Motivation
- Our task is to generate new images of flowers from text captions of flowers
- We were inspired by the idea that a machine could learn to imagine
- We hope this work opens avenues toward a more complete understanding of textual information

Data
- Oxford Flowers Dataset
  - 8199 images of flowers
  - 5 human-labeled captions per flower
  - 17 overall categories of flowers represented
- GloVe Word Embeddings
  - 300 dim representation of individual words
  - Trained on 6B words from Gigaword 5 + Wikipedia 2014
- Skipthoughts
  - 4800-dim learned representation of sentences
  - Trained on ~900 million words from BookCorpus

Baseline
- CLS-GAN w/ skipthoughts

Experiments
- CLS-GAN w/ LSTM
- DCGAN w/ LSTM & Upsample
- BEGAN w/ LSTM

Results
- Generated

Generative Adversarial Network (GAN) Architecture

Model Performances
- We perform a variety of architectural enhancements from our baseline, including Boundary Equilibrium (BEGAN), Deep Correlation (DCGAN)

Learned Embedding Representations
- We perform t-SNE dimensional reduction on learned caption features over the categories of flowers generated

Loss and Backpropagation
- L(Discriminator) = C_0 + 100g(y_0 + y) + 50g(1 + y) + C_l + 5g(x, f)
- L(Generator) = C_0 + 100g(y)
- L(LSTM) = C_l + 5g(x, f)

Future Work
- Explore different model architectures like StackGAN, AttnGAN
- Spend more time tuning architecture and hyperparameters to produce better images
- Video reconstruction from text

Training Techniques
- LSTM
  - Initialize hidden states to small random values
  - Batch size = 64
- GAN
  - Dataset augmentation with random crops + flips
  - Gaussian noise for generator
  - Normal initialization
  - Better resolution/eliminated checkerboard with upsampling
  - One-sided label smoothing
  - Leaky ReLU instead of ReLU
- Bi-directional, 1-layer
- Xavier weights initialization
- SGD instead of Adam for D
- Tanh for output
- Batchnorm
- Learning rate decay
- Average pool instead of max pool
- ELU instead of ReLU