Building Deep Learning Architectures to Understand Building Architecture Styles
Cole Thomson & Caroline Ho
(colet, cho19)@stanford.edu

**Introduction**
- **Motivation**
  - Explore whether deep neural networks can differentiate between architectural styles
  - Applications: geo-localization, urban surveying
- **Challenges**
  - Visually similar buildings differ in style (fig 1)
  - Visually different buildings have same style (fig 2)

![Figure 1: Georgian (Left) and Palladian (Right) Styles](image)

**Work**
- CNN to classify architectural style from image
- Conditional GAN to generate images of different architectural styles

**Data**
- Xu et al.¹: 4794 RGB images of buildings from 25 architectural styles (see figs 1 & 2 for examples)
- Different resolutions, dimensions, aspect ratios

**Preprocessing**
- Resize and center-crop images
- 224 pixels for ResNet and DenseNet, 299 for Inception, 64 for GAN
- 80/10/10 Train/Dev/Test split for classifier

![Figure 2: Two Deconstructivist Buildings](image)

**Motivation**
- Transfer learning trying both fine-tuning and fixed feature extraction
- Pre-trained architectures: ResNet152, DenseNet201, Inception v3
- Adam optimization with cross-entropy loss and learning rate decay

**Generator**
- Conditional GAN (cDCGAN): conditions on image class using a deep convolutional network
- Generator and Discriminator consist of five 4x4 strided conv layers
- Adam optimization and Batch Norm (except for output layer)

![Figure 3: Value Function for cGAN²](image)

**Results**
- **Classifier**
  - | Model       | Style       | Accuracy | Macro F1 Score |
  - |-------------|-------------|----------|----------------|
  - | DenseNet    | Fixed Feature Extraction | 0.795833 | 0.789431 |
  - | DenseNet    | Fine-tuning     | 0.675000 | 0.601738 |
  - | ResNet      | Fixed Feature Extraction | 0.747917 | 0.724015 |
  - | ResNet      | Fine-tuning     | 0.500000 | 0.410742 |
  - | Inception   | Fixed Feature Extraction | 0.654167 | 0.623107 |
  - | Inception   | Fine-tuning     | 0.722917 | 0.651555 |

![Figure 4: cGAN Architecture³ (x = training data, y = class info, z = random noise)](image)

**Discussion**
- Classifier outperforms previous benchmark accuracy of 0.4621
- Seemingly-disparate styles very occasionally confused (e.g. Achaemenid and Beaux-Arts)
- Generated images are abstracted but show their approximate stylistic forms (fig 7)
- Architectural style is easier to recognize from features than it is to create from them

![Figure 5: Classifier Performance Metrics](image)

![Figure 6: Confusion Matrix (DenseNet FFE)](image)

**Future Work**
- Interpretability on classifier
- Experiment with other GAN architectures including WGAN

**References**