Semi-supervised Learning of Galaxy Morphology using Equivariant Transformer Variational Autoencoders

Mizu Nishikawa-Toomey, Lewis Smith, Yarin Gal

Department of Computer Science, University of Oxford

The Galaxy Zoo dataset

- 250,000 of the brightest galaxies from the Sloan Digital Sky Survey were put on a website alongside a tree of questions
- Users logged on to answer questions based on features of the galaxies such as “Smooth, featured or artefact” or “Bar or no bar”
- The data set consists of the total number of responses for each answer to each question, and the corresponding galaxy image.

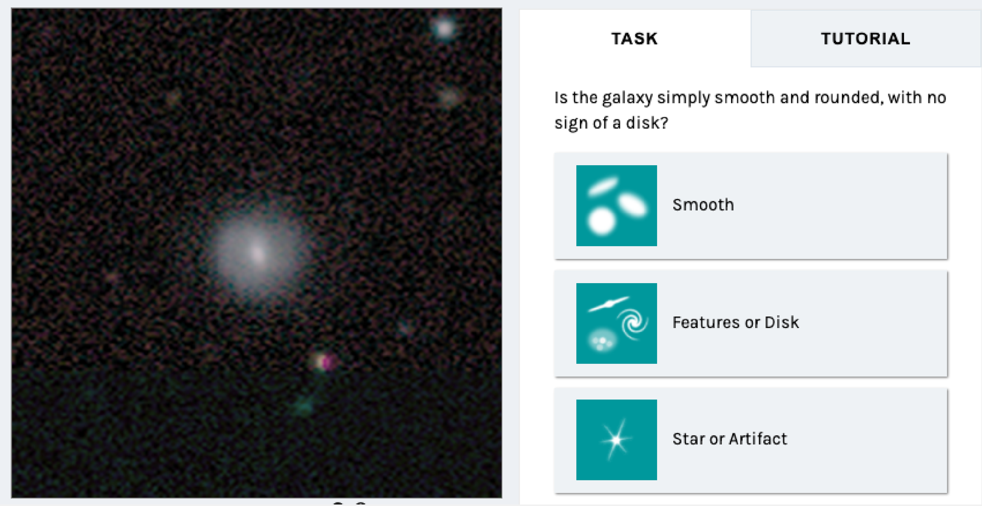

Figure 1:User interface on galaxy zoo

Why semi-supervised learning

- When new questions are introduced to the data set, we have zero responses to that question.
- Galaxy images continues to grow at a rate that is not possible to be classified by humans. It would take 5 years to collate 40 volunteer responses for each image in the Galaxy Zoo data set at the current response rate.

Variational Autoencoders

- VAEs learn the distribution of latent parameters of the image $p(z|x)$, and the generative model $p(x|z)$.
- Classification can be done from the latent representation which eliminates noise from the data and makes training more efficient.

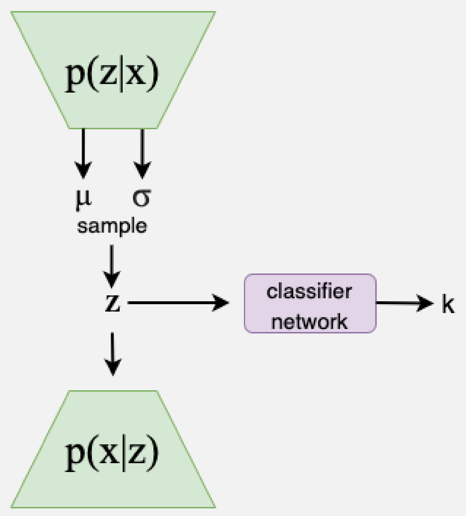

Figure 2:A VAE with a classifier from the latent space

The VAE (green) is trained using the ELBO objective using unlabelled data.

$$\mathcal{L}(x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(z,x)] - \mathbb{E}_{q_\phi(z|x)}[\log q_\phi(z|x)] \quad (1)$$

Eliminating redundancy in the data

There is redundancy in the Galaxy Zoo data set, as many galaxies are different transformations of a canonical galaxy image for those particular features.

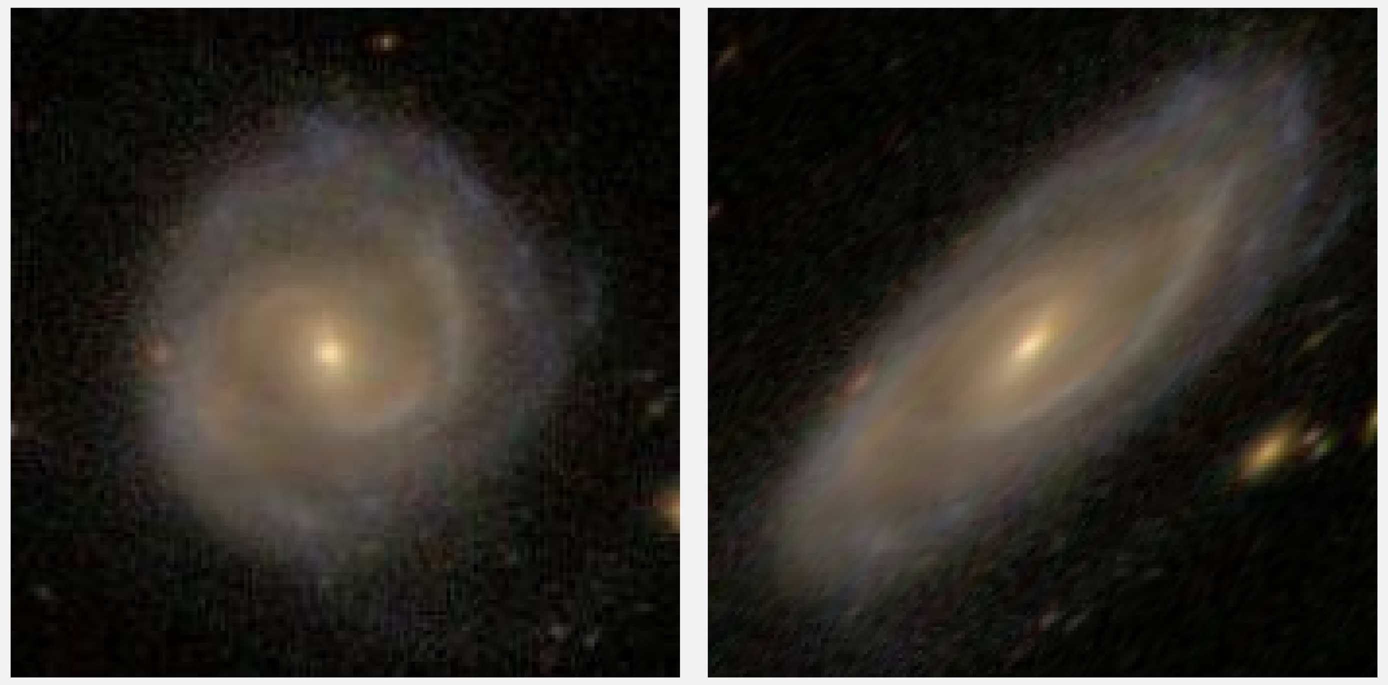

Figure 3:Same galaxy image, but viewed at different planes

Equivariant Transformer networks

- Assuming that each image ϕ is a transformation of the canonical image ϕ^* of that type of galaxy. The transformation T is governed by its pose parameters θ.
$$\phi = (T_\theta \phi^*) \quad (2)$$
We want to predict these pose parameters using a function f:
$$f(\phi) = \theta \quad (3)$$
- We want the function f to have a property that is called self consistency:
$$f(T_{\theta'}\phi) = f(\phi) + \theta' \quad (4)$$
How do we do this? Each transformation has an associated pose parameter. For each transformation, we have an associated mapping ρ that satisfies:
$$\rho(T_\theta x) = \rho(x) + \sum_{i=1}^{k} \theta_i e_k \quad (5)$$
which transforms from the cartesian coordinates to what we called the canonical coordinates for that image.
- We then apply a pose predictive function on the new coordinate system which is self consistent with respect to translation (such as a CNN).
- For example, the canonical coordinate system for the rotation transformation is the polar coordinate system. A rotation of angle θ in cartesian coordinates is a translation by θ in polar coordinates.

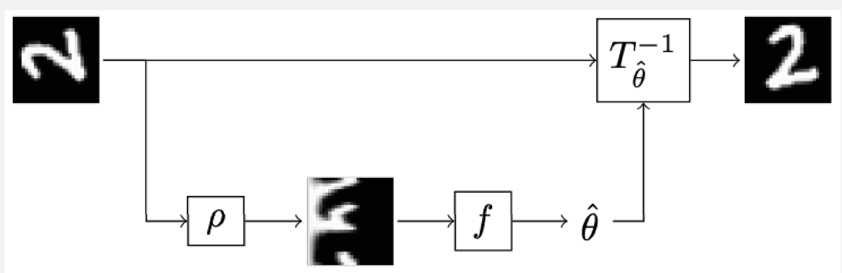

Figure 4:ET layer transforms and image and predicts a pose

Equivariant Transformer Variational Autoencoder

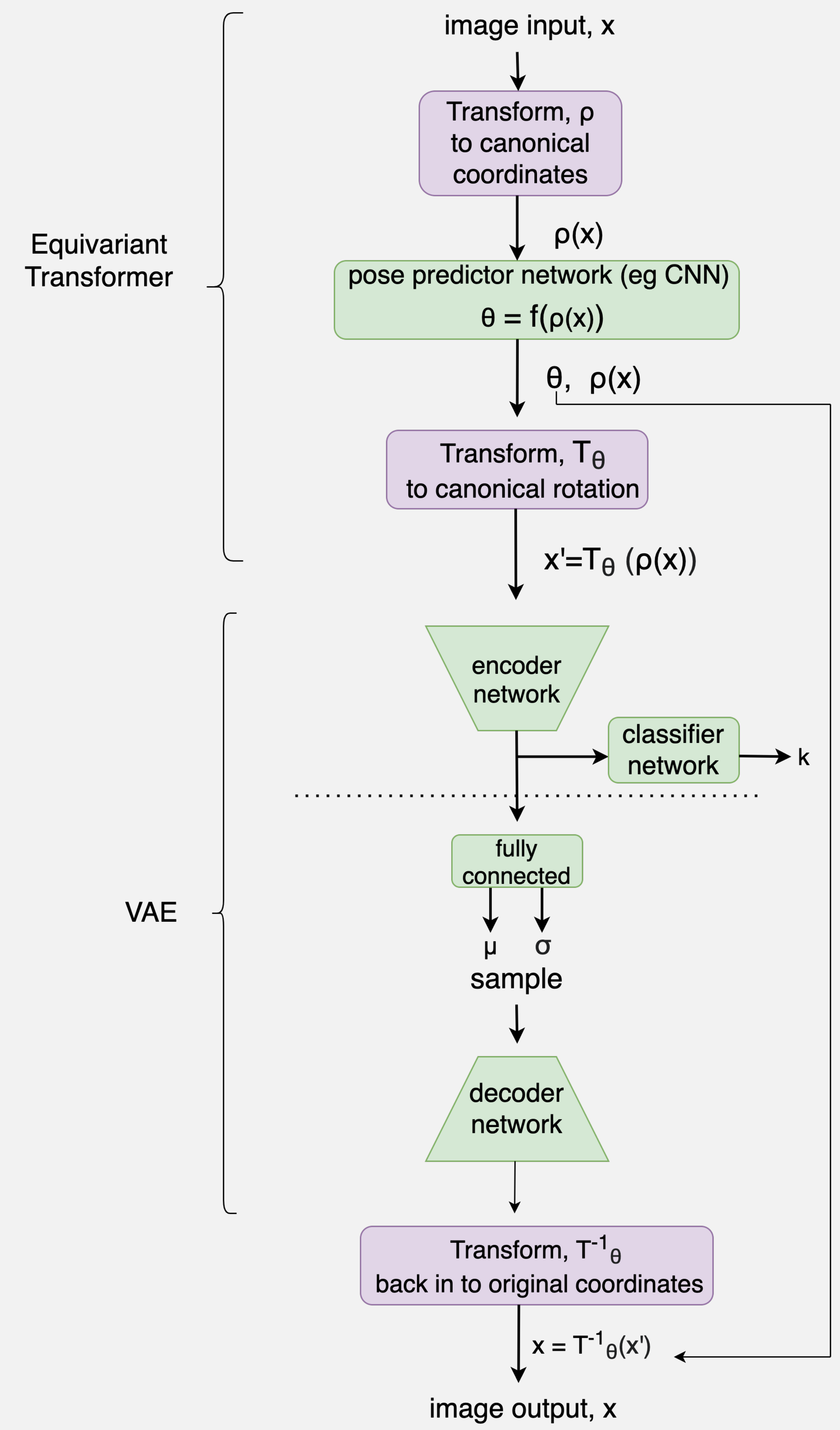

Figure 5:VAE using ET layers used to run experiments

Results

- Fully-supervised training consisted of the part above the dotted line in Figure 5. The objective function of the classifier was updated wrt weights of the classifier and the encoder.
- Semi-supervised training consistsed of alternately updating the weights of the VAE wrt the ELBO, and minimising the objective function of the classifier wrt the classifier and the encoder.
- A third experiment consisted of a two-step procedure of pre-training the VAE with unlabelled data then fine tuning the classifier weights using labelled data.

Number of labelled images	100	300	800	1200
Fully supervised	0.56	0.31	0.25	0.24
Semi-supervised, alternating steps of VAE and classifier	0.35	0.24	0.20	0.21
Semi-supervised 2-step training of VAE and classifier	0.37	0.28	0.25	0.25

Figure 6:RMSE for semi-supervised and fully-supervised training