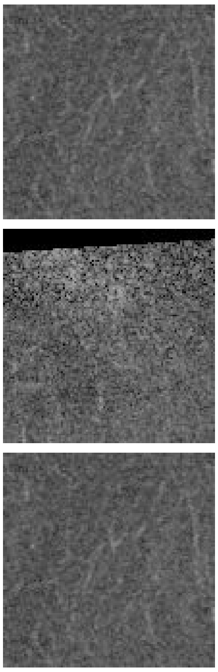
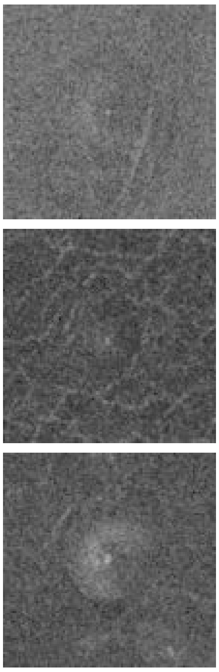
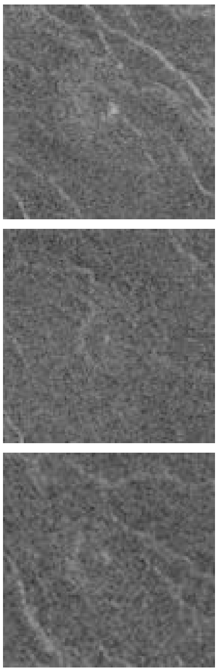
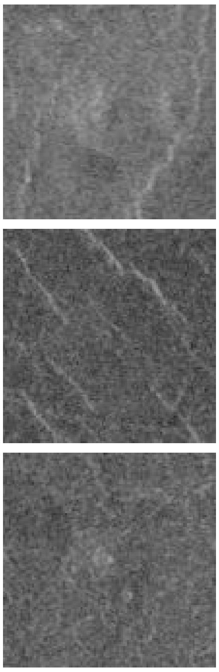
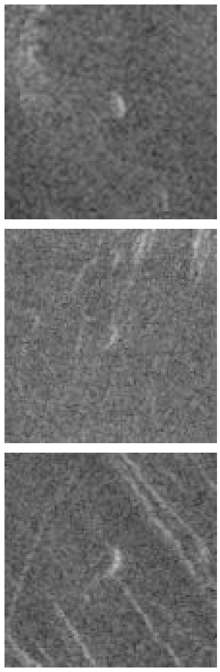
Volcanet

A CNN for classifying radar images of volcanoes.

The dataset

Shows a sample of the images in each class.

0	1	2	3	4
				

Class-0 can be discriminated from the rest by the lack of a small light circle/ring in the centre of the image indicating the “pit” of the volcano.
Classes 1-4 can be discriminated by the size/shape/presence of a “mound” surrounding the pit.

Shows the training support for each class (total 7000):

0	1-4	1	2	3	4
6000	1000	105	187	346	362
85.7%	14.3%	1.50%	2.67%	4.94%	5.17%

And the same for the test set (total 2734):

0	1-4	1	2	3	4
2300	434	35	88	148	163
84.1%	15.9%	1.28%	3.22%	5.41%	5.96%

Clearly the dataset is imbalanced; especially with many more negative/background (Class-0) samples than the rest.

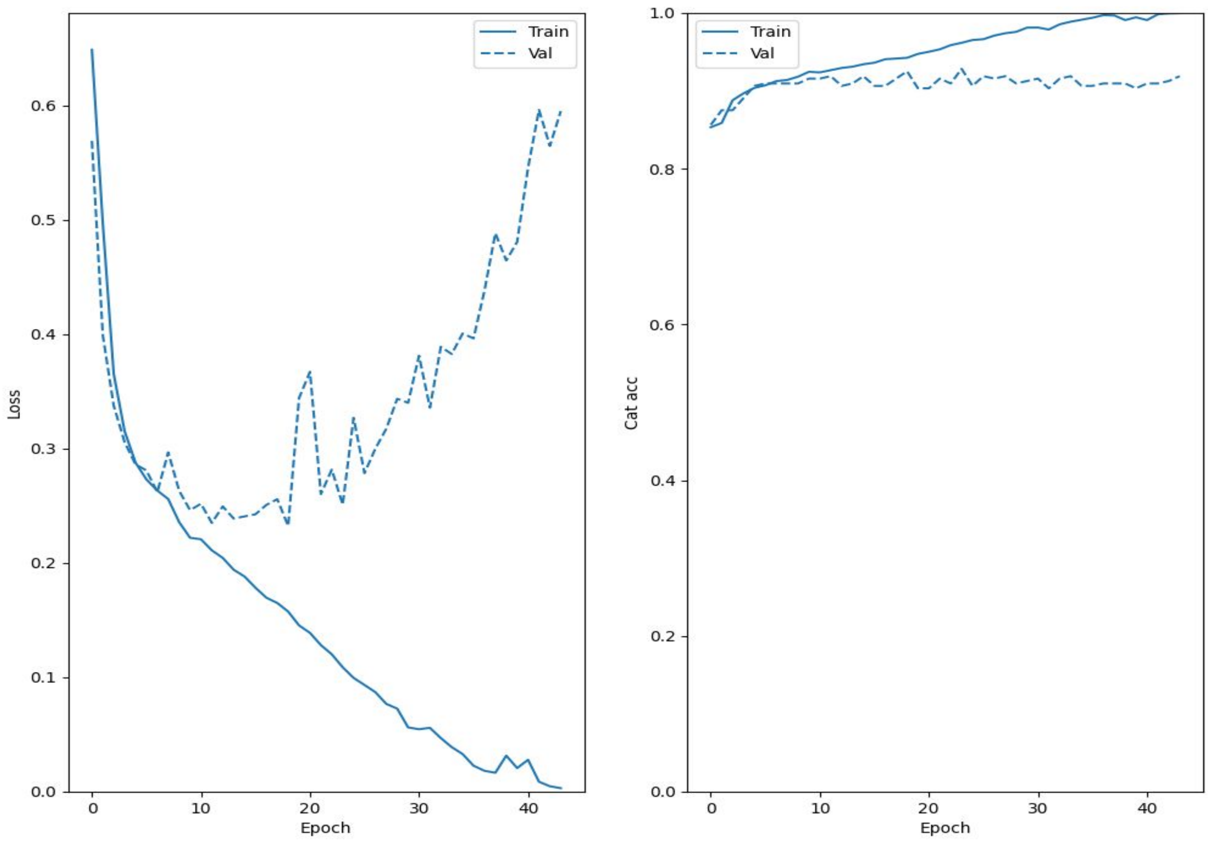
Model

The model consists of two convolutional blocks, followed by two fully connected layers and a fully connected output layer. Each convolutional block has a convolutional layer followed by a max pooling layer. The convolutional layers have filter size (3,3) and pooling layers have kernel size (2, 2). The two convolutional layers have 16 and 32 filters respectively and the fully connected layers 64, 32 and 5 units respectively. All the layers are relu activated, except the output layer which is softmax.

In front of the convolutional layer a preprocessing layer to crop the image to a central square was added in the final model.

Evaluation

Training history



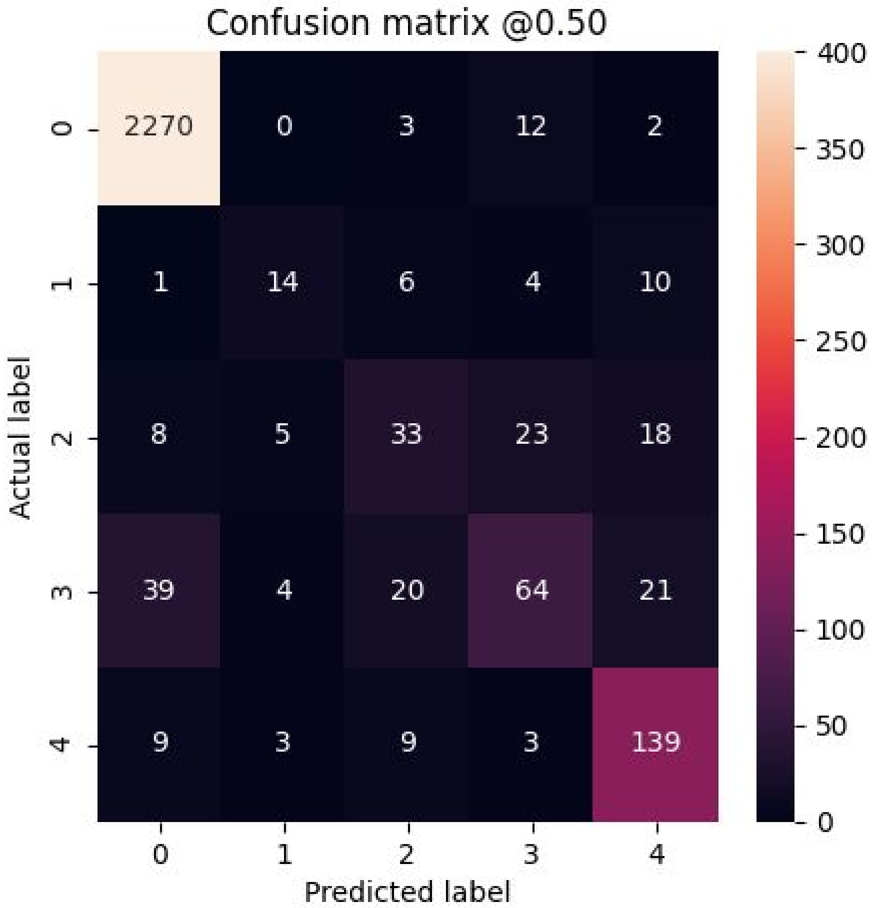
After around 7 epochs, the model goes on to overfit to the train data. In the first couple epochs the model learns the bias, seen as the steep drop in loss.

Accuracy

Test: **92.65%**
Train: **97.74%**

Confusion Matrix

On the test set:



Summary

Class	0	**1-4**	1	2	3	4
Precision	97.6%	**95.7%**	53.9%	46.5%	60.4%	73.2%
Recall	99.3%	**86.8%**	40.0%	37.5%	43.2%	85.3%

Despite being quite a simple approach, the best model performed reasonably well at volcano detection (which was the main target) with a recall of 86.8% and precision of 95.7%.

However, in the secondary target of volcano type classification, there was a lot of confusion, particularly between Class-1, 2 and to a lesser extent 3, (more data might be helpful here, since Class-1 and 2 were only supported by 1.28% and 3.22% of the dataset).

My next step would be to look at the incorrect predictions and try to ascertain if there is some label noise. For example, I would investigate why there was often confusion between Class-0 and Class-3 for around 50-60 samples in the test set.

Otherwise more data in the minority classes would be helpful. Assuming the labels are correct and no more data is available, I would focus on data augmentation and regularisation. Failing that, it might be better to use a different approach, like separate detection and classification stages.

Notes

Baseline

Overfitting
Early stopped at epoch 31 probably restored to 12
0.8956 test cat acc
0.9484 train cat acc

Oversampling

Overfitting a lot
Peak < 10 epochs
0.8846 test
0.9958 train

Central cropping to central 0.4

Not really overfitting - central cropping seems to regularise the baseline model.
Test acc peaks before 10 epochs
0.9180 test
0.9286 train

Oversampling with central cropping

Overfitting (due to oversampling) but still comparable acc on test to central cropping (at peak).
0.9029 test
0.9983 train
Ran for **81 epochs** before early stopping.

Clearly it's helpful to use central cropping - all experiments will from now on.

Increased validation set size from 192 to 320. (Mistakenly too small)

Augmentation

Flip left-right

With just central cropping (no oversampling): test **0.9070**; train: 0.9609 - causes cc to overfit.

Looking at training data again, it seems the images are maybe not flip/rotation (or translation) invariant.

Gamma compression/expansion

No improvements on test set over central_cropping+oversampling model.