

Photographic Text-to-Image Synthesis via Multi-turn Dialogue using Attentional GAN

MECE Thesis - Proposal Defense

Shiva K. Shrestha¹

¹ Department of Graduate Studies
Nepal College of Information Technology (NCIT)

4 January 2019



Overview

- 1 Introduction
 - Background
 - Problem Statements
 - Motivation
 - Research Objective
 - Scope of Thesis
- 2 Literature Review
- 3 Methodology
 - Proposed Solution
 - Proposed Solution (Contd.)
 - Datasets
 - Tools & Platforms
- 4 Expected Outcomes
- 5 Thesis Schedule
- 6 References



Background

- Based on AttnGAN, GAN Introduced by Ian Goodfellow
- GAN : 2 N/ws, D - Better Detective, G - Faking Artist
- After sufficient epochs, G generates surprisingly realistic image
- A unique solution exists, at that pt. D tells real for generated image.

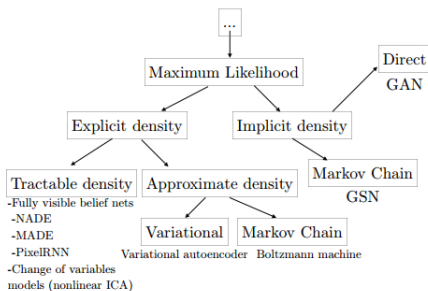


FIGURE – Taxonomy of Generative Models (GAN History & Overview CC-By-Ian)

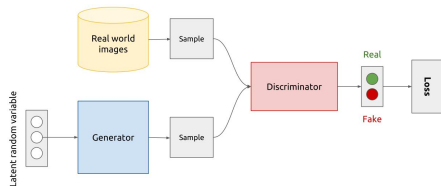


FIGURE – GAN Framework (SlideShare CC-By-Kevin McGuinness)



Problem Statements

- Most methods are based on GANs.
- Whole text-description encoded into a global sentence vector
- Stacked GANs [3] lack to use conversational text
- Proposed to solve the generation of photo-realistic images from the given text-description via multi-turn dialogue.



FIGURE – Flower Synthesis from : "The flower has red petals with yellow stigmas in the middle" [7]



FIGURE – Bird Generation by StackGAN++ [3]



Motivation

- GAN [5] is a recently introduced but evolving topic.
- Build a better model to generate or predict images, videos, anime character, object, etc.
- Image generation by user's conversation
- Imagination to computer-generated objects



Thesis Objective

- To generate photo-realistic image from given text description via multi-turn dialog using Attentional GAN



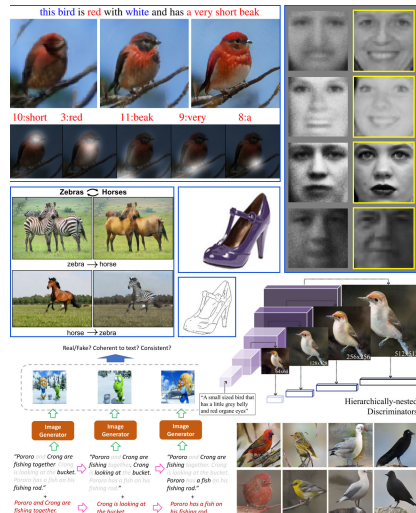
Scope of Thesis

- Text-to-Image Synthesis,
- High-resolution Image Synthesis,
- Intelligent Image Manipulation,
- Face Synthesis,
- Image Editing, etc.



Literature Review

- 1 Generative Adversarial Nets for Text-to-Image Synthesis [5]
 - Goodfellow et al. (2014)
- 2 Image Generation and Editing [1, 7]
 - AttnGAN, Xu et al. (2018)
 - Chen et al. (2017)
- 3 Image-to-Image Synthesis [4]
 - CycleGAN, Zhu et al. (2018)
- 4 Realistic Image Synthesis with Stacked Generative Adversarial Networks [3]
 - StackGAN++, Zhang et al. (2018)
- 5 Hierarchically-nested Adversarial Network for High-Quality Image Generation [2]
 - HDGAN, Zhang et al. (2017)
- 6 Story Visualization [6]
 - StoryGAN, Li et al. (2018)



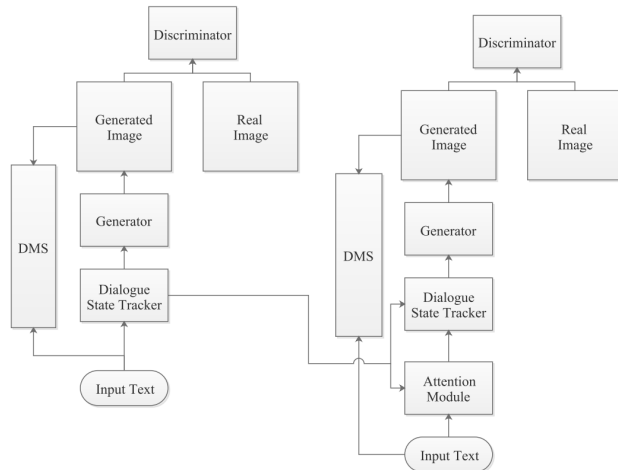
Proposed Solution

Generative research aims to learn a mapping from a semantic text space to a complex RGB image space.

Algorithm :

Steps

- 1 Start
- 2 Input Text/User Feedback
- 3 Text Encoder
- 4 Dialogue State Tracker
- 5 Context Feature
- 6 Generate Image
- 7 Similarity matching with image features
- 8 Again ? goto Step2.
- 9 Stop



Proposed Solution (Contd.)

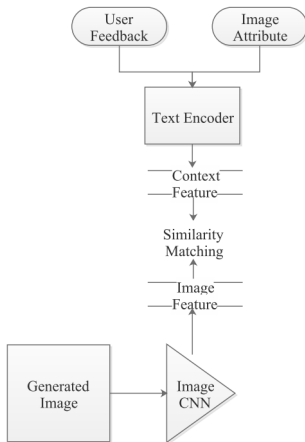


FIGURE – Deep Multimodal Similarity Regularize (DMS)
part of Proposed Work

DMS Regularizer

- Calculates Image description matching score in each step.
- Unconditional Loss : Image real or fake ?
- Conditional Loss : Image Sentence match or not ?
- DMS : Generate region-specific features
- Used for well descriptions & Visual diversity



Datasets

- 1 CelebFaces Attributes Dataset (CelebA)
 - 200K+ celebrity images
 - Each with 40 attribute annotations
 - 10,177 number of identities
- 2 Caltech-UCSD Birds-200-2011 (CUB-200-2011)
 - Number of categories : 200
 - Number of images : 11,788
 - Annotations per image : 15 Part Locations, 312 Binary Attributes, 1 Bounding Box
- 3 Common Objects in Context (COCO)
 - 330K images (>200K labeled)
 - 1.5 million object instances
 - 80 object categories
 - 91 stuff categories
 - 5 captions per image



Tools & Platforms

- 1 Python Programming Language
- 2 TensorFlow Framework
- 3 Python Packages
 - Pillow
 - SciPy
 - python-dateutil
 - easydict



Expected Outcomes

- 1 Photo-realistic image will be generated during conversational phases
- 2 Experiments will be on three datasets :

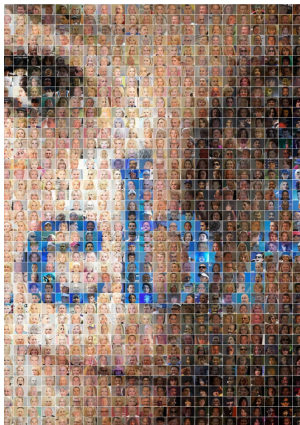


FIGURE GGGG



FIGURE GGGG

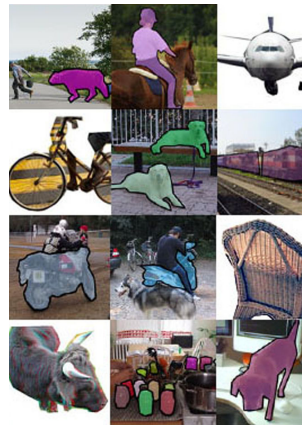


FIGURE GGGG

Thesis Schedule

Tasks	Month/Year 2018/2019					
	Dec	Jan	Feb	Mar	Apr	May
Preliminary Investigations	■					
Literature Review	■	■	■	■		
Proposal Defense	■	■				
System Design & Coding		■	■	■	■	
Mid-Term Defense			■	■		
Final Submission of Thesis					■	■
Documentation of Thesis		■	■	■	■	
Research and Experiments			■	■	■	■

FIGURE – Thesis Schedule



References

- 1 Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., and He, X. (2018). AttnGAN : Fine-grained text to image generation with attentional generative adversarial networks. CoRR, abs/1711.10485, 2017.
- 2 Chen, Q., Koltun, V. (2017, October). Photographic image synthesis with cascaded refinement networks. In IEEE International Conference on Computer Vision (ICCV) (Vol. 1, No.2, p. 3).
- 3 Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. N. (2017) StackGAN++ : Realistic image synthesis with stacked generative adversarial networks. arXiv : 1710.10916, 2017.
- 4 Zhu, J. Y., Park, T., Isola, P., Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint.
- 5 Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
- 6 Li, Y., Gan, Z., Shen, Y., Liu, J., Cheng, Y., Wu, Y., ... Gao, J. (2018). StoryGAN : A Sequential Conditional GAN for Story Visualization. arXiv preprint arXiv :1812.02784.
- 7 Chen, J., Shen, Y., Gao, J., Liu, J., Liu, X. (2017). Language-Based Image Editing with Recurrent Attentive Models. arXiv preprint arXiv :1711.06288.

