Multiple Granularity Descriptors for Fine-grained Categorization

Dequan Wang[1], Zhiqiang Shen[1], Jie Shao[1], Wei Zhang[1], Xiangyang Xue[1] and Zheng Zhang[2]
[1]School of Computer Science, Fudan University [2]New York University Shanghai

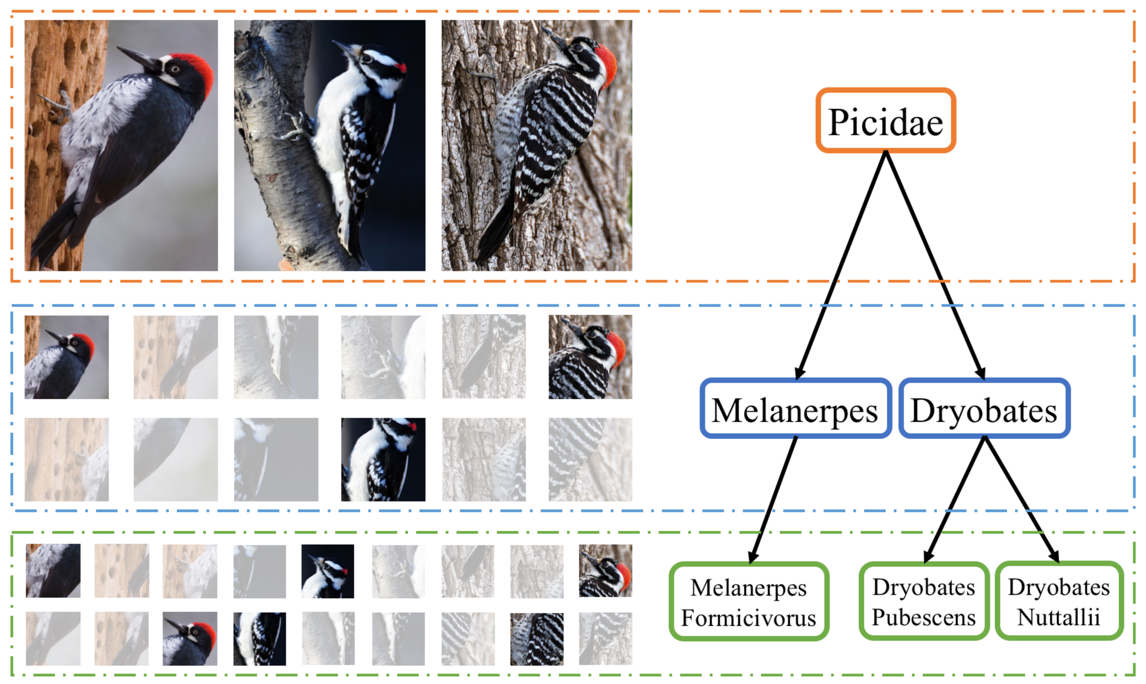
Introduction

Fine-grained categorization is challenging due to two main issues:

- how to localize discriminative regions
- how to learn corresponding features

Artic Tern Caspian Tern Common Tern Fosters Tern

We leverage the simple fact that a subordinate-level object already has other ancestor labels in its ontology tree. These "free" labels can be used to train a series of ConvNet-based classifiers, each specialized at one grain level.



Following the assumption that domain experts distinguish finer classes with visually distinctive features, hierarchies thus have embedded and latent knowledge. Multi-grained labels are free for extracting the corresponding discriminative patches and representations.

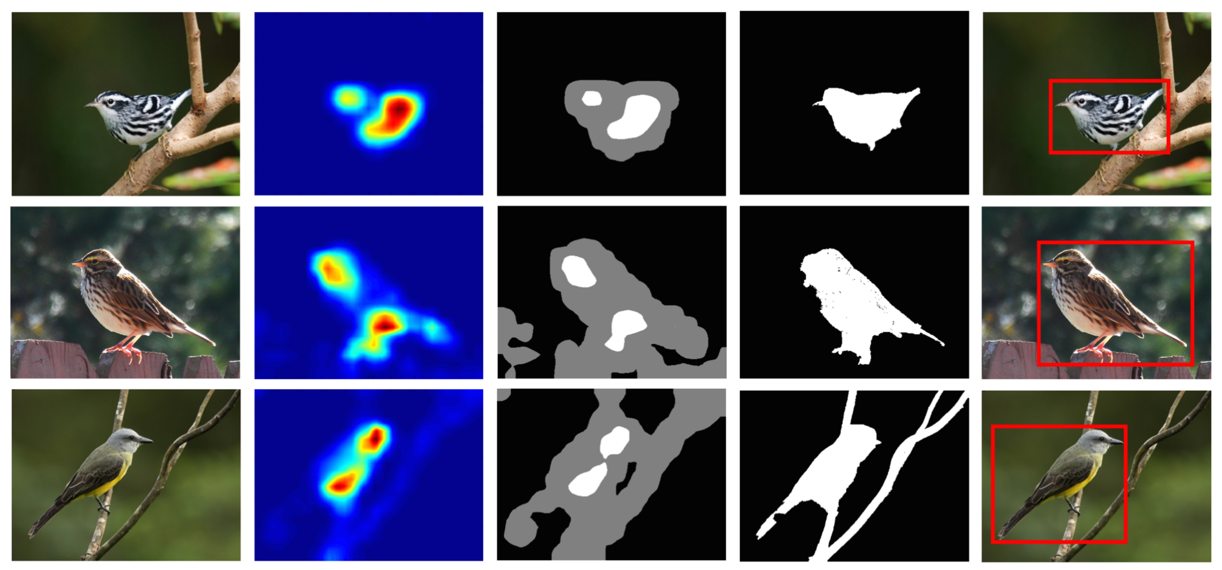
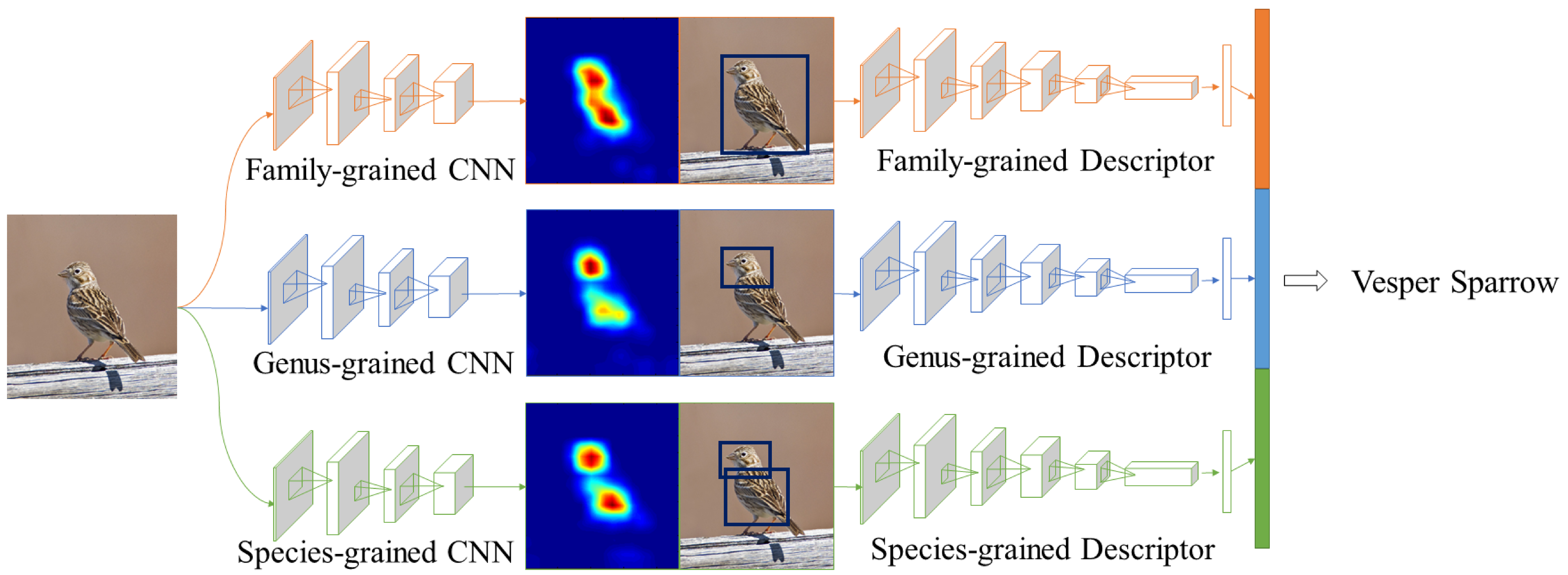


Image Heatmap Seedmap Segmentation ROI

Method

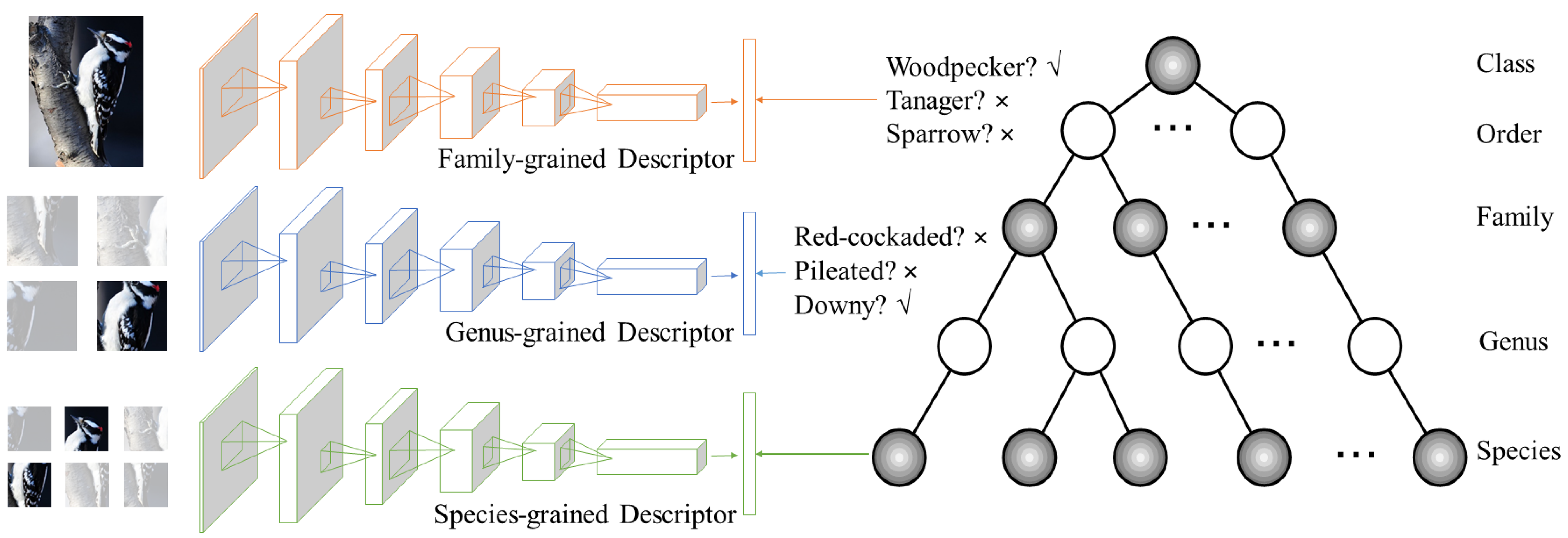


Our framework contains a parallel set of deep convolutional neural networks, each optimized to classify at a given granularity. In other words, it is composed of a set of single-grained descriptors.

Saliency in their hidden layers guides the selection of regions of interest (ROI) from a common pool of bottom-up proposed patches. ROI selection is therefore by definition granularity-dependent, in the sense that selected patches are results of the specific-grained classifier.

Meanwhile, ROI selections are also cross-granularity dependent: the ROIs of a more detailed granularity is typically sampled from those at the coarser granularities. This is built upon the intuition we discussed earlier, by emulating the process of multi-level attention.

Finally, per-granularity ROIs are fed into the second stage of the framework to extract per-granularity descriptors, which are then merged to give classification result.

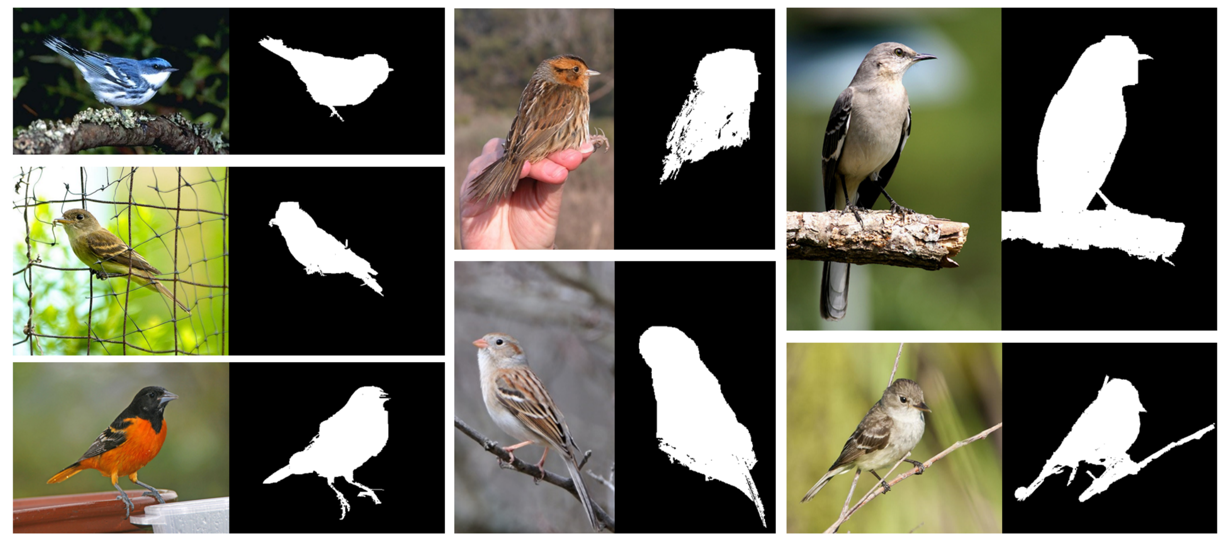


The top level of taxonomic tree, such as order-grained or family-grained labels, tells woodpecker from sparrow and tanager. The finer grained label, such as genus-grained or species-grained at the bottom level, provides more details about red-cockaded comparing to downy and pileated.

Contributions

1. We overcome the scarcity of labeled data by enriching a subordinate label with its ancestor labels in taxonomic hierarchy.
2. We derive a multi-grained learning framework that leverages hierarchical labels to generate comprehensive descriptors.
3. We propose a two-step fine-tuning mechanism consisting of salient region localization followed by classification of patches.

Experiments



Methods	Feature	BBox	Part	Oracle BBox	Oracle Part	Accuracy (%)
Zhang *et al.* [38]	KDES	√		√		51.0
Chai *et al.* [7]	Fisher	√		√		61.0
Gavves *et al.* [14]	Fisher	√		√		62.7
Zhang *et al.* [38]	KDES	√	√	√	√	64.5
Berg *et al.* [2]	POOF	√	√	√	√	73.3
Zhang *et al.* [36]	AlexNet	√	√			73.5
Branson *et al.* [5]	AlexNet	√	√			75.5
Zhang *et al.* [36]	AlexNet	√	√	√		76.7
Lin *et al.* [26]	AlexNet	√	√			80.3
Zhang *et al.* [36]	VGGNet	√	√			81.6
Krause *et al.* [22]	VGGNet	√				82.0
Zhang *et al.* [36]	VGGNet	√	√	√		85.0
Multi-grained	VGGNet	√				**83.0**
VGG-19[31]	VGGNet					67.0
Xiao *et al.* [35]	AlexNet					69.7
Xiao *et al.* [35]	VGGNet					77.9
Multi-grained	VGGNet					**81.7**

Table 1. Quantitative results on the CUB-200-2011 dataset [34] in comparison with state-of-the-art methods.

Methods	Annotation	Accuracy (%)
VGG-19[31]	BBox	62.3
Berg *et al.* [3]	BBox	66.6
Multi-grained	BBox	**74.8**
VGG-19[31]	None	51.7
Multi-grained	None	**65.9**

Table 2. Quantitative results on the Birdsnap dataset [3] in comparison with state-of-the-art methods.

Methods	Annotation	mA (%)
VGG-19[31]	BBox	63.2
Chai *et al.* [7]	BBox	75.8
Gosselin *et al.* [16]	BBox	81.5
Multi-grained	BBox	**86.6**
VGG-19[31]	None	56.6
Multi-grained	None	**82.5**

Table 3. Quantitative results on the Aircraft dataset [27] in comparison with state-of-the-art methods.

Methods	Annotation	Accuracy (%)
Single-grained	BBox	81.2
Double-grained	BBox	82.4
Multi-grained	BBox	83.0
Single-grained	None	79.5
Double-grained	None	81.0
Multi-grained	None	81.7
Detection CNN	BBox	77.3
Description CNN	BBox	81.2
Detection CNN	None	76.2
Description CNN	None	79.5

Table 4. Evaluation of individual components contributing to the overall performance on CUB-200-2011 dataset[34].

More Information

Visit Homepage!

http://goo.gl/SJLa15