A Deep Generative+Self-Supervised Approach that Uses Physical Quantities to Generate and Retrieve Solar Images

Subhamoy Chatterjee¹, Andrés Muñoz-Jaramillo¹, Anna Malanushenko² and the SEARCH team

(1) Southwest Research Institute, Boulder, CO, USA, (2) High Altitude Observatory, NCAR, Boulder, CO, USA

Funded by NASA HGIO Grant 80NSSC23K0416 and NASA HITS Grant 80NSSC22K1611

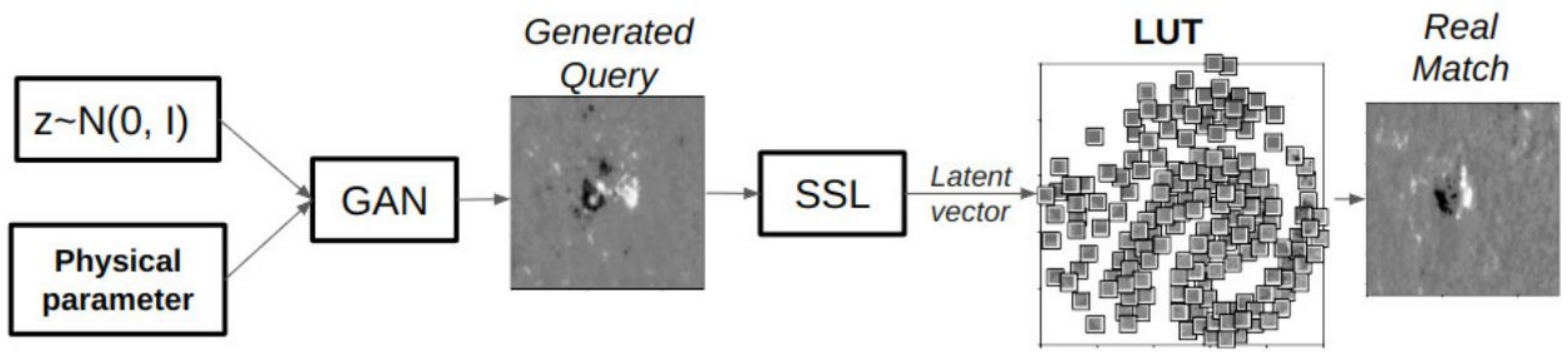
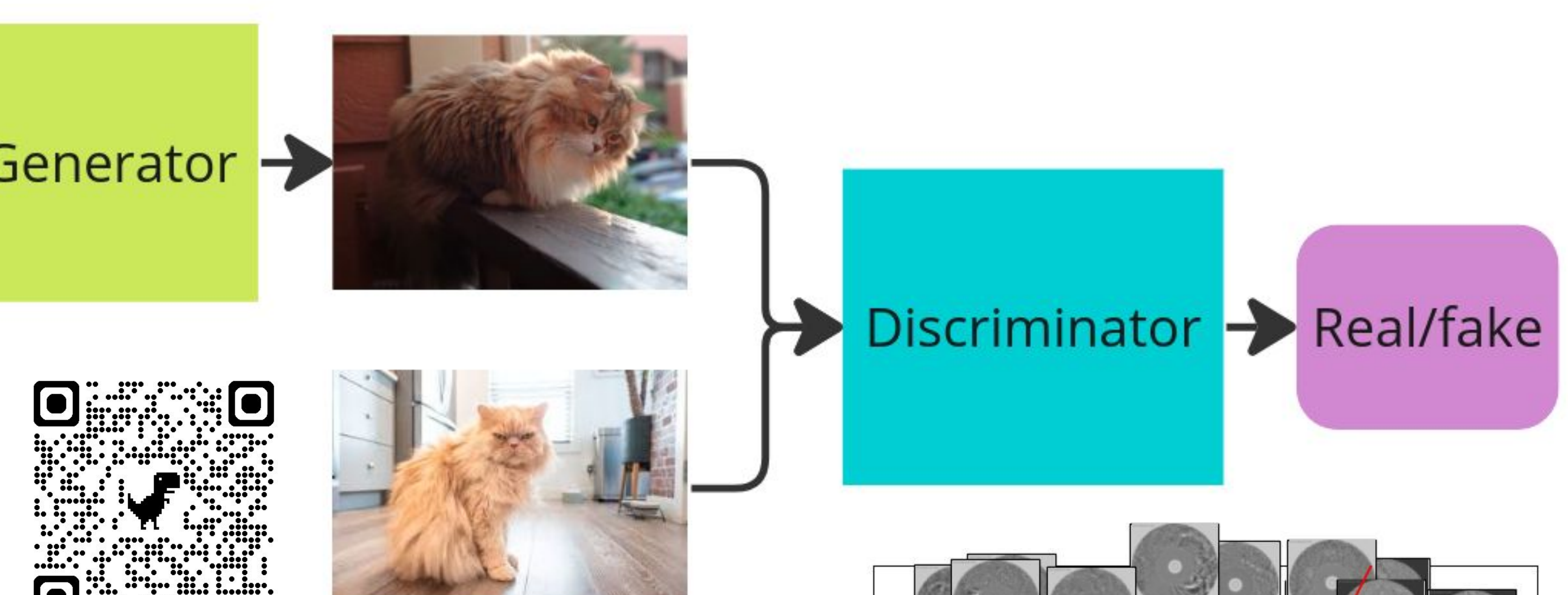


Figure 1

Motivation

- Deep generative models have shown immense potential in generating unseen data that has properties of real data.
- These models learn complex data-generating distributions starting from a smaller set of latent dimensions.
- Generative models have encountered great skepticism in scientific domains due to the disconnection between generative latent vectors and scientifically relevant quantities.
- ***In this study[1], we integrate three types of machine learning models to generate solar magnetic patches in a physically interpretable manner and use those as query to find matching patches in real observations. (see Figure 1)***
- We use the magnetic field measurements from Space-weather HMI Active Region Patches (SHARPs) to train a Generative Adversarial Network (GAN[2]). (see Figure 2)



Scan to access "Sun Generator"
https://sungenerator-subhamoy.streamlit.app/
This App that allows user to interact with latent space for tweaking solar magnetic patch generation

Figure 2

Potential Application on CME Data Mining

- ***Connecting CME properties to Generative Latent space will allow efficient exploration of CMEs***
- Latent vectors capture inherent structure of CME images (Figure 7)
- Generative models allow continuous changes in CME images by changing the input latent vectors
- Mapping latent vectors to CME properties allow generation of new CME images and mine rare cases that can be hidden in a large dataset

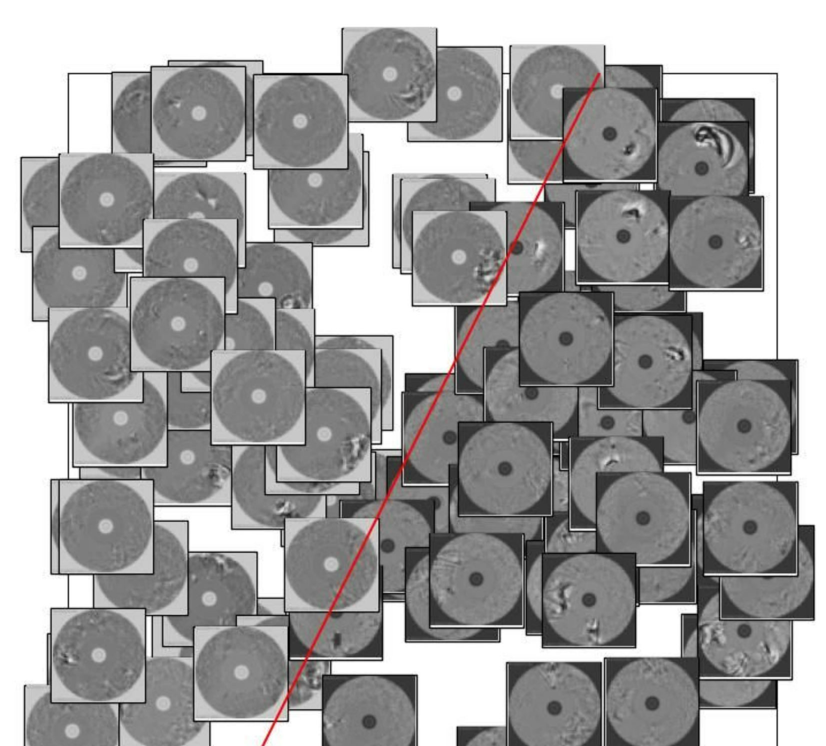


Figure 7

Supervised Learning on Generative Latent Space

- We calculate physical parameters from the generated images – Total Unsigned Field (TUF), Total Positive/Negative Field (TPF/TNF), polarity separation (PSEP) etc.
- ***We use supervised learning to learn decision boundaries (in red) that separate higher and lower values of physical parameters (see Figures 3, 4, 7)***
- Moving along direction normal to those decision boundaries makes smooth changes in those physical properties of the generated images[1]

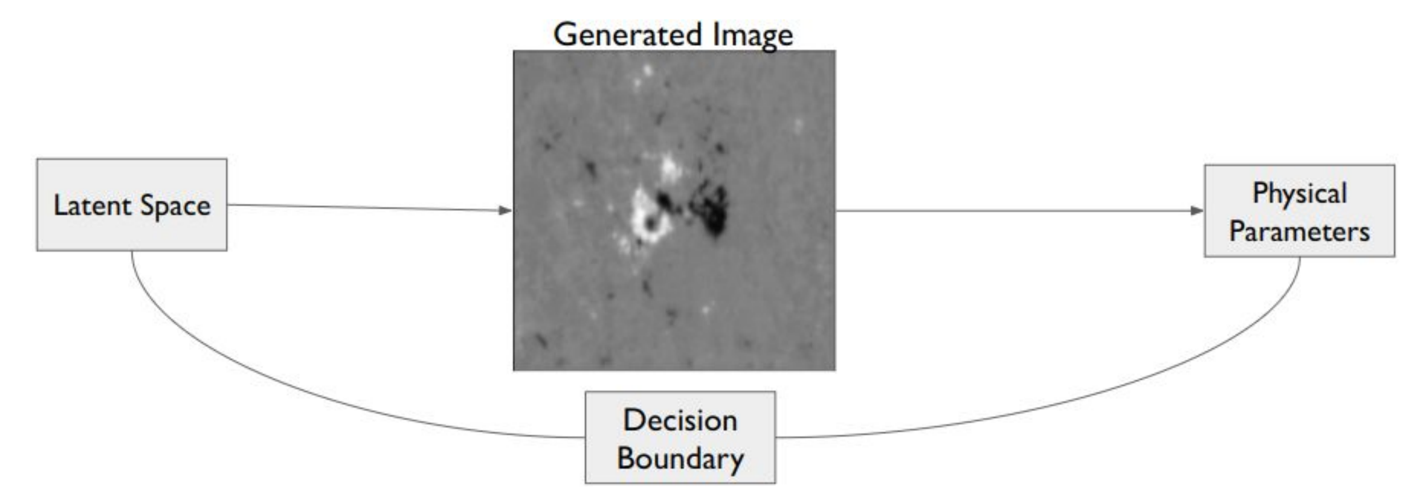


Figure 3

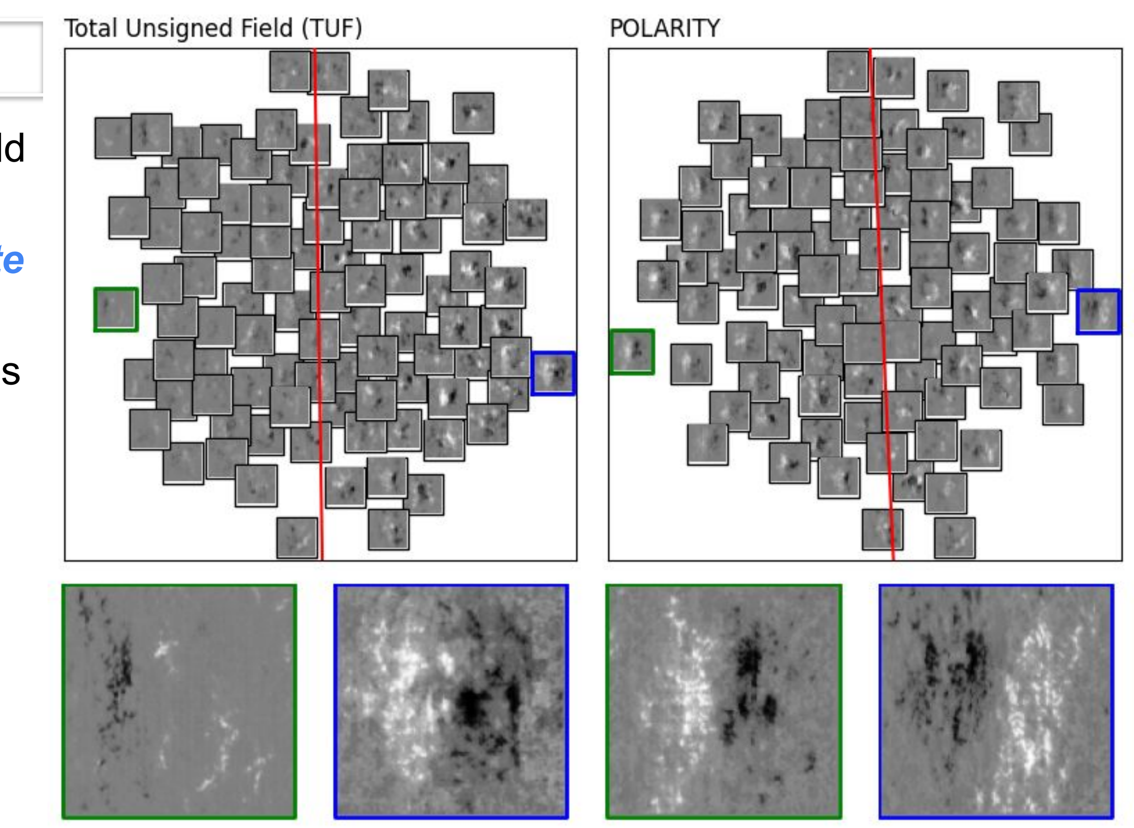


Figure 4

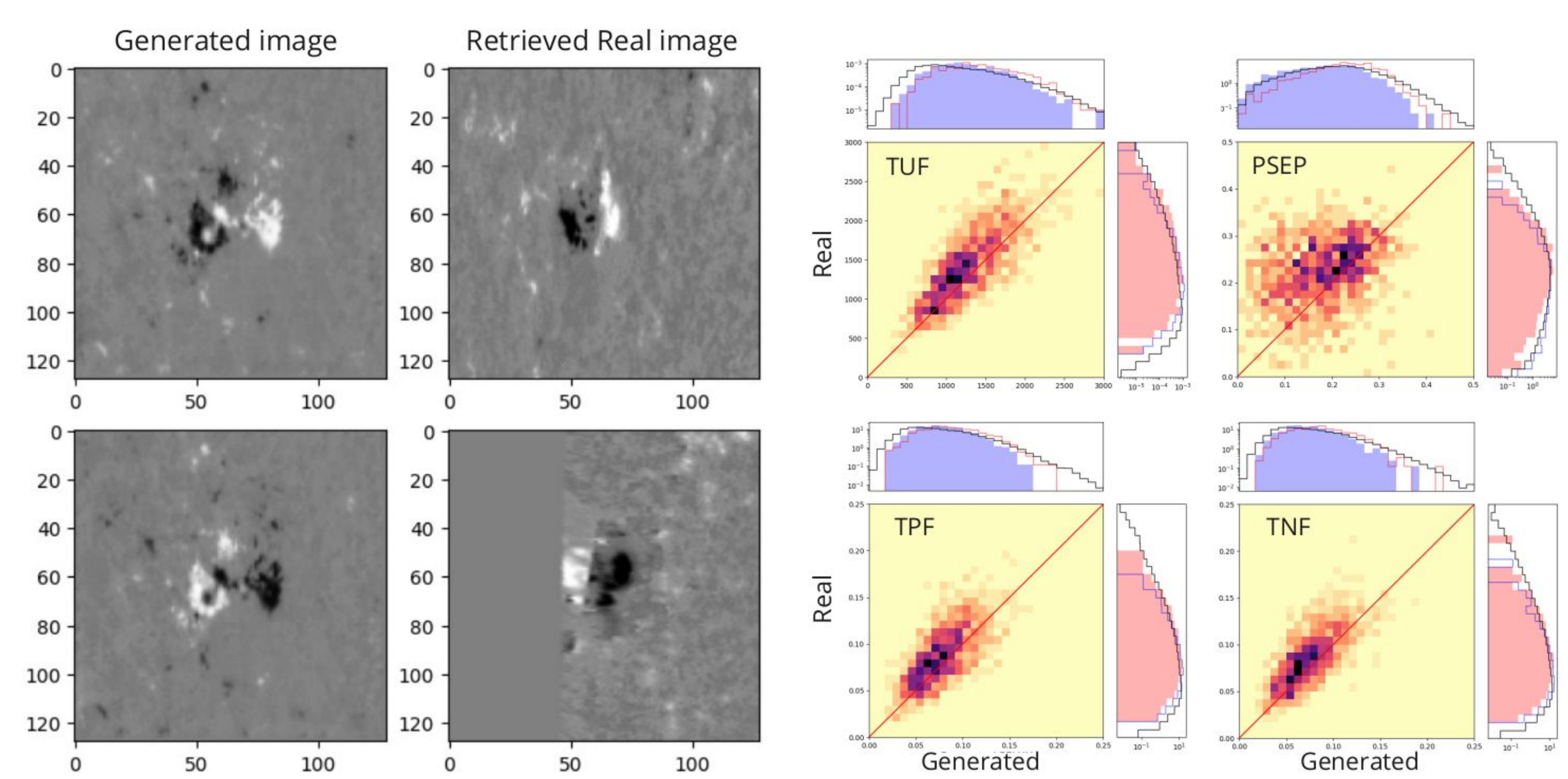


Figure 5

Self-Supervised Learning to find Nearest Observation

- We train a Self-Supervised Learning (SSL) model called SimSiam[3] on SHARP data to learn augmentation invariant latent representation
- We build a Look Up Table (LUT) that maps SHARPs to SSL-derived latent vectors. (see Figure 1)
- ***We use Generated Image as Query and find nearest observed SHARP using the SSL latent space. (see Figures 1, 5)***
- The physical properties of Generated Query and retrieved Real Match show high statistically significant correlation (2D histograms on Figure 5).
- Generated Queries match the physical domains of real images in terms of the derived physical properties. (marginal histograms on Figure 5).

SEARCH

- ***A community driven Self-Supervised Learning effort to build a reverse image search application on SDO Data***
- SEARCH = **S**DO **E**xploration **A**nd **R**esearch **C**ommunity for **H**eliophysics.
- Build SDO/AIA data downloader + AI/ML ready packager
- Trained SSL models on SDO/AIA mutli-wavelength images
- Built reverse image-search tool making use of SSL latent space (Figure 6)

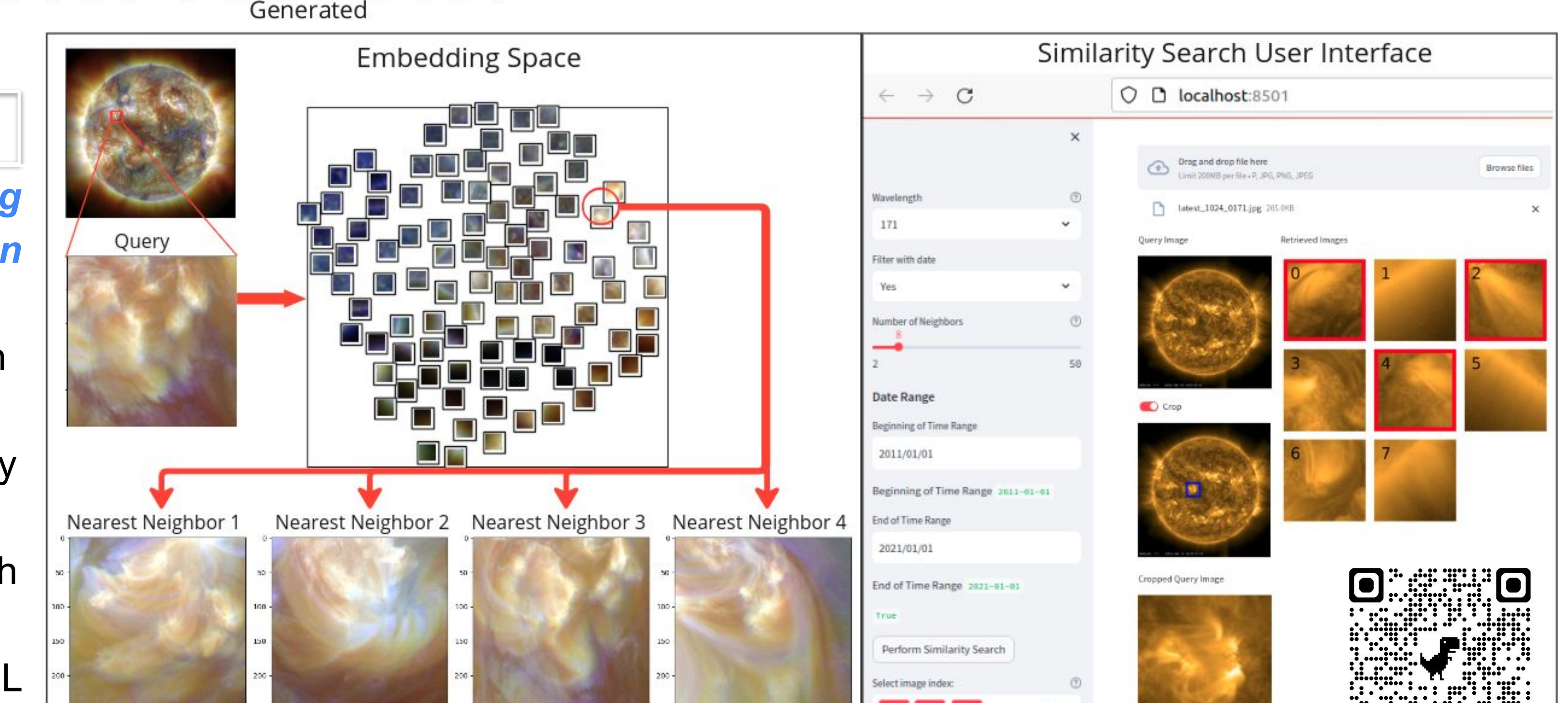


Figure 6

Conclusion

- We demonstrated that a Deep Generative model can be harnessed to generate scientifically meaningful queries to find matches in scientific datasets.
- This image generation can be tweaked along directions that reflect changes in any set of chosen physical quantities.
- We showed that the generated images can also be used as a query to the SSL-derived latent space to retrieve matches from solar images. These retrieved regions matched the query both visually and quantitatively.
- ***This approach thus elevates generative models from a means-to-generate-synthetic data to a novel tool for the efficient mining of real scientific data.***
- Even though we demonstrated this approach in a specific domain of solar astronomy, it can be easily adapted to any other field of astronomy dealing with big datasets of any modality and complexity.

References

1. Chatterjee, S., Munoz-Jaramillo, A., and Malanushenko, A., Deep Generative model that uses physical quantities to generate and retrieve solar magnetic active regions,arXiv e-prints arXiv:2502.05351 (2025). DOI 10.48550/arXiv.2502.05351
2. Goodfellow, I. J. et al. Generative Adversarial Networks. arXiv e-prints arXiv:1406.2661 (2014). DOI 10.48550/arXiv.1406.2661. 1406.2661
3. Chen, X. & He, K. Exploring Simple Siamese Representation Learning. arXiv e-prints arXiv:2011.10566 (2020). DOI 10.48550/arXiv.2011.10566. 2011.10566