Manga-to-Anime Translation Using Cycle-Consistency

Jonathan Griffin
Electrical Engineering, Stanford University

Motivation

The manga and anime industry adapt each others' work to produce content for their audiences. Unfortunately, artists are underpaid and overworked. Therefore, an automated system that handles the manga-to-anime translation task is much needed.

GANime is a Generative Adversarial Network (GAN) that performs on the manga-to-anime translation task using concepts of cycle-consistency.

Figure 1: Manga-to-Anime translation example.

GANime

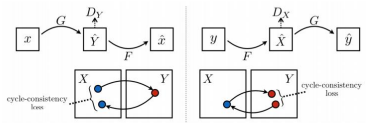

Figure 3: Cycle loss visualization.

Layer	Activation Size
Input	256x256x3
4x4x64 convolution, stride 2	128x128x64
4x4x128 convolution, stride 2	64x64x128
4x4x256 convolution, stride 2	32x32x256
5 4x4x512 convolution, stride 2	1x1x512
4 4x4x512 deconvolution, stride 2	16x16x1024
4x4x256 deconvolution, stride 2	32x32x512
4x4x128 deconvolution, stride 2	64x64x256
4x4x64 deconvolution, stride 2	128x128x128
4x4x3 deconvolution, stride 2	256x256x3

Layer	Activation Size
Input	256x256x3
7x7x64 convolution, stride 1	128x128x64
3x3x128 convolution, stride 2	64x64x128
3x3x256 convolution, stride 2	32x32x256
9 Residual block, 256 filters	32x32x256
3x3x128 deconvolution, stride 2	64x64x128
3x3x64 deconvolution, stride 2	128x128x64
7x7x3 convolution, stride 1	256x256x3

Layer	Activation Size
Input	256x256x3
4x4x64 convolution, stride 2	128x128x64
4x4x128 convolution, stride 2	64x64x128
4x4x256 convolution, stride 2	32x32x256
4x4x512 convolution, stride 1	32x32x512
4x4x1 convolution, stride 1	32x32x1

Table 2: Architectures; (left) UNet generator, (right) ResNet generator, (bottom) discriminator.

$$\mathcal{L}_{GAN} = \mathbb{E}_{x\sim X}[log(D_X(x)) - log(D_Y(G(x)))] + \mathbb{E}_{y\sim Y}[log(D_Y(y)) - log(D_X(F(y)))]$$
$$\mathcal{L}_{LSGAN} = \mathbb{E}_{x\sim X}[D_Y(G(x))^2 + (D_X(x) - 1)^2] + \mathbb{E}_{y\sim Y}[D_X(F(y))^2 + (D_Y(y) - 1)^2]$$
$$\mathcal{L}_{CYC} = \mathbb{E}_{x\sim X}[\,\| F(G(x)) - x \|_1] + \mathbb{E}_{y\sim Y}[\,\| G(F(y)) - y \|_1]$$

$$\mathcal{L}_1 = \mathcal{L}_{GAN} + \lambda\mathcal{L}_{CYC}$$
$$\mathcal{L}_2 = \mathcal{L}_{LSGAN} + \lambda\mathcal{L}_{CYC}$$

Data

Manga Anime

Figure 2: Example images from manga and anime domain.

Image Size	256x256
# Images per Domain	2,000
Split	80%-10%-10%
Sources	Manga109, Nico-Illust, Danbooru2017

Table 1: Dataset features.

Results

Parameter	Value
Learning rate	0.0002
Epochs	200
λ	10

Table 3: Hyperparameters.

Model	FID
LSGAN+ResNet	298.98
LSGAN+UNet	252.80
GAN+ResNet	287.24

Table 4: FID scores.

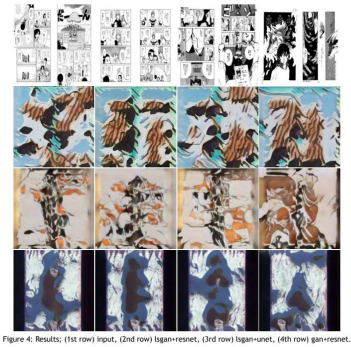

Figure 4: Results; (1st row) input, (2nd row) lsgan+resnet, (3rd row) lsgan+unet, (4th row) gan+resnet.

Discussion

Limitations: The manga domain has image panels while anime domain has one main image per sample. This makes the discriminators job really easy and causes generators to switch between panels and single images.

Future Work:

- Image processing of manga pages to extract panels to then feed into the network.
- Data collection for a professional anime dataset.
- CGAN + CycleGAN

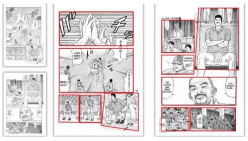

Figure 5: Manga panel extraction example.

References

[1] J. Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks.*CVPR*, 2017.
[2] Matsui et al. Sketch-based manga retrieval using manga109 dataset, multimedia tools, and applications, 2017.
[3] Ogawa et al. Object detection for comics using manga109 annotations.
[4] Danbooru community, Danbooru2018: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset, 2019.
[5] Pang et al. A robust panel extraction method for manga, *ACM*, 2014.