

# Final Project

# StackGAN

0752604 林亭家 0752212 劉冠睿 0750272 俞宗佑 0416106 彭敬樺

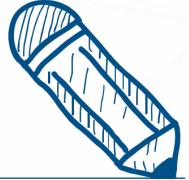


# OUTLINE

- 1) Motivation
- 2) Introduction
- 3) Problem Formulation
- 4) Proposed Model
- 5) Experiments & Results
- 6) Conclusion

# MOTIVATION

---

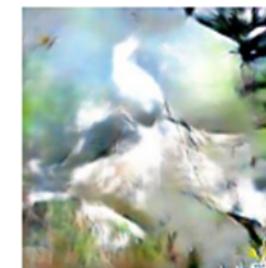


Generating images from text is important and has tremendous applications.

Problem:

Synthesizing high-resolution photo-realistic images from text.

This bird is white with some black on its head and wings, and has a long orange beak



This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face



This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments



# INTRODUCTION

---



- Propose a novel Stacked Generative Adversarial Networks
- Propose a Conditioning Augmentation technique
- Demonstrate the effectiveness of the overall model design

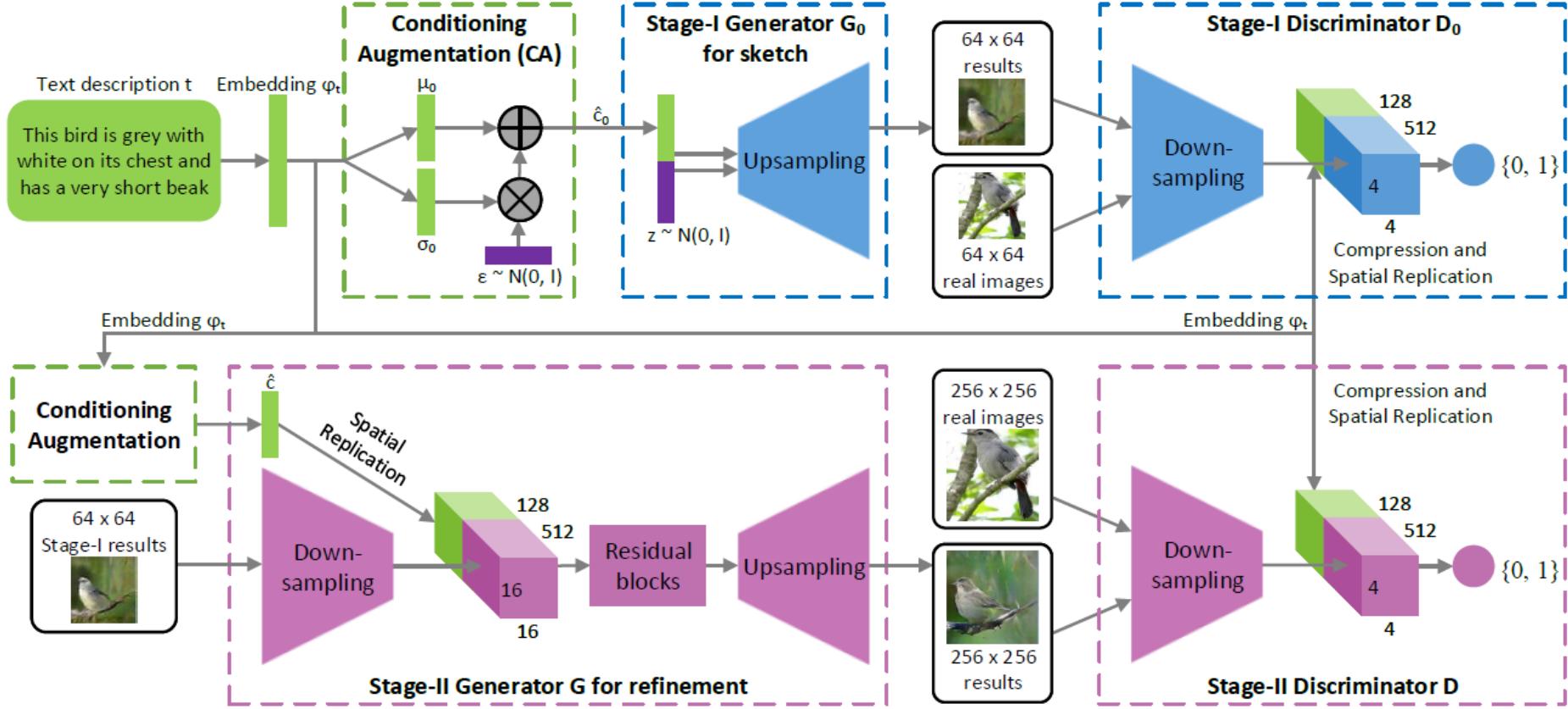
# PROBLEM FORMULATION



- Model input : Text description
- Stage 1 output :  $64*64$  Low resolution graph
- Stage II input:  $64*64$  Graph + text conditioning variable
- Model output :  $256*256$  High resolution graph



# PROPOSED MODEL

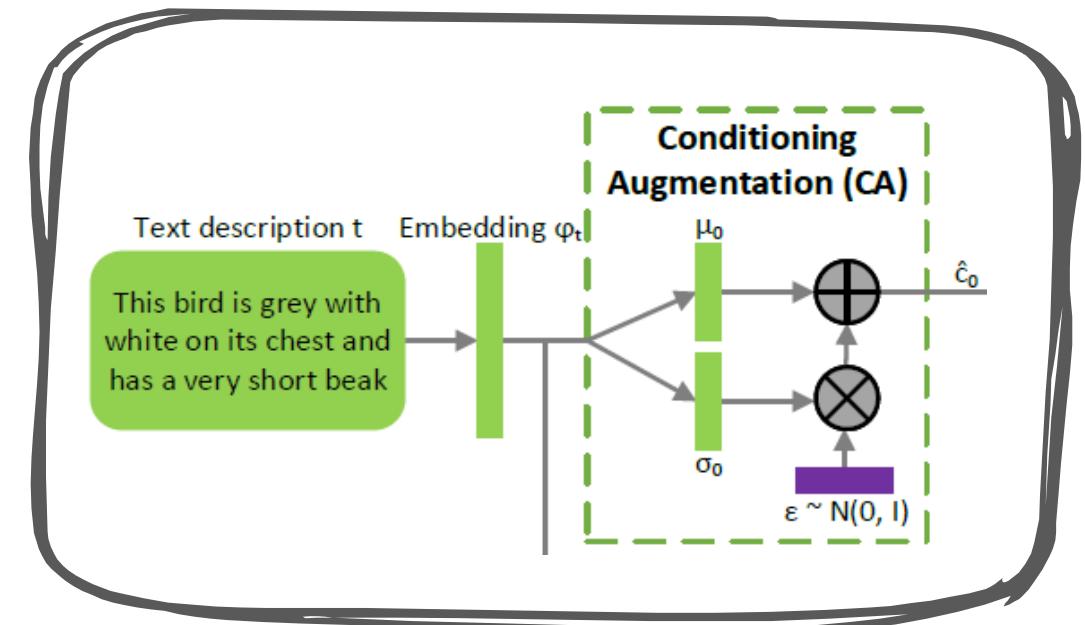


# PROPOSED MODEL

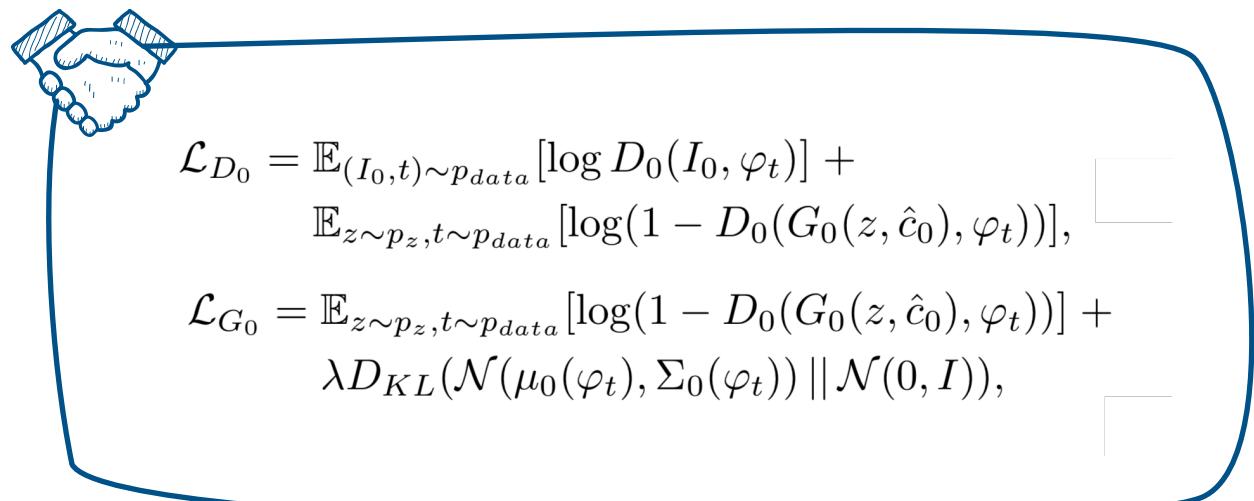
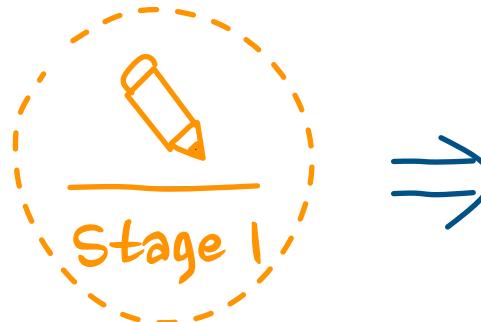
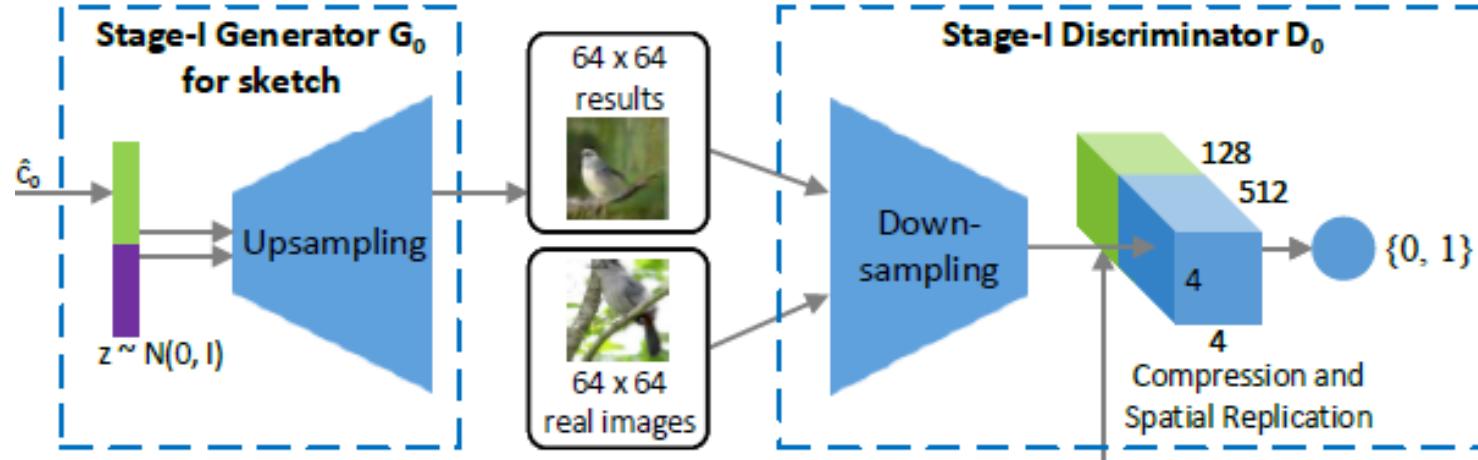
## Conditioning Augmentation



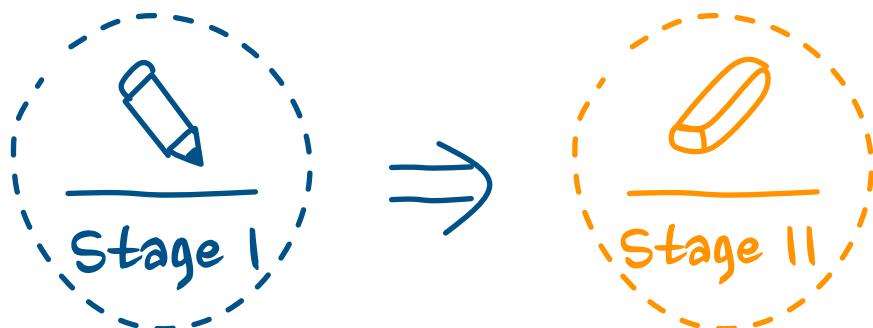
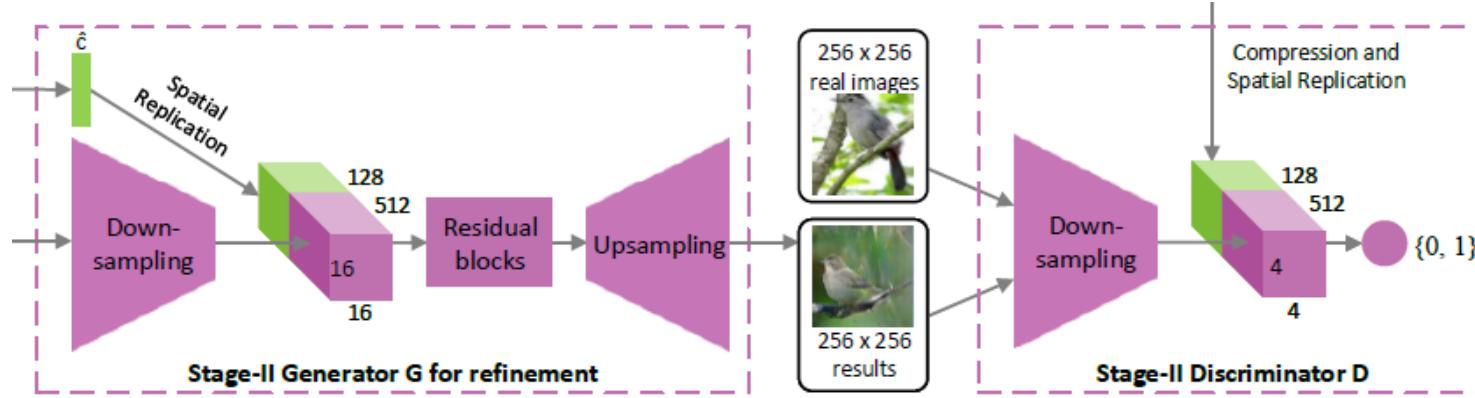
- Latent space for the text embedding is usually high dimensional .
- Conditioning augmentation
  1. Feed the input into a fully connected layer to calculate a Gaussian distribution.
  2. Sample a conditioning variables  $\hat{c}$  to be the input of Stage-1 GAN



# PROPOSED MODEL



# PROPOSED MODEL



$$\mathcal{L}_D = \mathbb{E}_{(I, t) \sim p_{data}} [\log D(I, \varphi_t)] + \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, \hat{c}), \varphi_t))],$$

$$\mathcal{L}_G = \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, \hat{c}), \varphi_t))] + \lambda D_{KL}(\mathcal{N}(\mu(\varphi_t), \Sigma(\varphi_t)) \parallel \mathcal{N}(0, I)),$$

# EXPERIMENTS & RESULTS

## Parameter Setting

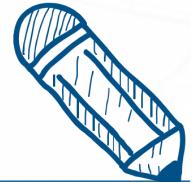


	train						
	Batch Size	Max Epoch	LR Epoch	Snapshot interval	D LR	G LR	KL
Stage 1	128	90	20	10	0.0002	0.0002	2
Stage 11	2	90	20	5	0.0002	0.0002	2

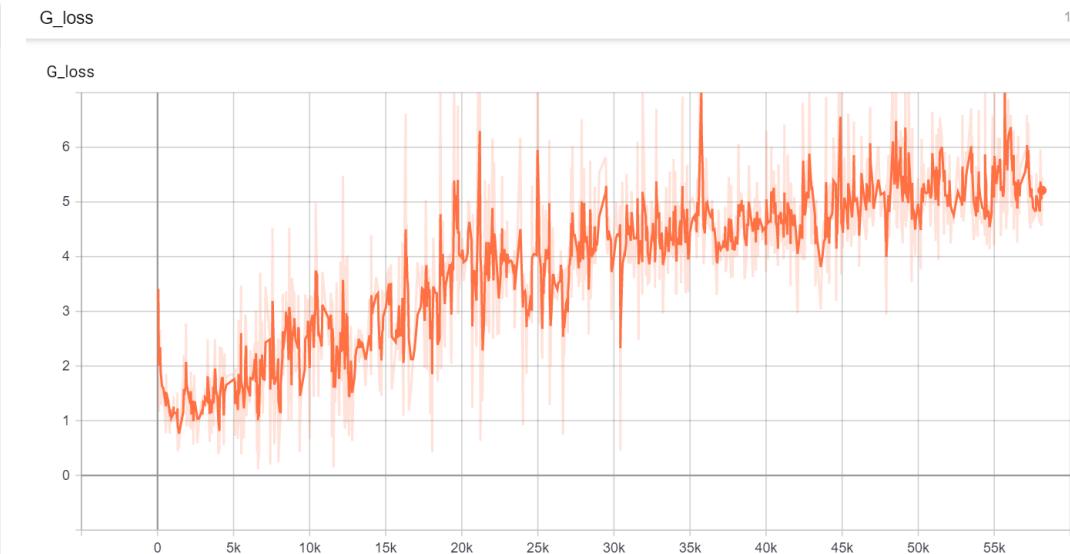
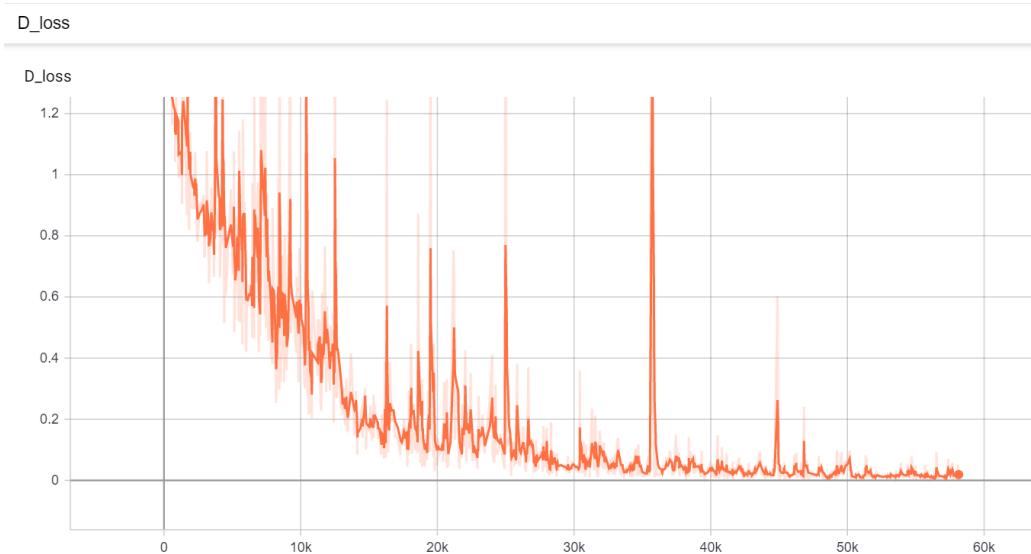
	GAN				Text
	Condition DIM	DF DIM	GF DIM	R num	DIM
Stage 1	128	96	192	X	1024
Stage 11	128	96	192	2	1024

# EXPERIMENTS & RESULTS

Learning Curve



## Stage 1

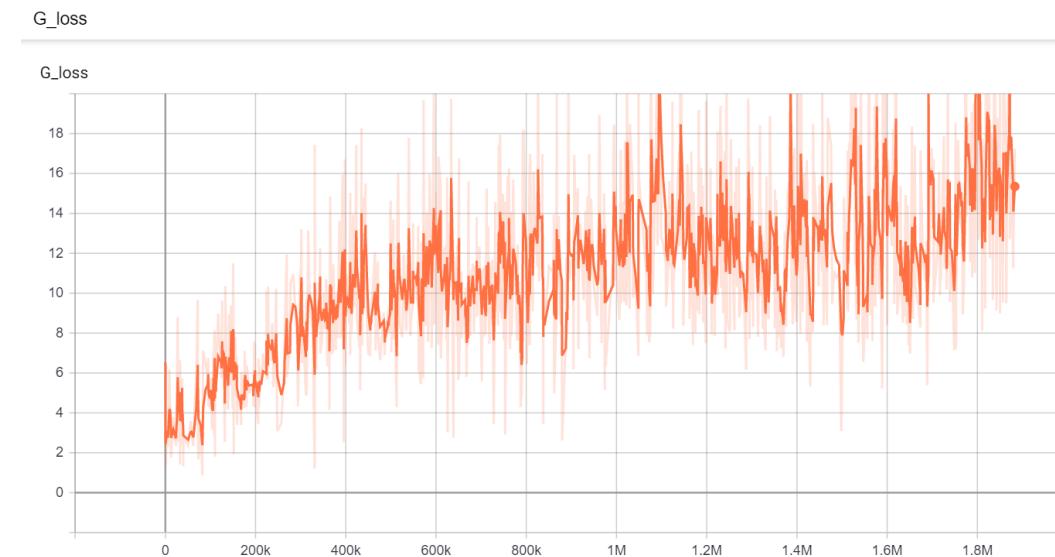
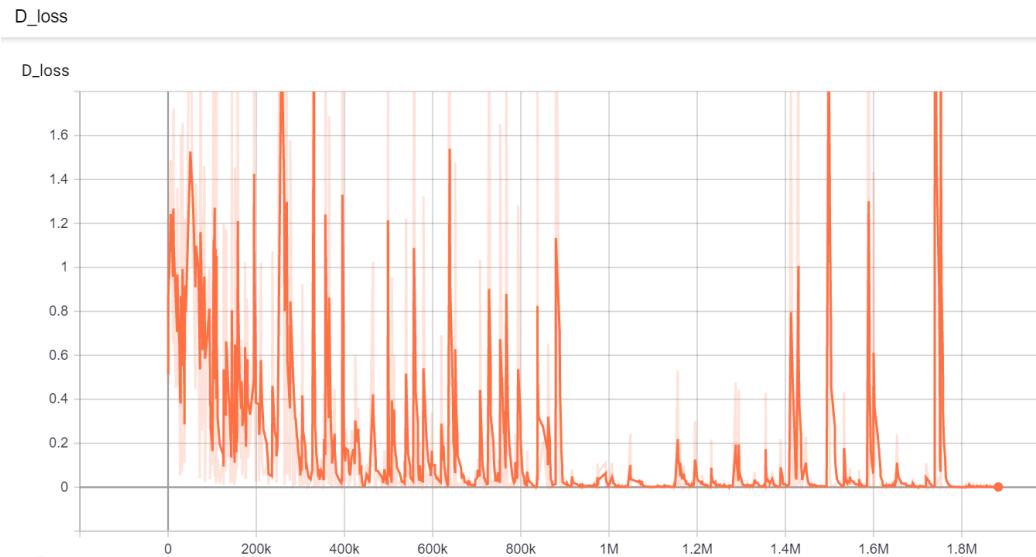


# EXPERIMENTS & RESULTS

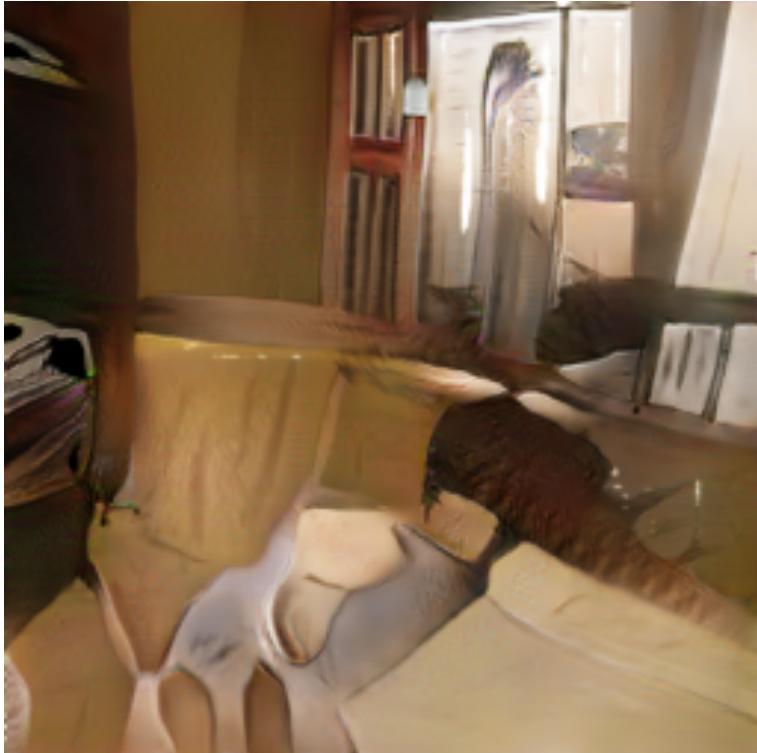
Learning Curve



## Stage 11



# EXPERIMENTS & RESULTS



a well lit living room with  
sofas and coffee table.



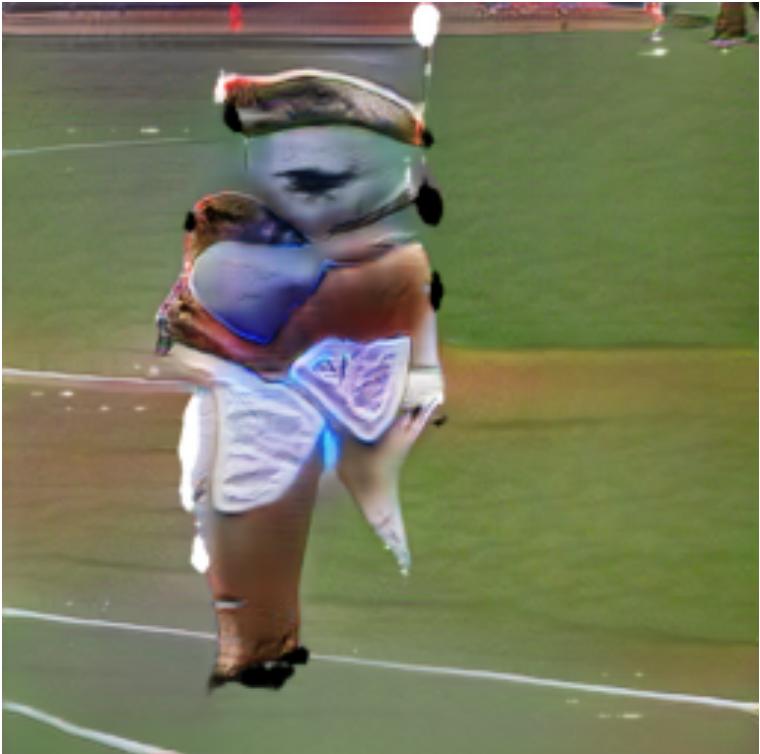
a man walking down a road  
with a horse and carriage  
coming at him.



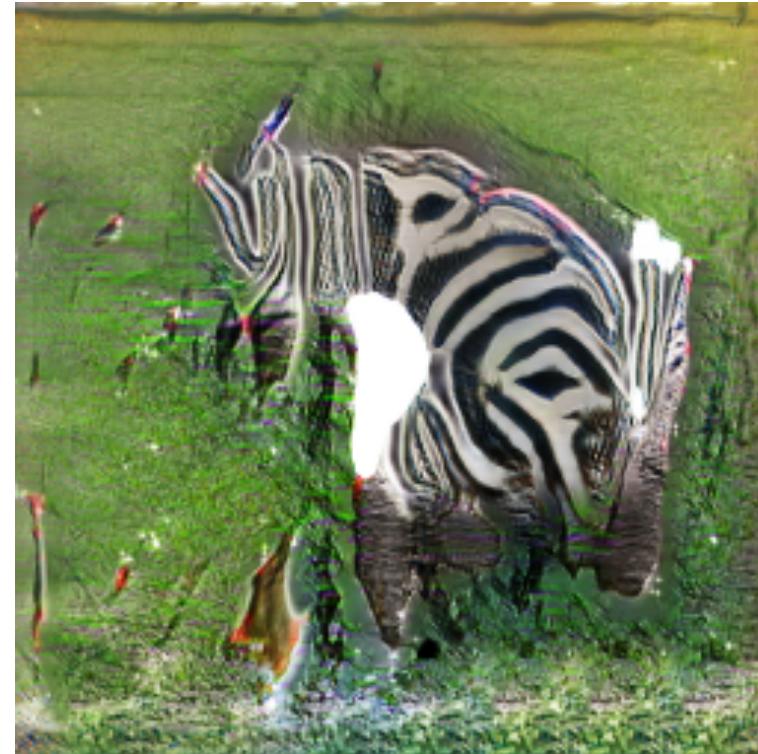
adult woman with yellow  
surfboard standing in water.

# EXPERIMENTS & RESULTS

---



a man hitting a tennis ball  
with a tennis racquet.



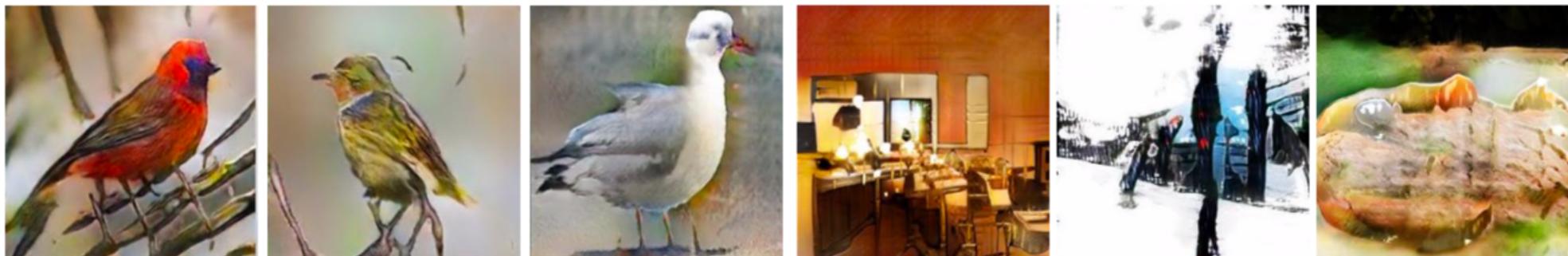
a zebra standing on a dirty  
rocky field.

# CONCLUSION

---



- Propose a text-to-image model StackGAN with Conditioning Augmentation
  - Stage-I GAN sketches the low resolution images
  - Stage-II GAN generates the high resolution images
- If we have more time or a powerful GPU



# Final Project



Thanks for Listening!