Unsupervised Machine Learning For Identifying Signatures Of Past Mergers In Galaxy Light Profiles

Joyo Smit (he/they)

Saint Mary's University , Halifax , Nova Scotia

Supervisors: I. Damjanov (SMU), H. Domínguez Sánchez (CEFCA) , J. Vega-Ferrero (CEFCA, UVa)

Project Background

Galaxy Interactions

- Interactions between galaxies result in stars and gas being pulled into different shapes
- Tidal features, *visible long after the interaction*, include streams and shells (Fig. 1)

Detection Difficulties

- Classification by eye
  - Large volume of galaxy images to be classified
  - Tidal features often faint

Solution – Machine learning (ML)

- Detects faint features and can deal with lots of data
- Supervised ML: not enough labels
- → Try unsupervised / contrastive learning!

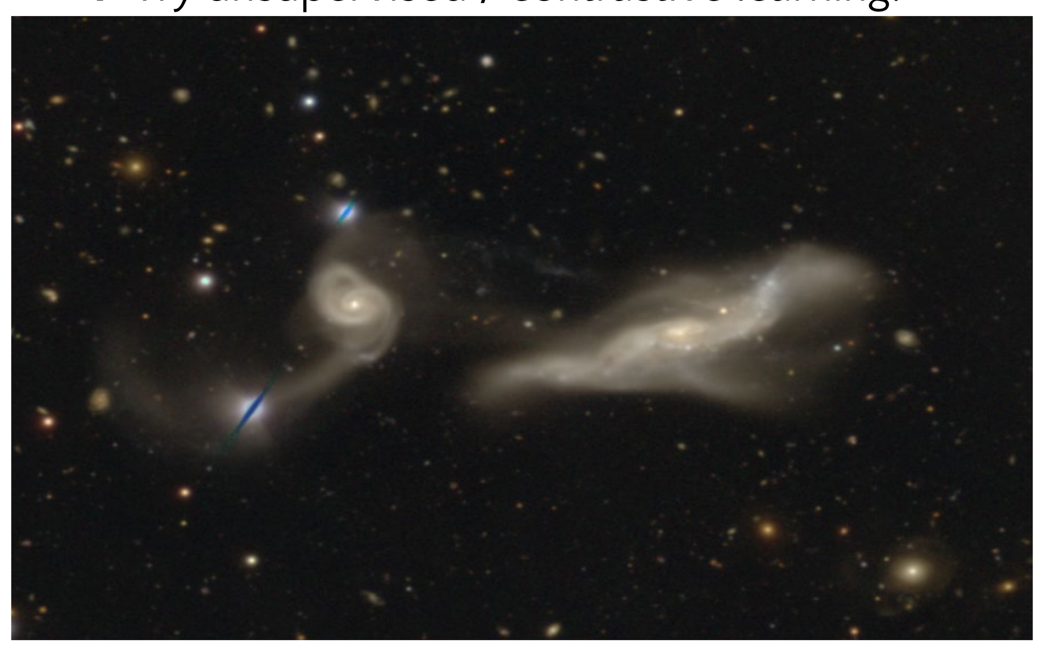


Fig. 1. Two galaxies interacting, producing tidal streams. Photo credit: HSC-SSP

Testing Methodology (1)

Images: Hyper Suprime-Cam Subaru Strategic Program PDR-3 512x512 pixels in g,r,i bands* *with labels for tidal features* (for method testing)

1. All images scaled ∝ galaxy radius (Fig. 2)

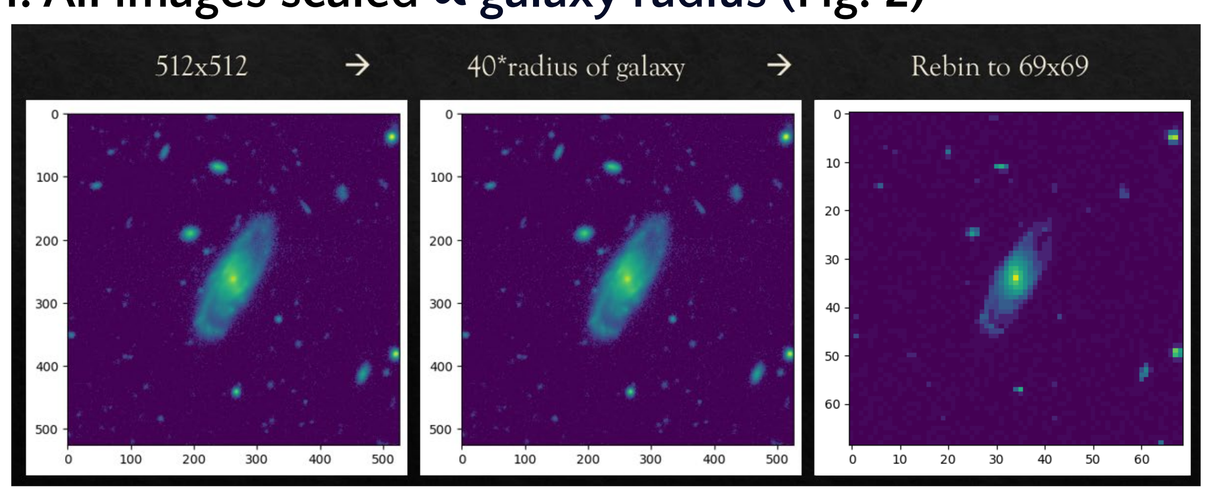


Fig. 2. Image scaling and rebinning. This a large galaxy and scaling does not change the image size

2. Augmentations of scaled images

- Center cropping and rotating while maintaining tidal features
- Result: image pair
  - one image scaled and augmented, one scaled and non-augmented

3. Convolution step (Fig. 3)

- Weights matrix scans across image pair

*https://hsc-release.mtk.nao.ac.jp/doc/

Testing Methodology (2)

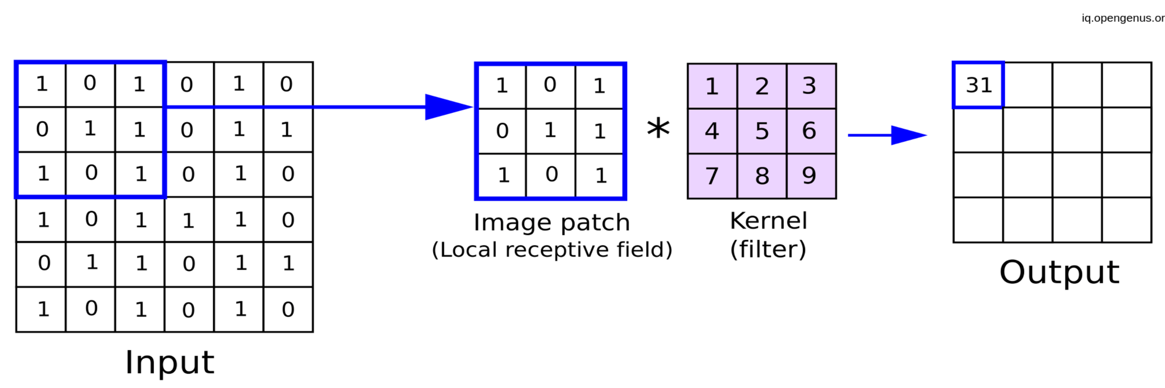


Fig. 3. The convolution of the input image with the weights matrix

4. Compute Similarity

- Loss function between augmented and non-augmented images

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

5. Adjust weights through iteration

- Goal: minimize loss function between images within pairs and maximize it between different pairs

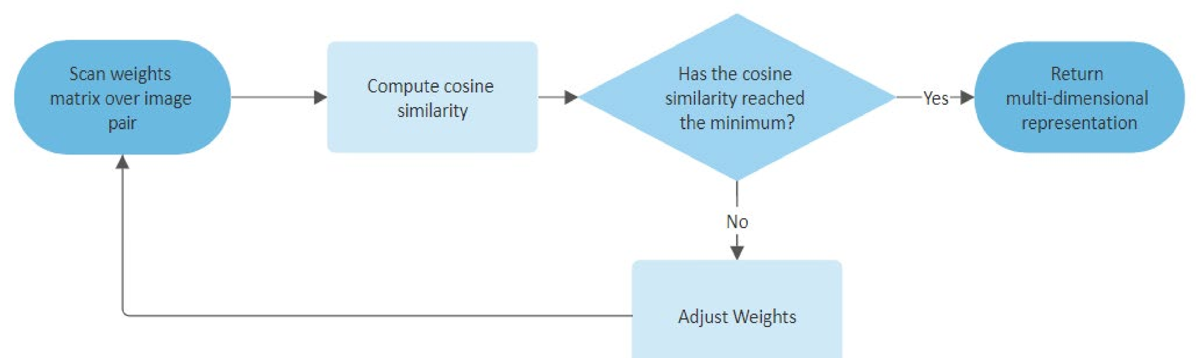


Fig. 4. Iterative loop to adjust weights and find global loss minimum

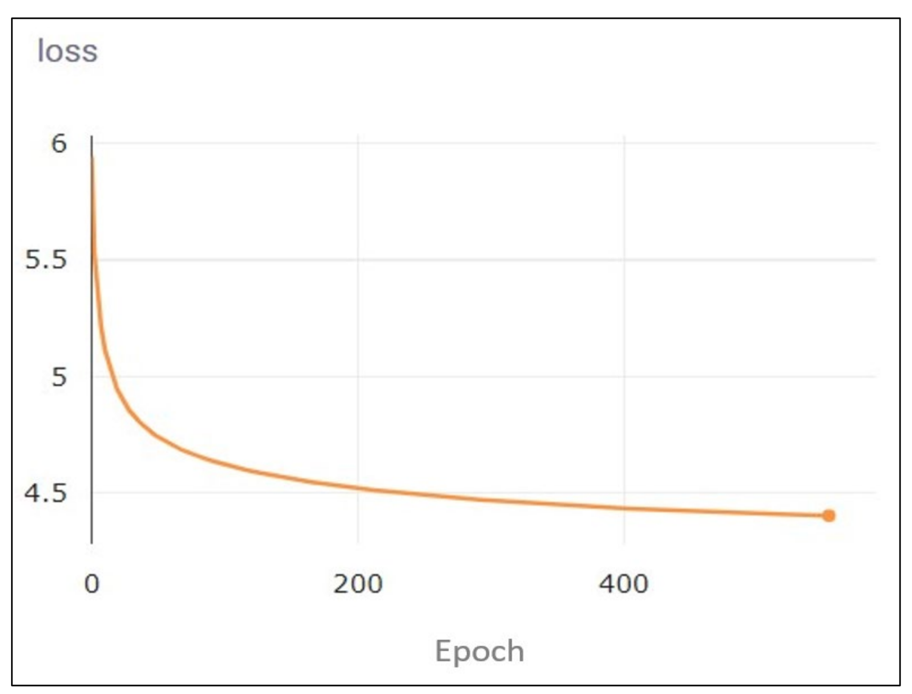


Fig. 5. Loss as a function of epoch (i.e. iteration)

Output

- Multi-dimensional representation of the dataset
- Similar images occupy the same region, far from dissimilar images.
- Scale down to two dimensions to make it readable.
- Plot various parameters as colour maps on the 2D representation to see what the machine has learned
  - Look for trends in colours when parameters plotted

Results

1st Method

- Algorithm inputs are regular 512x512 images

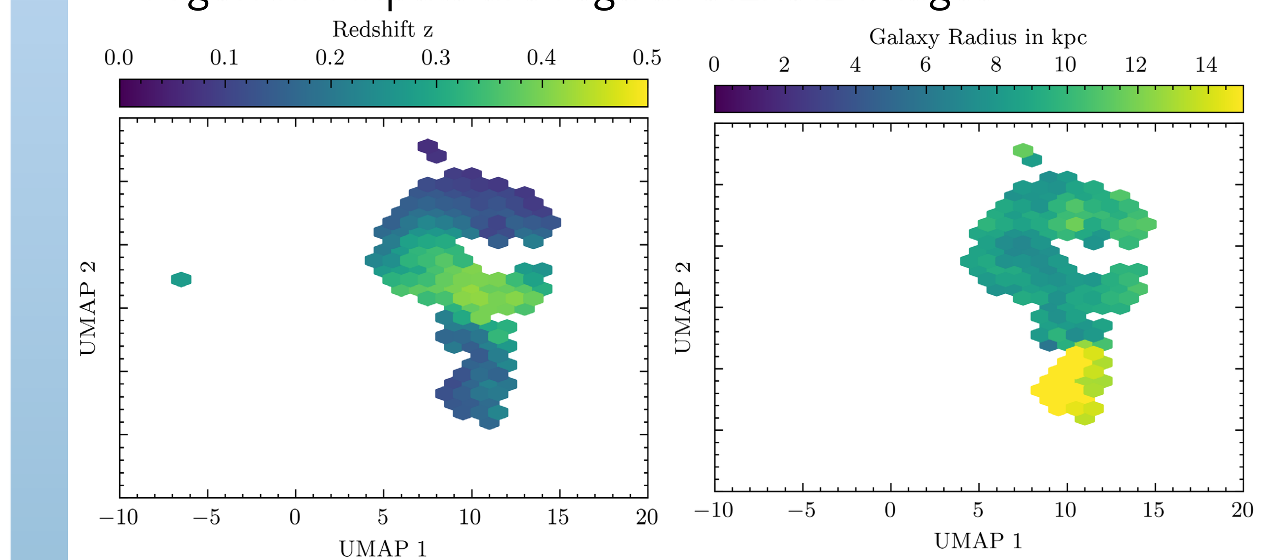


Fig. 6. *Left:* 2D representation of our sample with galaxy redshift as colour. *Right:* The same, but with galaxy size as colour

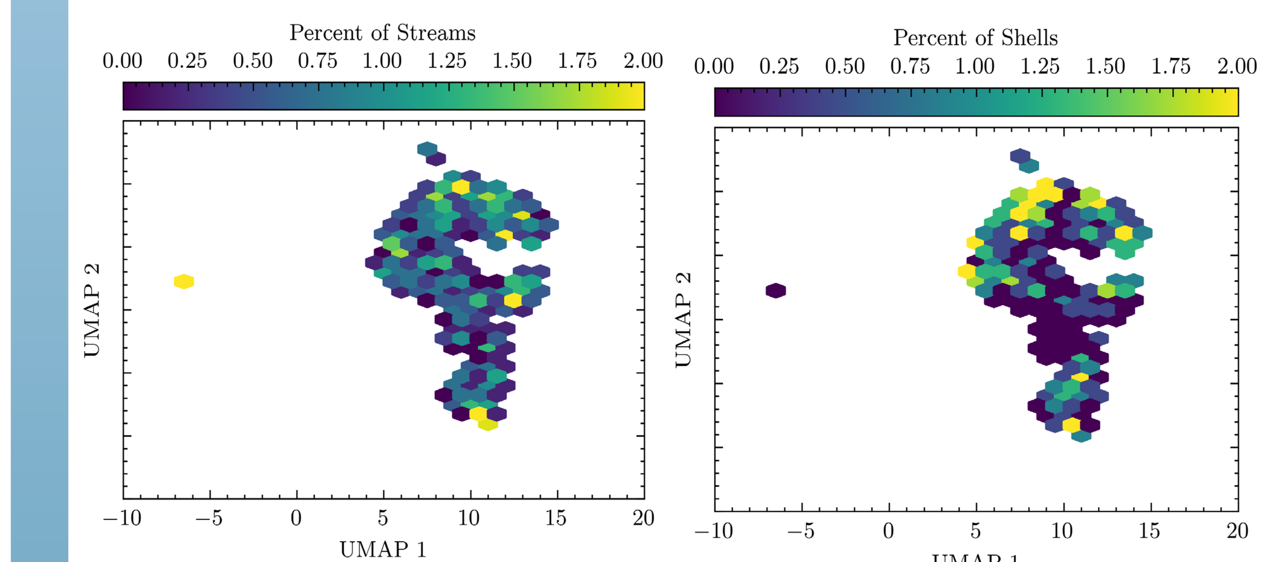


Fig. 7. *Left:* 2D representation of dataset with % of streams as colour. *Right:* The same, but with the % of shells as colour

- Fig. 6 shows clear trends in galaxy redshift and galaxy size
- Fig. 7 shows no trend in streams (left) and a weak trend in shells (right)

2nd Method

- Input images have their reflection subtracted (more focus on asymmetric data like tidal features)

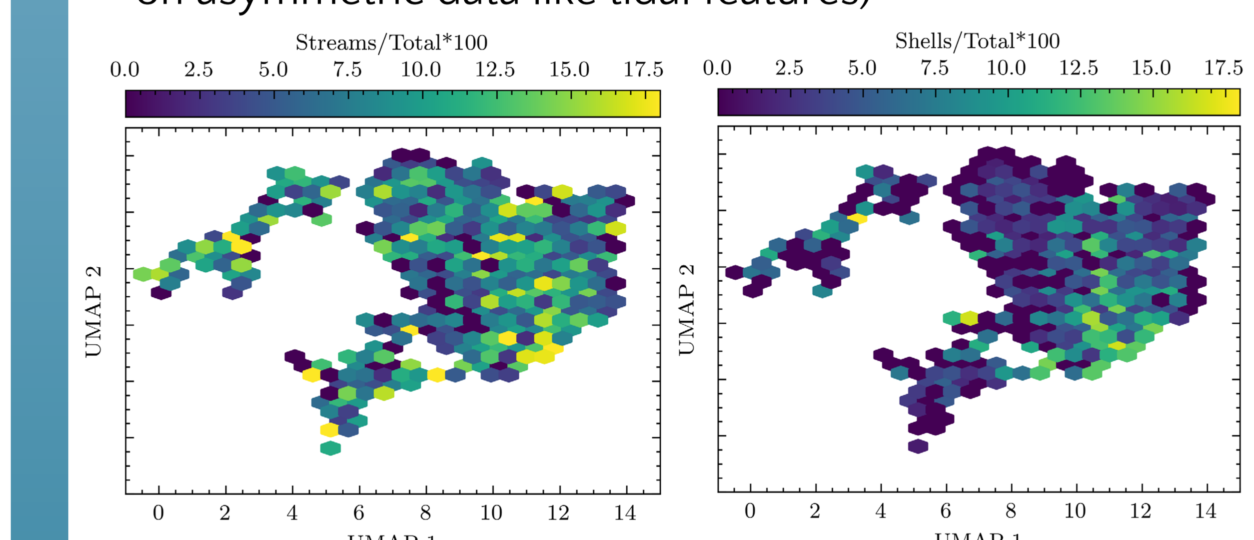


Fig. 8. *Left:* 2D representation for the second method with the % of streams as colour. *Right:* The same, but with the % of shells as colour

- A faint trend in streams is visible (left panel Fig. 8)
- The trend in shells has improved (right panel of Fig. 8)

Conclusions

- Contrastive Learning does separate galaxies based on their parameters (Figure 6).
- However, for now the ML algorithm does not separate sufficiently well galaxies based on their tidal feature, streams and shells (Figure 8).
- To continue improving the model, we need a larger set of galaxy images and to experiment with augmentations.

Future Work

In the next stage of this project, I will:

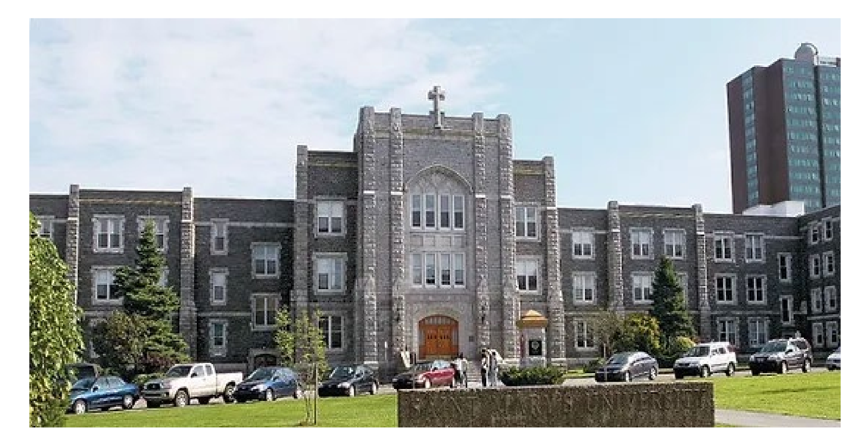
- increase the dataset size for better results (from 8K images to 20K images)
- move to the Digital Research Alliance resources, including utilizing GPU resources for shorter runtime
- find better augmentations to pick out the desired galaxy features

Internship

Sponsorship for this project was provided by the NTCO-CREATE program from May 2023 – August 2023. I was:

- working under Prof. Damjanov at SMU in Halifax
- coding in python, running programs in Google Colab, working with large datasets
- attending and presenting at weekly meetings with SMU Extragalactic Research Group

Overall, I learned a lot about the research environment in physics and that it suits my workstyle well.



St. Mary's University, credit: CUAC

Acknowledgements

Thank you to the NTCO program for supporting this project, as well as my supervisors and the SMU Extragalactic team for their assistance!