No Cap: Exploring Explainable AI in Galaxy Imaging with Capsule Networks and Semi-Supervised Learning

James Kostas Ray – UCL CDT in Data Intensive Science in Astronomy

1 Introduction

The new age of astronomy will be dominated by huge volumes of data from the new instrumentation. Naturally, the field has started to look towards data driven methods to process the incoming data.

- Number of papers in astronomy which involve **deep-learning** has risen exponentially over the last 15 years.
- Existing focus in astronomy is based on model **performance** rather than the **physics**.
- Many machine learning models rely on data from simulations or datasets from 10+ years ago.
- Is there a way to both **better utilise the data available** and make the models more **interpretable**?

Figure1: Left to Right: LSST, SKA, Euclid, DESI, JWST

2 Semi-Supervised Learning

Semi-Supervised Learning hasn't been explored much in astronomy and may be able to alleviate some of the data problems we face.

- Vast majority of data in astronomy is **unlabelled**!
- Labelling data is very expensive in astronomy, and **very few labelled datasets exist**.
- Semi-Supervised Learning allows a model to learn from **both** labelled and unlabelled data [3].
- Two common types are **Teacher-Student** models and **Pseudo Labelling**.

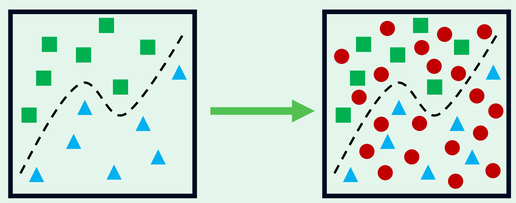

Figure 2: An example of semi-supervised learning by pseudo labelling a dataset

3 Capsule Network Architecture

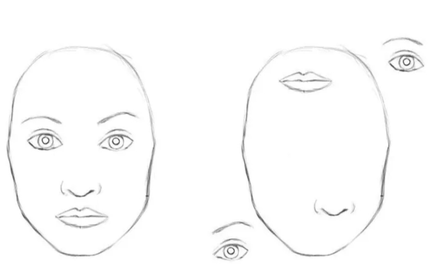

Figure 3: A 'standard' face on the left with a 'broken face' on the right. While the image on the right is very clearly not a face, but it does share all the same features as the image on the left. To a CNN, both these images would be identified as a face, due to the recognition of those facial features.

The modern Capsule Network was first described in Dynamic Routing Between Capsules [1]. Capsule networks have some key advantages over CNNs:

- CNNs **discard** information within an image. **Pooling** layers and multiple convolutions discard information within an image over many iterations.
- CNNs don't not capture the **spatial relationship between features** (see figure 1).
- CNNs lack the ability to **identify pose** (translational invariance).
- Capsules require **less data** to train than CNNs.

Capsules are still early on in their development phase; thus, they perform worse then existing CNNs but show excellent potential.

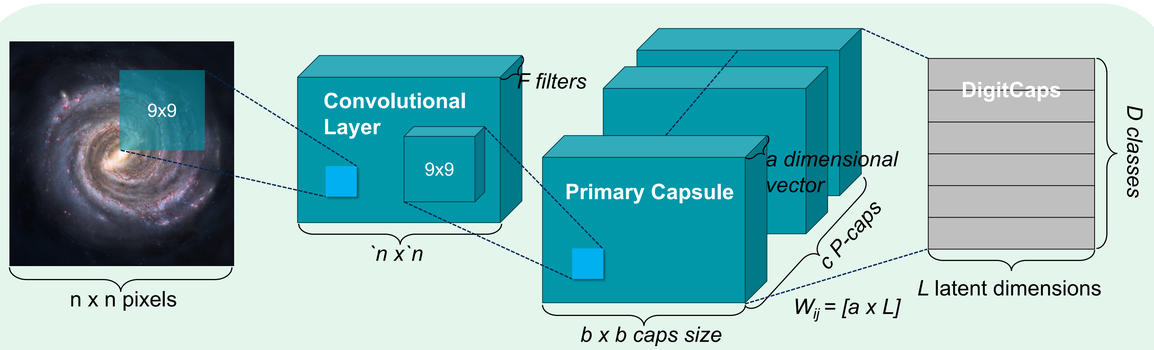

Figure 4: An example of a basic architecture showcased in Dynamic Routing Between Capsules [1]. The lengths of each activation vector in the DigitCaps is used to detect features present is each class label. The weighted matrix W is used to commune between the Primary Capsules and the DigitCaps.

The Primary Capsule has ['c' x 'b' x 'b'] outputs, with each output having 'a' dimensions. The final DigitCaps layer has a 'L' dimensional capsule per class. These DigitCaps can be treated as a latent space of each class, and can be probed as such, allowing for **explainability** methods to be used.

Keys:
- n: Pixels
- `n: Conv 1 dimensions
- a: Capsule vector dimensions
- b: Capsule dimensions
- c: Number of capsules
- U: Dense Units

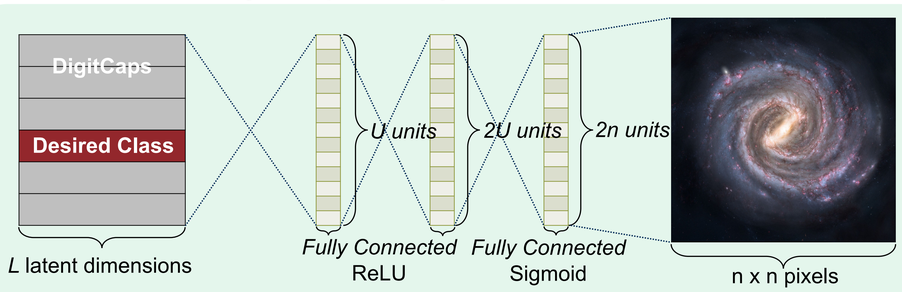

Figure 5: The decoder structure, which acts as a means of regularization by reconstructing the image based on the true label of the image.

4 Explainability

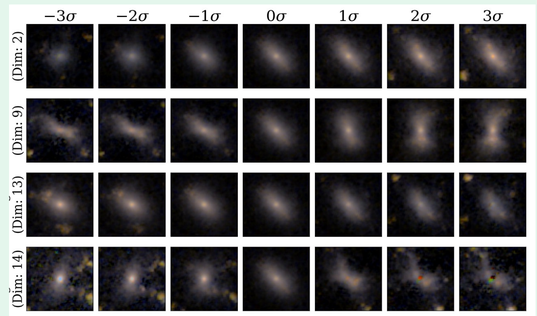

Figure 6: Figure taken from Dey B. et al [2]. The perturbations in the **DigitCaps reconstruction** highlights the network has captured physical features like size, surface brightness, rotation and the size of the nucleus.

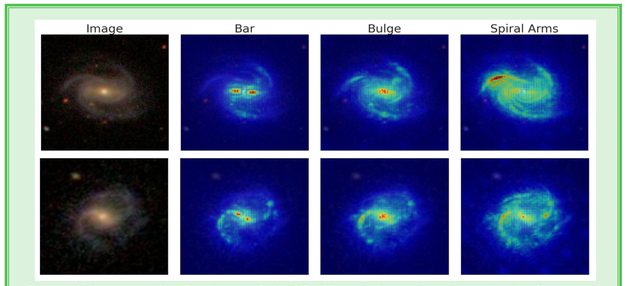

Figure 8: An example of using **explainable AI** techniques to extract physical characteristics from galaxy images. Bhambra, P. et al [4] (Explaining deep learning of galaxy morphology with **saliency mapping**.) showed how spiral galaxy bar lengths could be measured.

5 Proposal

- Utilise **semi-supervised** methods to combat the issue of **sparse labelled data**.
- Expand from existing datasets with these techniques (most labelled data comes from SDSS & DECaLS, while datasets from ViKINGS and KiDS remain relatively untouched in this area of astronomy).
- Attempt to use this leveraged data to **break the redshift limit (z>1)** in photometric redshift estimates.
- Explore **capsule networks** as a more advanced **galaxy classifier**.

References: [1]: Sabour, S., Frosst, N. and Hinton, G.E., 2017. Dynamic routing between capsules.

[2]: Dey, B., Andrews, B.H., Newman, J.A., Mao, Y.Y., Rau, M.M. and Zhou, R., 2022. Photometric redshifts from SDSS images with an interpretable deep capsule network.

[3]: Yalniz, I.Z., Jégou, H., Chen, K., Paluri, M. and Mahajan, D., 2019. Billion-scale semi-supervised learning for image classification.

[4]: Bhambra, P., Joachimi, B. and Lahav, O., 2022. Explaining deep learning of galaxy morphology with saliency mapping.

UCL CDT DIS