# EE599 Deep Learning – Initial Project Propsal

Haojing Hu, Zheng Wen

**Mentor: Jiali Duan**

April 18, 2020

**Project Title:** Object transfiguration with GANs

**Project Team:** Haojing Hu, Zheng Wen

**Project Summary:** In this project, we propose to do an unpaired object transfiguration focusing on some regions of interest (ROI), leaving other part of the image unchanged. There exists many well-organized recent works focusing on the image-to-image translation with different GAN pipelines. Instead of doing the translation on the whole image like cycleGAN[1], some new methods proposed in [2][3] use attention mechanism to guide the GAN to capture ROI and learn desired texture, making changes on color and texture simultaneously without changing the other part of the image. However, this algorithm is still limited to the color and texture changes of the object and can not change the shape or contours of the objects. For example, these algorithms are well done in the changes like zebra to horse, tiger to leopard, but have bad performance in tasks like dog to cat, where more significant changes in shape is contained. One method we find to address this problem is the DRIT[4] algorithm. The main problem of this algorithm is that it cannot fully disentangle the content and the background and the degree of shape changing is also not so obvious. In [5], the dilated discriminator and several loss like MS-SSIM are added to cycleGAN, forming GANimorph, making salient changes with unpaired images possible.
Based on these recent works, especially GANimorph proposed in [5], we want to propose a more versatile object transfiguration method which can localize the ROI as well as do the shape transformation using unpaired images. A desirable outcome would be a generated image which has a different object(with different shape, color and texture) but almost the same background compared to the original image.

**Data Needs and Acquisition Plan:** For this task, we need images of two types of objects (e.g. Apples vs. Oranges) in the train stage. In the test stage, we need some images containing only one type of the interested objects to change.
Candidate datasets:
**Fruit 360**[6]: including various kinds of fruits and vegetables
**ImageNet**[7]: choosing a subset including images of some classes
**Images crawled from BING**[8]: finding some images of different types of guns for formal training.

**Primary References and Codebase:** We propose to build on the method used in

- Attention aided GANs [2][3]
- DRIT[4]
- GANimorph[5]
- GitHub codebases: Attention Guided GAN code, Attention GAN code, DRIT code, GANimorph code

**Architecture Investigation Plan:** We plan to first implement the GANimorph code in its given dataset to make sure it works. Then we would train the model on the dataset collected and builded by us to see its performance and do some fine-tuning and some hyperparameter changes to get a relatively good result.
Next, we will seek to borrow and combine the insights from other GAN-based image translation methods to further improve the results, including adding or modifying the loss function, changing the network architectures and adopting some post-processing techniques. The ultimate goal of our model is to only transform the object of interest in the given image such as transforming the gun holded by a soldier but leave other background unchanged. Also, we will try to compare our final results with other proposed methods both quantitatively and qualitatively. Due to the time limit, outperforming other existing methods may be hard to achieve, but we will still provide some relatively good results generated by us in the final presentation.

Here are three millstones we set up to this challenging task:
1. Setting up GANimorph model and training on the given dataset.
2. Trying to improve the GANimporph method according to other reference
3. Comparing our results with other proposed methods

**Estimated Compute Needs:** Based on the data set size in the original paper, the structure of the neural network, and in this Lamnbda Labs Blog, we estimate that one training run for the GAN architecture with data set obtained by us will take 30 hours on a single Nvidia V100 GPU, which is the GPU resource in the AWS p3.2xlage instance. With spot pricing, which is roughly $1 per hour, we expect $30 per training run. We will train the network tentatively for several times to finetune the structure as well as the hyperparameters, which is estimated to cost $50. To get a better result, we expect to train approximately 2 full runs which brings our total estimated computing cost to roughly $110. To get the training faster, we could also use colab-pro to help our model out, where Tesla P100 GPU is available but kind of unstable.

**Team Roles:** The following is the rough breakdown of roles and responsibilities we plan for our team:

- Haojing Hu: Data collection, Data pre-processing, Video production.
- Zheng Wen: Model construction and modification, Data loader design.

All team members will work on the training and testing of the model, final presentation, slides and report.

**Requested Mentor with Rationale:** We request Jiali to be our team mentor because of his expertise in GANs and computer vision. We have a good codebase and few related thesis to fine-tune our own model, but we are not experienced in training GANs, so even though we are flexible regarding our mentor assignment, we expect someone expert in GANs as our mentor.

## References

[1] P. I. A. A. E. Jun-Yan Zhu, Taesung Park, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv:1703.10593v6 [cs.CV] 15 Nov 2018*, 2018.

[2] Y. A. Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," in *Advances in Neural Information Processing Systems*, 2018, pp. 3693–3703.

[3] X. Chen, C. Xu, X. Yang, and D. Tao, "Attention-gan for object transfiguration in wild images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 164–180.

[4] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 35–51.

[5] A. Gokaslan, V. Ramanujan, D. Ritchie, K. In Kim, and J. Tompkin, "Improving shape deformation in unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 649–665.

[6] [Online]. Available: https://www.kaggle.com/moltean/fruits

[7] [Online]. Available: http://www.image-net.org/

[8] [Online]. Available: https://www.bing.com/images