Applying Self-Supervised Representation Learning to UNIONS

Ashley Ferreira[1], Teaghan O'Briain[3], Spencer Bialek[2], Nat Comeau[3], Nick Heesters[4], and Sebastien Fabbro[2,3]

[1] University of Waterloo, [2] University of Victoria, [3] Herzberg Astronomy and Astrophysics Research Centre, [4] Ecole Polytechnique Fédérale de Lausanne

Introduction

- **U**ltraviolet **N**ear **I**nfrared **O**ptical **N**orthern **S**urvey (UNIONS) uses data from 3 key telescopes in Hawaii: Pan-STARRS, Subaru, and the Canada-France-Hawaii Telescope to answer big questions in astrophysics.
- There are many **M**achine **L**earning (ML) efforts underway to analyze the data but the models are trained from scratch using supervised learning to accomplish a specific task. However, the performance of these models is limited by the quantity and quality of labels (e.g. galaxy type).
- **S**elf-**S**upervised **L**earning (SSL) is a promising alternative to supervised learning where far fewer labels are needed and this has shown great potential when applied to data from a couple of astronomy surveys.

Task

- Our group hopes to exploit advances in SSL to create a model that generates meaningful lower dimensional representations of astronomy observations without the need for explicit labels.

Methods

- See Figure 1 for overview of methodology and Figure 2 for example reconstructions.

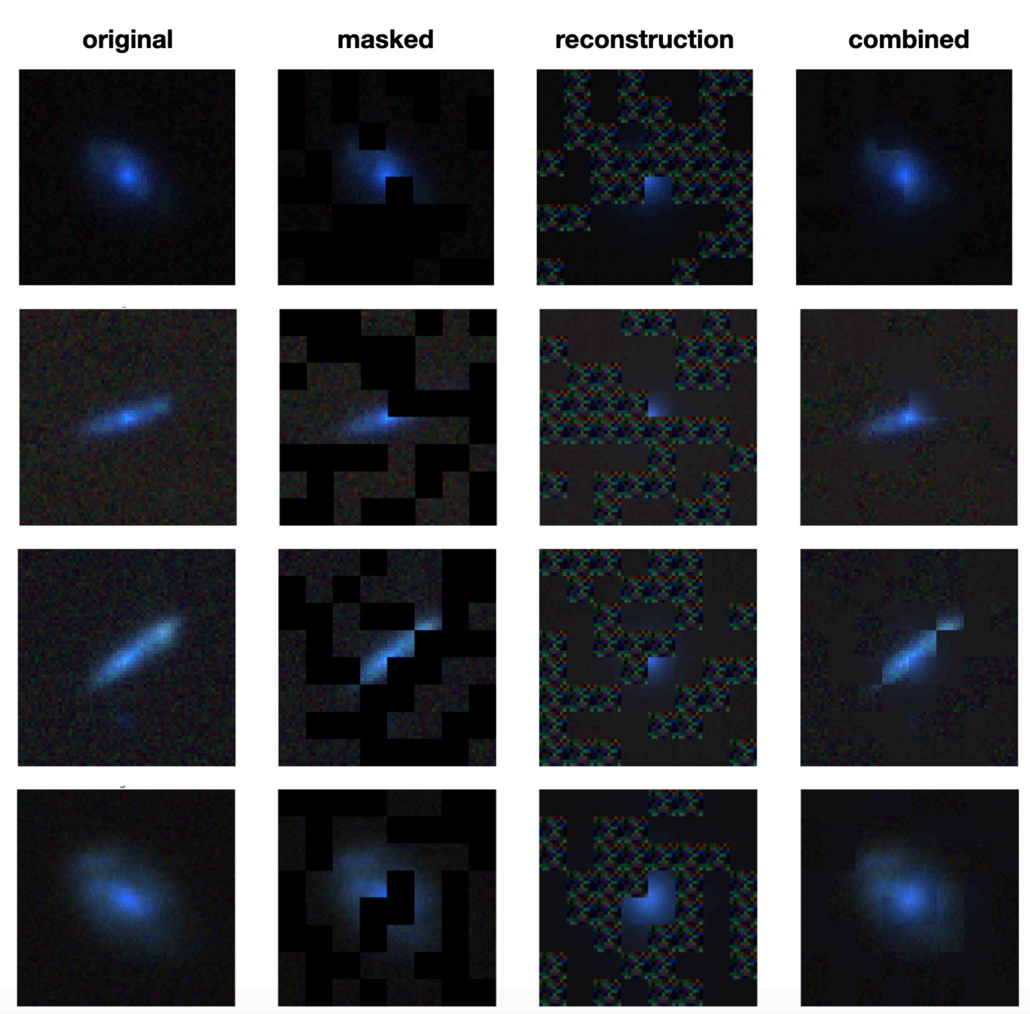

Figure 2: Example Pixel Reconstruction for 4 cutouts.
Key Takeaway: Model could be more expressive.

Can we train a machine learning **model** to understand **astronomy** data without **labels**?

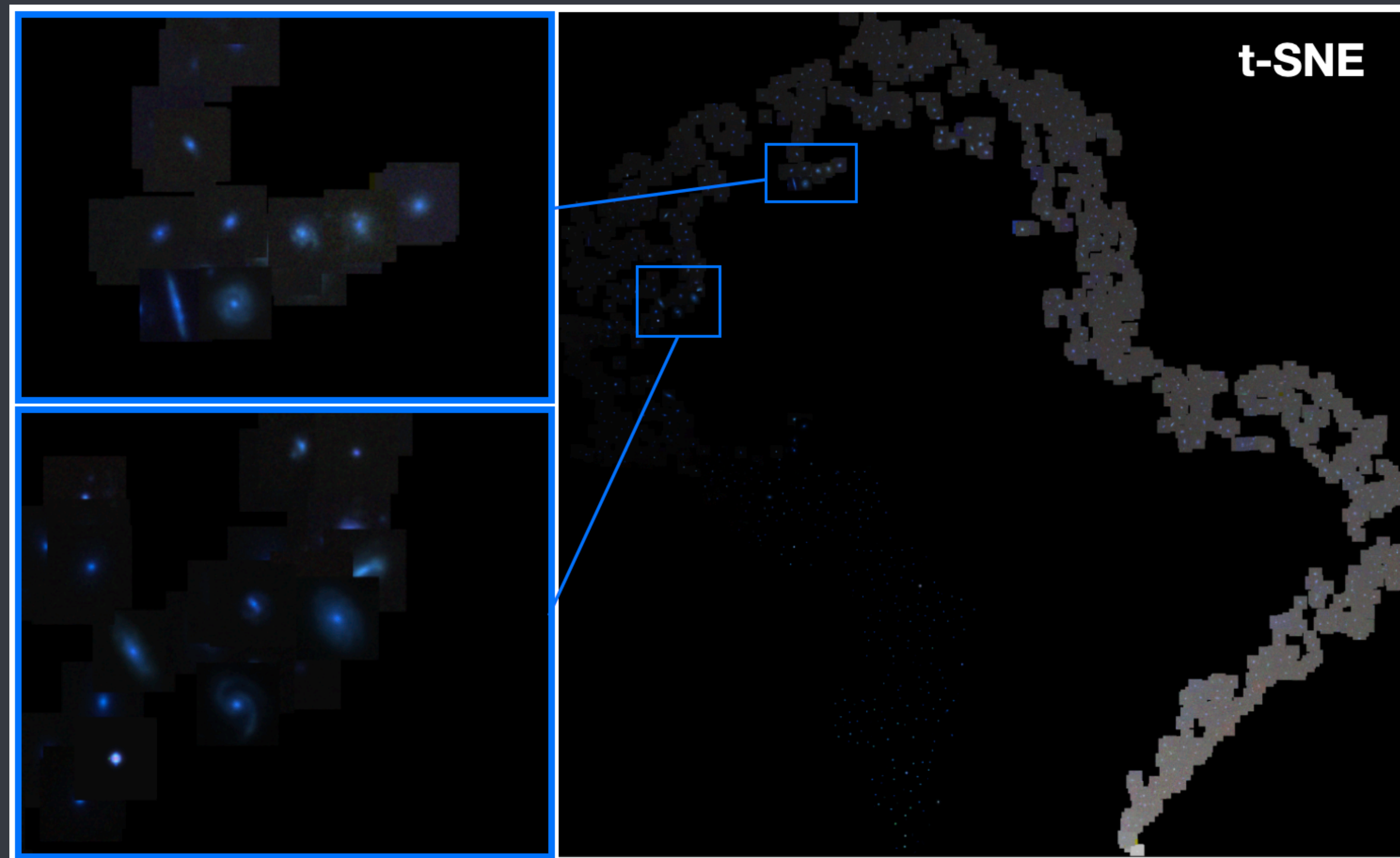

Figure 3: Visualization of Learned Representations via t-distributed Stochastic Neighbor Embedding (t-SNE).
Key Takeaway: There is semantically meaningful structure observed. Note these axes are meaningless.

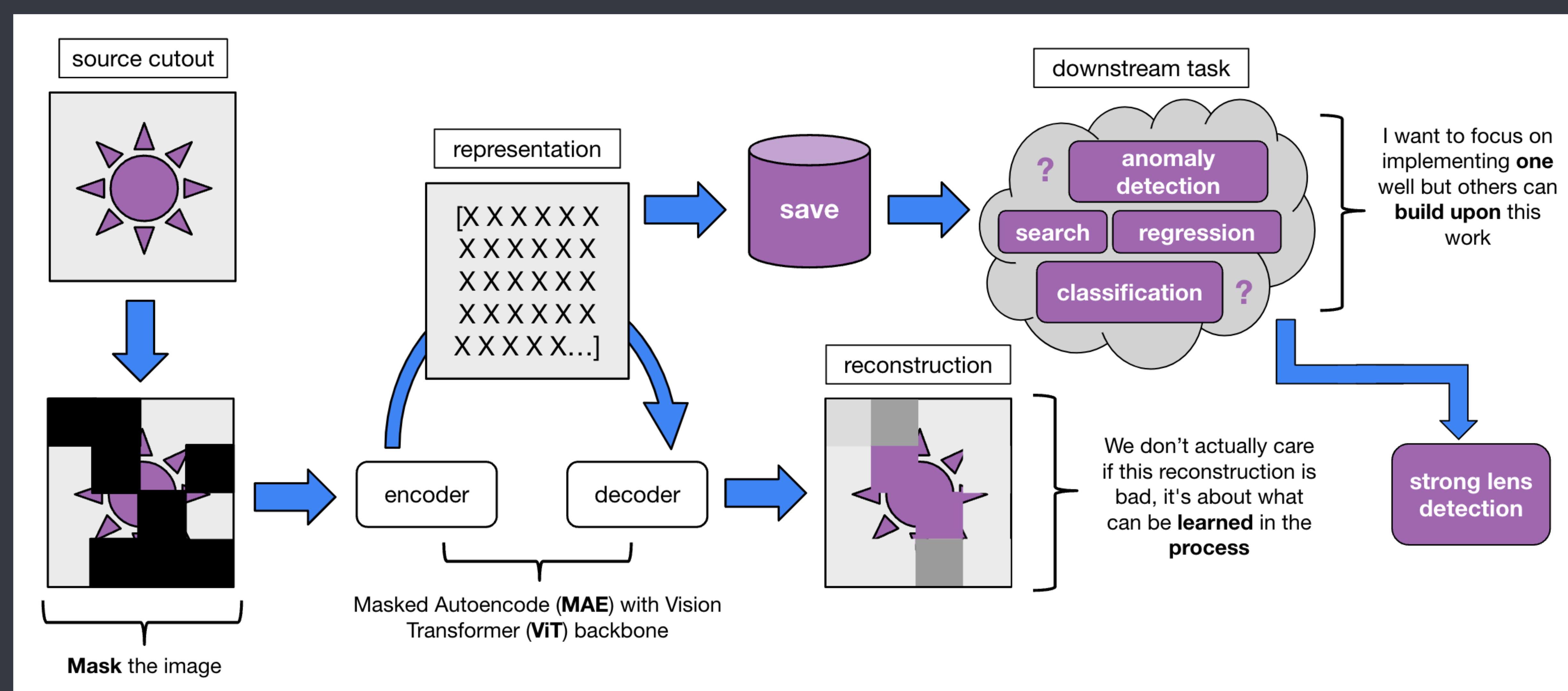

Figure 1: Overview of Methodology to use Masked Autoencoder for Pre-Training.
Key Takeaway: The focus will be to use the representations for strong lens detection.

Results

- See Figure 3 for a visualization of learned representations and Figure 4 for the use of these representations for 4 semantic search examples.

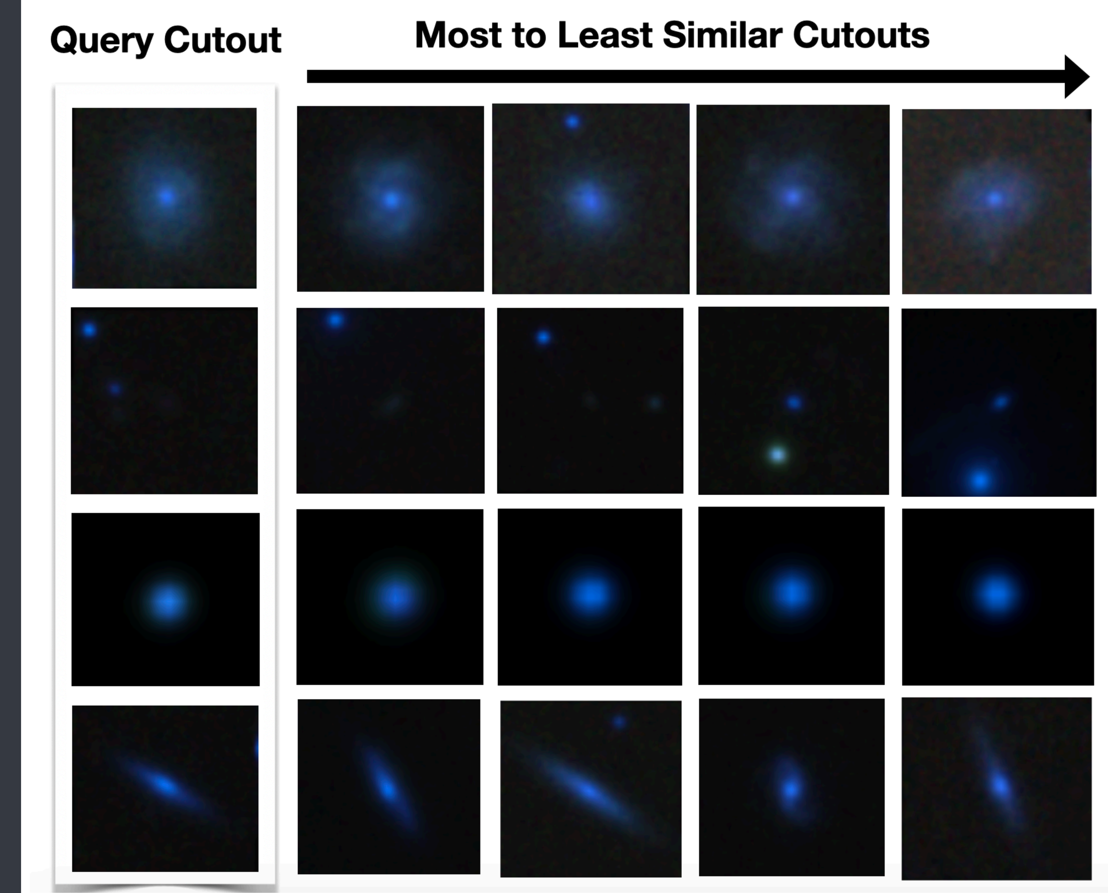

Figure 4: Example Similarity Search Application.
Key Takeaway: Model is able to return similar sources.

Future Directions

- **Scale:** use much more data for pre-training and downstream tasks.
- **Lenses:** use known strong lenses to find more.
- **Tune:** run experiments to optimize performance.

Summary

- SSL and astronomy can be a powerful duo.
- Initial results are promising but not perfect.
- If it works this model has the ability to make new discoveries in UNIONS and reduce the number of models that are trained from scratch.

Learn More

GitHub

Please scan this QR code to access the project: abstract, code, previous presentation slides, and mid-project report.

You are also welcome to contact: a4ferreira@uwaterloo.ca

Acknowledgements

All authors have been instrumental in the project and have taught me so much, for which I am very grateful. Additionally, my fourth-year research project co-supervisors Sebastien Fabbro and Mike Hudson have given me very valuable guidance on my presentations and reports.

CCUWiP 2024

NTCO

UNIVERSITY OF WATERLOO