From Categories to Subcategories: Large-scale Image Classification with Partial Class Label Refinement

Marko Ristin[1], Juergen Gall[2], Matthieu Guillaumin[1], Luc Van Gool[1,3]
[1]ETH Zurich [2]University of Bonn [3]KU Leuven

Problem setting. We are witnessing an exponential growth of digital imaging. The quanitity of data makes the semantic organization of images an imperative, but their manual categorization becomes excessively tedious at such a large scale. While automatic image classification has become accurate on smaller datasets, the research community has moved to more challenging, larger datasets, such as ImageNet [5] (thousands of categories and millions of images).

In such datasets, categories are often organized in a hierarchy. The deeper one descends in the hierarchy, the categories become finer and more numerous. Consequently, we dispose with fewer annotated samples per each one of them. In order to obtain training data for fine subcategories, a natural approach is to search for images of coarser categories and refine the labels. This can be expensive, especially if the subcategories require expert knowledge (*e.g.*, breeds of dogs, bird or flower species, etc.). In this work, we are interested in such a scenario where the training data comes in two subsets. One subset is annotated with fine subcategory labels, while the rest has only coarse category labels (*cf.* Fig. 1).

We investigate how the learning of subcategory classification can be improved by leveraging training data annotated with coarser labels. To this end, we build on the framework of NCM forests [3]. NCM Forests are multiclass classifiers that can be efficiently trained and have shown to perform well for large-scale image classification [3]. Moreover, each decision tree represents a hierarchy of classifiers such that the transfer of knowledge is implicitly effectuated [2].

Our contributions. Our first contribution is a principled approach to learn the hyper-parameters of the NCM forest during training. Namely, we regularize the number of class means used at each node by adding a regularization term to the objective function for training splitting functions.

The following contributions focus in more detail on the learning method that benefits from both coarse and fine-annotated training subsets. As a starting point, we adapt a state-of-the-art approach like Stacking [4] to our scenario. Next, as our main contribution, we introduce a novel model to boost the sharing of knowledge between coarse and fine categories by emphasizing their hierarchical relations in a new objective function.

Regularized NCM Forest. We base our methods on NCM Forests [3], *i.e.* Random Forests [1] which use the nearest class mean classifiers (NCM) as splitting functions at a node. During the training of each node, a pool of splitting functions is randomly generated and the one that maximizes the information gain U is selected. The conventional NCM Forests only sample the subset of class means with a predefined cardinality once, and then produce multiple random assignments of class means to children nodes. In contrast, we propose to vary the cardinality of the subset in the pool, and learn this parameter explicitly. By adding a regularization term based on the subset size to the objective function, splitting functions with fewer class means are preferred. Regularization not only caps the computational costs at the test time, but also yields better performance, as we empirically observed.

From categories to subcategories. In our scenario, two disjoint sets of training data S are provided: the first one, S_{coarse}, annotated with labels of *coarse* categories, and the second one, S_{fine}, that comes also with labels of *fine* subcategories. Our main goal is to classify images into fine classes by exploiting all the available data for which we consider the following strategies. **Stacking** the output of a forest trained to distinguish coarse classes to the input of the final forest trained for fine classes. **Joint** training of two independent classification tasks: the first task is to separate the coarse classes, and the second one is to separate the fine classes. The information gain U is

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

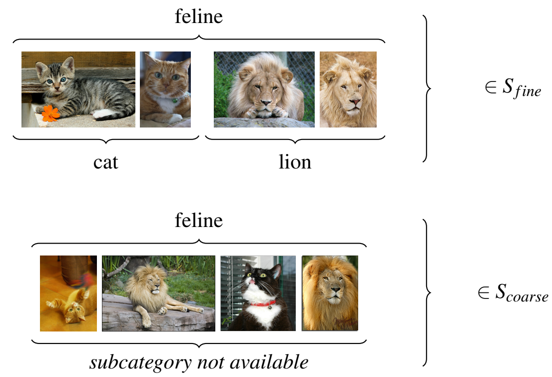

Figure 1: Our goal is to train the classification of fine subcategories. Yet, the refined labels are available only for a subset of the training data S_{fine}, while for the rest, S_{coarse}, only coarse category labels are available.

RNCMF (baseline)	12.04 (1.00)	14.35 (1.00)	16.96 (1.00)
Stacked	12.92 (1.07)	15.07 (1.05)	17.56 (1.04)
Joint	12.59 (1.05)	14.89 (1.04)	17.14 (1.01)
NN	13.99 (1.16)	15.48 (1.08)	17.93 (1.06)
Hierarchical	14.16 (1.18)	16.15 (1.13)	18.35 (1.08)
NN-H	**14.71 (1.22)**	**16.37 (1.14)**	**18.46 (1.09)**
$\lvert S_{\text{fine}}\rvert$	a) $0.1\lvert S\rvert$	b) $0.2\lvert S\rvert$	c) $0.5\lvert S\rvert$

Table 1: Average accuracy of our regularized NCM Forest trained on S_{fine} and the other methods trained on $S_{\text{fine}} \cup S_{\text{coarse}}$ with Bag-of-words features. The relative performance to the baseline is in brackets. We fix $\lvert S_{\text{coarse}}\rvert = 0.5\lvert S\rvert$ and set $\lvert S_{\text{fine}}\rvert$ to **a)** $0.1\lvert S\rvert$, **b)** $0.2\lvert S\rvert$ and **c)** $0.5\lvert S\rvert$. Our approaches improve the baseline, while the improvement is even more pronounced, when there are few fine-labeled samples. Our full model NN-H outperforms other methods.

optimized over both training sets disregarding their relations:

$$U^{\star}(f^n) = U_{\text{fine}}(f^n) + \lambda U_{\text{coarse}}(f^n) + \textit{regularization}. \tag{1}$$

Nearest neighbour (NN) approach to refine the coarse label of a sample in S_{coarse} based on the nearest sample within S_{fine}. The coarse labels of S_{coarse} are then replaced by the refined ones and we optimize U using only the labels of the finer categories. **Hierarchical** approach which derives for each sample in S_{fine} a category label based on its subcategory. Thus U_{coarse} is not computed only over the data $S^n \cap S_{\text{coarse}}$ with coarse-only labels, but over the entire S^n. In this way, a classification error of a subcategory in S_{fine} is more penalized if the labels also disagree at the coarser level. Our full model, **NN-H**, a combination of the previous two approaches, such that each sample is now assigned a fine as well as a coarse label, which we use to optimize 1. In this way, our model both respects the class hierarchy and enforces that coarsely labeled data match one of the finer subcategories.

Results. We perform a series of experiments on a subset of classes from ILSVRC10 [5] using different features (Bag-of-words, Fisher vectors, CNNs). Our full model successfully leverages the extra coarse-labeled data and outperforms the other models as well as the state-of-the-art. Compared to approaches trained solely with fine-labeled data, we achieve a relative improvement in subcategory classification accuracy of up to 22%, *cf.* Tab. 1.

[1] L. Breiman, Random Forests. *Machine Learning*, 45(1), 2001
[2] N. Razavi *et al.*, Scalable multi-class object detection. In *CVPR*, 2011.
[3] M. Ristin *et al.*, Incremental learning of NCM Forests for large-scale image classification. In *CVPR*, 2014.
[4] M. Rohrbach *et al.*, Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In *CVPR*, 2011.
[5] O. Russakovsky *et al.*, ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575*