Classifying plankton images with deep convolutional networks

Matt Graham[1,2], Gavin Gray[1,2], Scott Lowe[1,2], Finlay Maguire[3,4], Alina Selega[1,2] and Dragos Stanciu[1,5]

THE UNIVERSITY of EDINBURGH informatics

1. Neuroinformatics DTC, Sch. of Informatics, U. of Edinburgh, Edinburgh, UK. 2. ANC, Sch. of Informatics, U. of Edinburgh, Edinburgh, UK.
3. University College London, London, UK. 4. National History Museum, London, UK. 5. IPAB, Sch. of Informatics, U. of Edinburgh, Edinburgh, UK.

NATURAL HISTORY MUSEUM

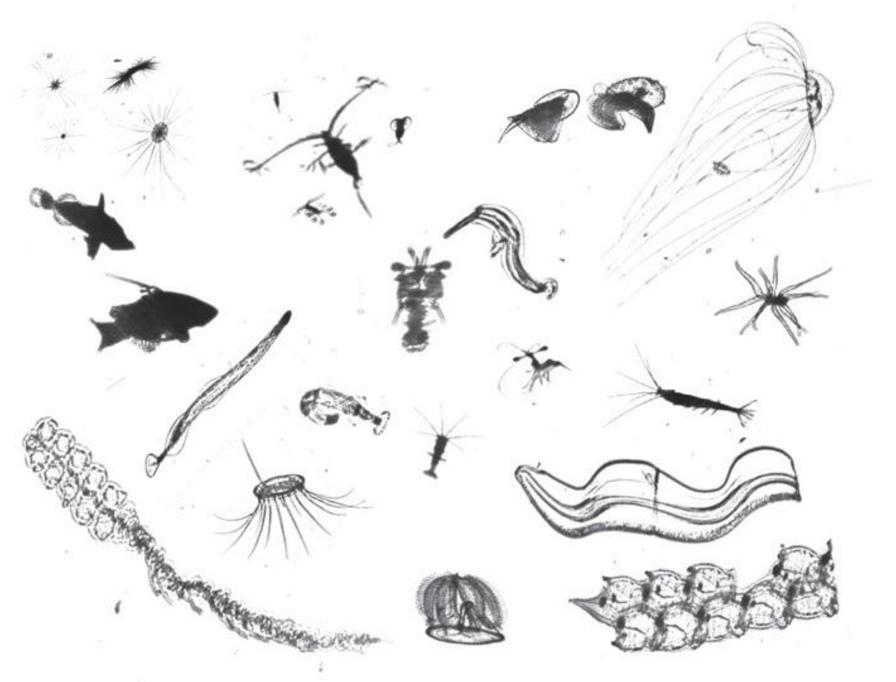
UCL

> INPUT

Plankton are an important part of the biosphere, comprising half of global carbon fixation. The population composition of plankton is a key indicator of ecosystem function in marine environments.

Marine biologists at the Hatfield Marine Science Center collect many thousands images of these microscopic organisms every day, but classifying these captured organisms remains time consuming. To automate this process, the researchers sought a solution through a machine learning competition, hosted on Kaggle.

Raw data: plankton images



The dataset contained 30,336 labelled and approximately 20,000 unlabelled images of plankton, classified into 121 different groups.

Problematically, the dataset featured a large number of classes, an imbalanced distribution of class sizes (9 to 2000 examples), and a wide range of image sizes (from <40 to >400 pixels in length).

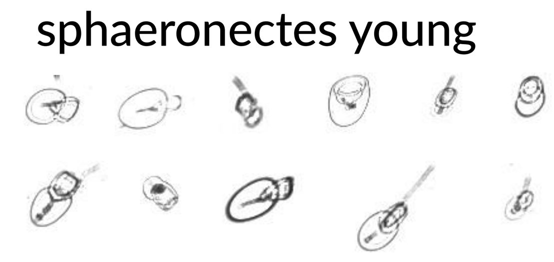
Test yourself!

Can you classify the plankton?

Given the two sets of example images as your training data, try to identify which group each organism in the test images belongs to (answers below).

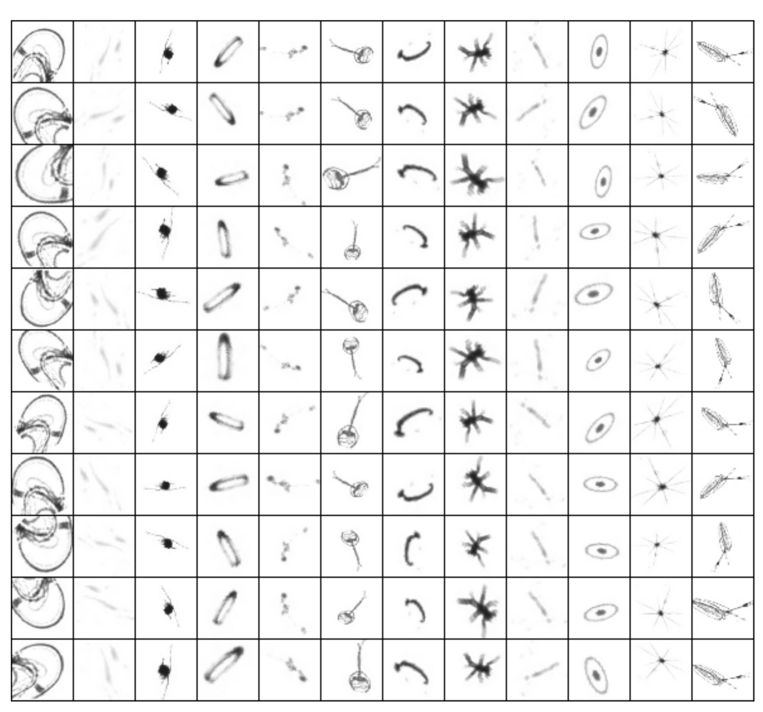
protist fuzzy olive

siphonophore calycophoran sphaeronectes young



combined: classify these examples

Augmentation



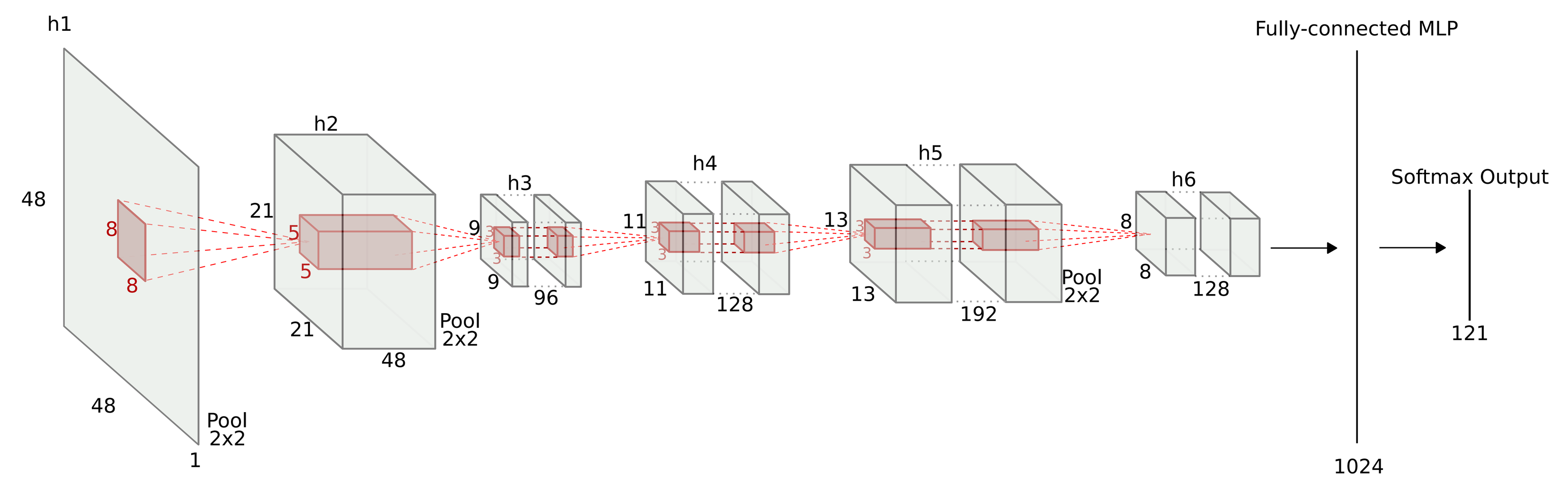
To help reduce overfitting to the training data, the original training dataset was augmented by applying random affine transformations to the images.

The augmentations were computed on the fly to minimise memory overhead, with the next augmented batch being processed on CPU cores while the current batch parameter updates were computed on a GPU.

> MODEL

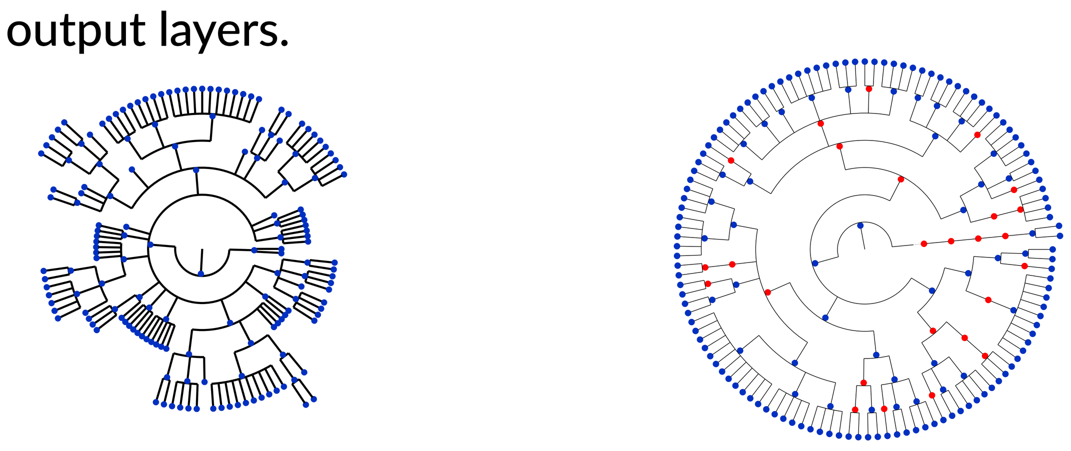
Convolutional network

For image classification tasks, deep convolutional networks are the current state of the art and so were an obvious choice of model. The initial structure used was based on the successful AlexNet design. After experimenting with a variety of structures, including siamese networks and significantly deeper architectures, the single best performing network differed from the original AlexNet based design by only one additional convolutional layer.



Hierarchical modelling

The classes had an intrinsic hierarchal structure which was provided by the competition organisers as a taxonomical tree (below left). This was used to provide additional information in the back-propogated error signal after converted the hierarchy into a series of six 1-of-k vectors (right) which could be supplied to six parallel softmax output layers.



This improved the initial learning rate, but seemed to have no impact on the final performances of the trained networks.

Additional visual features

In addition to the features learnt with convolutional networks, some attempt was made to use traditional computer vision techniques to extract further global and local visual features that might help in classification.

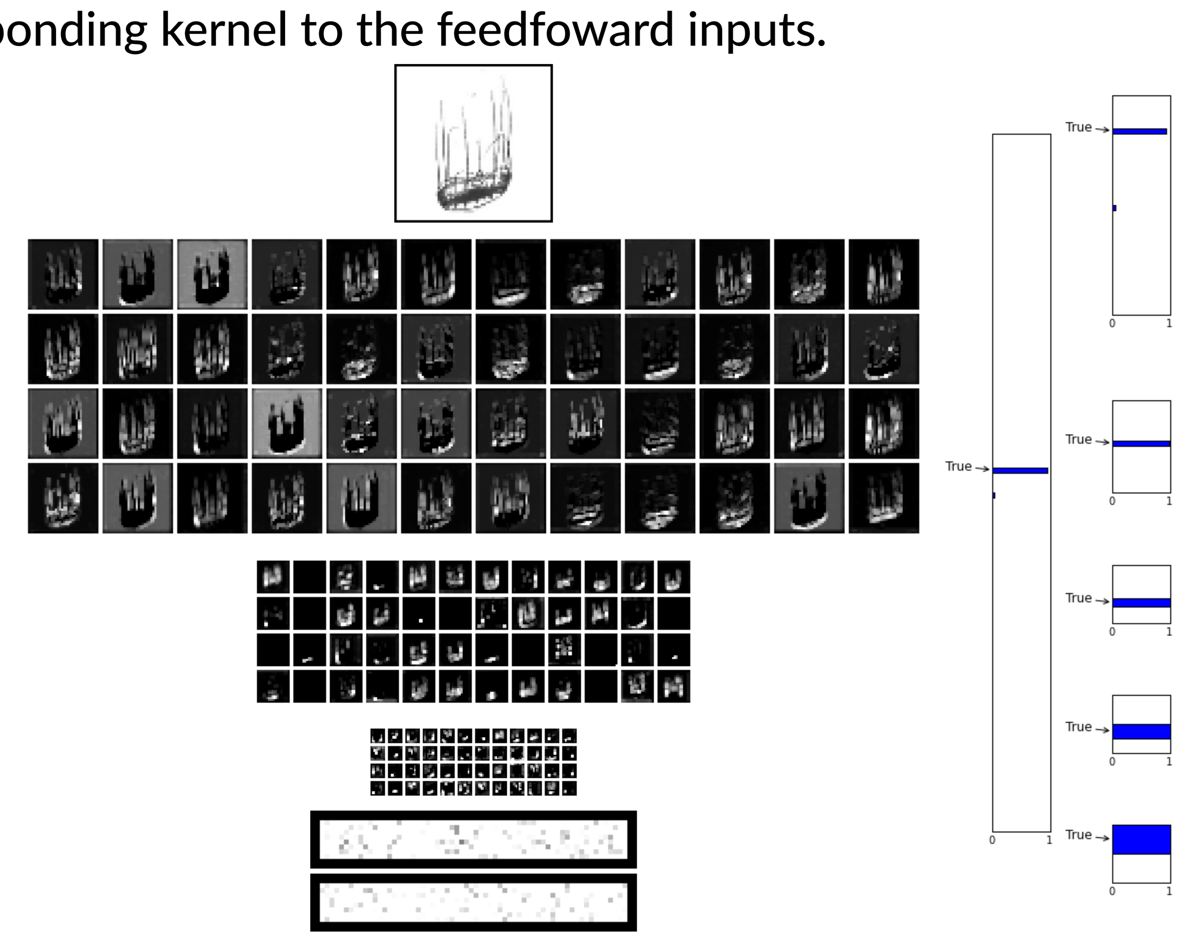
```
Global Features
- Haralick features
- Grey-Level Co-occurance Matrix attributes
- Zernike moments
- Parameter-Free Threshold Adjacency Stats
- Contour Moments and Hu Moments
- Within-contour histogram of grey-level
  intensity
```

```
Local Features
- ORB keypoints
- MSER keypoints
Keypoint descriptions were
clustered into words, or used to
directly classify the image, with
the collection of keypoints forming
a voting ensemble.
```

Classifiers built solely with these features unsuprisingly performed significantly worse than convolutional networks. However even when feeding in the extracted features to a convolutional network as extra inputs there was no measurable performance gain, suggesting that here the non-task-specific nature of these features made them largely redundant.

> OUTPUT

Below is an example of the convolutional network classifying an image. Each group of images represents a layer in the network with the individual panels showing the activation of the neurons in response to applying the corresponding kernel to the feedfoward inputs.



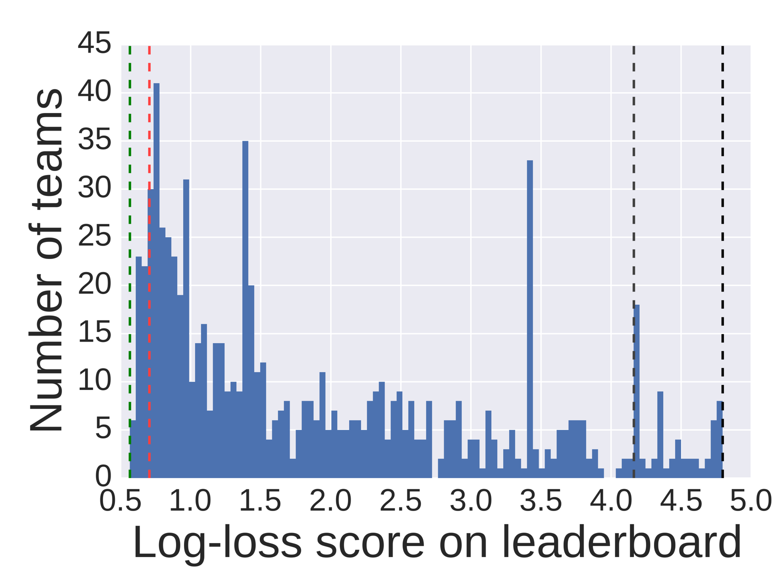
On the right, are the output probabilities for each of the classes and superclasses in the hierarchy, all correct in this (hand-picked) case.

> RESULTS

The final competition submission was an ensemble average of the output probabilities from three networks. Two had the same architecture but different training schedules, and the third was a hierarchically trained model.

The ensemble weightings were determined by performing gradient descent on heldout data.

Our log-loss score of 0.704 placed us 57th out of 1054 entrants on the leaderboard. Right: distribution of better-than-baseline leaderboard scores (green: leader; red: us; grey: prior distribution).



Acknowledgements

We would like to thank Kaggle, Booz Allen Hamilton and the Hatfield Marine Science Center for organising the competition, and thank Chris Williams, Bob Fisher and Amos Storkey for providing useful comments.

EPSRC Engineering and Physical Sciences Research Council · MRC Medical Research Council · BBSRC · kaggle · NATIONAL DATA SCIENCE BOWL