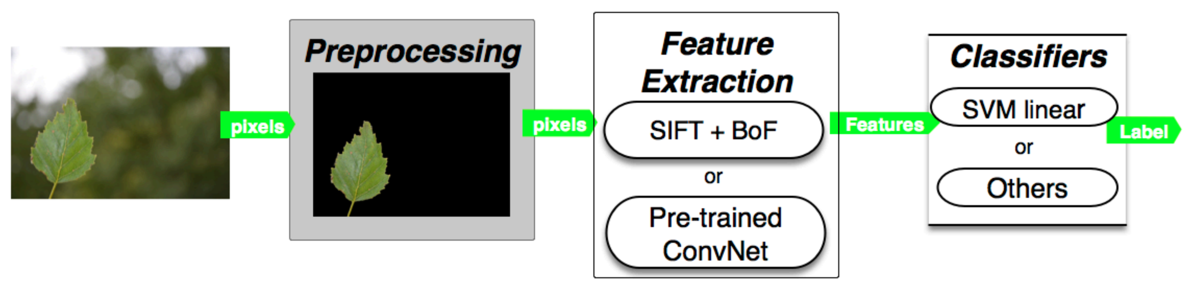
Plant Leaf Recognition

{ Albert Liu and Yangming Wang }@stanford.edu

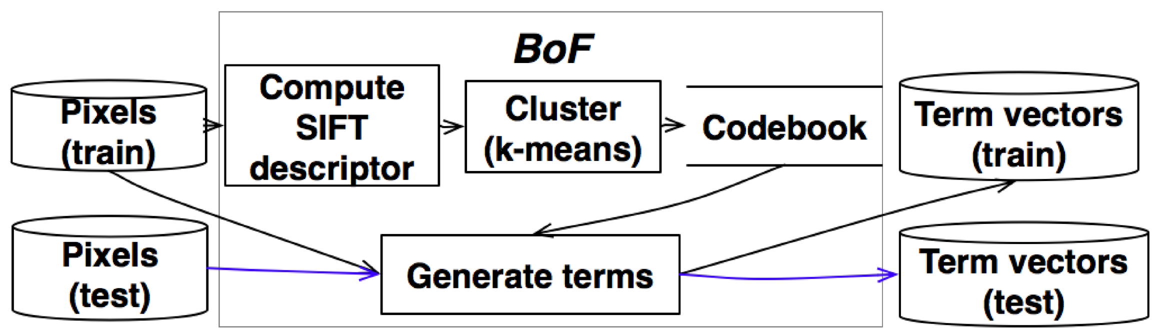
Problem

Fine-grained leaf recognition has important application in weeds identification, species discovery, plant taxonomy, etc. However, the subtle differences between species and the sheer number of categories makes it hard to solve.

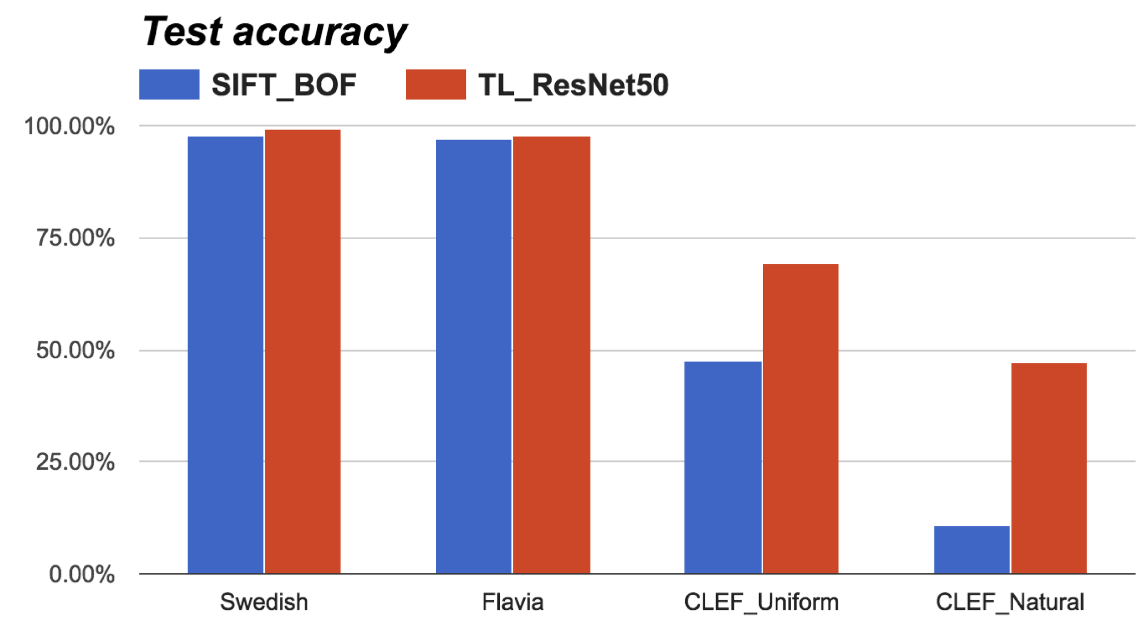
Method

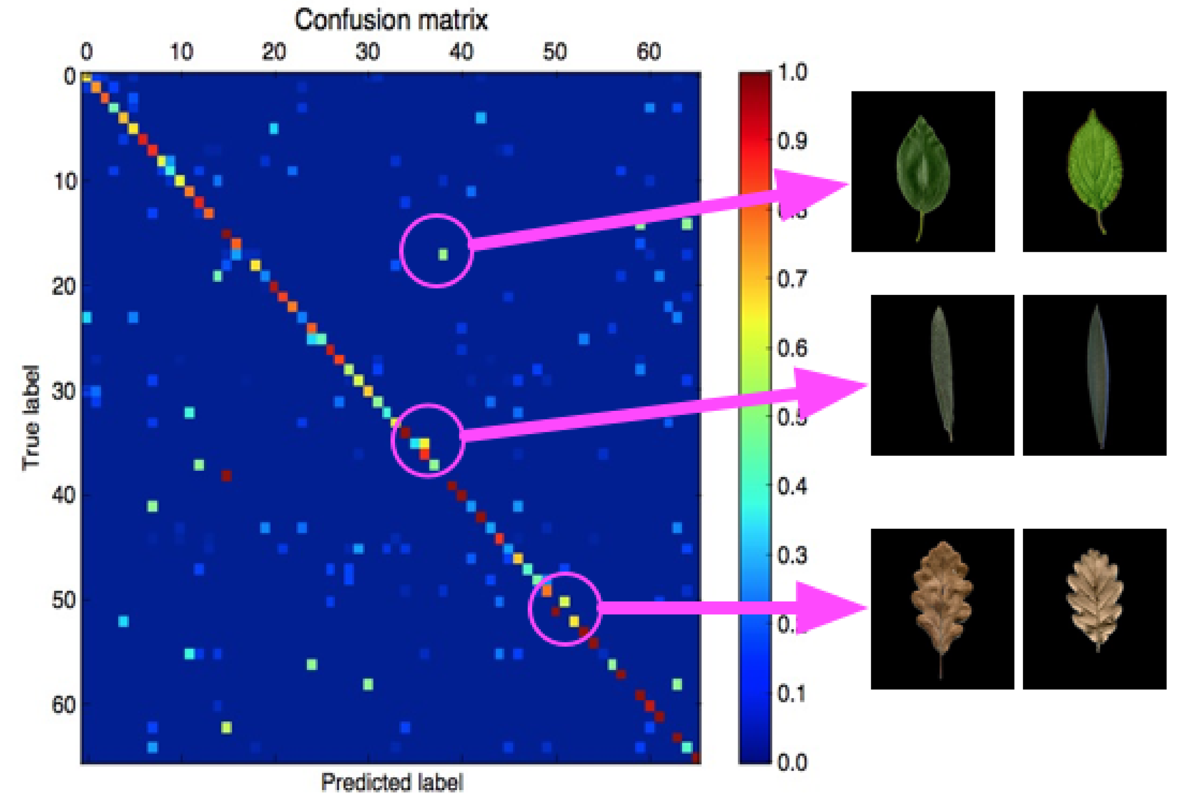


1. Preprocessing
   Apply CLAHE to reduce lighting condition variation and resize to fit the next layer. Selectively use K-means to remove background heuristically. For challenging dataset, find the convex hull containing the largest N contours and then use GrabCut to segment leaf out.
2. Feature extraction
   - ConvNets.
     Transfer learning approach. Specifically, we take a couple of ConvNets pretrained on ImageNet for ILSVRC object classification task, remove top layers and use it as generic feature extractor.
   - Traditional SIFT + BoF
     Key points are densely sampled. Size of the codebook (K) is fixed at 1000/3000.
3. Classification SVM Linear/RBF/Softmax/MLP/etc.



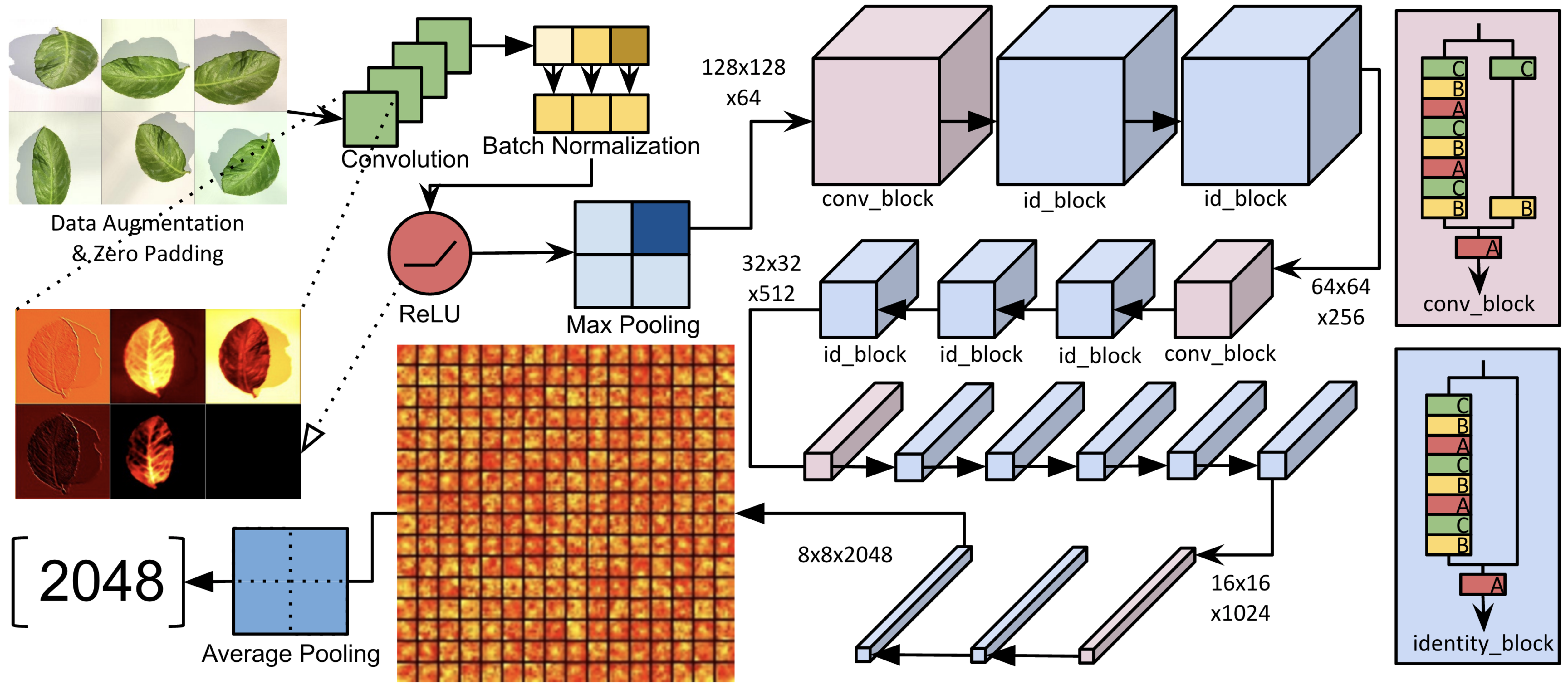
Experimental Results





Transfer Learning - ConvNet

The pre-trained weights of VGG16, VGG19 and ResNet50 are available from open source Keras Framework. After comparing, we choose ResNet50. Since ResNet50 uses a more aggressive Average Pooling with a pool size of 7x7, which makes the output dimension of the bottleneck layer much smaller than VGG architectures', it outputs a much smaller feature set which helps the prediction accuracy.

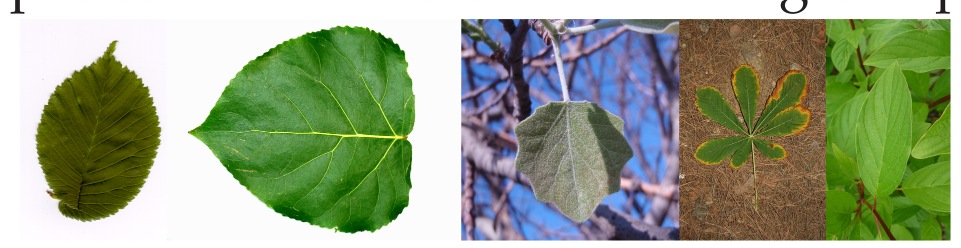


The ResNet is famous for it's deep layers, in our case, 50 layers, with 49 Conv layers and one Fully Connected layers on top. Except for the first Conv layer, the rest 48 composes 16 "residual" blocks in 4 stages. The block within each stage has similar architecture of same input & output shape.

The output of the stage 5 give 2048 features. Every cell in the grid shown above is a 8x8 filter, which will be a scalar after average pooling.

Dataset

1. Swedish/Flavia Leaf Dataset: clean images taken in controlled conditions. Use 20 samples per species for training.
2. ImageCLEF Plant: crowd sourced and noisy. Considerable variations on lighting conditions, viewpoints, background clutters and occlusions. Choose species with at least 20 training samples.



Discussions

1. As expected, CNN codes off the shelf yields similar or better accuracy, compared to SIFT+BoF. Particularly traditional method suffers with noisy datasets.
2. Error analysis shows that it helps greatly to reduce noise/variations on data.
3. Looking at the confusion matrix, we believe the main causes for misclassification
   - Very fine differences between species, which is hard even for human experts
   - Noisy and possibly non-representative train data lead to overfitting,
   - Since the ConvNet models are pre-trained for a different task, we speculate these features may not always generalize well.

Future Works

1. Acquire more data and fine-tune ConvNet to solve overfitting problem
2. Engage advanced techniques for image augmentation
3. Explore state-of-art method to detect and locate leaf for Image CLEF natural leaf dataset.