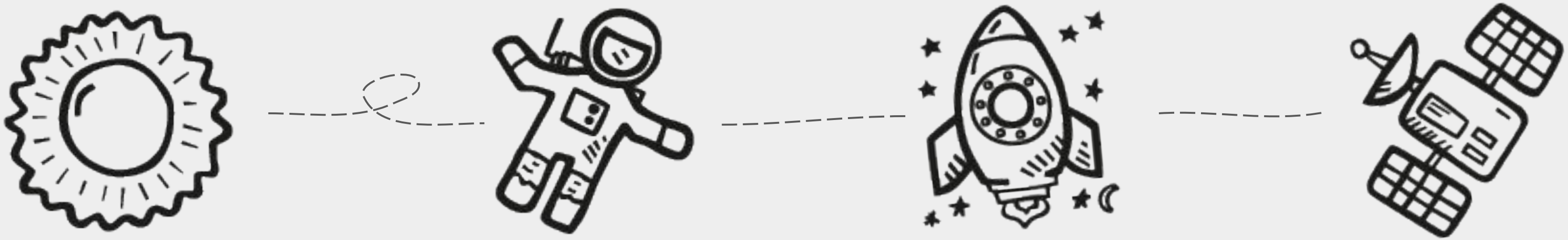
Data augmentation of magnetograms for solar flare prediction

Allison Liu
Mentor: Dr. Wendy Carande

Laboratory for Atmospheric and Space Physics
University of Colorado **Boulder**

Motivation



Solar Research

We care to characterize and understand the Sun…it gives us life!

Protecting Astronauts

High-energy solar radiation is harmful to the human body and can cause biological damage

Space Exploration

Accurate solar flare prediction is a concern that inhibits space travel

Communications

Large solar flares can disrupt critical infrastructure like the power grid, GPS, and radio communications

Source: NASA

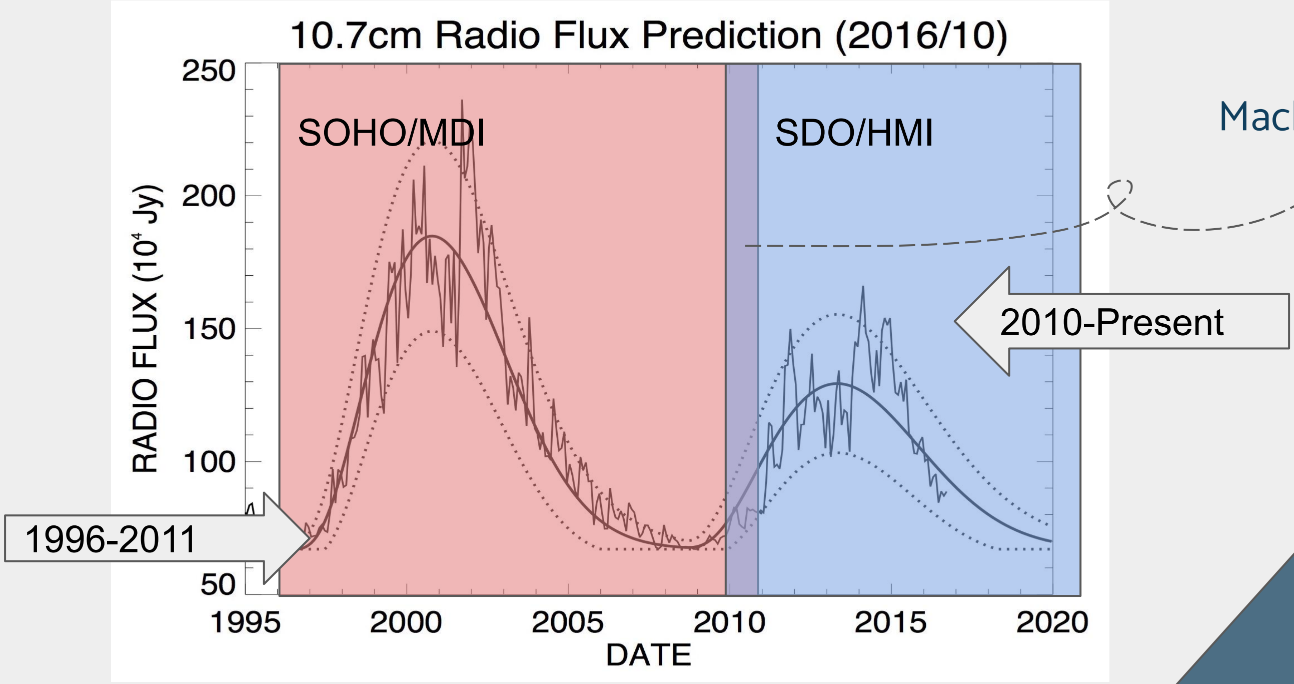
Source: NASA

Source: NASA

Source: NOAA

Introduction

Solar flare prediction is done largely by humans → Machine Learning ~2010



Machine Learning!

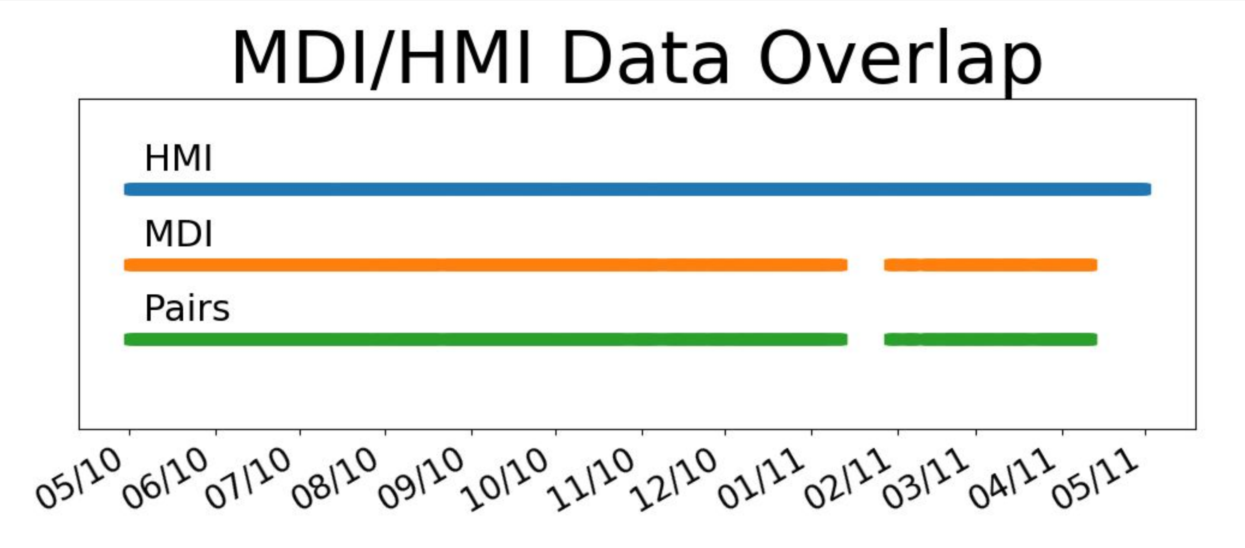
The goal:

Create a cohesive machine learning data set for solar flare prediction

Data

We use line-of-sight magnetograms:

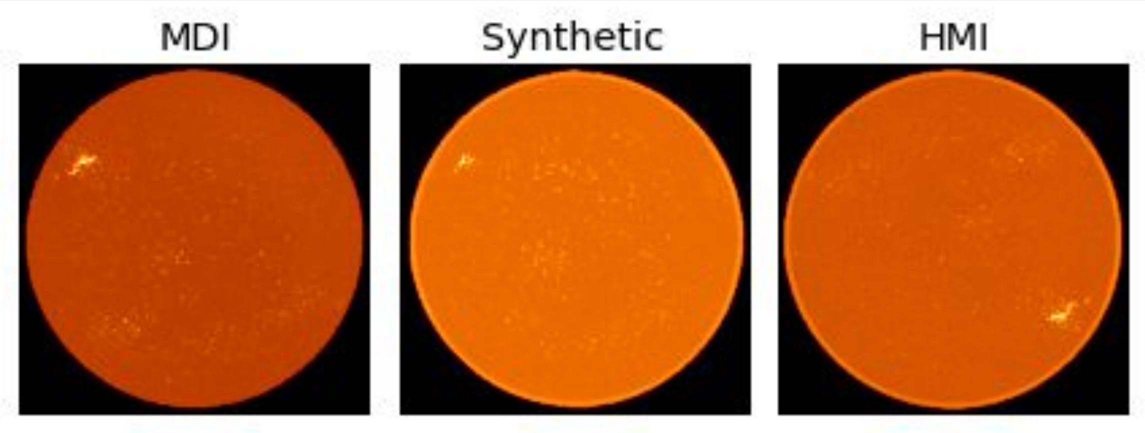
- from the SOHO/MDI magnetogram dataset (96m cadence)
- from the SDO/HMI magnetogram dataset (720s cadence)



Methods: Image Translation

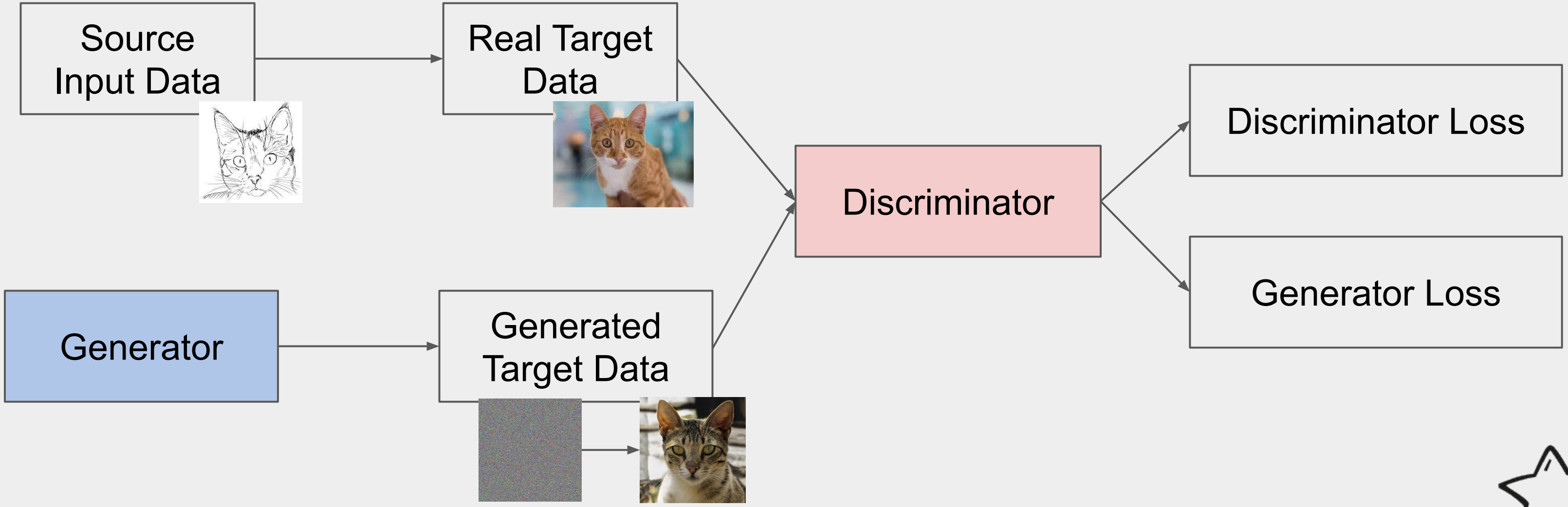
Image-to-Image Translation: generate a synthetic version of an given image with a modification

Good for super-resolution problems!



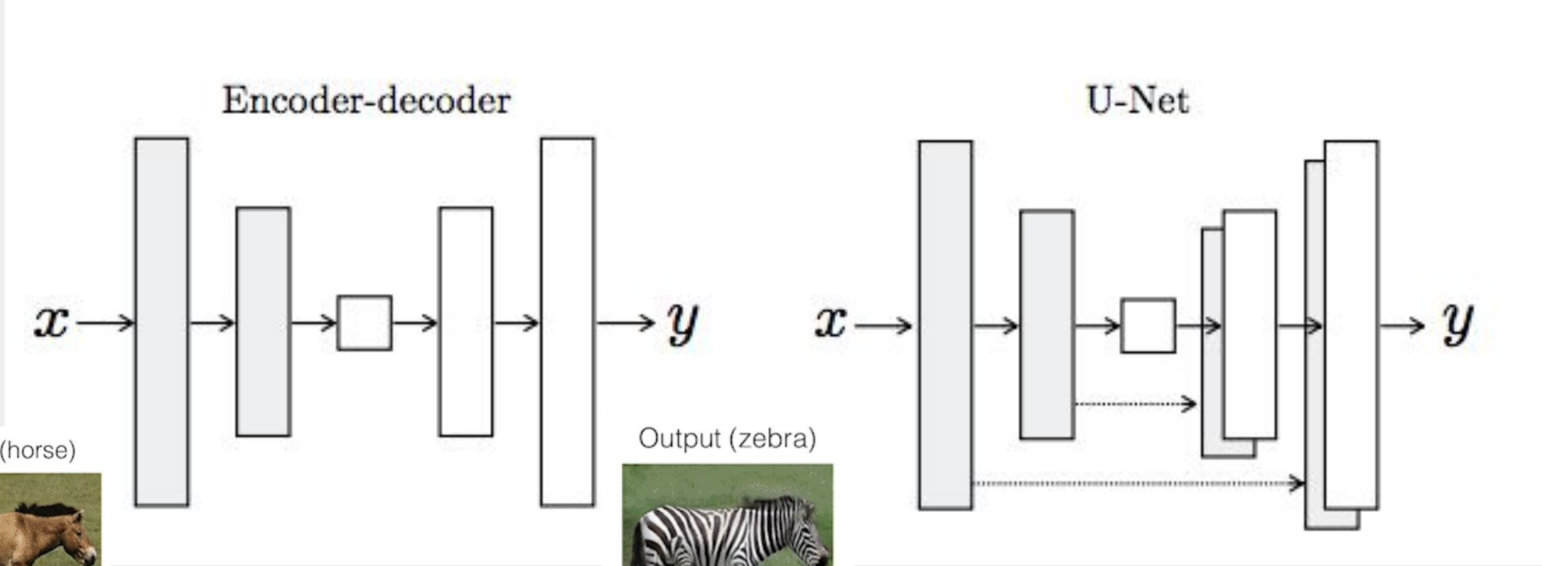
Methods

Generative Adversarial Network (GAN) - 2014



thiscatdoesnotexist.com

Generator Architecture



Input (horse)

Output (zebra)

Model exploration

Image Translation: Most models require INPUT→ OUTPUT image training pairs

Pix2Pix (2016)

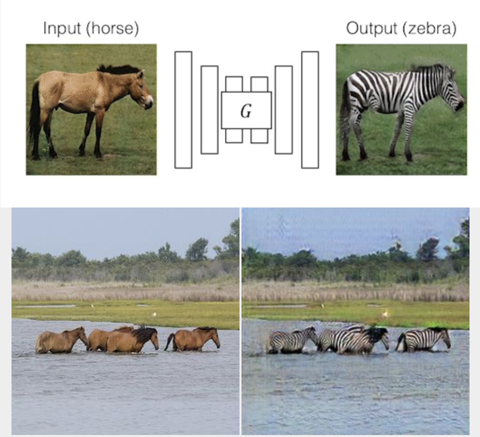
"General Purpose"

Paired

Isola et. al. 2016

CycleGAN (2017)

Unpaired

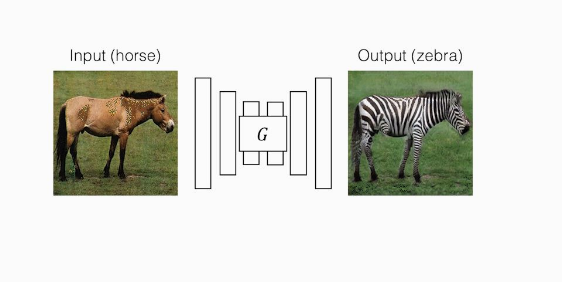


Zhu et. al. 2017

CUT (2020)

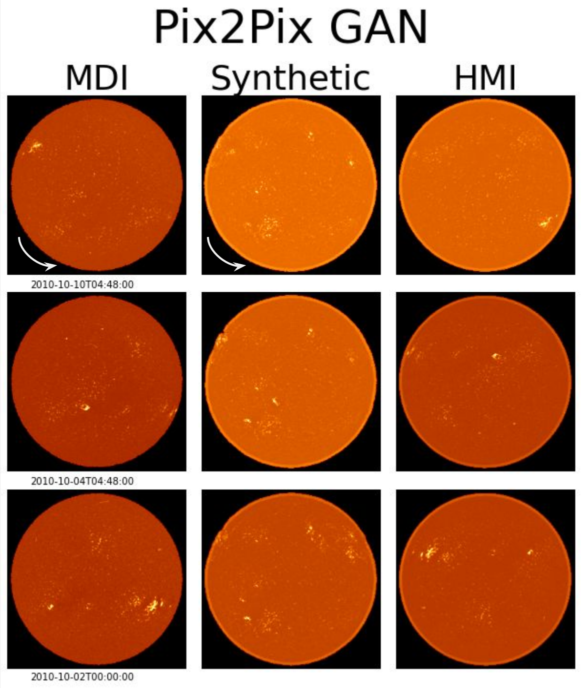
Model training is faster and less memory-intensive

Unpaired

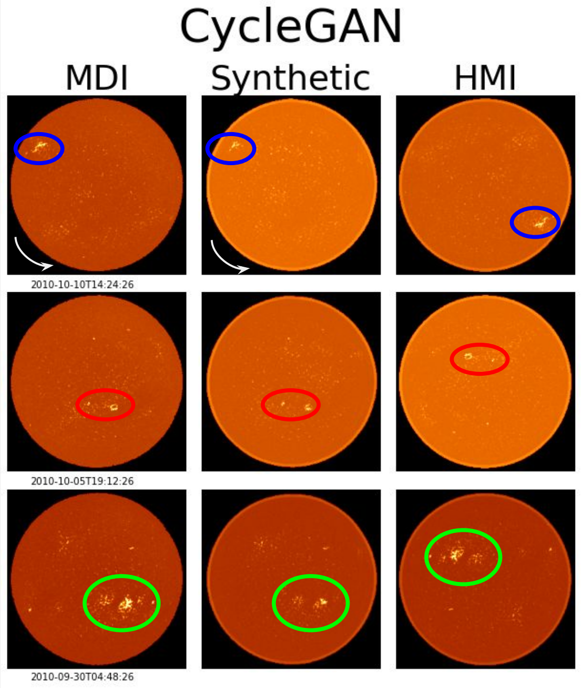


Park et. al. 2020

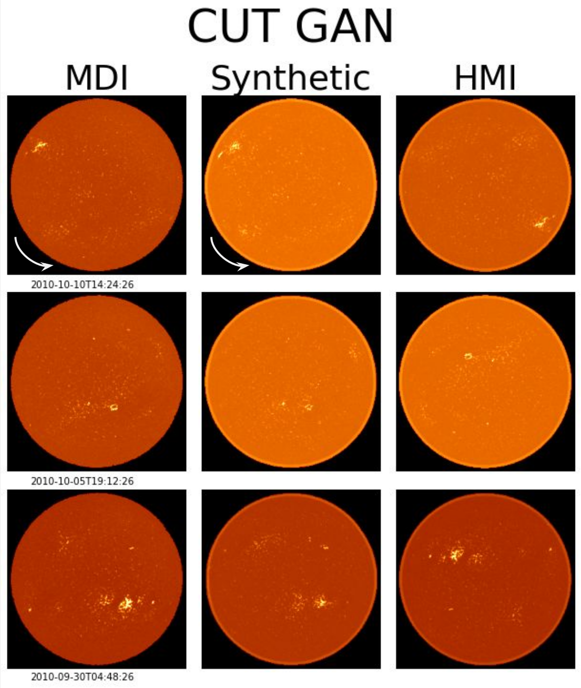
Results



200 epochs



200 epochs



400 epochs

Conclusion

Goal: Create a dataset of super-resolved SOHO/MDI images of SDO/HMI quality

- Identified data overlap
- Created image training pairs
- Tested 3 GAN models to upsample the MDI images
  - Pix2Pix introduced some strange image artifacts
  - CycleGAN and CUT worked well
- **Takeaways:**
  - Start simple
  - Modify existing and well-documented models
  - This technique shows promise for creating a high-quality combined solar magnetogram dataset!

Future Work

Next steps:

- Rotate MDI images in preprocessing
- Get a quantitative error calculation -- RMS Error
- Training, getting a new GPU!
  - More iterations
  - Training on full-sized images
  - Patchwise

Acknowledgements

Dr. Wendy Carande

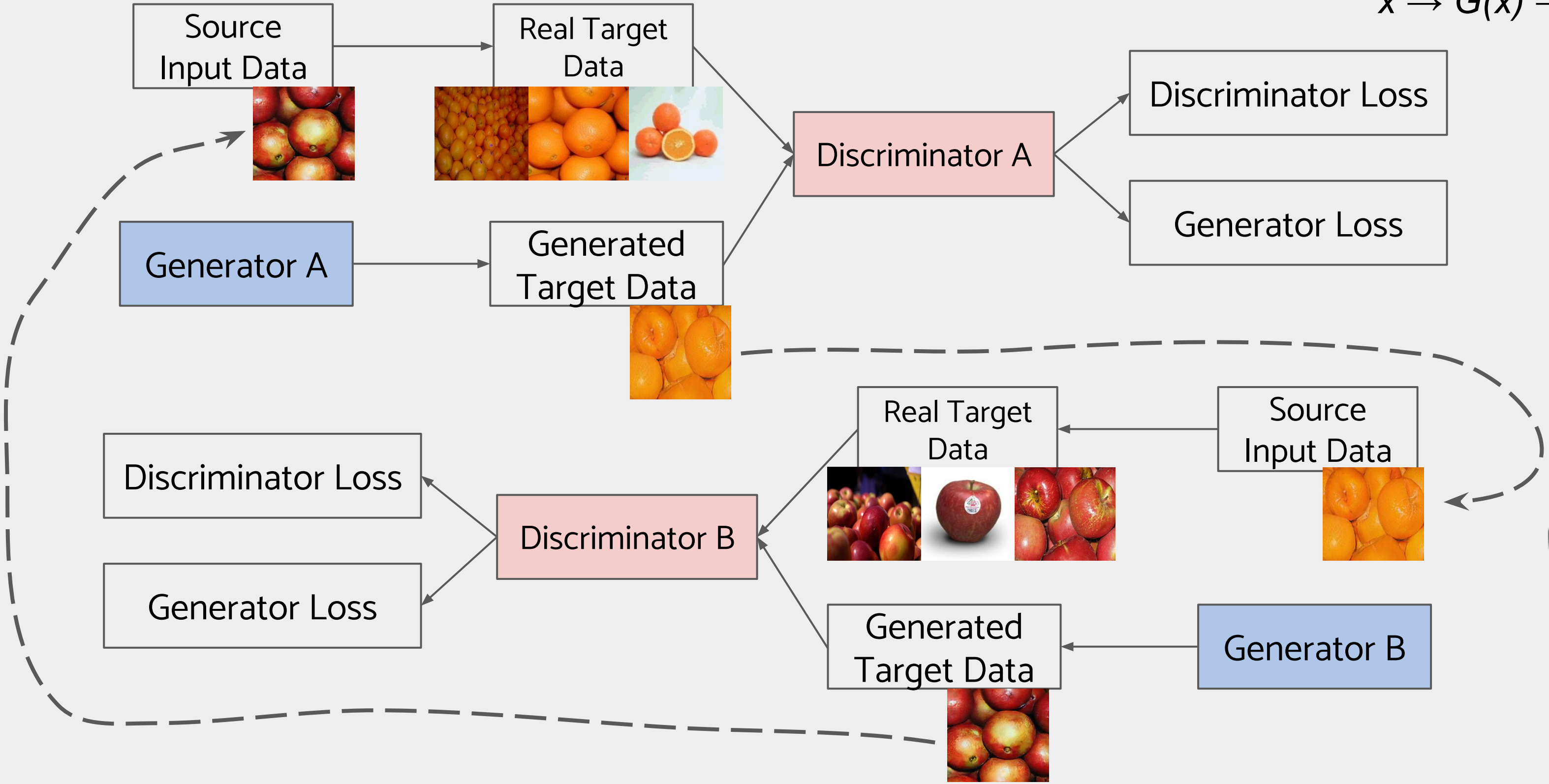
Dr. Andrés Muñoz-Jaramillo

Katy Luttrell

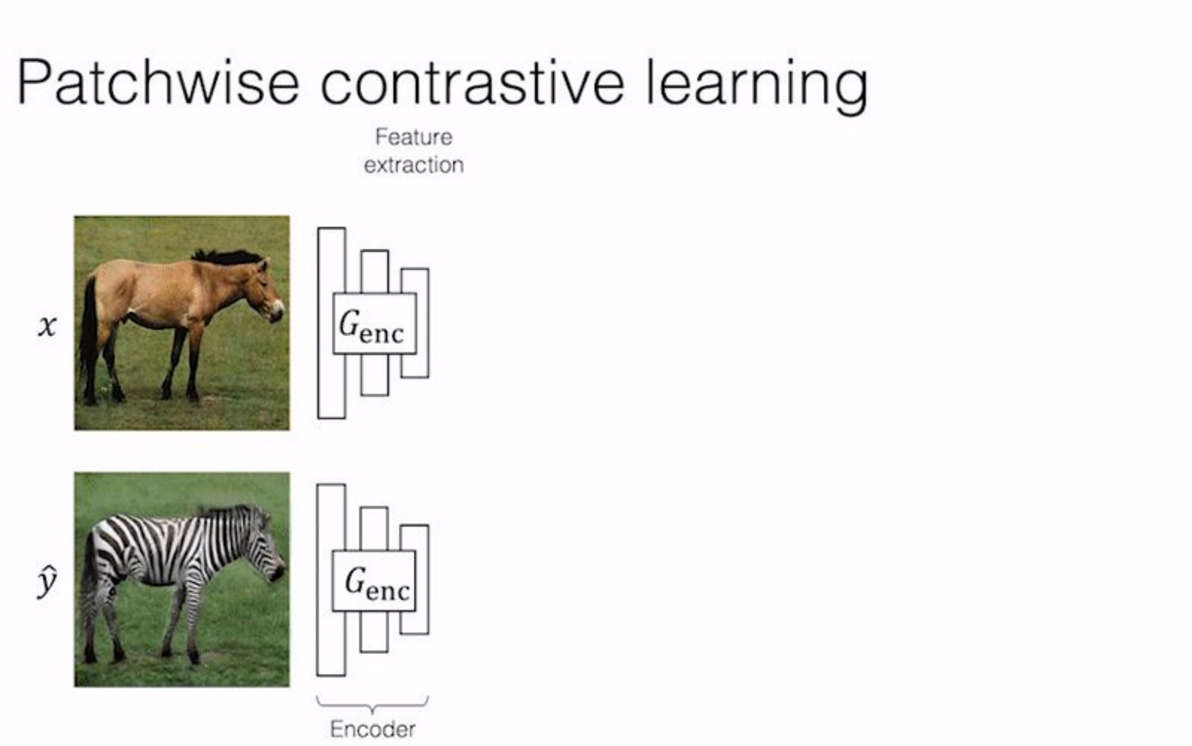
This research was funded by the SWx TREC Deep Learning Lab.

Appendix - CycleGAN

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$



Appendix - CUT



Park et. al. 2020

Appendix - Loss Functions

Discriminator Loss:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

GAN Loss:

$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) =& \mathbb{E}_{x,y}[\log D(x, y)] + \\ & \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z))],\end{aligned}$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1].$$

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

Isola et. al. 2016