Generative Adversarial Networks on Unpaired Image-to-Image Translation

Yue Kuang, Ziqiao Ma, Zhuowen Shen, Ruobing Wang

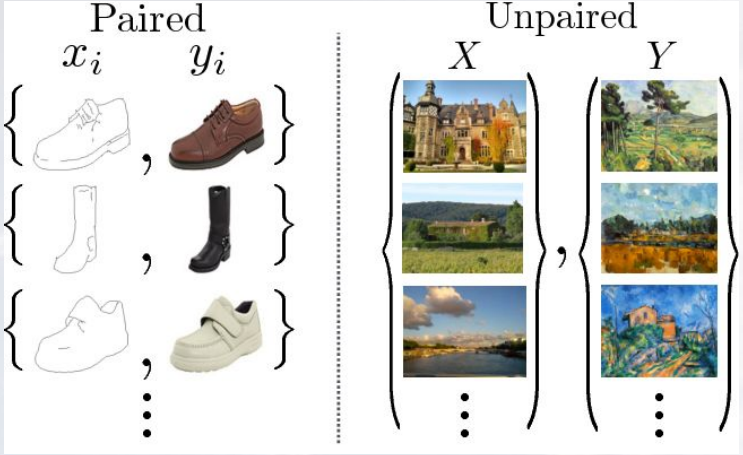
Background/Motivation

Problem With Paired Image-to-Image Translation

- Requires a dataset comprised of **paired examples**, which can be **challenging and expensive** to prepare, e.g. photos of different scenes under different conditions.

Unpaired Image-to-Image Translation

- Any **two collections of unrelated images** can be used.
- **Extract general characteristics** from each collection for the image translation process.



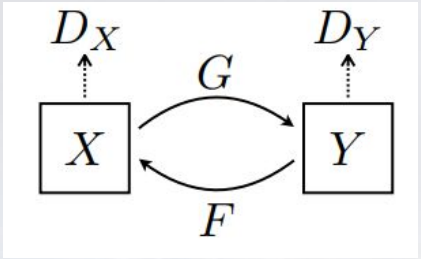
Method

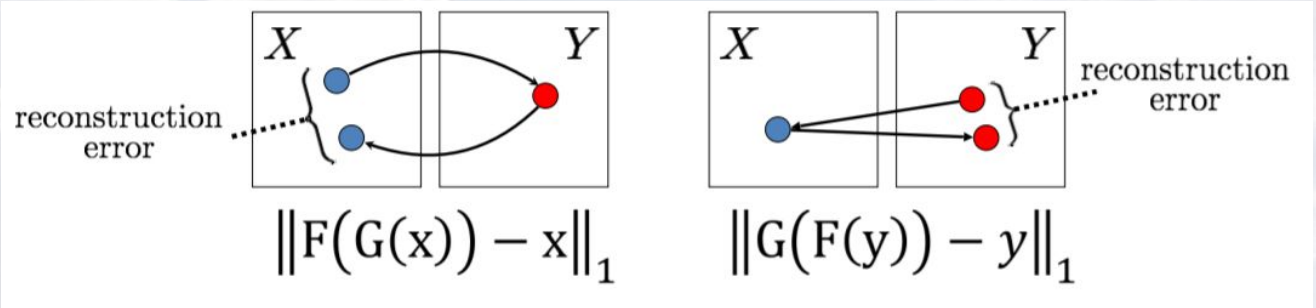
CycleGAN [1]

- Simultaneous training of 2 generator models and 2 discriminator models.
  - **Generators** G: X → Y, F: Y → X
  - **Discriminators**

    D_X learns to differentiate F(Y) from X

    D_Y learns to differentiate G(X) from Y
  - **Cycle Consistency Loss:**





[1] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

Method

StarGAN v1 [2]

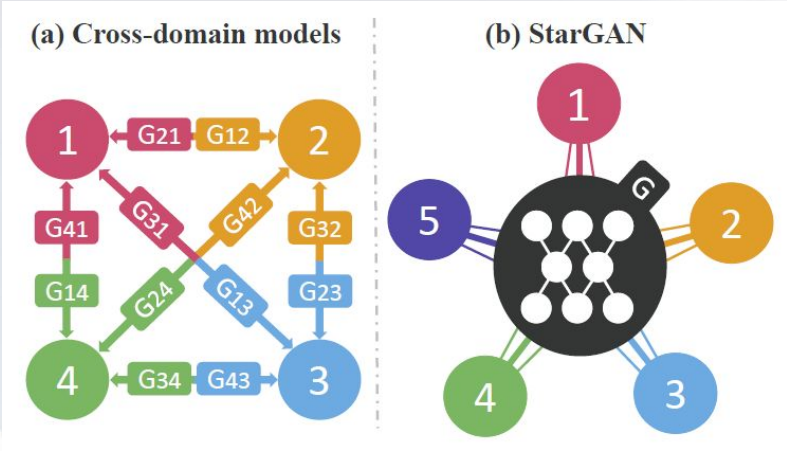
- Learning mappings among multiple domains using a single generator.
  - **Generators**

    Similar, but conditioned on the target domain label c.

    $$G: (X, C) \rightarrow Y$$

  - **Discriminator**

    distinguish real from fake images and classify real images to corresponding domain.



[2] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

Method

StarGAN v2 [3]

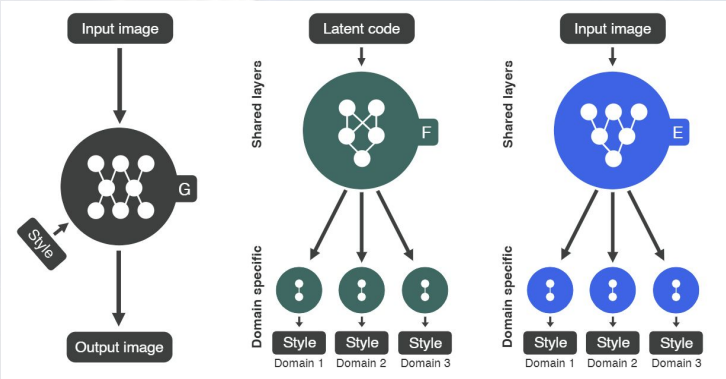
- Inherited from StarGAN v1 and learning diverse styles of a specific domain
  - **Generator**

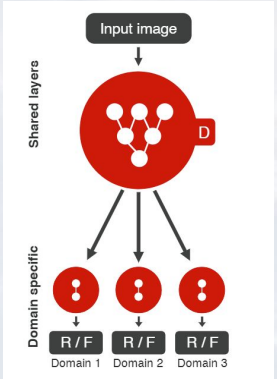
    G: (X, s) ⟶ Y (s: style code)

    F: z ⟶ s (z: latent code)

    E: Y ⟶ s
  - **Discriminator**

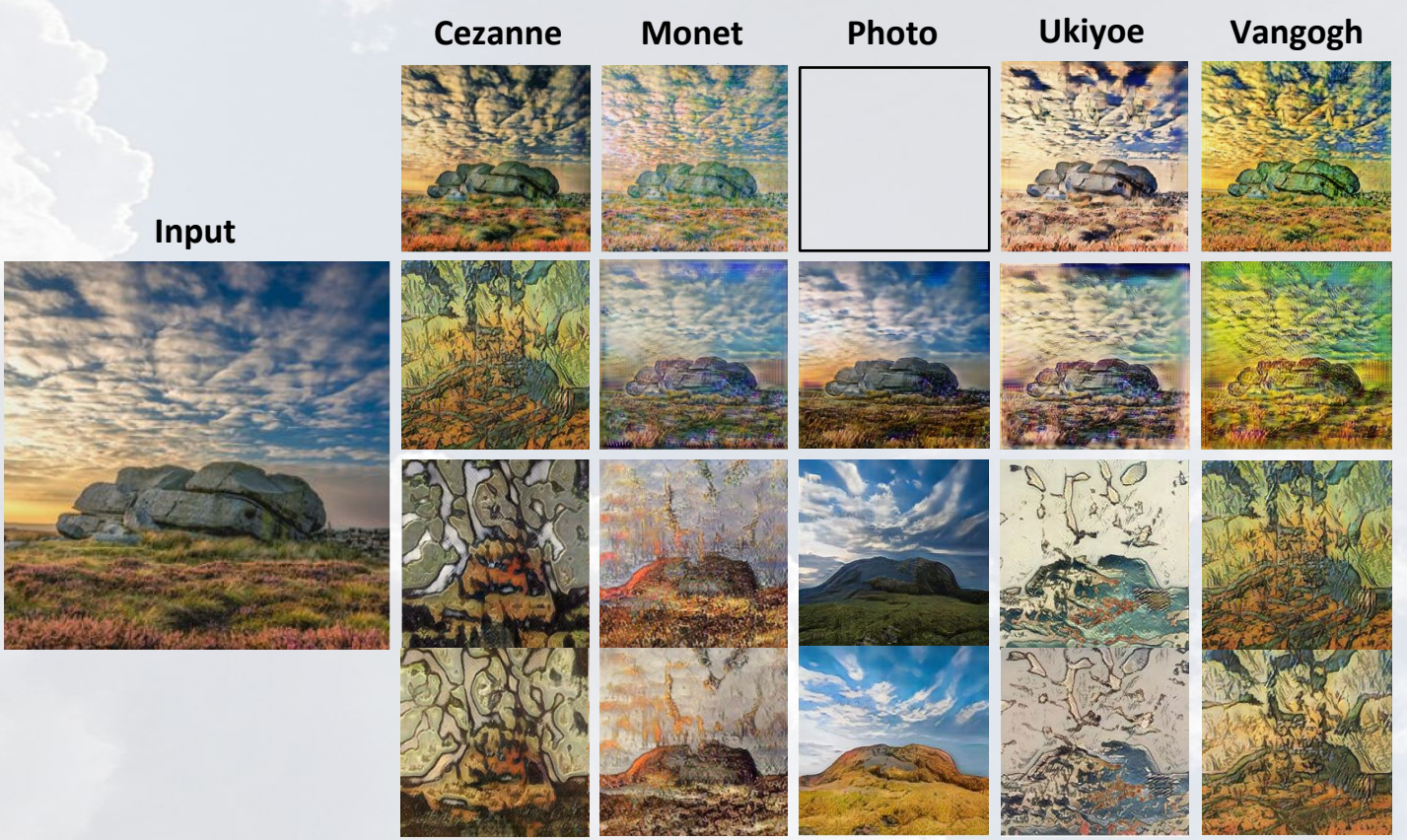
    Multitask, with multiple binary outputs indicating real or fake in each domain





[3] Choi, Y., Uh, Y., Yoo, J., & Ha, J. W. (2020). Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8188-8197).

Results


Cezanne
Monet
Photo
Ukiyoe
Vangogh
Input
CycleGAN
StarGAN v1
StarGAN v2

Results

Frechet Inception Distance (FID) Score [4]

- Evaluate the performance of GANs by the similarity of generated images compared to real images from the target domain.
- A lower FID indicates better-quality images.

	Original	CycleGAN	StarGAN v1	StarGAN v2
Cezanne	487.42	392.86	397.39	**363.15**
Monet	424.61	376.58	363.47	**241.03**
Ukiyoe	509.92	443.21	**438.93**	443.67
Vangogh	420.86	393.99	335.68	**312.55**

[4] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30, 6626-6637.

Results

Style Classifier

- RexNeXt model
- Pretrained on ImageNet and fine-tuned on the artists' paintings.
- Test accuracy on the dataset is about 98.5%.

Model	CycleGAN	StarGAN v1	StarGAN v2
Accuracy	67.11%	64.52%	**67.83%**

Conclusion

- CycleGAN and StarGANs both perform well for unpaired image-to-image translation tasks.
- CycleGAN is only capable of learning the relations between two different domains at a time.
- Both StarGANs solves the efficiency problem and StarGAN v2 brings diversity.
- Overfitting problem

References

[1] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

[2] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

[3] Choi, Y., Uh, Y., Yoo, J., & Ha, J. W. (2020). Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8188-8197).

[4] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30, 6626-6637.