is significantly worse than their reported one.

Deliverables: Attack and defense algorithms for each kind of adversarial budgets.
Supervised by: Chen
Bibliography: [8] This paper points out the existence of the adversarial attacks and propose adversarial training.
[12] This paper propose projected gradient descent (PGD), the current most popular defense algorithm against l_p attack.
[14] This paper introduces the one-pixel attack, which fool the classification model by perturbing just one pixel.
[17] This paper introduce denoising auto-encoders to construct robust features.

2.2 Microscopy Deblurring

Synopsis: Capturing a full scan of an eye, for medical examination, is very challenging. As visualized in Fig. 2, the cornia has a spherical shape, but an imaging system can only focus (at its maximum sharpness) on a given depth range in the scene. You might have witnessed this when capturing photos, all objects closer or further away from the object at which you focus your camera end up blurry/out of focus. While this can be used for artistic effects in photography, it is very detrimental in biomedical imaging. The goal of this project is to deblur the cornia scans, as much as possible while reconstructing faithfully the original signal. It is probably not possible to deblur the entire range of depth values, but the more we can extend the depth of field of the imaging system, the fewer shots need to be captured, and the less time patients need to sit in front of the imaging system. One challenge in this project is that you are not provided with a lot of data. Your approach should take that into account, and could possibly be developed as a method that needs retraining for new data (this can also be a good advantage of your method, since it means it can apply well to new imaging systems, or imaged content, without the need for a large dataset each time).
Deliverables: A method that takes as input single-focus cornea scans, with a certain depth range that is originally in focus, and extends the depth of field by deblurring as much as possible around that range. The deblurred result must be reliable (faithful reconstruction of the true underlying signal), as your results will be evaluated by a cell-counting algorithm and a biomedical imaging expert.
Supervised by: Majed
Bibliography: [5] presents a method to measure the extent of axial chromatic aberration of a given imaging system, [7] uses that phenomenon for the depth-from-defocus application, and [6] uses it to deblur an image channel with other channels as guide. They can give you a better understanding of chromatic aberration, and how to best work with multi-channel (RGB, or more) images.
[4] explains how to solve large-image optimizations very efficiently using the frequency domain. You can use these techniques to (1) estimate blur kernels, (2) deblur an image with a known degradation kernel (called non-blind deblurring), or (3) set up your own optimization functions and solve them efficiently.
If you work instead with deep-learning-based methods, the literature is very

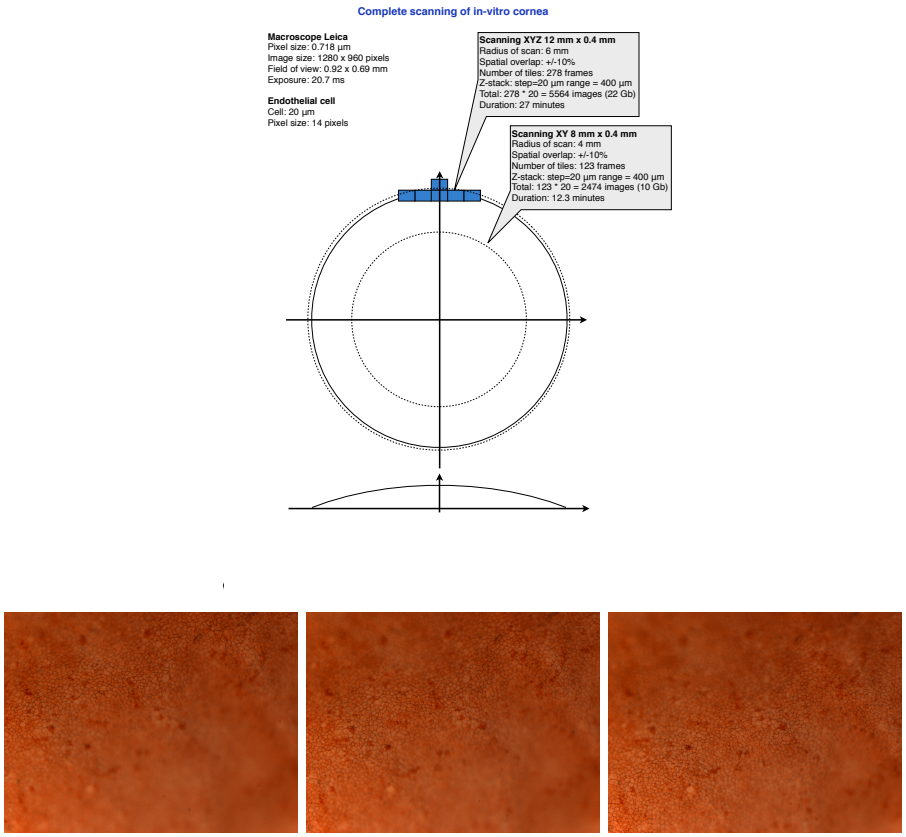


Figure 2: Setup illustration and some sample scans.

Figure 3: A representative indoor scene in the dataset. Each scene is captured under 40 unique lighting conditions, 12 of which are shown in the figure (4 color temperatures with 3 light positions).

broad and we can guide you during the project to make sure you follow best practices.

2.3 2D Scene Relighting

Synopsis: We used Unreal Engine to render high-resolution and realistic scenes to build a dataset that includes 396 different scenes (300 of which are used for developing a solution, and the rest for validation/testing), each with 40 predetermined lighting conditions: the combinations of 5 color temperatures (2500, 3500, 4500, 5500 and 6500) and 8 light directions (NW, N, NE, E, SE, S, SW, W). The dataset includes both indoor and outdoor scenarios, and miscellaneous objects with different surfaces and materials. All scenes were rendered with a resolution of 1024×1024, and we saved the raw (linear) and sRGB images. The lighting conditions are recorded with the rendered images. A scene example is presented in Fig. 3. Your objective is to develop a (possibly deep learning based) method to transform one scene illumination into another illumination (with a guide target image).
Deliverables: A method that takes in 2 different input scenes, with different illumination, and transforms the illumination of one scene image to match the other target image. The problem can potentially be divided into an illuminant

change subproblem, and a lighting direction change subproblem.
Supervised by: Majed
Bibliography: [13] This paper presents a dataset of interior scenes, mostly objects in small shooting areas. Each scene is captured with a flash coming from 25 different orientations. The illuminant itself is however not varied.
[18] This paper presents a dataset of pairs of underexposed images which is used to train a novel network. The dataset is used rather than the MIT-Adobe FiveK dataset, as the latter focuses on general photo enhancement, while the authors focus specifically on enhancing underexposed photos. The ground-truth enhanced images are created by photography experts using Adobe Lightroom.
[19] This paper, similar to [13], also builds a dataset of scenes with different light directions. The differences are that the images are rendered and that the light directions are randomized.
[15] This paper aims at relighting human portrait photos. The method is trained on a dataset consisting of 18 individuals captured under different directional light sources. The capture is under controlled settings, with the individual illuminated by a sphere with numerous lights. This is specifically targeted at face relighting, and does not extend to general scenes. Indeed, even the background is lost in the re-lit images.
IIW [2] and SAW [10] contain human-labeled reflectance and shading annotations, BigTime [11] contains time-lapse data where the scenes are illuminated with different lighting conditions.

2.4 Predicting Photographers' Retouchings with Deep Learning

this project can only be selected by a team if all other projects are already taken
Synopsis: In this project you will use a dataset[1] of professional photographers' image manipulations from [3] and we ask you to build a machine learning system that learns experts' edits from the dataset. Figure 4 shows two different manipulated images of the same raw-RGB image by two professional photographers. Where Figure 4.A shows the Expert A's preferred edits, Figure 4.B shows the Expert B's preferred edits. This is useful when automatically generating similar photographic manipulations based on the examples in a set of manipulated images. This is a well studied research area where you can find both traditional methods, e.g., [3], as well as recent deep learning based methods, e.g., [9].

As a baseline, all the teams are required to implement the methods in [3]. Leveraging the monotonicity properties of the retouching parameters the Team 3 will implement the approach proposed in [1]. The Team 2 will follow the approach proposed in [20]. Generalization between the two approaches will be tested on a set of images that will be provided at the end of the semester.
Deliverables: A report and a running prototype of a deep neural network that takes a raw image as input and produces an expert's preferred rendering. Comparison with the baseline methods [3], as well as the state-of-the-art [9]

[1] https://data.csail.mit.edu/graphics/fivek/

A. Photofinishing as applied by the expert A B. Photofinishing as applied by the expert B

Figure 4: Two different renderings of the same raw-RGB image that is manipulated by two different experts: A. Photofinishing as applied by the expert A, B. Photofinishing as applied by the expert B.

and improving it. In deep-learning approaches, you need to train the network for each different expert–for example, two different models are required for two experts.

Supervised by: Hakki

Bibliography:

[3] presents the dataset and employs the traditional baseline methods where you can compare your results.

[9] is a recent work where authors train a deep network for this task.

References

[1] Tensorflow lattice. https://www.tensorflow.org/lattice/overview. Accessed: 2020-03-02. 6

[2] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4):159, 2014. 6

[3] V. Bychkovsky, S. Paris, E. Chan, and F. Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *CVPR*, 2011. 6, 7

[4] M. El Helou, F. Dümbgen, R. Achanta, and S. Süsstrunk. Fourier-domain optimization for image processing. *arXiv preprint arXiv:1809.04187*, 2018. 3

[5] M. El Helou, F. Dümbgen, and S. Siisstrunk. Aam: An assessment metric of axial chromatic aberration. In *IEEE International Conference on Image Processing (ICIP)*, pages 2486–2490, 2018. 3

[6] M. El Helou, Z. Sadeghipoor, and S. Süsstrunk. Correlation-based deblurring leveraging multispectral chromatic aberration in color and near-infrared joint acquisition. In *IEEE International Conference on Image Processing (ICIP)*, pages 1402–1406, 2017. 3

[7] M. El Helou, M. Shahpaski, and S. Süsstrunk. Solving the depth ambiguity in single-perspective images. *OSA Continuum*, 2(10):2901–2913, 2019. 3

[8] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 1, 3

[9] Y. Hu, H. He, C. Xu, B. Wang, and S. Lin. Exposure: A white-box photo post-processing framework. In *SIGGRAPH*, 2018. 6, 7

[10] B. Kovacs, S. Bell, N. Snavely, and K. Bala. Shading annotations in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6998–7007, 2017. 6

[11] Z. Li and N. Snavely. Learning intrinsic image decomposition from watching the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9039–9048, 2018. 6

[12] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017. 2, 3

[13] L. Murmann, M. Gharbi, M. Aittala, and F. Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4080–4089, 2019. 6

[14] J. Su, D. V. Vargas, and K. Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, 2019. 1, 3

[15] T. Sun, J. T. Barron, Y.-T. Tsai, Z. Xu, X. Yu, G. Fyffe, C. Rhemann, J. Busch, P. Debevec, and R. Ramamoorthi. Single image portrait relighting. *ACM Transactions on Graphics (TOG)*, 38(4):79, 2019. 6

[16] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013. 1

[17] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008. 2, 3

[18] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6849–6857, 2019. 6

[19] Z. Xu, K. Sunkavalli, S. Hadap, and R. Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)*, 37(4):126, 2018. 6

[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 6