Very Deep Residual Networks with Maxout for Plant Identification in the Wild

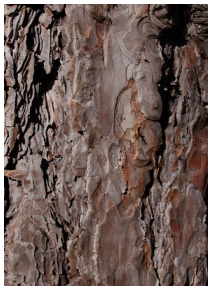
Milan Šulc, Dmytro Mishkin, Jiří Matas

Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

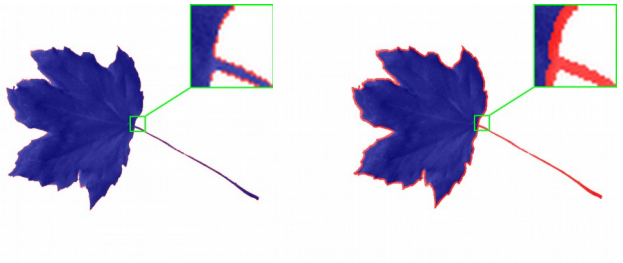
Plant Recognition at CMP before CLEF'16

We worked on narrower problems with hand-crafted features with state-of-the-art results:

- Bark recognition: textural description [1]

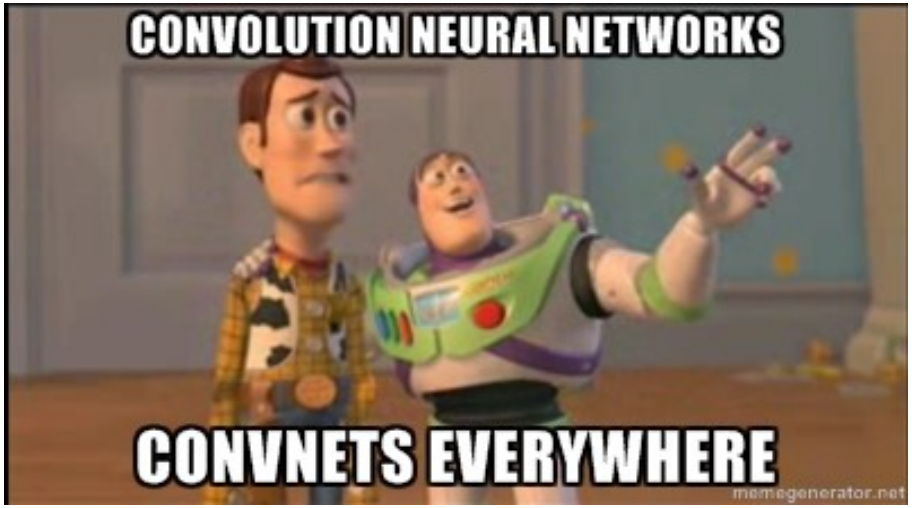
- Leaf recognition: describing texture of the leaf interior and border [2]

[1] Kernel-mapped histograms of multi-scale LBPs for tree bark recognition.
Milan Šulc and Jiří Matas. IVCNZ 2013.
[2] Fast features invariant to rotation and scale of texture.
Milan Šulc and Jiří Matas. ECCV 2014, CVPPP workshop.

Lessons from previous PlantCLEFs

- Best performing descriptors:



- Separate networks for different content types didn't help.

- Significant effect of bagging.

Lessons from CNN Evolution

- Residual Networks [3] (ResNet):
  Best results in ILSVRC 2015 and MS COCO 2015.

- Maxout [4] activation function looks promissing,
  when combined with dropout for better regularization.

[3] Deep Residual Learning for Image Recognition.
Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. CVPR 2016.

[4] Maxout Networks.
Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C. Courville, and Yoshua Bengio.
ICML (3) 28 (2013): 1319-1327.

Deep Residual Networks

- He et al. [3] showed that residual connections accelerate learning even for extremely deep networks.
- We build on the ResNet-152 model pre-trained on ImageNet.

- 8x deeper than VGG-19 [5], but still lower complexity.

VGG-19: 19.6 billion FLOPs.
ResNet-152: 11.3 billion FLOPs.

[3] Deep Residual Learning for Image Recognition.
Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. CVPR 2016.
[5] Very deep convolutional networks for large-scale image recognition.
Karen Simonyan and Andrew Zisserman. arXiv preprint arXiv:1409.1556 (2014).

The Maxout

- Maxout [4] unit: ~ network activation function.

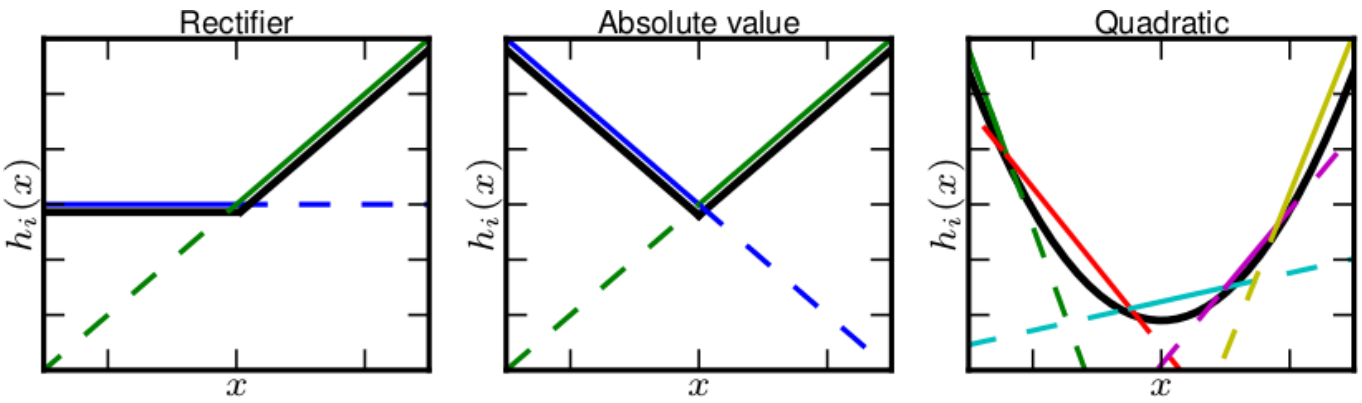
$$h_i(x) = \max_{j \in [1,k]} z_{ij} \qquad z_{ij} = x^T W_{\ldots ij} + b_{ij}$$

- Dropout is performed on x, before multiplication by weights.

[4] Maxout Networks.
Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C. Courville, and Yoshua Bengio.
ICML (3) 28 (2013): 1319-1327.

The Maxout

- A single Maxout [4] unit can be interpreted as making a piecewise linear approximation to any convex function.



[4] Maxout Networks.
Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C. Courville, and Yoshua Bengio.
ICML (3) 28 (2013): 1319-1327.

CMP Network: ResNet-152 with Maxout

- ResNet-152 pre-trained on ImageNet.
- FC layer replaced by 4 pieces of FC layers, 512 neurons each, followed with Maxout.
- Dropout is performed on the inputs of the FC layers.
- Another FC layer is added on the top for classification.

Preliminary Experiments

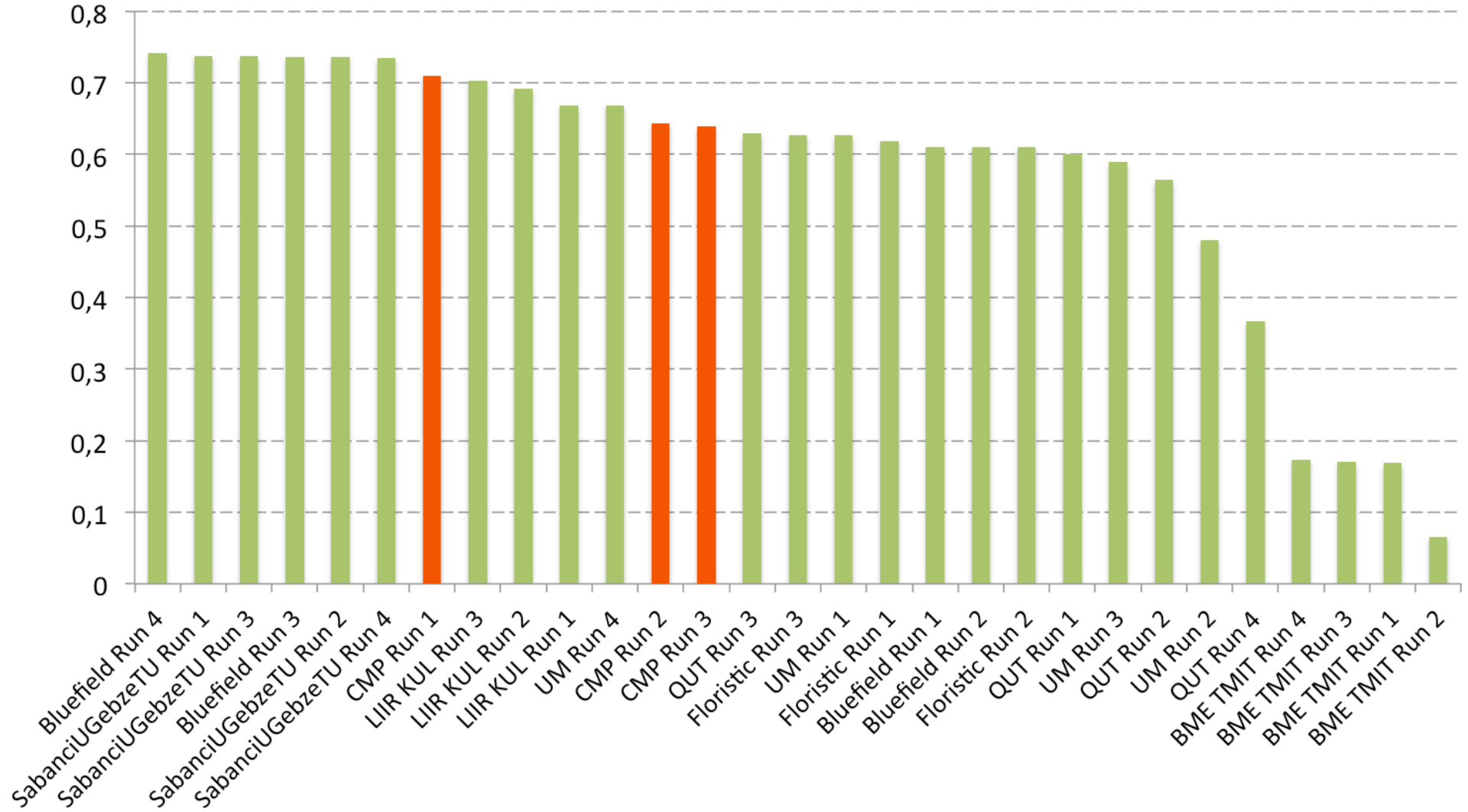
- Training set: 2015 training data.
- Validation set: 2015 test data.

Method	*mAP*
ResNet-50 (150K it.)	50.3%
ResNet-152 (150K it.)	52.2%
ResNet-152 (150K it.) + SVM	51.8%
ResNet-152 (150K it.) + sepSVM	50.6%
ResNet-152 + maxout (130K it.)	56.8%
ResNet-152 + maxout (130K it.) + 10-view test aug.	56.9%
ResNet-152 + maxout (130K it.) + fully convolutional eval.	55.9%
ResNet-152 + maxout (370K it.)	**57.3%**

CMP Submissions

- CMP Run 1 (main submission):
  - Bagging of 3 networks (ResNet-152 + MaxOut).
  - PlantCLEF 2016 training set divided into 3 folds, each network uses 2 folds for training.
  - Fine-tuning for 110K iterations (due to limited time).
- CMP Run 2:
  - Only one of the three networks.
- CMP Run 3:
  - Network fine-tuned in preliminary experiments on PlantCLEF 2015 training data, 370K iterations.

Official Score: Mean Average Precision



Results using Metadata

- The winning submission (Bluefield):
  - sums all scores with the same ObservationID = transforms the task from single-image recognition to multiple-image recognition.

- What if other pipelines used ObservationID?

Team	Single-image recognition [mAP]	Multiple-images recognition [mAP]
Bluefield Run 2 / 4	61.1 %	74.2%
SabanciUGebzeTU Run 1	**73.8 %**	**79.3 %**
CMP Run 1	71.0 %	78.8 %

- Should recognition from single image and from multiple images be evaluated separately?

Summary

- CMP recognition system:
  - state-of-the-art Residual Networks (ResNet-152)
  - added a Maxout layer
  - bagging of 3 networks (limited time)

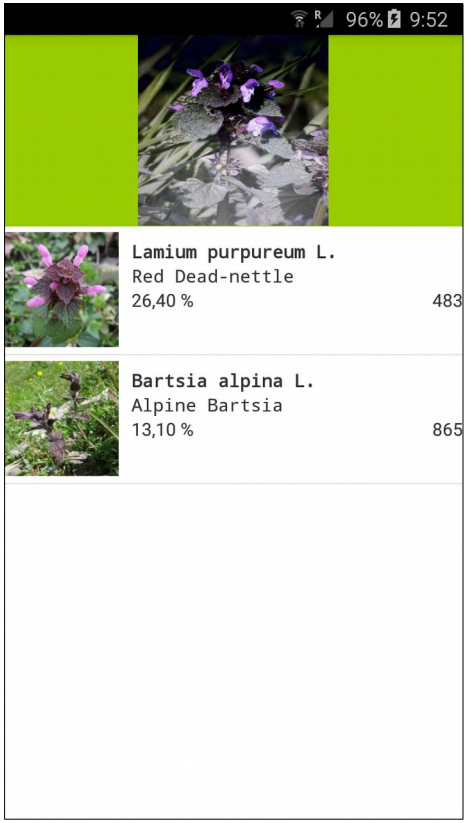
- 3rd best scoring team in the challenge,
  looking forward to PlantCLEF 2017!

- We are interested in other plant recognition tasks.

- Preparing an Android app deploying a CNN on the phone.

Mobile App

- We started developing an Android app deploying a deep CNN model on the phone.
  - no need for Internet connection :-) in the field
  - not a source of observations :-(
- Ask for a demo outdoors.



Thank you !

Questions ?

sulcmila@cmp.felk.cvut.cz