Generative Adversarial Networks Based Text to Image Generation

Abhinav Sharma, Lavanya Verma, Pranav Sharma

1 Problem Statement

Real-life automated image synthesis from a textual description by employing deep learning machines is a fascinating field that has a wide array of industrial and commercial applications. The project aims to employ Generative Adversarial Neural Networks (GANs) trained on Oxford Flower Dataset to generate real-life images of flowers from textual description.

2 Literature Survey

(Goodfellow et al.) showed that the GANs can be used to mimic a distribution by adversarial means. However, unsupervised cases (Radford et al.) presented a GAN architecture suited to generalize spatial dependencies(DCGAN). (Reed et al). presented novel optimization algorithms CLS and CLS_INT where the former focuses on the training of discriminator by inducing wrong examples whereas the latter induces an additional cost in generator minimization criteria. (F.Gong et al.) addressed the issue that the objective criteria of the CLS algorithm are not $f_g(y) = f_d(y)$ which is different from original GAN criteria (Goodfellow et al.), they also proposed a modified CLS algorithm to tackle this issue. (Mirza et al.) emphasized conditioning auxiliary information in GAN architecture and they showed that such conditioning improves the generator's ability to generate instances resembling multimodal distribution. (Reed et al.) extrapolated the oxford-102 flower dataset extrapolated with 5-caption per image, these captions are embedded using word-level LSTM(Reed et al.).

3 Baselines

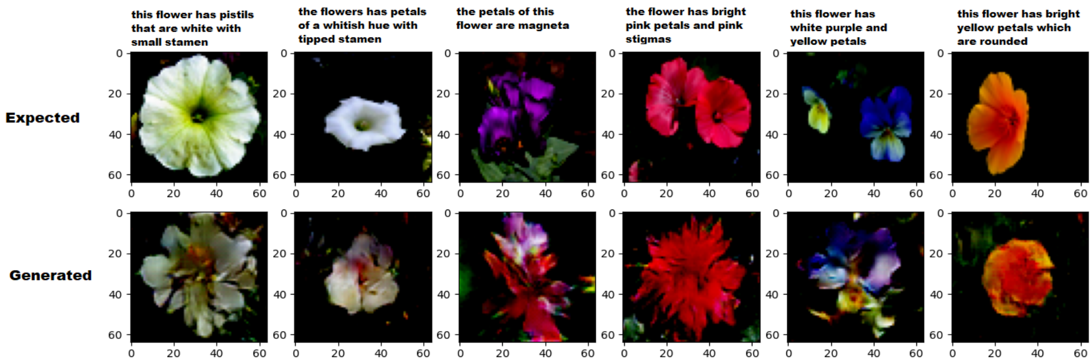
Architecture: **Generator**(Projector Layer $\rightarrow$ (ConvTranspose2D, BatchNorm2d, LeakyReLu)*4 $\rightarrow$ (ConvTranspose2D, Tanh)
Discriminator(Projector Layer $\rightarrow$ (Conv2D, ReLu) $\rightarrow$ (Conv2D, BatchNorm2D, LeakyReLu)*3 $\rightarrow$ (Conv2D, Sigmoid))
Epochs: 100 Image Size : (64, 64, 3) Optimizers = 'Adam'

3.1 Baseline 1

(*DCGAN*) Optimization Criteria

$$min_G max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{x\sim P_z(z)}[\log(1 - D(G(z)))]$$

The discriminator is trained to minimize V(D, G), it regard (real images, real embedding) as real instance and (generated images, real embedding) as a fake instance. On the same note, Generator is trained to minimize $E_{x \sim P_z(z)}[\log D(G(z))]$ with L1 regularization of (generated image, real image) and l2 regularization of Discriminator outputs(fake activation, real activation).activation: intermediary output from D(x) network (F.Gong et al).

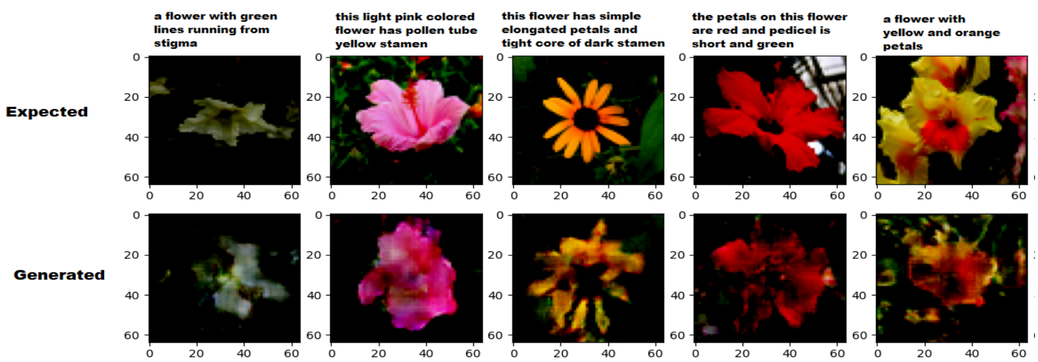


3.2 Baseline 2

($DCGAN\_CLS\_INT$) In addition to the above optimization criteria, generators in this has an additional cost to minimize

$$E_{t1,t2 \sim P_{data}(x)}[\log(1 - D(G(z, \beta t_1 + (1-\beta)t_2)))] + E_{t,x \sim P_{data}(x)}[\log(1 - D(x,t))]$$

- To implement the above criteria, the generator generates images based on interpolated embedding, regard the outputs as fake images in B.C.E. criteria.
- t_1, t_2 are text embeddings, z is drawn from noise distribution. x: different image from p_{data}, t:embedding of true image.
- D takes two arguments (image, embedding). Except for the last expression, input to D simplicity refers to the right image, corresponding embedding pair. However, to point the contrast of (wrong image, true embedding) pair, it was explicitly mentioned.



4 References

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In NIPS, 2014

2. F. Gong and Z. Xia. (2018). Generate the corresponding image from text description using modified GAN-CLS algorithm.

3. Reed S, Akata Z, Yan X et al. Generative adversarial text-to-image synthesis. In ICML, 2016.

4. Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks, 2014

5. Mirza M, and Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014

6. Reed, S., Akata, Z., Lee, H., and Schiele, B. Learning deep representations for fine-grained visual descriptions. In CVPR, 2016.

7. `https://pytorch.org/tutorials/beginner/dcgan_faces_tutorial.html`

8. `https://github.com/reedscot/icml2016`