

# Project Proposal (COMP-5900V 2021 Winter)

Ao Peng, 300145650

## 1 Introduction

The topic of this paper I chose is Image-to-Image Translation with Conditional Adversarial Networks [1], which was released in CVPR 2019. This paper proposed a new GAN model named conditional GANs (cGANs) for a specific image processing area (image-to-image translation) as a common framework in replace of traditional methods [2] [3] [4], etc. The results in this paper demonstrate that cGANs is a promising method for image-to-image translation tasks, especially for tasks involving highly structured graphical outputs.

## 2 Problem description

In fact, traditional methods for image translation tasks are actually based on the same setting: predict pixels from pixels. However, these methods are separate and special-purpose for each task. The coming of convolutional neural networks (CNNs) plays a significant role in this direction while it will tend to produce blurry results because Euclidean distance is minimized by averaging all plausible outputs. Fortunately, the generative adversarial networks (GAN) [5] will not tolerate this kind of blurry results as it mainly focus on classifying the image is real or fake which means that it can result in sharper and clear outputs.

In this paper, the author is trying to develop a common approach for tackling all problem in this area. As GANs learn a generative model of data, cGANs learn a conditional generative model of data which strength the ability of addressing image-to-image translation tasks. In addition, a new term named PatchGAN in the paper plays a key role in this framework on improving the the performance.

## 3 Project scope and objectives

- Implementing cGANs
- Training the model on Cityscapes Dataset [6] for semantic segmentation task and UT Zappos50K [7] [8] for edge to shoes task.
- Evaluating the results of cGANs

## References

- [1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” 2018.
- [2] A. A. Efros and W. T. Freeman, “Image quilting for texture synthesis and transfer,” in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 341–346.
- [3] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, “Removing camera shake from a single photograph,” in *ACM SIGGRAPH 2006 Papers*, 2006, pp. 787–794.
- [4] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2650–2658.
- [5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *arXiv preprint arXiv:1406.2661*, 2014.
- [6] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [7] A. Yu and K. Grauman, “Fine-grained visual comparisons with local learning,” in *Computer Vision and Pattern Recognition (CVPR)*, Jun 2014.
- [8] —, “Semantic jitter: Dense supervision for visual comparisons via synthetic images,” in *International Conference on Computer Vision (ICCV)*, Oct 2017.