Interclass Loss: Learning from Class Relationships

Albert Haque (ahaque@cs.stanford.edu)
Computer Science Department, Stanford University

Abstract

Existing convolutional network architectures are often trained as N-way classifiers to distinguish between images. We propose a method to incorporate interclass distance into our convolutional network for image classification. The rationale is that not all errors are equal. In some situations, misclassifying a dog as a cat may be more acceptable than misclassifying a dog as an airplane. We evaluate our method using interclass distance to classify plankton. Our method reduces the average interclass distance by 25%, compared to standard multiclass loss methods. For the Kaggle competition, we achieve a multiclass loss of 0.86755 with our method.

Background & Motivation

Why plankton?

Plankton are responsible for consuming 25% of CO2 released from burnt fossil fuels each year and are the foundation for several marine and terrestrial food chains. It is important to monitor their population levels to assess ocean and greater environmental health.

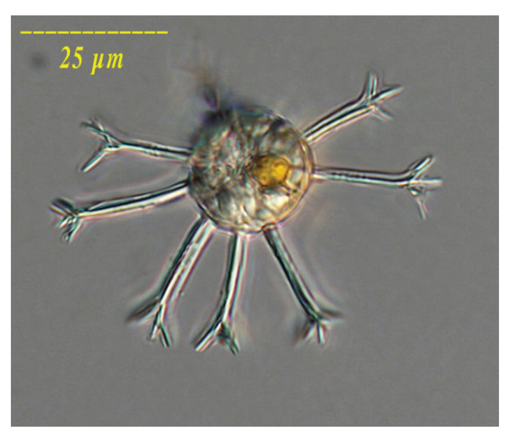

Figure 1: Image of a plankter (Cladopyxis Dinoflagellate)

Unmanned underwater vehicles have generated massive volumes of image data which are infeasible for manual human analysis. In this project, we leverage convolutional networks and employ a distance-based loss function [1] for image classification tasks to accelerate plankton classification.

Dataset

We use labeled images from the 2015 National Data Science Bowl [2] collected by the Hatfield Marine Science Center at Oregon State University. The training and test set consists of 30K and 130K grayscale images, respectively. There are 121 distinct class labels.

Figure 2: Sample images from the dataset

Methods & Models

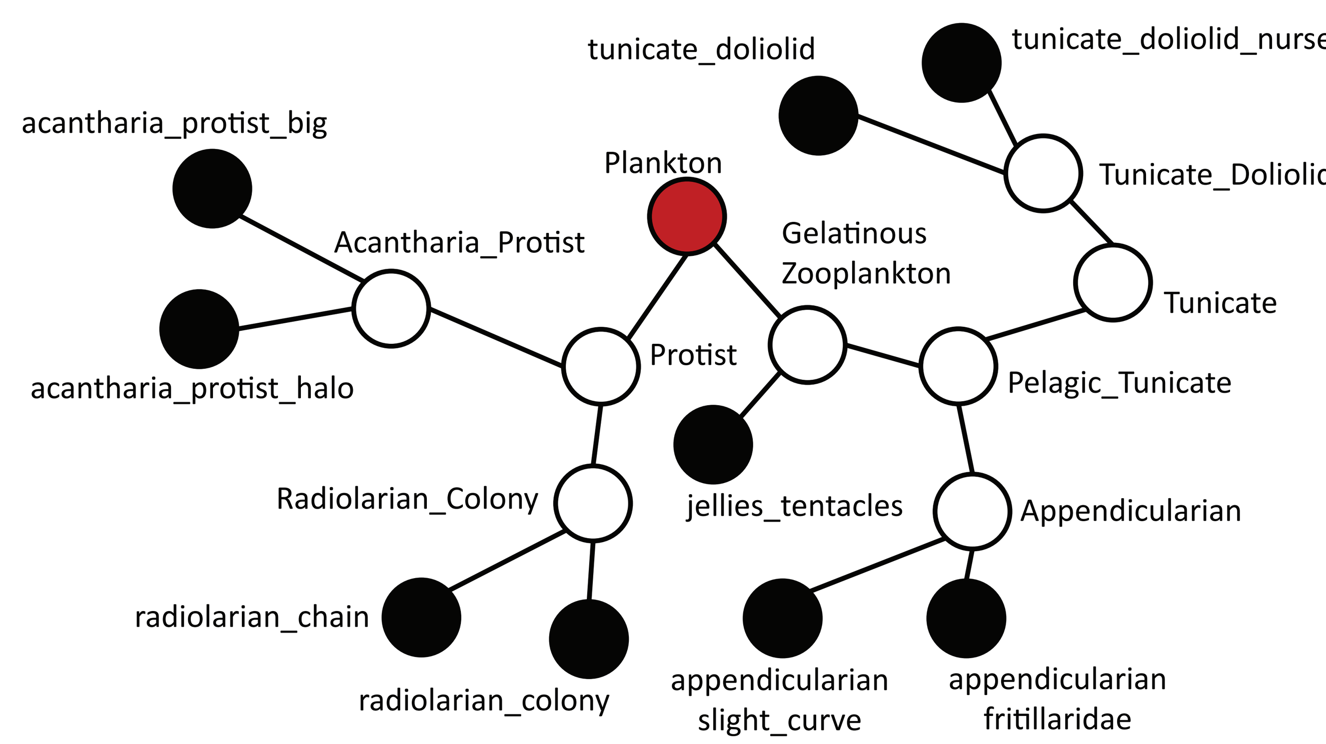

Figure 3: Graphical representation of the taxonomy. Red is the root and black nodes indicate leaf nodes (labels). This undirected graph contains weighted edges witch correspond to the phylogenetic distance between each label.

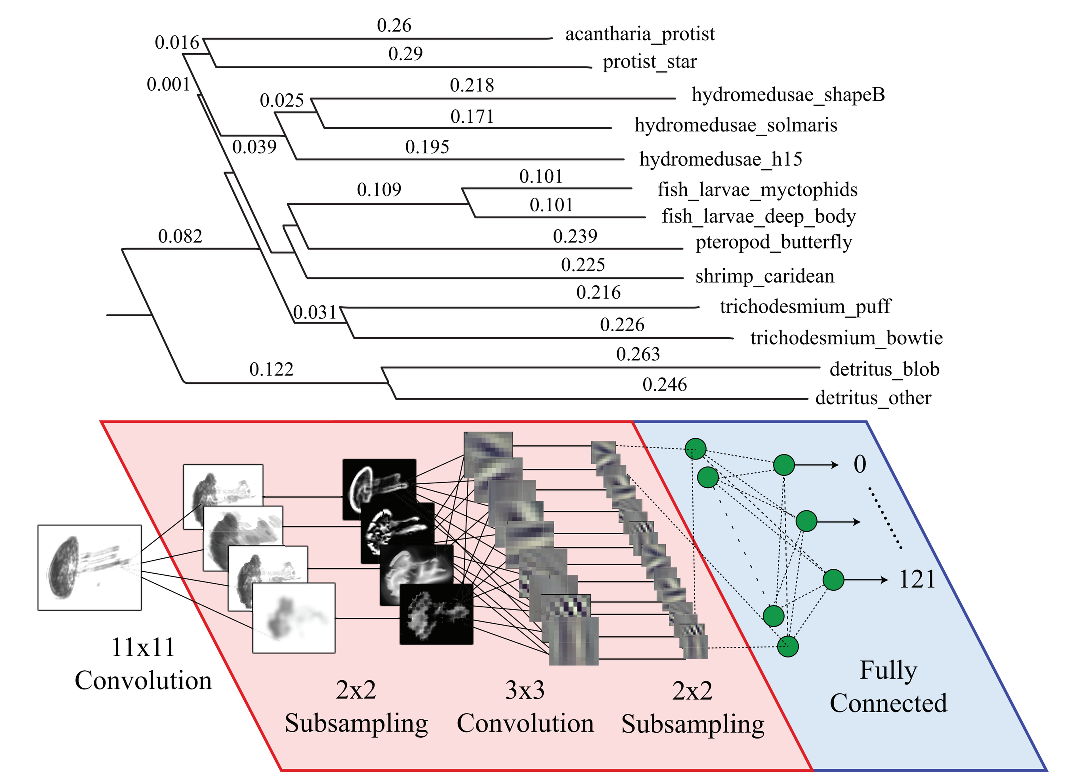

Figure 4: CNN overlayed with the phylogenetic tree of output classes. Output scores are then scaled according to the phylogenetic distance between the output node j and the correct class label y_i .

All living organisms are classified into a taxonomy. These structured trees can be represented in several forms. Figure 3 shows the graphical representation between the plankton class labels. Before training our network, we compute the pairwise distance between each leaf in our graph using a pairwise additive distance function. The distances are then stored in a matrix used during model training. We modify the multiclass log loss to factor interclass distance:

$$\text{Interclass Loss: } -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\left[1\{j=y_i\}\log(p_{ij})\left(1+\frac{\mathcal{D}(j,y_i)}{2\times \text{maxdepth}}\right)\right] \qquad \text{Average Interclass Distance: } \frac{1}{N}\sum_{i=1}^{N}\mathcal{D}(\hat{y}_i,y_i)$$

Where $\mathcal{D}(j,y_i)\in[0,\text{maxdepth}]$ denotes the distance from class j to the ground truth y_i and K denotes the number of classes.

Experiments & Results

Hardware

- Terminal.com GPU instance
- NVIDIA GRID K520, 3072 cores, 8 GiB GDDR5 memory

Software

- BVLC Caffe
- GraphLab Create (Deep Learning Module)

Preprocessing

Preprocessing is done in realtime with the exception of image resizing and centering; these steps are done before training and testing.

Performance Metrics

1. Classification accuracy (top-1 hit rate)
2. Average Interclass Distance (after all predictions are made)

Data Augmentation

- Rotation: random between 0 and 360 degrees
- Translation: random shift between -20 and +20 pixels in x and y direction
- Scaling: random scale factor between 1/1.25 and 1.25
- All transformations are performed in one operation, on the fly (see right)

$$P' = \begin{bmatrix}\cos\theta & -\sin\theta & t_x\\ \sin\theta & \cos\theta & t_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}s_x & 0 & 0\\ 0 & s_y & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix}$$

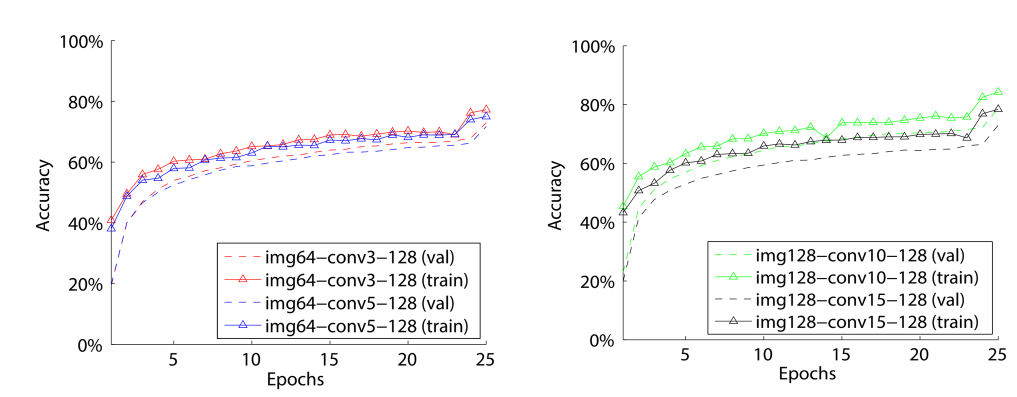

Figure 5: Accuracy Using Standard Multiclass Loss

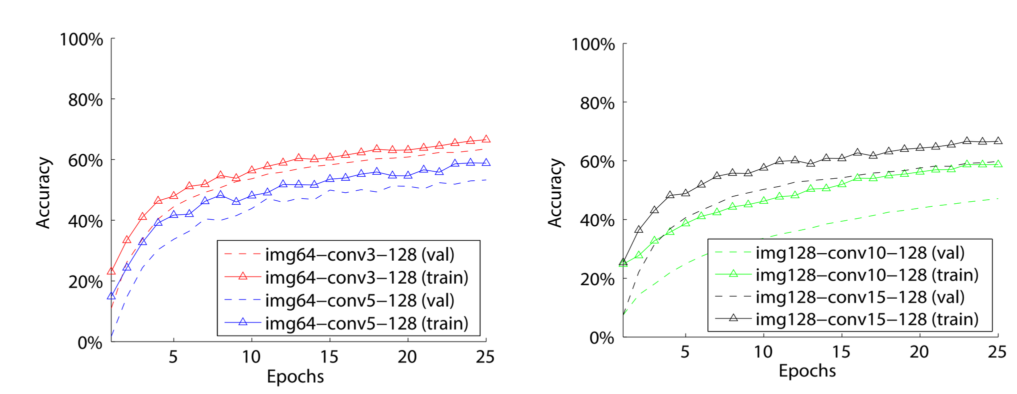

Figure 6: Accuracy Using Our Interclass Loss

Discussion

General Comments

Our interclass loss decreases classification accuracy from 82% to 65%. However, if we look at the average interclass distance, our method outperforms the standard multiclass loss.

Table 1: Class Structure

Depth	Count	Percent
0	2	1.1%
1	3	1.7%
2	16	9.1%
3	45	25.7%
4	36	20.6%
5	40	22.9%
6	26	14.9%
7	7	4.0%
Total	175	100.0%

Interclass Distance Comparison

Standard multiclass loss generated an average interclass distance of 0.821. Our interclass loss function generated an average interclass distance of 0.648. This means although our CNN misses more top-1 classifications, on average, it is closer to the correct label than the CNN with the standard multiclass loss.

With interclass loss, our network's softmax distribution has higher Kurtosis (i.e. the probability distribution has a sharper peak). This is because interclass loss penalizes flatter distributions with higher probability mass. Figures 7 and 8 show the output softmax probabilities for the same training image.

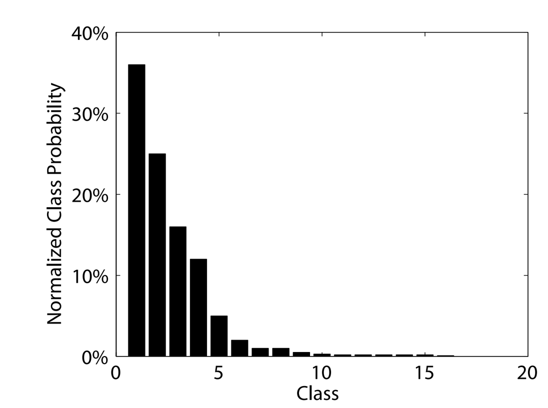

Figure 7: Standard multiclass loss contains more probability spread.

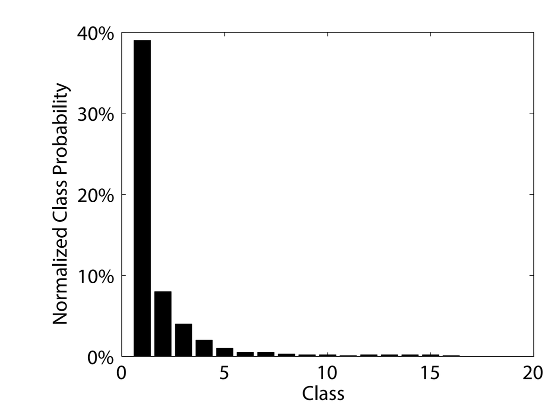

Figure 8: Our interclass distance loss causes high Kurtosis output.

Future Work

We hope to further test our interclass loss function on other standardized datasets such as Imagenet [3]. The degree of separation could be used as the distance metric, instead of phylogenetic distance. Additionally, we can visualize the model weights to identify the differences caused by our loss function.

References

[1] A. Vailaya, A. Jain, and H. J. Zhang. On image classification: City images vs. landscapes. Pattern Recognition, 31(12):1921–1935, 1998.
[2] Inaugural national data science bowl. http://www.datasciencebowl.com/. Sponsored by Kaggle and Booz Allen Hamilton. 2015.
[3] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575, 2014.

kaggle NATIONAL DATA SCIENCE BOWL 2015

Stanford University