Unsupervised Face-to-Manga Translation with CycleGAN

Marcella Cindy Prasetio (mpc21@stanford.edu)

Motivation
- Face-to-Manga translation can be useful in social application as image stylization.
- Lack of paired data makes this a novel and unique task.
- Solution: unsupervised translation with CycleGAN.

Approach
- **CycleGAN**: GAN Loss + Reconstruction Loss
  \[ L_{GAN}(G, D) = E_{p_{data}}[\log D_y(y)] + E_{p_{data}}[\log (1 - D_y(G(z)))] \]
  \[ L_{Cycle}(G,F) = E_{p_{data}}[\|G(F(z)) - y\|_1] + E_{p_{data}}[\|F(G(y)) - z\|_1] \]

  ![Architecture Diagram]

  - **PatchGAN**: 64×64 patches to produce robustness to crop failures.
  - **Self-attention Module**: Enhances the translation quality.

Evaluation for quality and diversity: Inception Score & FID Score
- **Model**: d: depth of PatchGAN
- **(1) ResNet + ConvNet**: 3.200, 0.497
- **(2) ResNet + PatchGAN d=3**: 3.800, 0.456
- **(3) ResNet + PatchGAN d=4**: 4.866, 0.523
- **(4) ResNet + PatchGAN d=4 + SN**: 4.520, 0.487
- **(5) ResNet + PatchGAN d=4 + Attn**: 3.960, 0.478
- **(6) ResNet + PatchGAN d=4 + SN + Attn**: 4.074, 0.523

Problem & Challenges
- **Dataset**: CelebA: 202,599 face images.
- **Manga109**: 10,619 pages with 26,602 characters.
  - Cropped, scaled to 64x64pxs, converted into grayscale.
  - 98% training, 2% test.

  ![Image 0x384 to 612x792]

**Problem**: Translating face images into Manga characters style.

**Challenges**:
- Unstable learning in GAN.
- Unavailable paired dataset for supervised learning, making evaluation challenging.

Discussion - Next Steps
- Model learns decent face-to-manga translations.
- Spectral normalization makes the images stable and more realistic.
- Self-attention highlights more facial features.
- Pair manga images are not all centered and aligned.

Future work:
- Make the discriminator simpler to allow more freedom for the generator to style the images.
- Experiment with other generators that can highlight more facial features.
- Train best model longer.

References