Fast Bird Part Localization for Fine-grained Recognition*

Yaser Souri and **Shohreh Kasaei**, Sharif University of Technology, Tehran, Iran

Fine-grained recognition

Fine-grained recognition task: Recognize objects of different sub-classes in images.

CUB-200-2011 Dataset: 12k images, 200 different bird species.

Due to the similarity between different classes this is a hard problem. Also discriminative features are highly local.

Blue headed vireo

Red eyed vireo

State-of-the-art** method of [1] uses a modified R-CNN [2] to find parts. Then uses a Convolutional Neural Network to extract features from bird parts (head, body) and SVM to classify.

- Due to the use of R-CNN, this method is very slow. (needs O(1000) forward passes of the convolutional neural network for localizing each part)

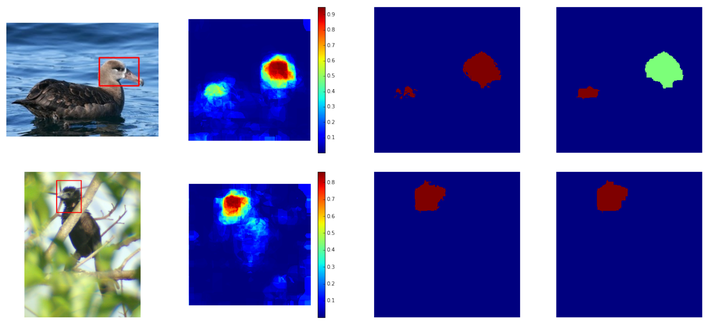
Localizing parts as classifying pixels

We pose the problem of finding a part as localizing the set of pixels that belong to that part.

Below are example results of our system classifying pixels whether they belong to the head of a bird. Output of the part localizer would be something like this:

From these classification scores of each pixel we can easily extract the bounding box around each part:

- Threshold the pixel probability map (cleverly!)
- Do some morphological filtering to remove the noise
- Find the largest connected component
- Find the box around that connected component

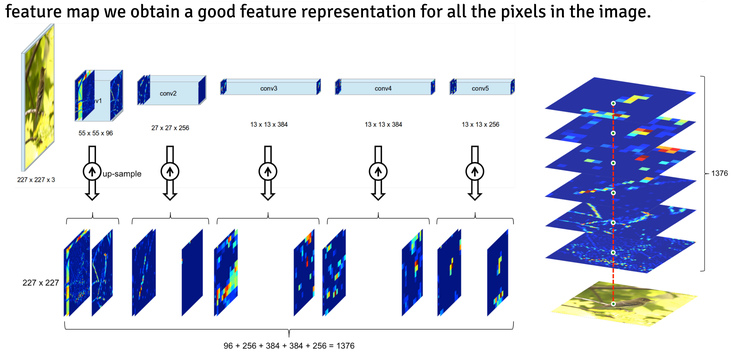


Part localization pipeline

To train each part localizer we need three things:

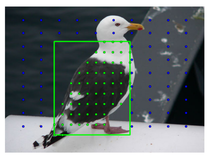
- feature representation for each pixel
- a set of positive and negative example pixels
- a classifier

Feature representation for each pixel (similar to [3]):
We use an off-the-shelf Convolutional Neural Network which has been pre-trained to classify ImageNet and only use a single forward pass of its convolutional layers. After up-sampling each feature map we obtain a good feature representation for all the pixels in the image.

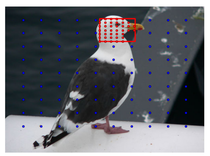


Positive and negative training pixels:
We mine a large set of positive and negative pixels for each part that we want to localize from the training set of CUB-200-2011 dataset.

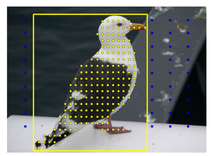
Points for body localization



Points for head localization



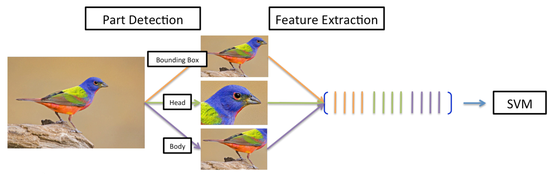
Points for bounding box localization



Classifier:
We use a Random Forest classifier to classify pixels. (very fast when testing)
At test time we densely classify all pixels.

Classification pipeline

We use a similar classification pipeline as [1] to perform classification on CUB-200-2011 dataset.

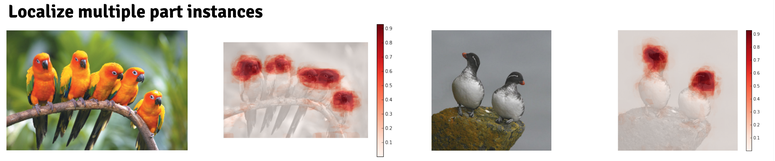


Classification results:
We achieve comparable accuracy with a method which is at least 2 orders of magnitude faster.

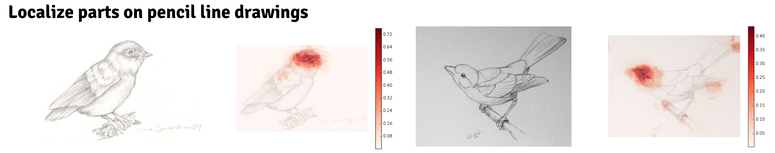
Method	Mean accuracy
[1]	73.89%
Ours	72.02% (+- 0.33)

Qualitative results and properties

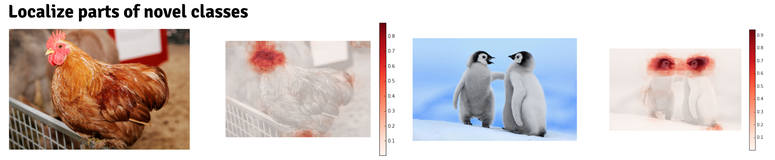
Localize multiple part instances



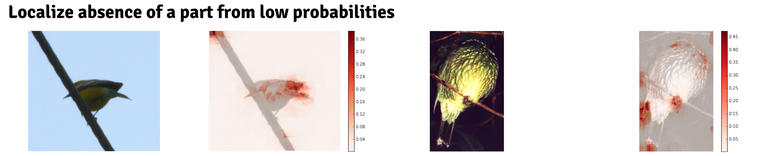
Localize parts on pencil line drawings



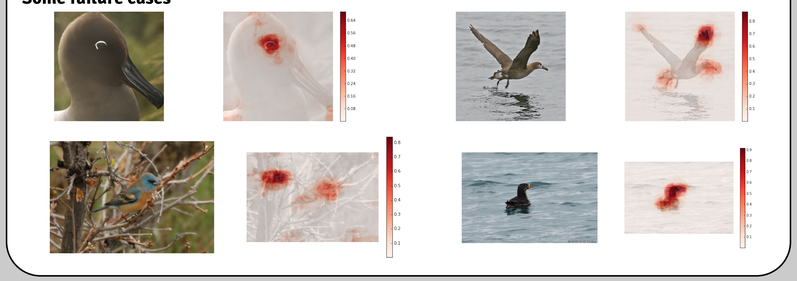
Localize parts of novel classes



Localize absence of a part from low probabilities



Some failure cases



Takeaway

- We have proposed a fast method for localizing parts of objects. And with the help of them we can achieve comparable accuracy to the state-of-the-art for the task of fine-grained recognition.
- Feature maps of Convolutional Neural Networks are very powerful representation for pixels of an image.
- With similar pipelines we can perform many other pixel-wise labeling tasks (e.g. edge detection, semantic segmentation, pose estimation, etc.)

References and Notes

* This work was accepted to the FGVC workshop in conjunction with CVPR 2015.
** Some facts might be old, since few months have past since this work was published.

[1] Zhang, Ning, et al. "Part-based R-CNNs for fine-grained category detection." ECCV 2014.
[2] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." CVPR 2014.
[3] Hariharan, Bharath, et al. "Hypercolumns for object segmentation and fine-grained localization." CVPR 2015.