Object recognition and computer vision: Image classification challenge

Mohammed Amine Kheldouni
Ecole des Ponts ParisTech
mkheldou@ens-paris-saclay.fr

Abstract

We present a model for multi-class image classification using instance segmentation, data augmentation and deep architectures of neural networks combined with transfer learning and fine-tuning. The aim of this assignment is to classify birds images within 20 different classes. We achieve a final score of 73.55% *on the test set.*

1. Setting up the model

We used a standard ResNet-34 that serves as a feature extractor. This model has the advantage to be easily trainable and is less likely to overfit the data compared to a deeper structure. The dataset provides us with 1080 training examples of 20 different species of birds. The test set is composed of 517 images and includes harder examples which explains the accuracy difference with the validation set. The difficulty lies in image segmentation and birds location as well as in the classification between similar species of birds such as the *American crow* and the *fish crow*.

Figure 1. Examples of images ”hard” to classify in the test set

2. Transfer Learning

In view of the models complexity, it would genuinely be too hard to train it from a totally random distribution of parameters. Therefore, we acknowledged the importance of taking pre-trained weights (using ImageNet dataset) into account in our experiments and fine-tuning the last layers according to the number of classes.

3. Data Augmentation

To increase the capability of our CNN, we used two refinements of the standard ResNet-34 model:

- **Data transformation -** We used the transformation module of PyTorch to resize the images to $(244, 244)$ RGB images and allow some randomness during the training (rotations, flips, ...)
- **Mask R-CNN**[1] **-** This system uses computer vision techniques (NMS, RoI pooling ...) to extract proposals and boxes which helped us crop the data and broaden our training set to more than 2000 images.

Figure 2. Using a Mask R-CNN for cropping our region of interest and augmenting the training and validation sets

4. Results and comments

The accuracy increases notably with pre-trained weights and helps the SGD to converge towards small values of cross entropy loss. Data augmentation helps the algorithm converge faster and induces a more robust accuracy of 73.55% on the test set.

There are improvements to keep in mind such as taking Deeper residual networks for better results. Unfortunately, due to a limited computational capacity of the hardware, we couldn’t go deeper in residual networks. We also could have thought of refining the data augmentation by removing the background in cropped images and optimizing parameters.

5. Conclusion

We successfully managed to build a fine-tuned ResNet34 model for birds classification and used many properties for accuracy improvement. However, due to limited computational capacity, we couldn’t use Deeper architectures but still managed to reach 86% accuracy on validation set and 73.55% accuracy on the test set.

References

[1] K. He, G. Gkioxari, P. Dollar, and R. B. Girshick. Mask R-CNN. (https://github.com/facebookresearch/maskrcnn-benchmark).

[2] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition.