Fine-Grained Image Classification

Debabrata Sengupta
dsgupta@Stanford.edu

Motivation

- Discriminating between sub-classes of an object (e.g. breeds of dogs)
- Harder problem than image classification
- Large intra class variation
- Small inter-class variation

Models

Objective : Minimize cross entropy loss.
Optimizer : Adam.

Tried training three different models.

My own CNN

- Trained from scratch.
- Insufficient training data => Overfits.

VGG16

- Transfer learning, data augmentation, Regularization.
- Best model I got.

ResNet50

- Overfits a lot even with transfer learning, data augmentation, regularization.

Results

- VGG16 model performed best.
- Achieved 40.70% test accuracy.
- Authors of the Stanford Dogs dataset achieved 22% accuracy.
- Current state of art : 88.9% accuracy.

Table 1 : Classification Accuracy (%)

Model	Training Set	Validation Set	Test Set
My own CNN architecture with regularization	93.66	30.67	31.55
VGG16 (Transfer Learning + Augmentation + Dropout)	43.87	40.20	40.79
ResNet50 (Transfer Learning + Augmentation + Dropout)	94.14	23.83	24.72

Future work

- Explore ensemble methods => Training K expert classifiers and aggregating their decisions

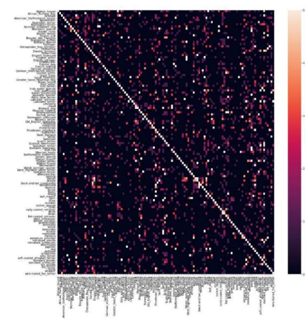
Dataset

- Stanford Dogs - more than 22,000 images from 120 breeds of dogs.
- Partitioned into 12,000 training and 8580 test images.
- Split training set into train and validation (80% : 20%)
- Breed label and Bounding boxes annotated.
- Background clutter, occlusion, variation in color, poses.
- Images of different sizes.

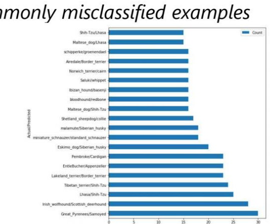
Error Analysis

Confusion Matrix Heat Map

- Diagonal has bright spots.
- Accuracy not very high => some off-diagonal elements also have bright spots.



Commonly misclassified examples



Great Pyrenees (left) that is often misclassified as Samoyed (right)