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

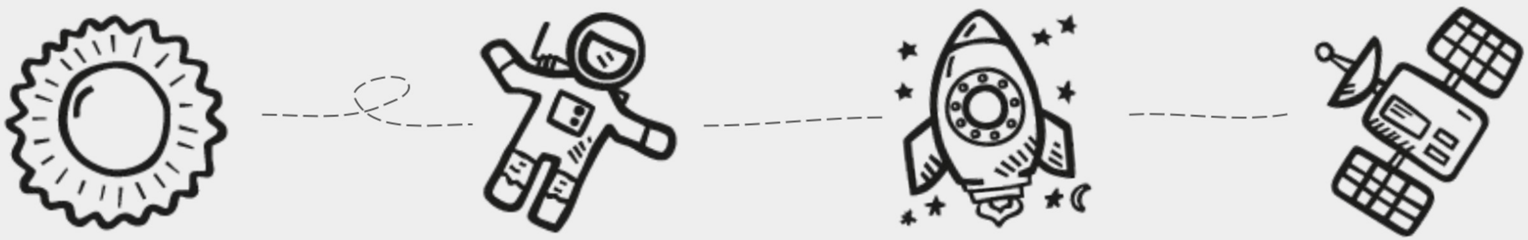
Data Augmentation of Magnetograms for Solar Flare Prediction using GANs

Allison Liu[1], Wendy Carande[1]

[1] Laboratory for Atmospheric and Space Physics

allison.liu@lasp.colorado.edu

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

BBC | Sign in | Home | News | Sport | Reel | Worklife | Travel

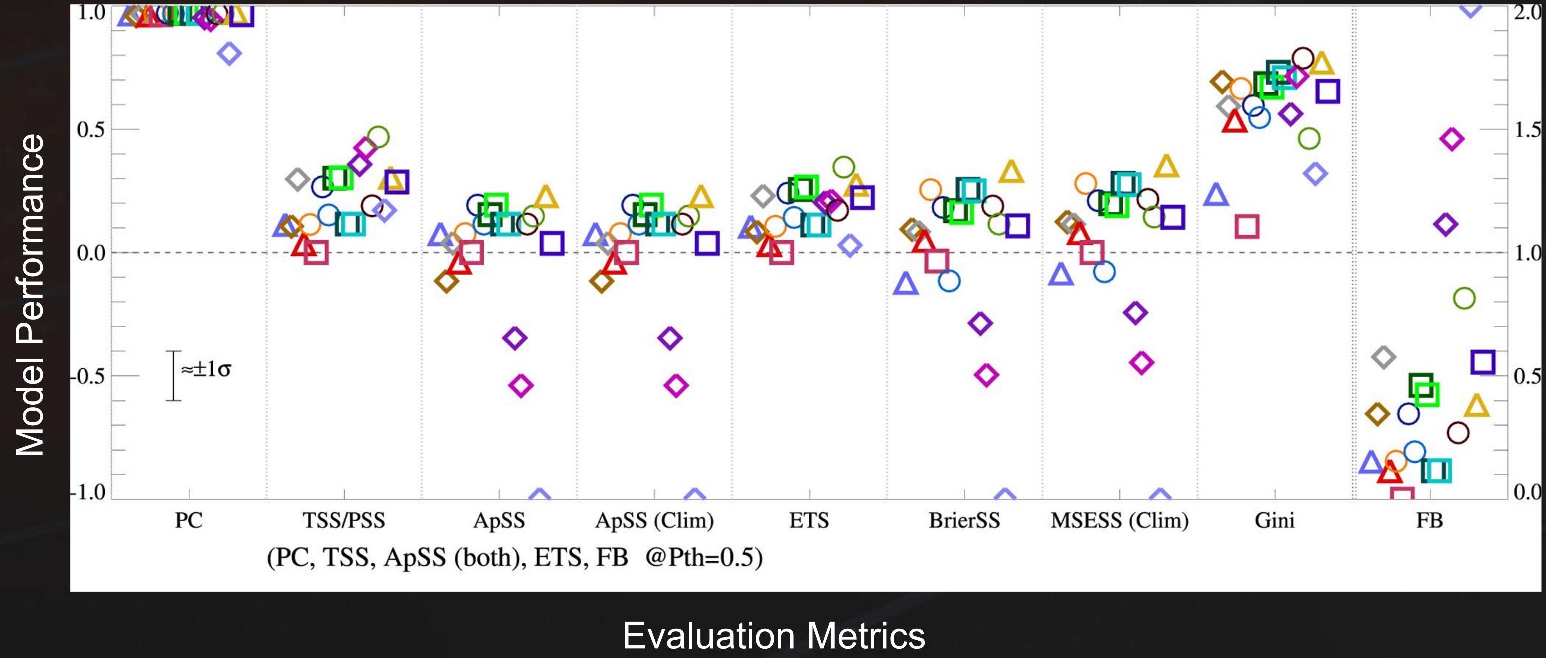
NEWS

SpaceX loses 40 satellites to geomagnetic storm a day after launch

9 February

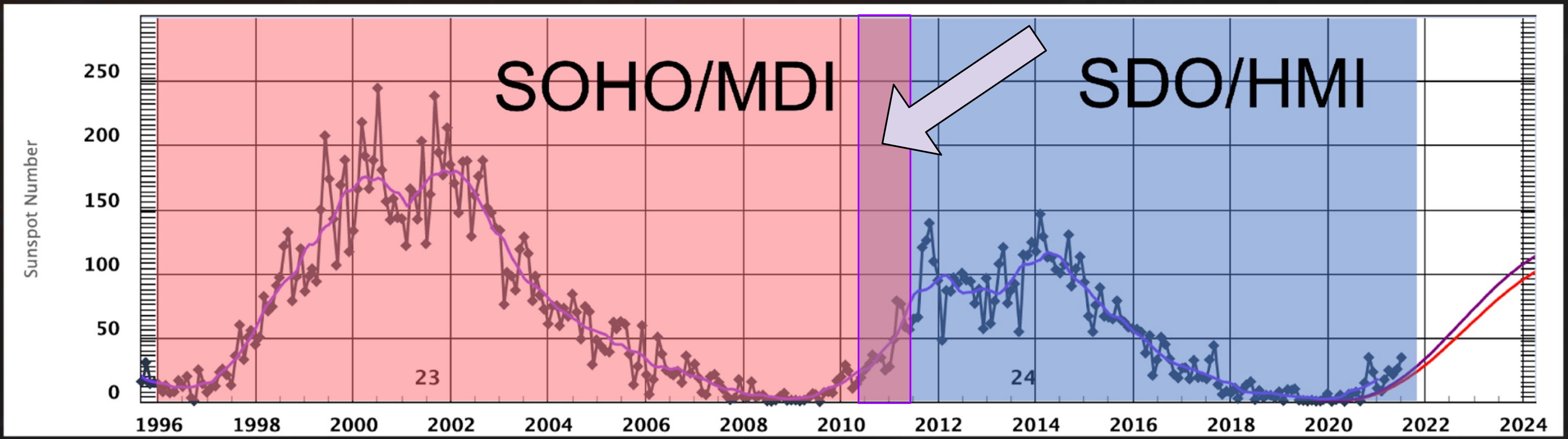
Background

- Solar flare prediction is done largely by humans → Machine Learning ~ 2010
- The operational model used by the Space Weather Prediction Center (SWPC) is a human-in-the-loop climatology-based forecast model

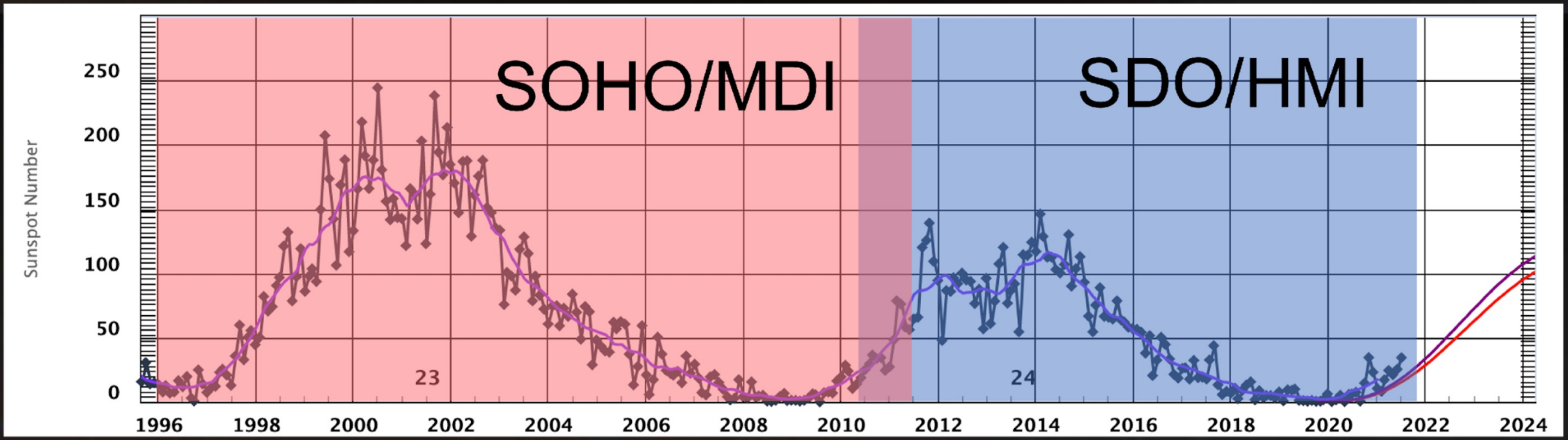


Goal

- **Problem:** the two magnetogram datasets used for solar flare prediction differ in resolution and field of view, so the older SOHO/MDI dataset is often unused in training of solar flare prediction models
- **The goal of this project is to create a combined dataset that could improve the accuracy of solar flare prediction models by incorporating data which spans an additional solar cycle.**



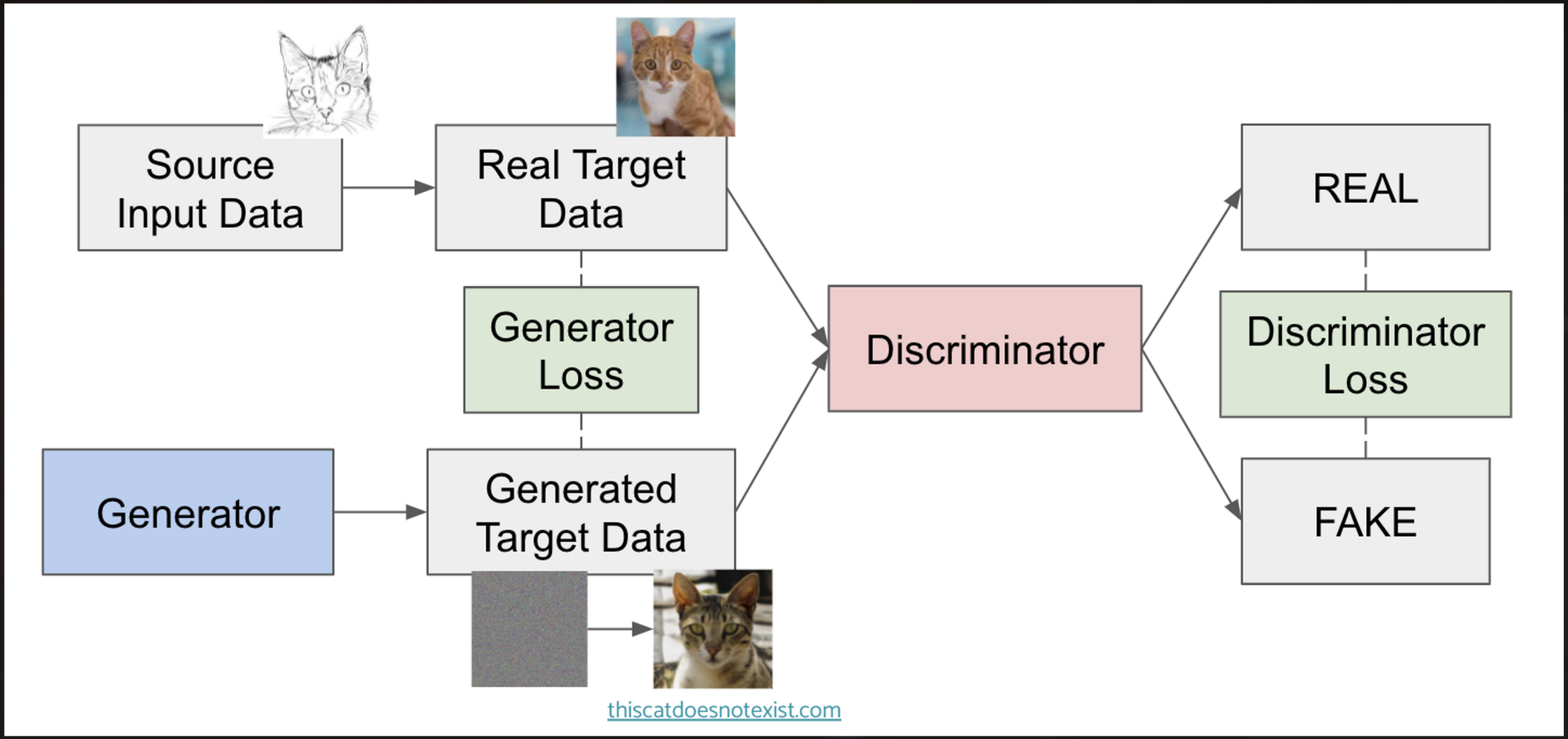
Data and Preprocessing



- We use line-of-sight, full-disk magnetograms from:
  - the NASA Solar Dynamic Observatory/Helioseismic and Magnetic Imager (SDO/HMI), 720 sec cadence.
  - the Solar and Heliospheric Observatory/Michelson Doppler Interferometer (SOHO/MDI), 96 min cadence.
- Preprocessing: Images with holes or missing header files removed

Generative Adversarial Networks (GANs)

GANs are a class of generative models, which are useful for creating new data instances.



Model Exploration

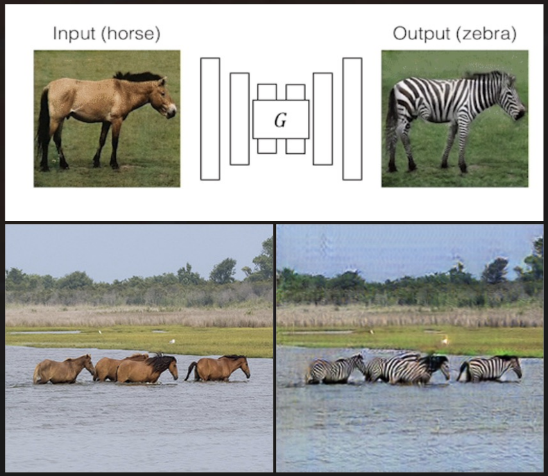
Image Translation: Most models require INPUT → OUTPUT training pairs

Pix2Pix
(Isola et al. 2016)

Paired

CycleGAN
(Zhu et al. 2017)

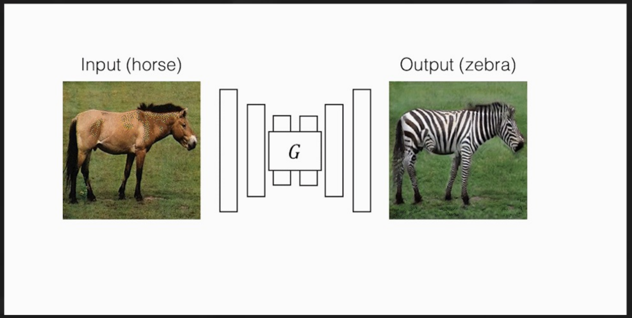
Unpaired



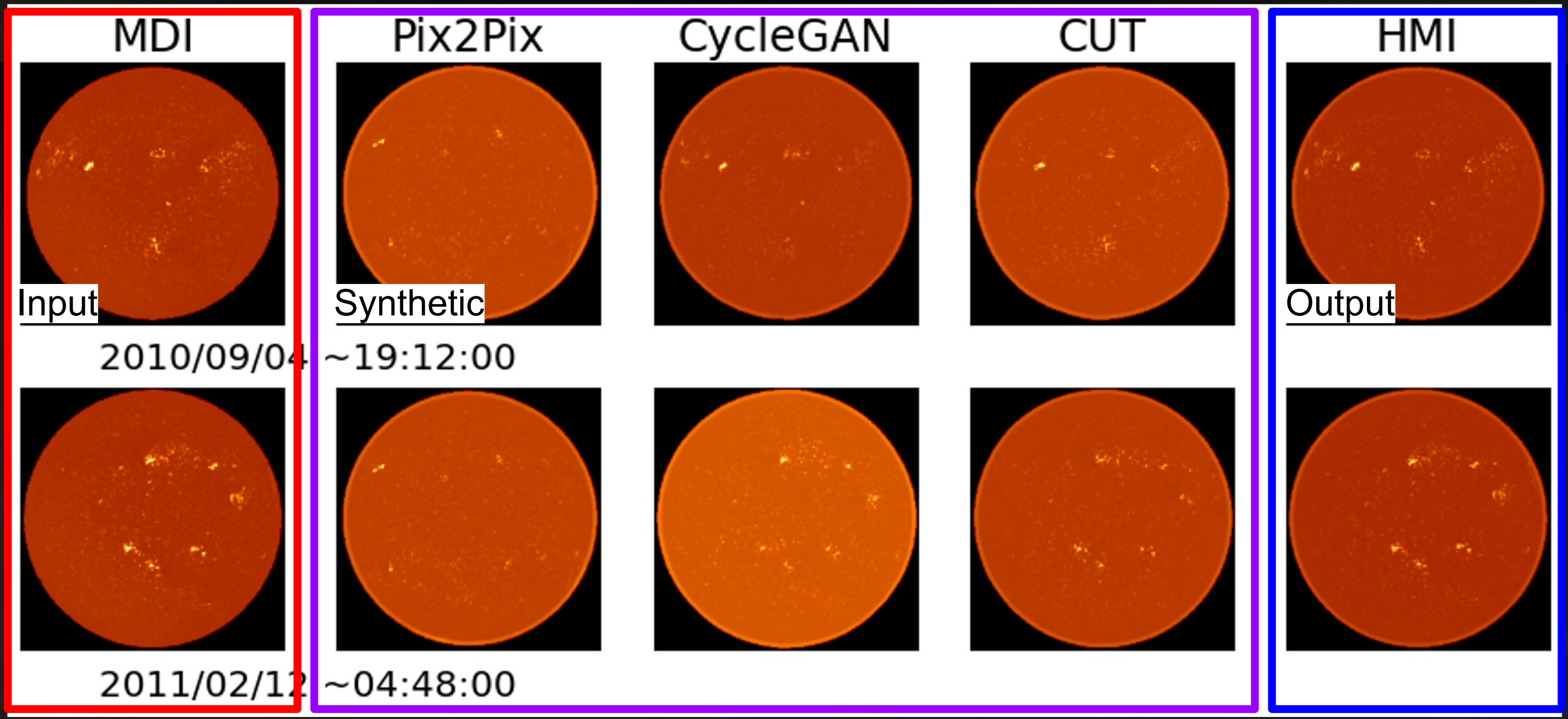
CUT
(Park et al. 2020)

Unpaired

Model training is faster and less memory-intensive

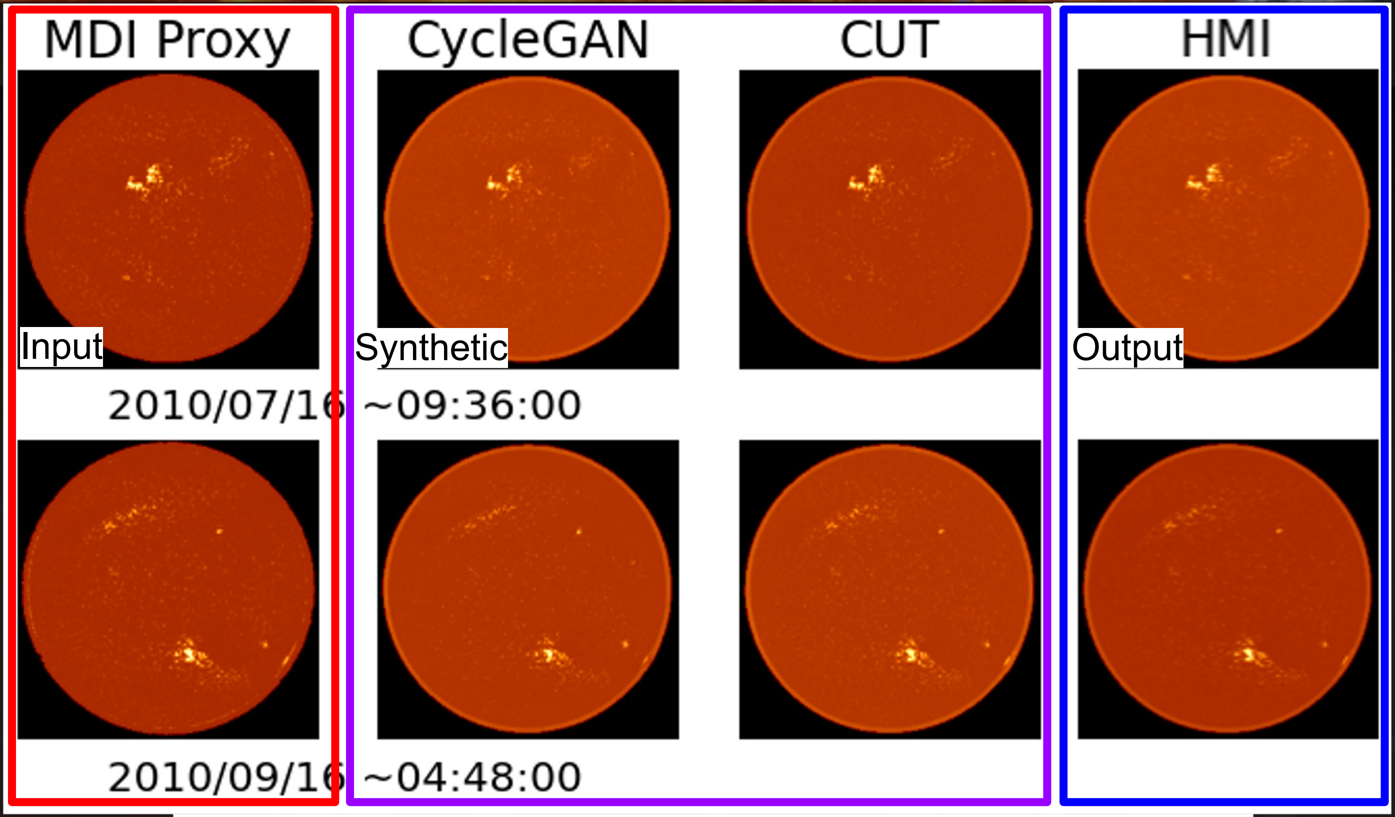


Preliminary Results - Magnetograms



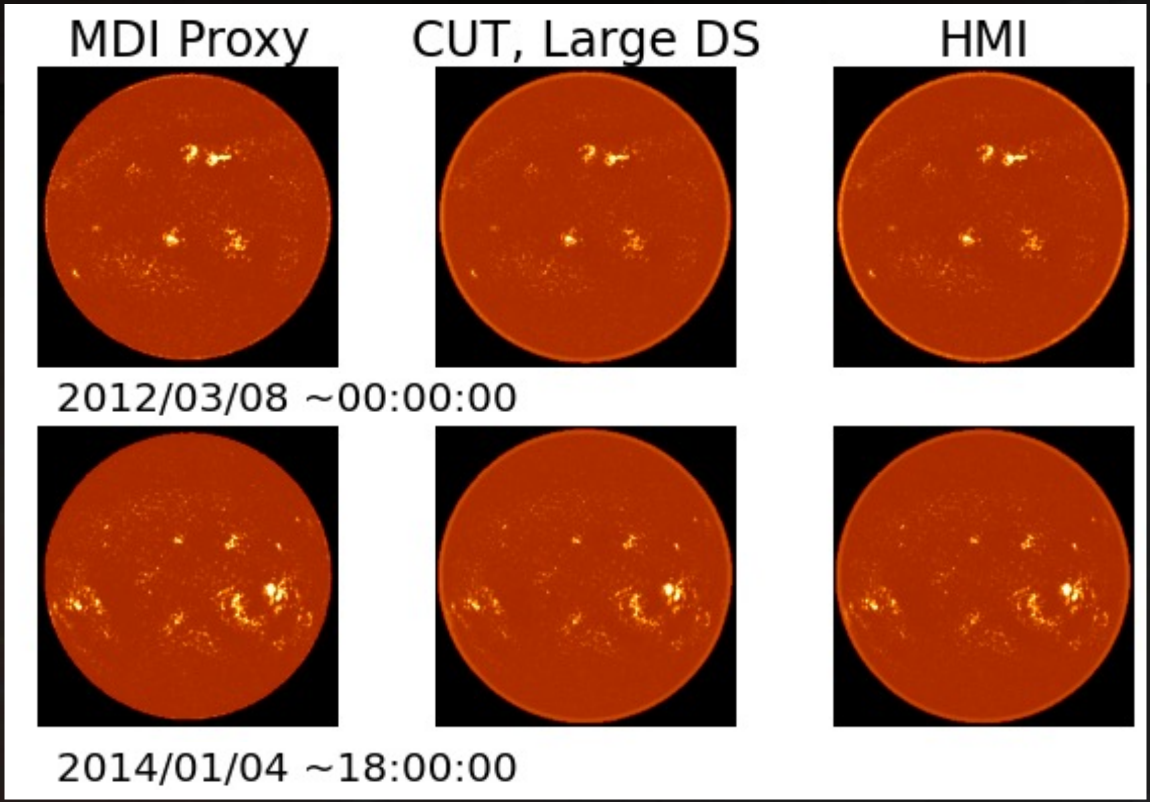
	Pix2Pix	CycleGAN	CUT	HMI vs MDI
MSE	0.001597	0.000782	0.000807	0.001247
RMSE	0.039666	0.027545	0.027956	0.034912
PSNR	28.096537	31.330650	31.219764	29.240296
SSIM	0.831117	0.898249	0.893415	0.775676
FID	45.366197	15.864683	18.880193	126.748518

Preliminary Results - Magnetograms



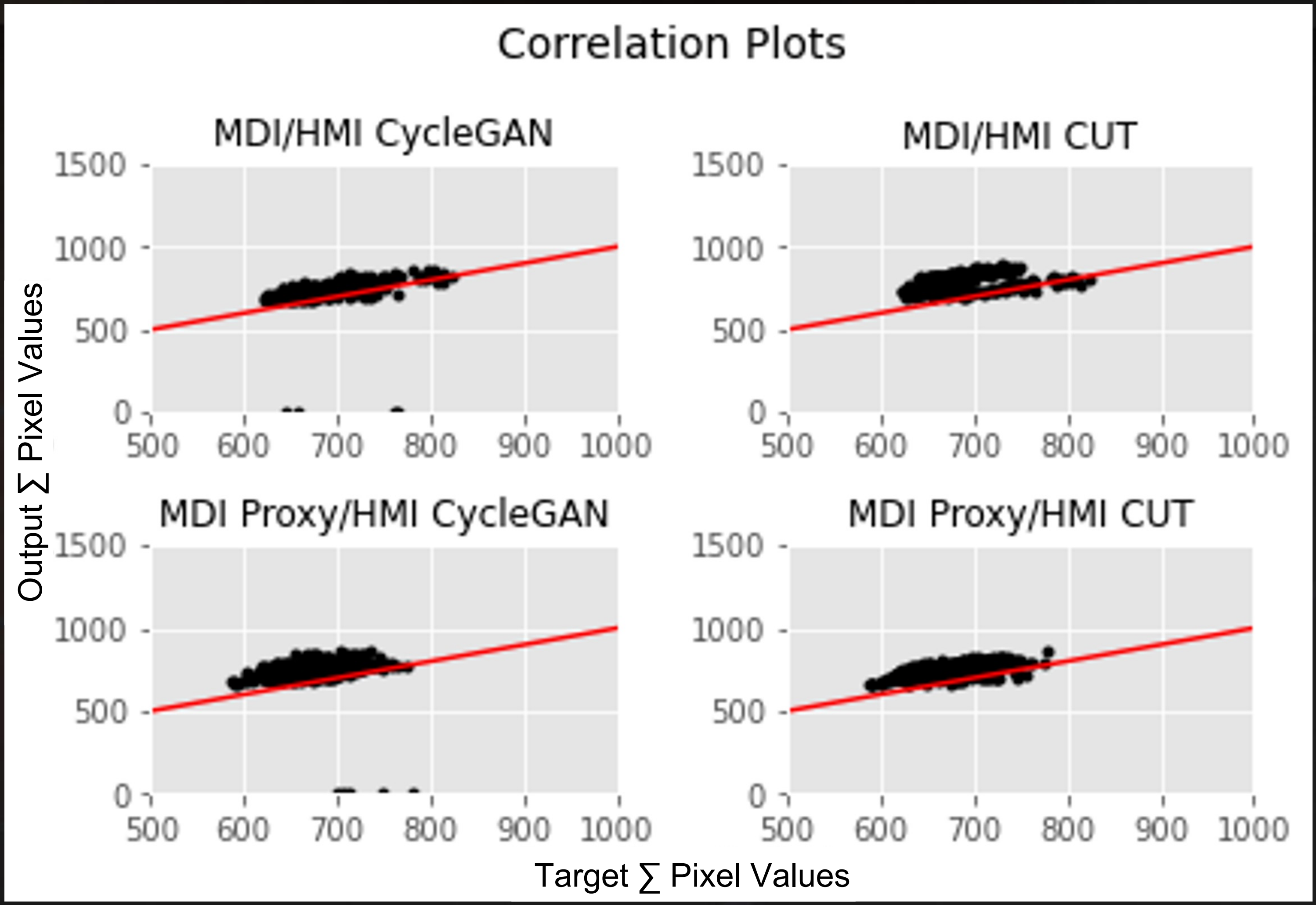
	CycleGAN	CUT	CUT (Large)	HMI vs MDI Proxy
MSE	0.001116	0.001160	0.002058	0.000873
RMSE	0.032977	0.033492	0.045368	0.029362
PSNR	29.759957	29.658709	26.865056	30.702266
SSIM	0.873809	0.875324	0.795206	0.774147
FID	5.663433	2.275642	0.972010	81.722521

Preliminary Results - Magnetograms



	CycleGAN	CUT	CUT (Large)	HMI vs MDI Proxy
MSE	0.001116	0.001160	0.002058	0.000873
RMSE	0.032977	0.033492	0.045368	0.029362
PSNR	29.759957	29.658709	26.865056	30.702266
SSIM	0.873809	0.875324	0.795206	0.774147
FID	5.663433	2.275642	0.972010	81.722521

Preliminary Results - Analysis



Conclusions

Unpaired models like CycleGAN and CUT are promising for translating SOHO/MDI magnetograms to SDO/HMI quality.

- Both models perform similarly, with CUT having faster training times and appearing to resolve finer features more accurately

Next Steps:

- Feature alignment and per-pixel accuracy analysis
- Try running models on full-resolution magnetogram data

Acknowledgements

- Dr. Wendy Carande
- Katy Luttrell
- Tom Berger

Contact

Allison Liu
allison.liu@lasp.colorado.edu
Wendy Carande
wendy.carande@lasp.colorado.edu

This project was funded by the SWx TREC Deep Learning Laboratory and NSF REU Program, award #1659878

Appendix: Downsampling Image Pairs

Downsampling:

- Using Gaussian filter with FWHM 4.7 HMI px and truncated at 15 HMI px.
- Downsizing from 4096x4096 px to 1024x1024 by averaging using a bicubic interpolation over a 4x4 px neighborhood (using cv2 implementation of resize).
- Correcting for pixel values using the equation MDI = -0.18 + 1.4*HMI
- This was the procedure done in Y Liu 2012, comparing HMI and MDI data

	MDI vs HMI	MDI Proxy vs HMI
MSE	0.001247	0.000873
RMSE	0.034912	0.029362
PSNR	29.240296	30.702266
SSIM	0.775676	0.774147

MDI vs MDI Proxy FID: 6.753753003016939

Comparison of LOS Magnetograms Taken by SDO/HMI and SOHO/MDI. Y Liu et al (2012)