Synthesizing Comics via Conditional Generative Adversarial Networks

Darwin Burkard Morris
Supervised by Prof. Lydia Chen, Dr. Zilong Zhao
Delft University of Technology

Background

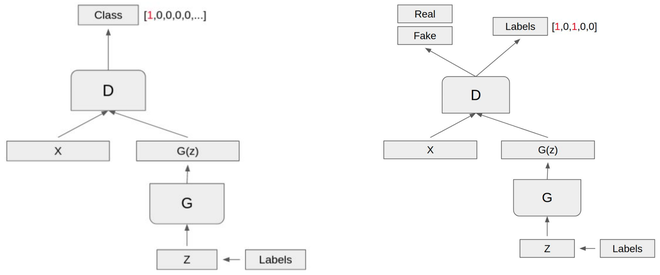
- **Generative Adversarial Networks** (**GANs**) [1] excel at image synthesis.
- **Deep Convolutional GAN** (**DCGAN**)[2], **Wasserstein GAN with Gradient Penalty** (**WGAN-GP**)[3], and **Stability GAN** (**SGAN**)[4] represent the state-of-the-art.
- **Conditional GANs** can be conditioned in order to generate output that matches a class label.

Research Question

Can conditional Generative Adversarial Networks Synthesize Images that accurately match preconditions?

- How does the unconditional performance of DCGAN, WGAN-GP, and SGAN compare in the comics domain?
- Can a multi-label and multi-label version of the best performing architecture conditionally generate semantically accurate panels?
- How does the performance compare between the two networks?

Method



MC-SGAN ML-SGAN

1. **Empirical Analysis**
   Compare DCGAN, WGAN-GP, and SGAN on comics domain. Assess prominence of conditions.
2. **Conditions**
   - Most prominent conditions were determined to be background color and character presence.
3. **GAN Architectures**
   - **Multi-Class SGAN:** ResNet based architecture using multi-class classification via LP transformation.
   - **Multi-Label SGAN:** ResNet based architecture using multi-label auxiliary classification technique [5].
4. **Evaluation**
   - Frechet Inception Distance [6]
   - Accuracy of label occurrence
   - Network loss

Empirical Analysis

Results from DCGAN, WGAN-GP 64, and WGAN-GP 128

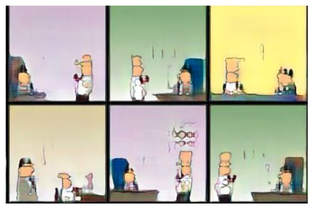
Results from SGAN 128

Example of vanishing gradients seen in DCGAN

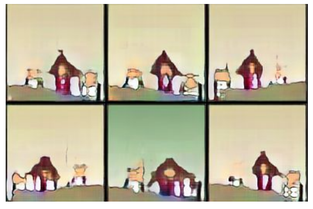
Through empirical analysis it was determined that the ResNet based SGAN architecture synthesized superior comics to both DCGAN and WGAN-GP. SGAN also proved to be extremely stable.

Results

Network	Color	Two-Character	Four-Character
MC-SGAN	100%	96%	84%
ML-SGAN	100%	92%	63.10%



[1, 1, 0, 0] -> 4



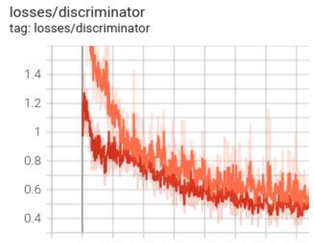
[0, 0, 1, 1] -> 7

Both ML-GAN and MC-GAN performed exceedingly well at generating semantically correct panels during experiments with color and with two-characters. MC-SGAN outperformed ML-SGAN in output accuracy when conditioned on four characters.

FID was similar between ML-GAN and MC-GAN during although it was lower than unconditional SGAN. This is likely due to differences in data.

Conclusion

- Generative Adversarial Networks provide a great platform for conditional synthesis of comics
- SGAN outperforms both WGAN-GP and DCGAN when applied to comics
- MC-SGAN generates comics with high semantic accuracy although it is limited by growth in class number
- ML-SGAN struggles on more complex problems due to collapse of the auxiliary classifier



Result of auxiliary classifier collapse in ML-SGAN. A viable solution could be a pre-trained classifier.

Contact

Name: Darwin Morris
Email: dmorris@student.tudelft.nl
Responsible Professor: Dr. Lydia Chen
Supervisor: Dr. Zilong Zhao
Other Researchers: Ben Provan-Bessel, Maciej Styczen, Bartlomiej Kotlicki, Krzysztof Garbowicz

References

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," arXiv preprint arXiv:1406.2661, 2014.
[2] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.
[3] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," arXiv preprint arXiv:1704.00028, 2017.
[4] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for gans do actually converge?" in International conference on machine learning. PMLR, 2018, pp. 3481–3490
[5] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in International conference on machine learning. PMLR, 2017, pp. 2642–2651.
[6] A. Mathiasen and F. Hvilshøj, "Fast fr\'echet inception distance," arXiv preprint arXiv:2009.14075, 2020.