CSCI 481/597J Project Literature Review

Michael Albert, Jonah Douglas, Ethan Lindell, Archan Rupela, River Yearian

May 6, 2020

Paper 1 The key techniques used were skip connections with a convolutional neural network (CNN). The CNN was used for multi class classification using feature extraction from selected parts of a bird, considering size, shape, and color of the part. The data used was obtained from the Internet of Birds, a citizen science platform and dataset for bird species identification. Also obtained images from the internet, which needed to be augmented in a number of ways to conform to to the dataset. The model was trained to identify 27 bird species endemic to Taiwan. The overall accuracy achieved by their model of identifying bird species was 98.70%.

Paper 2 This paper describes in detail popular deep learning architectures for image recognition and tests which one is the most accurate while remaining quick. The architectures described were LeNet, AlexNet, GoogLeNet, ResNet, and Inception-ResNet; which all consisted of 5+ convolution layers with each successive architecture built upon the results and shortcomings of the others. Each of the neural networks were trained from scratch with training sets of over 60k+ images. The experiment uses two different datasets: one of flowers and another with human faces. Upon testing each of the models with both datasets, they concluded that Inception-ResNet had the highest accuracy while also keeping the computational requirements in-check. Although it was the best of the other models, Inception-ResNet's shortcoming is that it only works well with large datasets and often overfits smaller datasets.

Paper 3 This paper, Learning to Compare: Relation Network for Few-Shot Learning, goes into the use of using deep learning from a very limited dataset by using a technique called few-shot. Many practical problems, the bird identification task included, have datasets that are too small and will result in deep learning overfitting for the specific cases rather than getting the correct generalizations. Current solutions to the few-shot approach are too complex with heavy duty inference algorithms or time consuming with extreme fine tunings. The researchers of this paper used a two-branch Relation Network (RN) instead, which cuts down the previous solutions (Fine-tuning, Recurrent Neural Network, Metric Learning, and Siamese Network) significantly. The RN achieves this by "learning-to-learn" on a related very large data set, then transferring the metrics onto the target data. The RN uses a training , support, and test set where all sets have labels but the training set is disjoint from the labels used by the support and test sets. During the learning phase, there are episodic trainings to tune the embedding model and metrics. As a result we get a model that outperforms all other current few-shot meta-learning approaches and has a very high accuracy across one-shot and five-shot models (with the lowest at an accuracy rate of 97.6% $\pm$ 0.1%).

Paper 4 In this paper, Divya Meen and L. Agilandeeswari built a semi-supervised learning based Multi-part convolutional Neural Network that classifies from their base of 35992 images. The technologies and techniques used were the Fisher vector, inception-V3, modified hellinger Kernel classifier, multi-part based convo-lutional neural network, pseudo-labels, semi-supervised learning and stacked auto encoder. Classifying an animal as a dog or cat is a primary level classification where as a secondary level classification would be identifying different species of dogs or animal, where you may discriminate within the primary level. This is also known as fine-grained classification. In application they foresee implementing thermal IR camera's, sending the images to a classification processing units where it pre processes, does feature extraction and classification before it decides whether the animal is dangerous or not. Their process is as follows. Generic classification with Tensorflow, those classification results are then analyzed with Tensorboard. If the class of animal has high chance of miss-classification

pseudo-labels are assigned to them and retrained. The second state is to MP-CNN to do fine grained classification. This method had a 98.17% accuracy rate.

Paper 5 This paper describes the use of pose normalization by prototype regions and a convolutional neural net to classify birds by species. First, the object's pose is estimated and then warped to match a prototype image. From this normalized image, features are extracted by an eight layer CNN and used for classification. This is useful because the normalized images allow for easier extraction of useful features from a given region. The model is pre-trained on a 1.2 million ImageNet dataset and then trained/tested on a dataset of 200 bird species with 11,788 images. The results of this are a neural network that can classify images with an accuracy of 85.4%, reducing the error rate on the dataset they used by 30% compared to previously used models.

References

1. Y. Huang and H. Basanta, "Bird Image Retrieval and Recognition Using a Deep Learning Platform," in IEEE Access, vol. 7, pp. 66980-66989, 2019,

2. P. Shah, V. Bakrola, S. Pati, "Optimal Approach for Image Recognition using Deep Convolutional Architecture,".

3. F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr and T. M. Hospedales, "Learning to Compare: Relation Network for Few-Shot Learning," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 1199-1208.

4. S. Divya Meena and L. Agilandeeswari, "An Efficient Framework for Animal Breeds Classification Using Semi-Supervised Learning and Multi-Part Convolutional Neural Network (MP-CNN)," in IEEE Access, vol. 7, pp. 151783-151802, 2019.

5. Branson, Steve & Horn, Grant & Belongie, Serge & Perona, Pietro. (2014). Bird Species Categorization Using Pose Normalized Deep Convolutional Nets. BMVC 2014 - Proceedings of the British Machine Vision Conference 2014.