

指导教师: 杨涛

提交时间: 2013/3/20

## CVPR2015 Paper Translation



No: 01

姓名: 贾双江

学号: 2013302479

班号: 10011301

## 从类到类：用分部类标号细化大型图像分类

Marko Ristin<sup>1</sup> Juergen Gall<sup>2</sup> Matthieu Guillaumin<sup>1</sup> Luc Van Gool<sup>1,3</sup> 1ETH Zurich 2University of Bonn 3KU Leuven

### 摘要

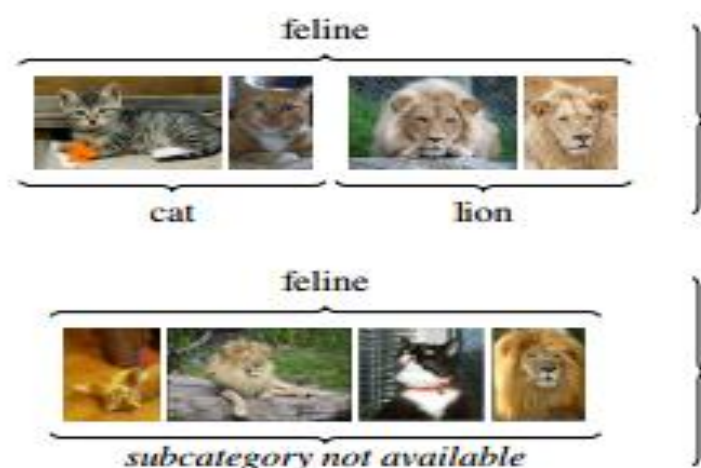
数字图像的数量正在迅速增长，因此需要对其进行分类。但是，随着越来越多的图像预先定义的类别可用，他们也变得更加多样化同时也覆盖更精细的语义差异。最终，类别本身需要分类别因为语义重新细化。在过去几年中，图像分类一般有明显的改善，但它仍然需要大量的手工注释数据。细分类到小类的倍数的标签数量，加重标注问题。因此，我们可以对预期的注释进行细化，只用于已标记的数据的一个子集，并利用粗糙的标记的数据，以提高分类。在这项工作中，我们研究了粗分类标签可以用来提高小类的分类。为此，我们采用随机森林的框架，提出了一个正规化的目标函数，考虑到类别和子类别的帐户之间的关系。方法相比不特粗的标记数据，我们在我们的大规模图像分类实验类的分类精度有了较大的改进实现了 22% 的增长。

### 简介

在过去的几年里，我们见证了数字成像的指数增长。可能是通过社交媒体共享视觉内容，如 Facebook 和 Flickr 和负担得起的高品质摄像头让图像在我们的生活中无处不在。因此，在他们在一个有明确的目标下如果一个人想要毫不费力地访问图像，图像的语义组织成为一个必要的。对这样的增长，手工分类是非常繁琐和昂贵的。随着计算机变得更强大和更好的算法，图像自动分类技术已成为精确的分类小数据集的类别和数十万图片。

研究团体已经从此面临更大的挑战，更大的数据集，如 ImageNet [ 28 ]，其中包含的类别和数以百万计的图像。在这样的数据集，类别往往是组织的层次结构。更深层次的层次，更精细的类别和注释的训练数据变得罕见。为了获得好的子类别的训练数据，

一个自然的方法是寻找大类图像细化的标签。这可以是非常昂贵的，特别是如果子类需要专家知识（如狗、鸟、花卉等品种）。在这项工作中，我们有兴趣在这种情况下训练这种只有训练数据的一个子集注释与细子标签，而剩下的只有粗分类标签。



在这种情况下，我们特别感兴趣的是对亚分类学习可以通过训练数据注释标签同时改善粗糙。为此，我们建立在 NCM 森林的框架，这是基于最近的类平均分裂功能分类的森林（NCM）分类。NCM 的森林多类分类实验，可以有效地训练和显示每一个形成大规模图像分类 [ 26 ]。我们的第一个贡献是自动在培训学习的 NCM 林超参数的最佳值的原则性方法，即类的数量是指利用各节点。这是通过增加一个正则化项的目标函数的训练分裂的功能。我们在第 3 节提出这一改进。在我们的实验中，我们证明了我们的正则化 NCM 树优于 NCM 树。

在 4 节中，我们提出的学习方法更详细。我们展示了如何分类的最先进的方法，如堆叠可以适用于我们的场景，然后描述我们所提出的模型，用于共享知识之间的粗和细类别。我们的方法的关键是引入一个目标函数，考虑到类别和子类别的层次关系。

正如我们在第 5 节中的实验，我们所提出的方法在训练森林与一个额外的训练集合中包含的类别标签，分类精度相对提高达 22%。我们优于也训练了这种额外的训练集的其他方法，包括 27 的叠加方法，并达到 99% 的性能，达到了一个基线训练与完整的监督在整个训练集与精致的标签。最后，我们表明，一个额外的改进，可以通过结合森林与公制学习方法提出的。

## 相关工作

大型数据集上的图像分类在

最近几年得到了很大的关注[ 2, 13, 28 ]。而深卷积网络（美国有线电视新闻网）实现高分类准确率[ 18 ]，他们是计算密集型的，需要几个星期的训练。简单的分类，如最近的类平均分类器（NCM）结合学指标[ 23 ]，被证明是一个可行的替代方案，更短的运行时间和接近零成本整合新课。在[ 26 ]，新的被集成在一个随机森林结构[ 9 ]，增加的性能和功能进行了大规模的快速增量学习。在这项工作中，我们解决的问题是学习一个分类的最优秀的类别级别时，只有一部分的训练数据是注释的，而其他训练样本只有一个粗糙的水平标签。这与半监督或转移学习的方法有关。

在图像分类中的知识转移，减少所需的标记的数据，以学习新的图像类，利用现有的其他类的标签。一个简单的方法是扩展的训练数据与未标记的样本，其注释是诱导或基于邻域结构中的特征空间或类别之间的语义距离。基于支持向量的方法，从语义上相关的类的样本可以在学习过程中用来作为约束和正则化。如果模型是基于形状，功能或部分可以共享。类之间的关系也可以被配制成一个神经网络的约束，或作为一个概率框架

的一部分，也可以基于属性相似性或语言知识库[。类层次结构也可以隐式学会分享模型参数之间的父亲和孩子结或通过层叠分类器的分类速度。

类之间的关系，也可用于建立各种级别的类层次结构的分类器。在假设上较细级别的类是比类分类更困难在一个较高的水平，这将导致准确性和特异性之间的折衷。这种其中特异性通过给定的类层次结构中定义的折衷在被处理，而它的作者试图基于口语人类反应以另外学特异性。图像标签还可以与外观，位置分布和上下文一起利用来注释一个大的数据通过附加的注释设置如边界框或分割掩码。

随机树，它已被用于许多计算机视觉应用中，包括图像分类也已在半监督或转移学习的上下文中使用。它的作者提出用图像标签作为regularizers 为叶统计信息和树结构来揭开图像分割掩码。当数据采样每而不是单个实例的包装袋标记，随机森林已被改编为多个实例学习那里的树木运用确定性退火]训练。从弱相关视频了解到时空一致性在[20]用于改善对象检测器。准确的标签的训练数据能提高与数据缺失或嘈杂标签。这可以

通过最大化的加权信息增益而这需要训练数据的两个来源考虑在内来实现。在这项工作中，我们不考虑额外的数据嘈杂或弱相关，只利用类别和子类别之间的关系，以提高子类别的分类精度。

### 正则 NCM 森林

西北中心的树已被证明是大型图像分类提供培训的时间和精度之间的良好折衷，我们首先简要讨论在 3.1 节 NCM 森林。在 3.2 节中，我们提出了培训 NCM 森林一个新的目标函数，它考虑到了信息增益和分离功能的计算成本来自动设置 NCM 森林的重要参数。在我们的实验中，我们表明，这种修改可以提高性能。在第 4 节中，我们然后进行演示如何仅训练被细化标签的一小部分使他们的分类精度可以提高。

在一个大规模的设置分类 NCM 森林属于家庭随机森林[9]，其结合组成的  $\check{T}$  决策树以分布式方式独立地训练的分类的。每棵树  $t$  被递归从根节点开始，上，其中每个图像由  $d$  dimensional 特征向量  $\tilde{x} \in \mathbb{R}^D$  所表示的标记的像  $S$ 。在节点  $n$  传入数据  $S_n$  由一个分

裂函数  $F_N$  分成两个不相交的集合， $\mathbb{R}^D \rightarrow \{0, 1\}$ 。两套， $S_n$  中  $F=0$  且  $S_n$  中  $F=1$ ，然后被传递到节点，分别的左和右子树，训练继续递归地为每个子。在 NCM 森林，分裂函数基于最近的类均值分类器 (NCM) [23]。这包括在计算  $\kappa \in K_n$  的  $C_K$  的数据  $\{S_n$  中  $\kappa\}$  到达节点  $n$  中观察到的类的子集的装置  $\{C_{n\kappa}\}$ ：

每个类意味着为  $C_{n\kappa}$  是由一个二进制值  $e_{\kappa} \in \{0, 1\}$  的装置分配给两个子节点中的一个  $\{0, 1\}$ 。因此，该分离功能被定义者： $F_N(\tilde{x}) = e_{\kappa} * (\tilde{x})$ ，其中  $\kappa * (\tilde{x}) = \arg \min_{\kappa \in K_n} \|\tilde{x} - \bar{x}_{\kappa}\|$ 。 (2) 换句话说，最靠近类  $\kappa$  的平均数据点被传递到基于  $e_{\kappa}$  左侧或右侧孩子。

对于训练，通过随机抽样的类  $K_n$  的和二进制值  $\{e_{\kappa}\}_{\kappa \in K_n}$  所述子集生成分裂函数  $F_n$  的池。分裂函数  $f$  的  $n \in FN$  最大化的信息增益则  $U$  被选择和存储在节点  $n$ ：

其中  $H$  是类熵和  $P(\kappa | S_n)$   $S_n$  中内的类  $\kappa$  的经验概率。训练结束时，对所选择的分离功能， $|S_n$  中  $F_N=0| \leq \mu$  或  $|S_n$  中  $F_N=1| \leq \mu$ 。然后，节点  $n$  被转换成叶  $L$  和观察到的类分布  $P_L(\kappa)$  被存

储。值得注意的是， $P_L(\kappa)$  是基于至少  $\mu$  样品计算的。

对于分类，图像  $\tilde{x}$  被通过每个树  $T$ ，直到它到达一个叶子  $L_T$  传递 ( $\tilde{x}$ ) 的。然后，单株响应平均，最可能的类  $\kappa^*(\tilde{x})$  返回作为输出：

### 3.2 正则 NCM 森林

正如[26]所示，NCM 林的训练和测试时间在很大程度上取决于多少级。这意味着  $|K_n|$ 。较大的  $|K_n|$ ，图像特征  $\tilde{x}$  和类装置之间的更多的比较  $\{C_{n\kappa}\}_{\kappa \in K_n}$  需要计算。据计算， $|K_n| = P|K|$  结果在效率和准确性之间的良好折衷。但是，固定  $|K_n| = P|K|$  对于树的所有节点可能不是最佳的。事实上，人们期望类的最优数量实际上意味着取决于到达特定节点的训练数据。因此，我们建议使用可变大小的样品拆分功能  $|K_n|$  并使用有利于信息增益和惩罚大计算量的目标函数。要做到这一点，我们添加一个稀疏术语(3)：

其中， $\lambda_{reg}$  是加权参数。在实践中，我们品尝  $K_n$  的满足  $2 \leq |K_n| \leq c|P|K|$ 。值得注意的是，

如果  $c=1$ ，我们用  $P|K|$  作为上限为多少级意味着  $|K_n|$  的分离功能  $F_N$ 。注意， $\lambda_{reg}$  微调的在节点分裂功能的选择，而参数  $c$  只影响采样过程。我们用正则 NCM 森林 (RNCMF) 的 NCM 森林与 (5) 培训。在第 5 节，我们证明了正则 NCM 跑赢 NCMFs [26]。

### 从分类到小类

如前所述，我们认为，我们正在提供两个不相交组训练数据  $S$  的第一组， $S_{coarse}$ ，都标注有粗类  $K_{coarse}$  (如“犬”，“猫”) 的标签的场景，而第二组  $S_{fine}$  还附带了细化粗那些子类别的标签，即细类  $K_{fine}$  (如“狼来了”，“狮子”)。我们系统的主要目的是通过利用所有可用的数据进行分类的图像分成细类。这种情况下，可以看作是在几个级别监督学习的一个特例。首先，我们可以考虑粗，细类作为完全独立的 (第 4.1 和 4.2)。我们还可以看到标有粗类别作为一个不确定的观测或部分缺失标签 (第 4.3 节) 的数据点。最后，因为我们知道在设计类之间的关系，我们还可以看到粗到细的环节上的知识转移 (第 4.4 和 4.5) 编码的约束。

#### 4.1. 堆叠 RNCMF

在对知识共享的方法[27]认为，基于堆叠 SVM 分类之一取得了显著的权衡在简单和性能之间。在这样的设置中，一组源分类器的输出被连接到输入特性为学习的目标类的 classifiers。在我们的分层场景中，我们使用粗类作为源分类以及更精细的人作为目标。堆叠策略也容易适应 RNCM 森林。我们培养的 Scoarse 第一 RNCM 森林 T1 粗类分类。Sfine 的样品，然后通过 T1 推压而得到的概率  $P$  比粗类归一 ( $P_0 = \alpha P$ ) 和级联到特征向量  $\sim X$ ，得到的新的特征向量  $\sim X_1 = (\sim X, P_0(\kappa 1), P_0(\kappa | K_{coarse}|))$ 。

第二个森林 T2 然后在 Sfine 培训，以区分使用增强的功能  $\sim X_1$  精类。在测试时间，图像遵循相同的过程。首先，它被送入 T1 到获得  $\sim X_1$ ，然后通过 T2 被推到得到在细类的概率。堆叠具有的缺点是两个森林需要训练和评估分类，并且其性能在很大程度上取决于源分类器的质量。相反，我们建议在 4.4 节的学习单 RNCM 森林与分层的类信息的新的培训方法。

#### 4.2. 联合 RNCMF

类似于[10]，我们考虑一个扁平标签集，我们简单地尝试学习一个 RNCMF 共同执行两个独立分类任务的情况下：第一任务是分开的粗类，第二个是分开的细类。因此，我们表示这样的森林作为联合 RNCMF，或 J-RNCMF。为了培养出 J-RNCMF 上 Scoarse 和 Sfine，正则化信息增益 (5) 可以计算为既有收获的总和，每个人对自己的类级别操作：

换句话说，我们考虑的类 Kfine 和 Kcoarse 是无关，并且仅更喜欢在平行分类以及两组类分裂功能。因为 Kfine 和 Kcoarse 样品商都存在于一个节点，类装置可以上计算这两组的类。类的手段更丰富的提高性能，我们将在第 5 节看到的，并允许我们以直接的方式整合粗类的额外的知识。

#### 4.3. NN-RNCMF

[14]类似的，在 Scoarse 各样品的标记物可以基于简单的分类器被细化，如最近邻 (NN) 上训练 Sfine。虽然这种方法是作为堆叠作为通用的，它可以很容易地适应利用我们的分层设置。这是通过对在 Scoarse 样品的最邻近搜索约束只在 Sfine 其相应的子类别的样本。Scoarse 的粗标签然后由精制所取代和 RNCMF 可以 Scoarse 仅使用更细的类别标签进行训练，精

$U_{S_{fine}}$ 。我们把这种方法称为 NN-RNCMF。

#### 4.4. 分层 RNCMF

上述联合 RNCMF 方法忽略类别  $K_{coarse}$  和子类别  $K_{fine}$  之间的关系，通过处理层次结构的层次作为两个独立的分类任务。作为事实上，我们可以轻松地添加一个层次的味道 J-RNCMF。我们得出在  $S_{fine}$  的每个样品按层次结构是根据它的子类别的类别标签。使用该附加信息，仅在术语  $U_{coarse}$  (6) 需要被重写为：

公式。。。

即  $U_{coarse}$  不仅是在数据  $S_n$  中  $\cap$   $S_{coarse}$  粗类中计算的标签，而且在整个  $S_n$  中。

以这种方式，在  $S_{fine}$  子类别的分类错误，如果错误地预测的子类别属于同一类别的真子类别，或者，换言之，更惩罚标签较粗级别不同的较少缺陷。我们把这种模型分层 RNCMF，或 H-RNCMF。

4.5. 我们的全部型号：NN-H-RNCMF  
NN-RNCMF 和 H-RNCMF 是互补的，可以组合。事实上，尽管 NN-RNCMF 提供有粗有细标签标记的数据，H-RNCMF 本质上是对面，利用了精细的标签数据派生粗标签。这导致我们的整个模型，其中我们称之为 NN-H-RNCMF。首先，

$S_{coarse}$  的训练样本被分配在  $S_{fine}$  最接近的邻居的标签为 NN-RNCMF。因此， $S_n$  中每个样品现已被分配一个类别标签  $\kappa_{coarse} \in K_{coarse}$  以及  $\kappa_{fine} \in K_{fine}$  其提炼  $\kappa_{coarse}$ 。对于这样的学习 NN-H-RNCMF，(6) 通过使用  $U_{coarse}$  如在改性 (7) 和计算  $U_{fine}$  超过  $S_n$  中使用估计的细标签  $S_{coarse}$ ：

以这种方式，我们的模型利用两者的类层次结构和该粗标记的数据应该匹配一个观察的更细的子类别。

### 实验

我们于 2010 年 05 ILSVRC[28]，一个完善的和具有挑战性的用于大规模图像分类数据集的子集执行我们的实验。2010 ILSVRC 的层次结构被给定为基于 WordNet 有向非循环图（每个节点代表一个类，也称为同义词集合）。在这项工作中，我们着重对具有独特的父类的类。更确切地说，我们收集了 2010 ILSVRC 的所有叶子同义词集作为我们的子类别和他们的父母类别和重叠被完全忽略类子树。其余父同义词集构成了我们的粗分类  $K_{coarse}$ ，而其对应的子女代表的子类别  $K_{fine}$ 。因此，我们得到  $|K_{coarse}|=143$  粗类和  $|K_{fine}|=387$  精细的。



2010 ILSVRC 的原始训练, 验证和测试集减少到  $K_{fine}$ 。降低训练集由我们有图像 1.4K 和 9.8K 之间的每个粗级和 668 之间 2.4K 图像的每个细类 487K 的图像。有每个细类 50 和 150 的图像分别在验证和测试集。如果没有另外说明, 性能如顶-1 平均准确度  $K_{fine}$  测量。

因为在我们的场景中, 我们假设子类别标签只能用于训练数据的一个子集, 我们随机在两个分离的分割训练数据  $S$  为每个细类。而第一组  $S_{coarse}$  只有类别标签, 第二组  $S_{fine}$  包括子类别的标签为好。如果没有另外说明,  $|S_{fine}| = |S_{coarse}| = 0.5|S|$ 。类的详细清单的补充材料和培训作为考核的一部分是测试图像上线发布的。

在我们的实验中, 我们使用公开可用  $1kdim$  的视觉字这是由[5]提供了一种基于密集采样 SIFT 特征 (弓)。该特征基于训练数据白化。

### 5.1. 正则 NCM 森林

我们评估的验证集所有参数

的影响。如果没有另外说明, 我们按照 [26], 并在叶执行至少  $\mu = 10$  个样品, 并限制  $|K_n| \leq |K|$ 。森林包括 50 棵树。除了准确性, 我们还测量由平均每棵树和图像比较的测试运行时, 如 [26]。

首先, 我们研究的是被我们 RNCMF 参数的影响下对完整的训练集  $S$  练就精类分类。RNCMF 的性能取决于分裂函数  $F$  训练的  $n$  个生成的数目。为此目的, 我们产生 10 和 1000 之间的类设置为  $K_n$  的和, 对于每个组, 我们在 10 和 100 之间进行采样

表 1: 对细类分类  $K_{fine}$  测试集的平均精确度。拟议的转正 NCMFs 和 NCMFs 和其他方法。使用度量学习 (MET) 如在 [23] 进一步提高精度。

分配结构  $\{e_k, k \in K_n\}$ 。这两个参数的结果示于图图 2a) 和 b)。在这些实验中, 默认值是 1000 套, 50 分配和  $\lambda_{reg} = 0.01$  (5)。图。图 2a) 示出了, 如果更多组  $K_n$  的被采样的准确性增加并且测试运行时减小。试验运行时的降低是由该调整装置引起的。随机任务数量  $\{e_k, k \in K_n\}$  没有显著影响。图。图 2c) 示出了稀疏术语的在 (5) 的影响。对于  $\lambda_{reg}$  较大的值降低测试运行时, 还分类精度。我们固

定  $\lambda_{reg} = 0.001$ ，因为它降低了测试运行，而不损失精度。

在表 1 中，我们比较 RNCMF 与 NCMF [26] 等多类分类器，即 NCM, k 近邻，和多分类 SVM [3]。对于所有的方法，参数的验证集进行了优化。评价是在测试集执行。结果表明，所提出 RNCMF 优于 NCMF 和其他方法。此外，我们还培养了距离度量 MET 在 [23] 训练集，并将其应用到的功能，从而降低他们的空间的维度 512 和额外提高性能，比照。表 1b)。

### 5.2. 从分类到小类

为了评价的第 4 节中所讨论的方法的性能，我们使用训练集  $S_{fine}$  和  $S_{coarse}$  以及第一评价对验证集这些方法的参数。

#### 5.2.1 参数

堆积 RNCMF。叠 RNCMF 由两个林，其中第二森林的培训和应用上的图像的特点和在类别的概率  $K_{coarse}$  由第一森林估计。由于  $|K_{coarse}| = 143$ ，功能总共有 1143 尺寸的第二个林中。该概率可被归一化，以提高性能。我们评估归一化系数  $\alpha \in \{0, 1, 2, 3\}$ ， $\alpha = 10$  导致了最佳的性能。

J-RNCMF 和 H-RNCMF。两者的 J-RNCMF 和 HRNCMF 取决于加权参数  $\lambda$ ，其中操纵  $U_{coarse}$  在冲击 (6)。图 3 示出的精确度

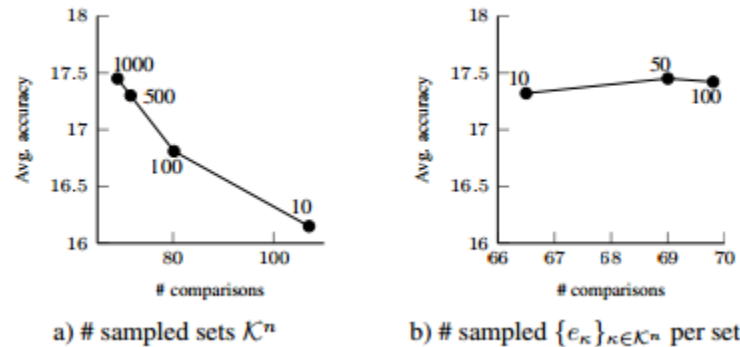


图 2: a) 的影响) 产生的等级的平均子集  $K_n$  的数量 (10, 100, 500, 1000), b) 该生成的分配  $\{e_{\kappa}\}_{\kappa \in K_n}$  每套 (10, 50  $\kappa \in K_n$  的) 数目, 100, 和三) 的加权参数  $\lambda_{reg}$  (0.1, 0.01, 0.005, 0.001, 0) 中的正则化项的 (5) 上的精度和测试时间效率, 通过平均每个树和图像比较的测量。同时将产生的任务的数目不是一个非常敏感的参数, 生成类装置的许多子集, 并通过正则宁愿较小的子集同时提高效率在测试时间和分类精度。

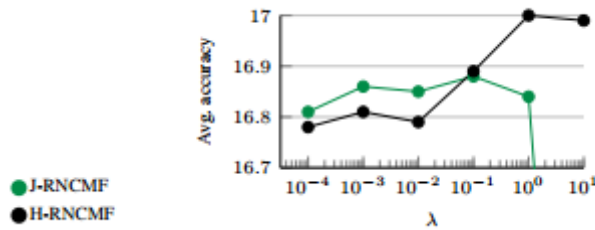


图 3: 加权因子  $\lambda$  其中操纵  $U_{coarse}$  的冲击的影响。H-RNCMF 优于 J-RNCMF, 因为它也认为粗细类之间的等级关系。

$\lambda$  的不同的值。由于  $U_{coarse}$  不同的 J-RNCMF 和 H-RNCMF, 最佳  $\lambda$  不同了。而 J-RNCMF 实现为  $\lambda = 0.1$  的最佳精度 (16.86), H-RNCMF 达到最佳的性能 (17.00) 用  $\lambda = 1.0$  和优于 J-RNCMF, 其中突出类关系的重要性。在这个实验中, 我们只从  $K_{fine}$  品尝  $K_n$  的。然而, 从  $K_{coarse} \cup K_{fine}$  采样  $K_n$  的当平均精度为 H-RNCMF 进一步增加至 18.10。在下文中, 我们使用  $\lambda = 0.1$  J-RNCMF,  $\lambda = 1.0$  的 H-RNCMF 和 NN-H-RNCMF, 和从  $K_{coarse} \cup K_{fine}$  采样  $K_n$  的。

NN-RNCMF。我们可以以各种方式预测  $S_{coarse}$  的缺少精细的标签。一种可能性是在训练的  $S_{fine}$  RNCMF, 并用它来预测  $S_{coarse}$  精细的标签。这样就可以得到 14.02 的平均准确度。我们还可以使用  $S_{fine}$  最

近邻居的罚款标签, 忽视了阶级关系, 达到 14.23 更好的精度。如果我们放弃近邻匹配的是子类别, 类别不匹配, 精度提高到 17.03。当近邻只对应子类别的样品中搜索的最佳精度 (17.52) 的实现。这表明, 类别和子类别之间的关系也可用于最近邻方法很重要。我们使用第三种策略在下面。丢弃基于在特征空间中的距离最近的邻居匹配不会任何进一步改善性能。

i)	NCM	7.14	14.49	16.30
	k-NN	9.45	12.33	12.58
	Multiclass SVM [3]	17.22	16.24	16.41
	RNCMF (baseline)	16.96	17.43	18.54
ii)	Stacked RNCMF	17.56	17.80	18.83
	J-RNCMF	17.14	17.44	18.39
	NN-RNCMF	17.93	18.26	19.29
	H-RNCMF	18.35	17.93	19.20
	NN-H-RNCMF	<b>18.46</b>	<b>18.48</b>	<b>19.55</b>
		a)	b)	c)

表 2: 准确的比较时) 没有指标, B) 指标  $MET_{coarse}$  上  $S_{coarse}$  了解到, 和 c) 指标  $MET_{fine}$  上  $S_{fine}$  学会应用。在我的方法) 只培训了  $S_{fine}$ , 而在第二方法) 都在  $S_{fine}$  和  $S_{coarse}$  培训。我们完整的模型 (NN-H-RNCMF) 优于基准和其他方法, 达到精度接近使用整个数据如表 1 报告。

运行。在 RNCM 森林的培训花费 800

内核的群集上约 2 小时。发现从 Scoarse 在 Sfine 样品的最近的邻居花费一小时。通过一棵树的图像的分类需要大约  $50 \mu s$  的为 1K 维预先计算的功能。

RNCMF (baseline)	12.04 (1.00)	14.35 (1.00)	16.96 (1.00)
Stacked RNCMF	12.92 (1.07)	15.07 (1.05)	17.56 (1.04)
J-RNCMF	12.59 (1.05)	14.89 (1.04)	17.14 (1.03)
NN-RNCMF	13.99 (1.16)	15.48 (1.08)	17.93 (1.06)
H-RNCMF	14.16 (1.18)	16.15 (1.13)	18.35 (1.08)
NN-H-RNCMF	<b>14.71 (1.22)</b>	<b>16.37 (1.14)</b>	<b>18.46 (1.09)</b>
$ S_{fine} $	a) 0.1 $ S $	b) 0.2 $ S $	c) 0.5 $ S $

表 3: 作为基线 RNCMF 平均准确性上训练 Sfine 和训练有素的上 Sfine  $\cup$  Scoarse 的其他方法。到基线的相对性能是在括号中。我们固定  $|S_{coarse}| = 0.5|S|$  并设置  $|S_{fine}|$  到) 0.1 $|S|$ , B) 0.2 $|S|$  和 c) 0.5 $|S|$ 。我们的方法提高基准, 即使  $|S_{fine}|$  和  $|S_{coarse}|$  不平衡。的改善更为显着, 当有几个细标记的样品。我们 NN-H-RNCMF 优于其他方法。

### 5. 2. 2 比较

现在, 我们评估在第 4 节提出的方法, 这是两个 Sfine 和 Scoarse 培训, 并把它们比作是受过训练的只有 Sfine 显示额外的训练数据 Scoarse 的利益而 RNCMF。结果列于表 2 中。因为我们已经看到, 度量学习提高了准确性, 我们评估两

种情况。在第一种情况下, 度量 METcoarse 被训练上 Scoarse 和在第二种情况下 METfine 上 Sfine 训练。表 2 i) 表明 METcoarse 略微提高了精度, 但是 METfine 更加提高它是最好的选择。表 2 ii) 表示, 该办法忽视阶级关系 (叠放, J) 提高 Sfine 训练的只有轻微的 (基线) RNCMF, 同时利用层次结构 (NN, H, NN-H) 的方法跑赢基准和其他方法。我们完全模型 NN-H-RNCMF 与 19.55 的精度性能最佳。值得注意的是, 这几乎与 RNCMF 与全程监督, 即培训了 S 采用子类别标注为所有训练样本, CF 的性能。表 1b), 用细标签减少 50%。

### 5. 2. 3 培训尺寸大小的影响

到目前为止, 我们已经使用  $|S_{coarse}| = |S_{fine}| = 0.5|S|$ , 但 obviously 那些集的相对大小关系。因此, 我们现在评估 Scoarse 和 Sfine 训练大小的影响。我们先留着  $|S_{coarse}|$  固定的, 但有所不同  $|S_{fine}| \in \{0.1|S|, 0.2|S|, 0.5|S|\}$  观察方法是如何应付一些精细标记的数据时  $|S_{fine}|$  和  $|S_{coarse}|$  不平衡。作为基准, 我们训练 RNCMF 上 Sfine 并没有使用公制。如表 3 所示, 精度提高了所有方法时  $|S_{fine}|$  增加。当有一些细

微标记的样品，基线的改善变得更大。对于  $|S_{fine}| = 0.1|S|$ ，我们完全模型 NN-RNCMF 提高从基线到 12.0414.71，22% 的相对增加。精美的标签是凤毛麟角，并获得比粗更贵，这是相关的。在第二个实验中，我们固定  $|S_{fine}| = 0.5|S|$  并改变  $|S_{coarse}|$ 。结果示于表 4 中的所有方法从较大的训练集获益  $|S_{coarse}|$  粗标签。

	a) 0.1 S	b) 0.2 S	c) 0.5 S
RNCMF (baseline)	16.96 (1.00)	16.96 (1.00)	16.96 (1.00)
Stacked RNCMF	17.18 (1.01)	17.21 (1.01)	17.56 (1.04)
J-RNCMF	16.89 (1.00)	16.80 (0.99)	17.14 (1.01)
NN-RNCMF	17.27 (1.02)	17.51 (1.03)	17.93 (1.06)
H-RNCMF	17.59 (1.04)	17.90 (1.06)	18.35 (1.08)
NN-H-RNCMF	<b>17.87 (1.05)</b>	<b>18.13 (1.07)</b>	<b>18.46 (1.09)</b>

表 4: 作为基线 RNCMF 平均准确性上训练  $S_{fine}$  和训练有素的上  $S_{fine} \cup S_{coarse}$  的其他方法。到基线的相对性能是在括号中。我们固定  $|S_{fine}| = 0.5|S|$  并设置  $|S_{coarse}|$  到 a)  $0.1|S|$ , b)  $0.2|S|$  和 c)  $0.5|S|$ 。精度粗标记的数据的数量增加，并且提高我们的 NN-H-RNCMF 最基线的性能。

RNCMF (baseline)	24.54 (1.00)
Stacked RNCMF	22.35 (0.91)
J-RNCMF	21.99 (0.90)
NN-RNCMF	24.19 (0.99)
H-RNCMF	23.95 (0.98)
NN-H-RNCMF	<b>25.15 (1.02)</b>

表 5: 平均准确的粗分类类别比较  $K_{coarse}$ 。基线是  $S$  上训练  $K_{coarse}$  的分类，而其他车型进行培训，以精细分类的类  $K_{fine}$ 。到基线的相对性能是在括号中。实验是在测试，每类粗 150 图像设定进行。公制学习不使用，我们固定  $|S_{coarse}| = |S_{fine}| = 0.5|S|$ 。我们 NN-HRNCMF 表现最佳，甚至优于基线由于子类别的更多的知识。的 22% 的相对增加。精美的标签是凤毛麟角，并获得比粗更贵，这是相关的。在第二个实验中，我们固定  $|S_{fine}| = 0.5|S|$  并改变  $|S_{coarse}|$ 。结果示于表 4 中的所有方法从较大的训练集获益  $|S_{coarse}|$  粗标签

### 5.3. 类别分类

- A)  $|K^n| \leq \sqrt{|K|}$
- B)  $|K^n| \leq 5\sqrt{|K|}$
- C)  $|K^n| \leq 10\sqrt{|K|}$
- D)  $|K^n| \leq 20\sqrt{|K|}$

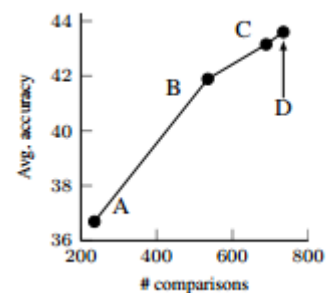


图 4: 上限的冲击  $|K^n| \leq c p |K|$ 。报告在通过比较 (每棵树和

图像)的数量来衡量测试时间平均准确率和效率。实验是在用 Fisher 向量验证组执行和度量 METfine 上 Sfine 教训。

虽然这不是我们最初的目标,我们的子分类,也可用于对类别进行分类。这可以被看作是利用附加信息用于学习类的分类器的方法。为此,我们只需要储存在叶子 Kfine 到 Kcoarse 类的概率转换。到获得合适的,平衡测试组,我们随机子采样原始测试设置为包含每个类别 150 的图像,并测量其上平均准确性。我们设置  $|S_{coarse}| = |S_{fine}| = 0.5 |S|$  没有使用任何指标。表 5 给出了其中利用层次结构的方法 (NN, HNN-H) 达到基准的准确性和超越其忽略等级关系 (堆叠, J) 的其他方法。值得注意的是, NN-H-RNCMF 优于基准,显示出子类别的训练数据的 50% 的额外知识有助于提高类别的分类,我们的模型很好地利用这些额外的信息。

#### 5.4. 高维特征

费舍尔向量高维特征表明,在

图像分类有良好的性能。因此,我们也评估我们使用机器学习同费舍尔载体子分类,而不是低头功能在[23]的方法。度量 METfine 进行训练与 Sfine 费舍尔载体。与此相反的功能, RNCMF 的精度可进一步提高的上限为类的数目在每个节点装置的采样改善  $|K_n| \leq c$   $p|K|$  增加。我们培养森林与  $C \in \{1, 5, 10, 20\}$  上 Sfine, 随着  $|S_{fine}| = 0.5 |S|$ , 并在验证组对其进行评估。图的结果。4 表明,增加的上限  $|K_n|$  提高了精度。在下面的实验中,我们用  $c = 10$ , 达到接近最优的性能与测试时间复杂度的增益。

最后,我们用我们的方法比较正则的 theart 的 NCM[23]。如第 5.1 和 5.2, 我们优化的验证集为 RNCMF, H-RNCMF, NN-RNCMF 和 NN-H-RNCMF 的参数。参数不只是  $K_n$  的改变被采样,每个节点 100 倍,而不是 1000 倍和加权参数  $\lambda$  (6), 它操纵  $U_{coarse}$  的影响,增加至 100 有关的测试数据的评价,粗的量标记的数据被固定为  $|S_{coarse}| = 0.5 |S|$ , 而  $|S_{fine}|$  而变化。我们应用培训了相应的 Sfine 度量 METfine。结果示于表 6 中相反的功能,最近邻居表现不佳,并减少 NN-RNCMF 和 NN-H-RNCMF 的性能。

这与[23]其中，表示出了一个度量高维特征到不一定与最近的邻居接近一致。对于高维特征，H-RNCMF 因此性能最佳。

我们也进行了初步的实验与 4096-CNN 的功能[31]。由于功能上细分类 2012 ILSVRC 的预训练的，结果是不能直接媲美的其它实验。RNCMF 如表 1 中实现了 74.18 的精度和优于多分类 SVM[3] (71.67) 和 NCM (66.02)。使用的协议，如表 6 中，NN-H-RNCMF 达到 69.95 (0.1 | S |)，71.41 (0.2 | S |) 和 73.43 (0.5 | S |) 的准确性，同时 RNCMF (基线) 的准确性是 68.49, 70.49 和 73.07

## 结论

在本文中，我们已经解决了学习的子类别分类的时候只训练数据的一小部分标有精美的标签，而其余的只有粗糙的类别标签的问题。为此，我们提出使用基于最接近类均值分类随机森林的方法，并通过引入一个正则化的目标函数为训练延长该方法。我们还通过实验表明，在一个大规模的设置分类与仅类别-标签的额外训练数据提

高子类别的高达 22%。最后，我们已经提出了实验证据，考虑到类别和子类别之间的层次关系的方法进行办法比忽略这些关系更好。

致谢。作者承认从 CTI 项目 (15769.1 PFES-ES), DFG 埃米诺特程序 (GA1927 年/ 1-1), DFG 项目 (GA1927 年/2-21505) 和丰田汽车的财政支持。

## References

- [1]  
[http://www.vision.ee.ethz.ch/datasets\\_extra/mristin/ristin\\_et\\_al\\_cvpr15\\_data.zip](http://www.vision.ee.ethz.ch/datasets_extra/mristin/ristin_et_al_cvpr15_data.zip). [Online; accessed 19-Mar.-2015].
- [2] P. Agrawal, R. Girshick, and J. Malik. Analyzing the performance of multilayer neural networks for object recognition. In ECCV, 2014.
- [3] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Good practice in large-scale learning for image

- classification. In ECCV, 2014.
- TPAMI, 2013.
- [4] Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. In ICCV, 2011.
- [5] A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. <http://www.image-net.org/challenges/LSVRC/2010>, 2010. [Online; accessed 13-Nov.-2014].
- [6] A. Bosch, A. Zisserman, and X. Munoz. Image classification ~ using random forests and ferns. In ICCV, 2007.
- [7] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. In ACCV, 2012.
- [8] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101: Mining Discriminative Components with Random Forests. In CVPR, 2014.
- [9] L. Breiman. Random forests. Machine Learning, 45(1):5-32, 2001.
- [10] A. Criminisi and J. Shotton. Semi-supervised classification forests. In Decision Forests for Computer Vision and Medical Image Analysis, Advances in Computer Vision and Pattern Recognition. Springer, 2013.
- [11] J. Deng, J. Krause, A. Berg, and L. Fei-Fei. Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition. In CVPR, 2012.
- [12] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In CVPR, 2013.
- [13] J. Deng, D. Nan, J. Yangqing, F. Andrea, M. Kevin, B. Samy, L. Yuan, N. Hartmut, and A. Hartwig. Large-scale object



- classification using label relation graphs. In ECCV, 2014.
- [14] K. Driessens, P. Reutemann, B. Pfahringer, and C. Leschi. Using weighted nearest neighbor to benefit from unlabeled data. In Advances in Knowledge Discovery and Data Mining, volume 3918 of Lecture Notes in Computer Science. Springer, 2006.
- [15] R. Fergus, H. Bernal, Y. Weiss, and A. Torralba. Semantic label sharing for learning with many categories. In ECCV, 2010.
- [16] M. Guillaumin and V. Ferrari. Large-scale knowledge transfer for object localization in ImageNet. In CVPR, 2012.
- [17] M. Guillaumin, D. Kuettel, and V. Ferrari. ImageNet Autoannotation with Segmentation Propagation. IJCV, 2014.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.
- [19] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In CVPR, 2009.
- [20] C. Leistner, M. Godec, S. Schulter, A. Saffari, M. Werlberger, and H. Bischof. Improving classifiers with unlabeled weakly-related videos. In CVPR, 2011.
- [21] C. Leistner, A. Saffari, and H. Bischof. Miforests: Multipleinstance learning with randomized trees. In ECCV, 2010.
- [22] B. Liu, F. Sadeghi, M. Tappen, O. Shamir, and C. Liu. Probabilistic label trees for efficient large scale image classification. In CVPR, 2013.
- [23] T. Mensink, J. Verbeek,

- F. Perronnin, and G. Csurka. Distance-based image classification: Generalizing to new classes at near-zero cost. TPAMI, 2013.
- [24] V. Ordonez, J. Deng, Y. Choi, A. C. Berg, and T. L. Berg. From large scale image categorization to entry-level categories. 2013.
- [25] P. Ott and M. Everingham. Shared parts for deformable partbased models. In CVPR, 2011.
- [26] M. Ristin, M. Guillaumin, J. Gall, and L. Van Gool. Incremental learning of NCM forests for large-scale image classification. In CVPR, 2014.
- [27] M. Rohrbach, M. Stark, and B. Schiele. Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In CVPR, 2011.
- [28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet large scale visual recognition challenge. arXiv:1409.0575, 2014.
- [29] R. Salakhutdinov, A. Torralba, and J. Tenenbaum. Learning to share visual appearance for multiclass object detection. In CVPR, 2011.
- [30] J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the Fisher vector: Theory and practice. IJCV, 2013.
- [31] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.
- [32] M. Stark, M. Goesele, and B. Schiele. A shape-based object class model for knowledge transfer. In ICCV, 2009.
- [33] T. Tommasi, F. Orabona, and B. Caputo. Safety in

numbers:

Learning categories from few examples with multi model knowledge transfer. In CVPR, 2010.

[34] A. Vezhnevets and J. M. Buhmann. Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. In CVPR, 2010.

[35] B. Yao, A. Khosla, and L. Fei-fei. Combining randomization and discrimination for fine-grained image categorization. In CVPR, 2011.

[36] A. Zweig and D. Weinshall. Exploiting object hierarchy:

Combining models from different category levels. In ICCV, 2007