Photo to Doodle generator using GAN

Talbot Morris-Downing (talbotmd@stanford.edu), Manu Agarwal (amanu@stanford.edu)

Video: https://www.youtube.com/watch?v=6SUxwUy3k8A

**Introduction**
*Drawn by people, not our generator

- **Goal**: Generator that converts photo to a doodle. Prior art only covers search not doodle generation.
- **Challenge**: High variation in doodles made by people. No simple loss function for training a generator.
- **Solution**: GAN - Train classifier that matches photos to doodles, and make a generator “fight” it
- **Result**: Good early results, more training and model iteration needed.

**GAN Architecture**

- **P**: Input photo
- **D**: Human doodle (ref. P)
- **d**: Generator doodle

\[ J_{hd} = \frac{1}{m_{o}} \sum_{i=1}^{m_{o}} \log(\mathbb{H}_{D}(P_i, D_i)) - \frac{1}{m_{g}} \sum_{i=1}^{m_{g}} \log(1 - \mathbb{H}_{D}(P_i, \mathbb{H}_{O}(P_i))) \]

\[ J_{ho} = \frac{1}{m_{o}} \sum_{i=1}^{m_{o}} \log(\mathbb{H}_{D}(P_i, \mathbb{H}_{O}(P_i))) \]

**Network Architecture**

- **Generator**
  - Encodes (Conv2D) to features, then decodes (Conv2DTranspose) to full size.
  - Forward (FF) paths are like ResNets, but forward info to layers with same sizes.
  - FF paths let's model choose to skip “over-encoding” and keep detail.

- **Discriminator**
  - 6 input channels, 3 each for P & D.

**Training**

- **Disc Loss**: \( J_{D} \)
- **Gen Loss**: \( J_{G} \)

**Generated Doodles**

- **Over Epochs**: Oscillating between eroded edges and smudges

- **Note**: Only 1/3 of sketchy data used for due to time/compute constraints

**Discussion & Future Work**

- **Model Performance**
  - Model confused by input detail (background and subject)
  - Over iterations, didn’t hit the right tradeoff between capturing edges and filtering extraneous detail.

- **Future work**
  - More compute and training!
  - Architect separate block to separate subject & background
  - Increase generator model complexity to identify and render key edges in subject and filtering out detail

**Dataset**

- **Source**: Sketchy, http://sketchy.eye.gatech.edu

- **Details**
  - 125 object categories (e.g. cat, guitar, airplane)
  - Average 100 photos per object category
  - 5-7 human drawn doodles per photo
  - Sketches labeled for wrong context, errors, wrong pose or other ambiguity

- **Overall**: 125 categories, 12,500 photos, ~75,000 doodles (sketches)

**Introduction**

- **Drawn by people, not our generator**