Advanced Machine Learning: Image-to-Image Translation

The aim of the last coursework is to gain a deeper understanding into unsupervised image-to-image translation techniques and put them into a unified context. Each student will 1) critically analyze and 2) experimentally compare existing work.

1 Tasks

Task 1: Discussion and analysis of existing work. This will focus on recent advances in image(-to-image) translation: Each student should discuss at least five academic publications including CycleGAN [7] (or DiscoGAN [4])[1] appeared in the past four years. There are no constraints on the nature of the contributions the papers make, e.g. they can present new theories, algorithms or applications, but they should be published in reading machine learning or computer vision conferences or journals, e.g. NeurIPS, ICML, ICLR, JMLR, ML, CVPR, ICCV, ECCV, TPAMI, and IJCV. The students are expected to demonstrate an understanding of image translation by conveying the main ideas and contributions, and putting existing work in context, not just provide a list of papers and describe each individually.

Task 2: Empirical evaluation. Among the (minimum) five algorithms that are analyzed in Task 1, each student should perform experiments with and experimentally compare at least three different image translation techniques including either CycleGAN [7] or DiscoGAN [4]. Examples include but are not limited to [2, 5, 3, 1, 6] and the methods listed in `https://paperswithcode.com/task/image-to-image-translation`. Students do not have to implement these techniques by themselves: They can use code provided by the authors of the respective publications. We will consider two image translation datasets:

- *Map* $\leftrightarrow$ *Aerial photo* dataset provides collections of maps and aerial photos downloaded from Google Maps. The dataset is available for download at the website of a CycleGAN author: `https://people.eecs.berkeley.edu/~taesung_park/CycleGAN/datasets/maps.zip`. Here, the two domains are *inherently* related. Figure 1 shows example image translation results.[2]

- *OSR* $\leftrightarrow$ *Pubfig* dataset consists of outdoor scene images (*OSR*) and face images (*Pubfig*). This task (dataset) is much more challenging to translate than *Map* $\leftrightarrow$ *Aerial photo* as there's no underlying connection between the two domains. The accompanying 'OSRToPubfig.zip' file contains images of the respective domains: 'OSR.zip' and 'Pubfig.zip' therein provide images of the respective domains.

2 Assessment and deadlines

2.1 What to hand in for assessment

This task will be assessed based on a report. Students can be asked to submit their network weights and the translated images later: Please keep the trained network weights and coding environment (e.g. venv) such that the experiments described in the submitted report can be reproduced. If a student is asked to submit their code and weights, and they fail to do so or if the submitted code cannot reproduce the results presented in the report, the final grade will be kept at the 50% of the original grade.

Students will be assessed based on:

[1] CaycleGAN and DiscoGAN are very similar to each other. Machine learning is a rapidly evolving and expanding field. It is often the case that multiple groups independently contribute with the same or similar ideas.

[2] Kindly provided by Ms. Insoo Kim.

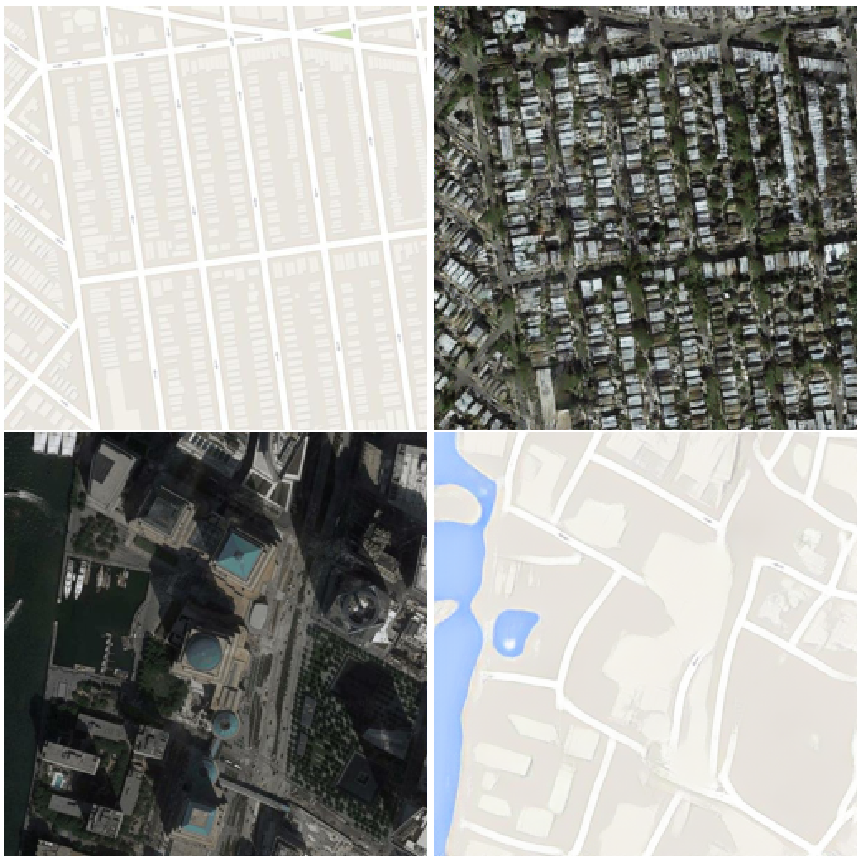

Figure 1: Example image translation results. Top: source map image (left) and the corresponding translated aerial photo (right). Bottom: source aerial photo (left) and the translated map (right).

- Task 1 (20/100)
  - Understanding of the majority of the contents of the papers;
  - Identified and explained the key contributions and how the respective authors convincingly support their claims (e.g. via experiments or theoretical analysis);
  - Put each paper in the context of others.
- Task 2 (20/100):
  - How convincingly they communicate their findings via, e.g. trying multiple experimental settings and an in-depth analysis of experimental results including the effect of varying hyper-parameter and loss choices.
  - Scientific conclusions and how they are drawn from the experiments. In the report, please discuss how the translation quality differ in *Map* $\leftrightarrow$ *Aerial photo* and *OSR* $\leftrightarrow$ *Pubfig* cases.
- Task 1 and 2 (20/100)
  - Quality of written content: You are encouraged to use figures (of course!). Please make your report self-contained: The submitted report should include Introduction and Conclusion sections and it should define the problem.
- Additional effort (40/100): The remaining marks are allocated for any additional effort that you make in your experiments and report. Examples include
  1. Evaluating (performing experiments with) more than three techniques.
  2. Ablation study, e.g. adding new loss functions, new training (weight update) scheme, and network architectures.
  3. Experiments with and critical analysis on additional datasets.

Submit a single pdf file via Black Board: Please format the submission as in **'StudentID_Name.pdf'**. If your submission does not meet this file name requirement, the final mark will be kept at 95% of the initial mark.

Report format. We will follow the 2020 Conference on Neural Information Processing Systems (NeurIPS) format. The format kit can be downloaded from NeurIPS2020 website: `https://nips.cc/Conferences/2020/PaperInformation/StyleFiles`

The recommended section structure of the report is

1. Introduction
2. Analysis of existing work
3. Experiments
4. Conclusions

with additional sections (if necessary) and subsections.

Your report should be no longer than 12 pages excluding references and appendix.

2.2 Deadline

The submission deadline is December 23 2020, 5pm (with a 30-minute margin to accommodate possible Internet connection issues): If a CW is uploaded later than November 19 2020, 5:30pm, the maximum possible mark will be kept at 40% of the full mark. If work is submitted more than three working days after the submission date, student will receive zero mark.

3 Plagiarism

Plagiarism is the attempt to pass-off someone else's work as your own, with the intention or expectation of receiving credit for it. It diminishes the degree you are on, the university you are in, and yourself; it is wholly unprofessional. Plagiarism can be intentional or un-intentional, but it is still plagiarism. You are free to use any resource you like as part of this coursework. But if you do use anything at all from anyone else, or anywhere else, then you must give *full credit* inside your document.

References

[1] D. Bhattacharjee, S. Kim, G. Vizier, and M. Salzmann. Dunit: detection-based unsupervised image-to-image translation. In *Proc. IEEE CVPR*, 2020.

[2] Y.-C. Chen, X. Xu, and J. Jia. Domain adaptive image-to-image translation. In *Proc. IEEE CVPR*, 2020.

[3] A. Gonzalez-Garcia, J. van de Weijer, and Y. Bengio. Image-to-image translation for cross-domain disentanglement. In *NeurIPS*, 2018.

[4] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proc. ICML*, 2017.

[5] Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim. Image to image translation for domain adaptation. In *Proc. IEEE CVPR*, 2018.

[6] Y. Zhao, R. Wu, and H. Dong. Unpaired image-to-image translation using adversarial consistency loss. In *Proc. ECCV*, 2020.

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