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.

<table>
<thead>
<tr>
<th>Model</th>
<th>VGG16</th>
<th>ResNet50</th>
</tr>
</thead>
<tbody>
<tr>
<td>My own CNN</td>
<td>Trained from scratch.</td>
<td>Trained from scratch.</td>
</tr>
<tr>
<td></td>
<td>Insufficient training data =&gt; Overfits.</td>
<td>Overfits a lot even with transfer learning, data augmentation, regularization.</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Model</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>My own CNN architecture with regularization</td>
<td>95.06</td>
<td>30.67</td>
<td>31.55</td>
</tr>
<tr>
<td>VGG16 (Transfer Learning + Augmentation + Dropout)</td>
<td>43.87</td>
<td>40.20</td>
<td>40.79</td>
</tr>
<tr>
<td>ResNet50 (Transfer Learning + Augmentation + Dropout)</td>
<td>94.14</td>
<td>23.83</td>
<td>24.72</td>
</tr>
</tbody>
</table>

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)

Future work

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