TEXT-TO-IMAGE GENERATION

Milestone 1

1. PROBLEM STATEMENT

2. RELATED WORK

3. CONTRIBUTION

4. PROGRESS

5. NEXT STEPS

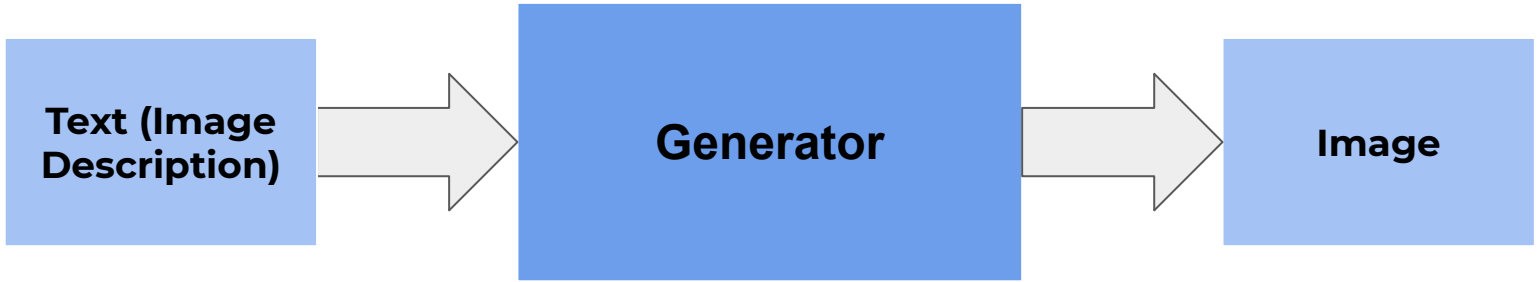
Text to Image

1

PROBLEM STATEMENT

Problem

Translate text in the form of human-written description into image that is indistinguishable from realistic one



Examples[1]

a flower with long pink petals and raised orange stamen.

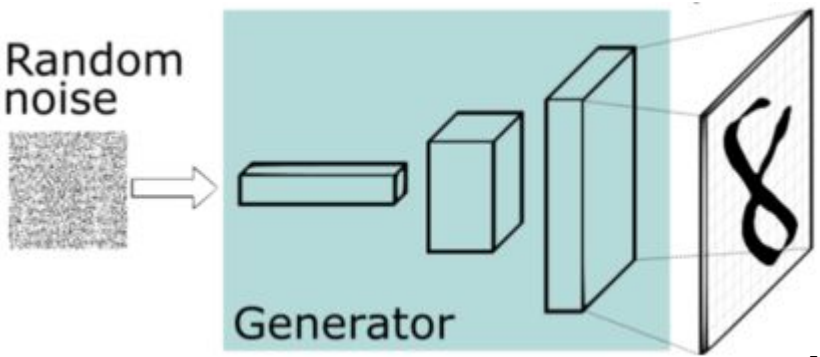
a sheep standing in an open grass field.

2

RELATED WORK

Background: Generative Adversarial Network (GAN)[2]

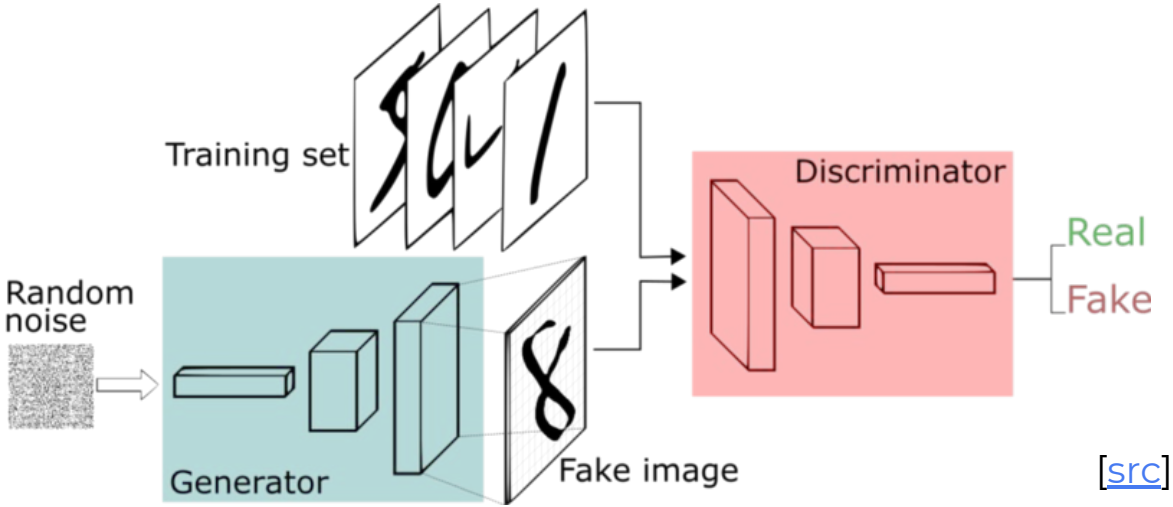
- Want to sample from a complex distribution (Hard)
- Sample from a simple distribution (e.g. random noise) then **learn the transformation** using a neural network (Easy)



[src]

Background: Generative Adversarial Network (GAN)

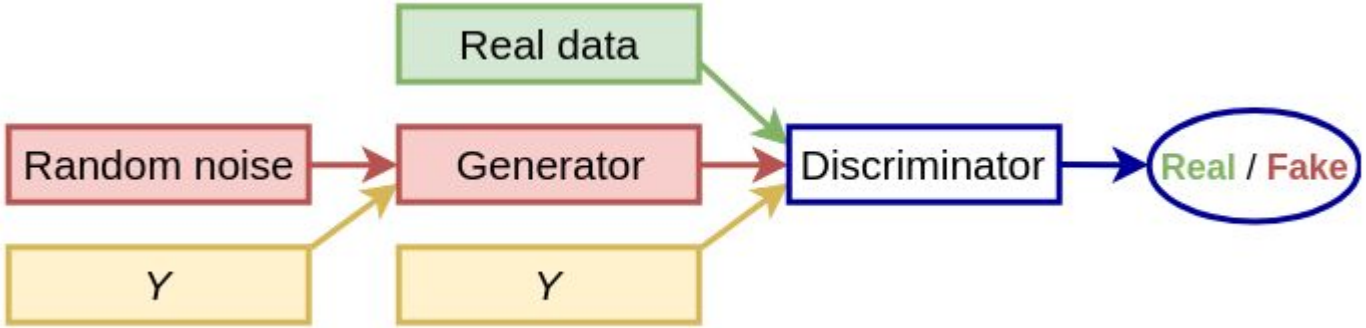
- How to train? **Minimax game**
  - **Generator**: fool the discriminator by generating pseudo-real images
  - **Discriminator**: distinguish between real and fake images



[src]

Background: Conditional GAN

- Adding a vector of features to control the output of the generator
- Make the output conditioned on the input features



[src]

DC-GAN (Reed et al., 2016)[1]

- Train a DC-GAN conditioned on text features encoded by a hybrid character-level Convolutional RNN

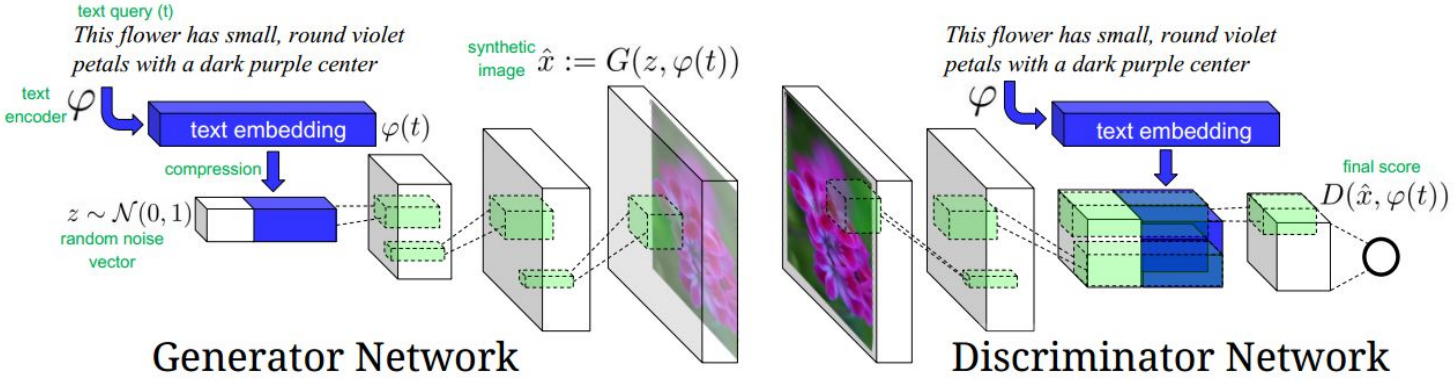


Figure 2. Our text-conditional convolutional GAN architecture. Text encoding $\varphi(t)$ is used by both generator and discriminator. It is projected to a lower-dimensions and depth concatenated with image feature maps for further stages of convolutional processing.

StackGAN (Zhang et al., 2017)[3]

- A **low resolution image** is generated using **Stage-I GAN**. It sketches the primitive shape and basic colors of the object.
- By conditioning on the generated image and the text again, **Stage-II GAN** then yields a **high resolution image**.
- This decomposition allows Stage-II GAN to capture text information that is omitted by Stage-I GAN and draw more details

3

CONTRIBUTION

Contribution

Build on **DC-GAN**[1]

Hyper-parameters

Try to tune to increase performance

Architecture

Try to modify (e.g. introduce progressive augmentation)

2-in-one image

Filter COCO to get images with birds and flowers in same text description, then train the model on them

4

Progress

Building simple model[4]

- Understand and train a simple conditional GAN
- Fashion-MNIST: 60,000 28×28 pixel grayscale images of items of 10 types of clothing
- Generate images based on the class label

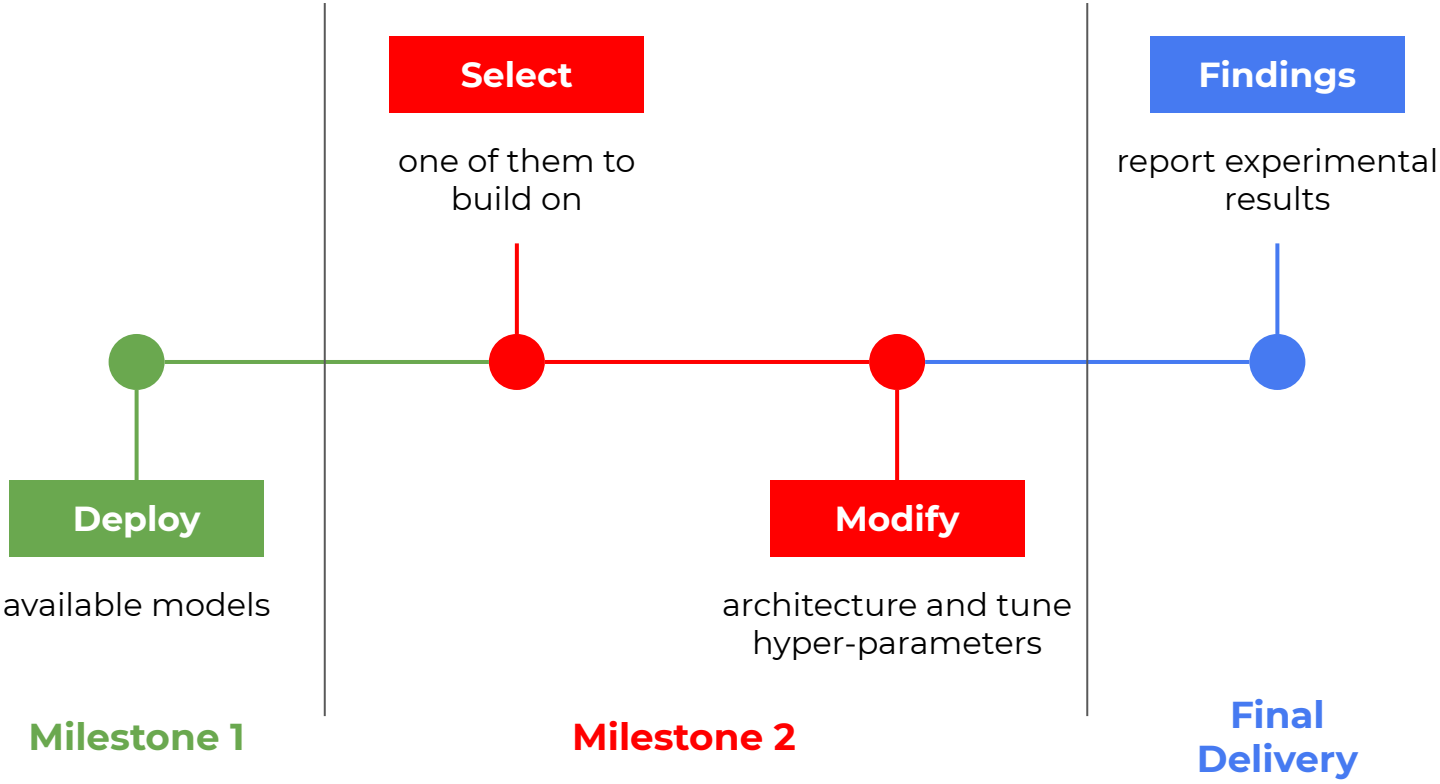
Deploying original model

- Found 2 implementations of the paper in TensorFlow
  - https://github.com/paarthneekhara/text-to-image
  - https://github.com/zsdonghao/text-to-image
- Try to deploy them, but found some errors

5

Next Steps

Next Steps


Deploy
available models
Select
one of them to build on
Modify
architecture and tune hyper-parameters
Findings
report experimental results
Milestone 1
Milestone 2
Final Delivery

References

1. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016b). Generative adversarial text to image synthesis. Proceedings of the International Conference on Machine Learning (ICML). Available: https://arxiv.org/pdf/1605.05396.pdf
2. Generative Models, CS231n Stanford course. Available: http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture11.pdf
3. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, Z., Huang, X., and Metaxas, D. (2017b). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In IEEE International Conference on Computer Vision (ICCV), Venice, pages 5908–5916. Available: https://arxiv.org/pdf/1612.03242v2.pdf
4. How to Develop a Conditional GAN (cGAN) From Scratch, Jason Brownlee. Available: https://machinelearningmastery.com/how-to-develop-a-conditional-generative-adversarial-network-from-scratch/

OUR TEAM

Mohamed Sharaf (54)

Mahmoud Tarek (63)

Thanks!