Stanford University

Patch-based Classification on Low Resolution Images

Weizhi Cheung, Xiaojun Sun, Tian Du
cheungwz@stanford.edu, sunxj@stanford.edu, tian99@stanford.edu
CS229: Machine Learning, Stanford University

Abstract

The state-of-art image classification methods can learn subtle details between visually similar classes, however low-resolution data source poses a significant challenge. Meanwhile, low-resolution images exist in various industry fields given the actual physical conditions as well as the limited information capturing devices. In addition, classifying low resolution images is critical for developing robust, resolution-agnostic classification systems. This project attempts to explore the BagNet architecture on low-resolution images, since it has been proven to be effective on ignoring the spatial ordering of the image features, which might help when classifying low-resolution imagery.

Dataset

- tiny ImageNet Dataset is created for Stanford CS231N.
- it has 200 classes, and each class has 500 training samples, 50 validation samples and 50 test samples.
- images are of size 64 x 64.
- for the purpose of this project we are only aiming to predict the class of the images.

Methods

- (1) Extract features from patches of an image
- (2) Apply linear classifier on top of local feature representation
- (3) Training model with different combination of optimizer and loss function
- (4) Compare performance from different approach

Experiments

Below are the heat map generated using the trained model (since 64 x 64 pixel is small to visualized clearly, we use trained model to generate heatmap on regular ImageNet images).

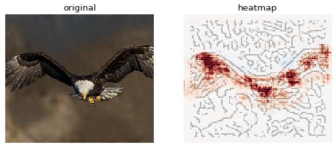

Figure 1

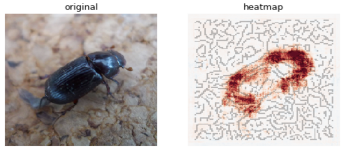

Figure 2

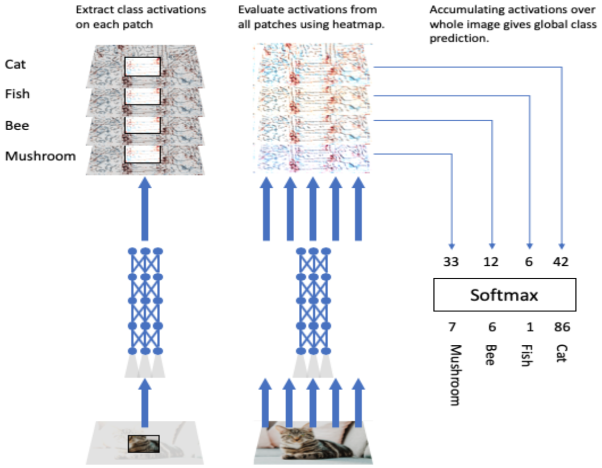

Figure 3

Our modeling approach consists of two steps:

- we use multiple stacked ResNet blocks to extract feature representation in the form of a 2048-dimension vector from each patch (of size p x p pixels) in a sample image, and then
- we apply a linear classification model over these features to make a global prediction on the whole image.

Comparison & Tuning

To better explore the utilization of BagNet on Tiny ImageNet, we conducted control experiments to evaluate the impact of patch size, optimizer, and inference mode on model performance as a guide for future implementation of our model on other applications.

Here is a sample loss trakcing graph generated for one set of optimizer and loss function combination:

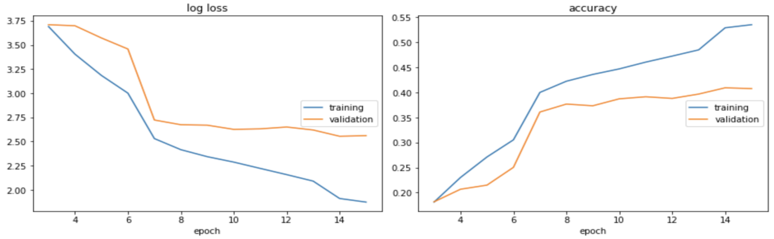

Figure 4

We conducted set of experiments where we compared SGD, Adadelta, and Adam on BagNet-9, BagNet-17 and BagNet-33. This set of experiments (4, 5, 6, 8 10 12, 14 16, and 17, 14) fixed number of epochs for training, initial learning rate, and the decay learning rate. The models were all trained from scratch

- Adadelta and Adam both outperforms SGD significantly and yields best validation accuracy of 43%
- Best validation accuracy increases with the increase of BagNet patch size
- Best validation accuracy can be further improved if BagNets are pre-trained with ImageNet dataset.

The graph below visualized the comparison:

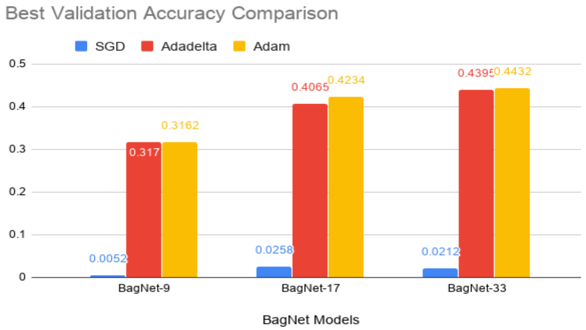

Figure 5

Future Work

The future work on this topic will be on fine-tuning the current architecture to perform classification on both low and high resolution images. Or in other words, a resolution-agnostic network architecture. The challenge of designing a multiple-resolution image classifier is to maximize the performance in both the target and source domains. There are several attempts we could try:

- use super-resolved / super-resolution image to enhance the image feature before feed the data into the network
- use adversarial training approach that has been proven to be resistant to small changes in the input image that are imperceptible to the human eye

References

[1] Wieland Brendel, Matthias Bethge, Approximating CNNs with Bag-Of-Local Features Models Works Surprisingly Well on ImageNet, 2019