指导教师：杨涛　　提交时间：2015 年 3 月 29 日

The task of Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名：秦泽群

学号：2012302498

班号：10011205

通过不共享的方法消除视觉语义属性的相关性

摘要

现有的学习视觉属性的方法都很容易学习到错误的属性——即在训练集中和我们感兴趣的属性相关的属性。虽然，许多已经提出的基于能够学习到正确语义的应用确实和每个属性都是对应的。我们计划通过消除联合学习的相关性和离散属性模型来解决这些问题。利用语义相关性的辅助信息，我们开发出了一种多任务的学习方法，这种方法使用结构化稀疏化来造成不相关属性的竞争和相关属性的共享。在三个非常有挑战性的数据集上，我们的研究表明在视觉属性空间对于结构的描述是学习属性模型的关键，这种模型可以保留语义，提升了的泛化能力。这种泛化能力在识别和发现未知物体类别方面很有帮助。

1. 介绍

视觉属性是人类可以描述的中级语义属性。这种描述既包括整体的描述，例如“有毛发”，“黑色的”或者“金属的”，也包括局部的属性，例如“有轮子”或者“有鼻子”。最近的研究表明属性在低等级的图像特征和高等级实体（例如物体或者场景类别）提供了一个有效的桥梁[5， 14， 17]。基于属性学习的方法一般地都会遵循标准的有判别力的学习途径，这已经成功地被应用到了其他视觉识别问题上面。使用被其自身所呈现出的属性标记过的训练图片，低等级的图片描述会被提取出来，并且独立地为每一个孤立属性训练一个判别分类器

而问题就在于这种标准方法很容易学习到与目标图片属性相关联的属性，而不是目标属性本身。图片 1 有助于说明这是为什么。假定你的任务是学习前三幅图片的属性，其他的图片不出现。即使你限制使用“可描述的”属性，也会有很多合理的假设：灰色的？有毛发？有耳朵？陆栖动物？

图 1：什么属性是在前三张图片中出现的，而没有在后两张出现的？尝试从这些图片中学习“有毛发”属性的标准方法很容易就会学习到“灰色的”，或者是其他一些相关的属性。我们提出了一种多任务的属性学习方法可以消除那些语义不同但是经常一起出现的属性之间的特征分享。

一个潜在的挑战是对于属性学习来说可能的范围太大了。当一个实例恰巧在训练集中既有积极又有消极方面，一个标准的判别模型会把它的属性绑定到特征空间的任意方向上，这就导致在学习属性时经常学习到与目标属性相关的属性上去。而许多一张图片中可以描述的属性常常在同样的空间出现，这更加剧了上述问题。例如，一个“棕色的”物体很可能是“圆的”而且“有光泽”。相反地，当学习物体类别时，每一个像素都被唯一一个目标物体占据。进一步说，即使我们尝试更强大的训练标记，对于大规模属性的标记会更加困难并且对于物体本身会产生更多歧义。想象一下，例如，在图片 1 的图像中怎么可能标记如此毫无意义的大规模属性。

但是就算是我们不小心学习到了相关的属性，这会对结果造成很大影响吗？毕竟，长久以来弱监督物体识别系统都会利用目标物体周围的相关背景特征，就像“上下文”一样。然而对于属性学习来说，在两方面这都会是一个问题。首先，伴随着相当多可能的属性组合（对于 k 对属性多至 2^k），在训练时我们只能得到一小部分合理的结果，这很可能会使相关的线索被当做有用的信号。事实上，相比于物体类别，语义属性在可扩展性方面是被过分夸大了的，相关模式可能会很容易地从训练数据中观察到的模式区别出来。第二点就是许多属性的应用——例如图像搜索[14, 12, 22]，小样本学习[17]，文本描述生成[5]——要求确定的属性要与图像内容有意义地匹配。例如，一个图像搜索用户查询“尖头”皮鞋时，如果由于训练数据的相关性系统（错误的）将“尖头”与“黑色”这两种属性合并到一起，那么用户将会是非常沮丧的。与这不同的物体识别装置，物体类别自身和相关的属性集合可能会被认为是共同出现的。学习如何识别一个物体因此毫无疑问

要涉及到学习那些相关的东西。

鉴于这些问题，我们的目标是在学习的时候消除属性之间的相关性。为了达成这个目标，我们提出了一种多任务学习框架，能够鼓励每一个属性分类器使用互不联合的图像特征集来进行预测。这种特征之间相互竞争的思想是我们方法的核心。由于一般的模型会独立地训练每一个属性分类器，所以很容易就会重复使用相关属性的图像特征，我们的多任务方法会抵抗属性之间的共享。相反地，对于很独特的属性，它致力于使独特的低等级特征孤立起来。在图片 1 的例子中，和颜色直方图每一分量相对应的数量（颜色比例）可能会被用来检测“灰色的”，然而和图像中心区域纹理相对应的那些则可能会被保留用来检测“有毛发的”。此外，由于有些属性天然地分享特征，我们通过利用辅助信息来鼓励在关联性很强的属性之间的特征共享，这些辅助信息是关于属性语义的关联性的（例如，考虑一下“红色”和“褐色”就很可能共享特征）。

我们的方法将根据每一个属性的出现与否标记过的图像，还有一系列语义相互联系的属性“组”作为输入。至于输出，它为每一个属性都生成一个二元分类器。在同一个组内的属性会被鼓励去分享低等级特征维度，而不相关的属性则会竞争。我们通过使用结构化归一稀疏性来进行参数选择，这会在标准化的特征选择的一个多任务分类上学习。

我们发现，我们的方法可以帮助消除歧义属性从而保留了更好的语义——通过像属性局部化和小样本辨识，和一个新的语义视觉属性应用这样的标准测试。我们的结果在三个数据集上一致表明，该方法有助于“学习正确的事情。”

2. 相关工作

属性语义特征 可视属性是一个用于图像的二进制描述，指示属性是否存在[14, 5, 17]。最近的研究侧重于属性作为人机交流的语义的桥梁。例如，使用属性的图像搜索允许用户指定精确语义查询（“寻找微笑的亚洲男子”）[14, 12, 22]；如果使用它们来增强标准的训练标记方法，这会提供一种新的方式来训练对象视觉系统（“斑马是有条纹的”，“这只鸟有一个黄色的肚子”，等等）[17, 3, 23]。从属性的预期构造产生的差异可以用来产生人类易于识别的文字描述[5, 21]。在所有这些应用中，不经意间学习相关的视觉特性是一个必须面对的问题；为了使人机交流是有

意义的，系统和用户对内容的解释必须调整。然而，尽管所有的注意都在关于属性的应用程序上，但在如何保证学习属性准确，并保持其语义方面，我们只做了很少的工作。

属性相关性 虽然大多数方法独立地学习属性，一些初步的步骤已经被采用来对它们之间的关系进行建模。对属性之间的共生现象进行建模，有助于确保预测遵循通常的相关性，即使缺乏一些特定属性的图像证据（如，“有耳”通常意味着“有眼”）[30, 25, 17, 24]。我们的目标基本上与这些方法相反。我们并不是将共生与真实语义关系的关联划等号，我们认为避免将很多属性混为一谈是一个学习算法最重要的部分。这将防止似然函数会过度倾向于训练数据，因此，在遇到陌生的配置，未出现的属性设置时会有一个更好地处理。

区分属性 据我们所知，在以前的工作中，唯一一个试图明确地消除相关的语义属性是[5]。对于每个属性，他们的方法为**每个物体类别**选择不同的图像特征，然后将它们聚集在一起训练属性分类器。例如，它首先找到易于区分车辆的特征——有没有“轮子”，然后是公交车有没有“轮子”，等等。这种方法的思想是：同一个类的实例会使得感兴趣的属性隔离开来。然而，在每一个类别样本数目都很少时，这种方法很容易受偶然相关性的影响，而且需要非常昂贵的质量很好的属性标注。我们的方法克服了这些问题，正如我们与[5]的结果上的大量比较所显示的。

虽然以前这是在去相关语义属性方面唯一的的工作，无监督的一些方法试图将发现的（未命名/非语义）“属性” 多元化[31，18，5]——例如通过设计生成不相关属性的类别分类[31]，或将多余的语义属性转化为差别很大的属性[18]。与此相反，我们联合地学习特定语义属性词汇。

多任务学习（MTL） 多任务学习由于多任务使用了联合训练预测函数，往往通过选择特征尺寸（“支持”）的每个函数使用，以满足一些标准。大多数方法强调各类别中的分享 [1, 19, 11]。例如，特征之间的共享能产生更快的探测器[27]，而对象及其属性之间的共享可以独立出适合各个任务的属性[29, 8]。一些工作已经开始探索对不好的相关性建模[33, 15, 7, 20]。例如，在分层分类器，特征竞争是通过不相交稀疏或“正交转移”而被鼓励，这样可以在父子节点分类器之间移除冗余[15, 7]。这些方法利用对象标签中固有的相互排斥性，而在我们的属性设置中，并不

会包含这些排斥性的内容。不同于任何这些方法，我们使用多任务组来对在目标空间中的语义结构进行建模。

虽然大多数 MTL 方法在所有的任务中都强制执行共同学习，一小部分 MTL 方法会寻找发现任务组间共享属性的方法 [9, 10, 13]。我们的方法涉及到分组过得任务，但有两个重要的区别：（1）我们明确组与组之间竞争模型和组内共享模型来实现组内部的相关性消除，和（2）我们把额外的语义组的信息作为学习期间的监督。与此相反，在现有的方法[9, 10, 13]从数据中寻找任务组，这就和单任务学习一样容易受到相关性的影响。

3. 方法

我们的目标是只有在正确的语义属性存在的时候学习属性分类。特别是，我们希望这种方法能够自己学会归纳，然后可以测试共生属性的模式与训练集中不同的图像。我们做法的关键是，在一个词汇表上共同学习的所有属性，同时实施结构化稀疏，将特征共享模式和语义相近的属性，以及特征竞争和语义相去甚远的属性联系起来。

接下面，我们首先描述我们算法的输入：属性之间的语义关系（3.1 节）和低级别的图像描述（3.2 节）。然后，我们介绍我们的学习目标和优化框架（3.3 节），算法最后会输出词汇表中每个属性的分类器。

3.1. 语义属性分组

假定我们在为词汇 M 学习分类器。为了表达属性的语义联系，我们使用 L 属性分组，L 包含 $S_1,\ldots,S_L$，每一个 $S_i=\{m_1,m_2,m_3,\ldots\}$ 包含着每个组中每一个具体的属性的指数，并且 $1\le m_i\le M$。虽然我们的方法中没有任何“无联合”的限制，为了精确，我们的实验中每个属性只出现在一组中。

如果两个属性都在同一组中，这反映了他们有一些语义先关性。例如，在图 2 中，S_1 和 S_2 分别对应于纹理和形状属性。对于属性说明的颗粒度类别，就像鸟类，一个组别可以专注于特定领域方面的固有分类中——例如，一组为喙的形状（钩状，弯曲的，匕首状等），而另一组为腹部颜色（红腹，腹黄等）。尽管这

一的分组可以令人信服地自动收集数据（从文本数据，WordNet，或其他来源），但我们依靠现有的人工定义的组[17, 28]，在我们的实验中。

正如我们将在下面看到，组间共同成员的信号对于我们的学习算法来说，它的属性更容易分享。对于空间本地化属性组（例如，喙的形状），这可能会导致算法集中在描述同一个对象部分；对于全局属性组（例如，颜色），这可能导致算法集中在相关的特征信道的一个子集。我们并不强调存在一个单一的“最佳”的分组；相反，我们期望这样的部分的辅助语义信息能够帮助智能地决定什么时候允许共享。

我们关于属性标签维度分组，这种分组可以用来利用任务间的关系，这种方法有别于描述唯独分组，而且不会与描述维度分组代表的特征空间结构搞混，就像在单任务“group lasso”[32]中相混淆的那样。同时利用特征空间结构，可以想见，这将进一步改善我们的方法的结果，在本文中我们将重点放在建模和利用任务关系上。

3.2. 图像特征表达

当我们将分类器的学习空间指定为低级别的图像特征空间时，我们注意到一个主要标准：我们会想要将学习算法空间局部化和频道局部化特征暴露出来。通过在空间局部化，我们想使不同局部区域内部图像的内容在不同维度的图像特征向量出现。类似地，通过频道局部化，我们想使不同类型的描述（颜色，纹理等）占据不同的维度。通过这种方式，学习算法可以挑选一个稀疏化的集合，这个集合对于空间和特征都是稀疏化过的，这样在一个语义组中可以很好的区别不同的属性。

为此，我们提取了一系列的直方图特征，在多尺度网格单元内汇聚多个特征频道。我们减少了每个使用 PCA 的直方图（对应于特定的窗口和特征）的维度。这样的缩减使得我们从两方面获益，一方面丢弃了低方差的维度，而且分离了对于特定属性特征的选择。由于我们每个通道都执行 PCA，所以在最后的表述中我们保留了原本的模式和原本的互联关系。更多的数据集及具体细节都在第 4 节。

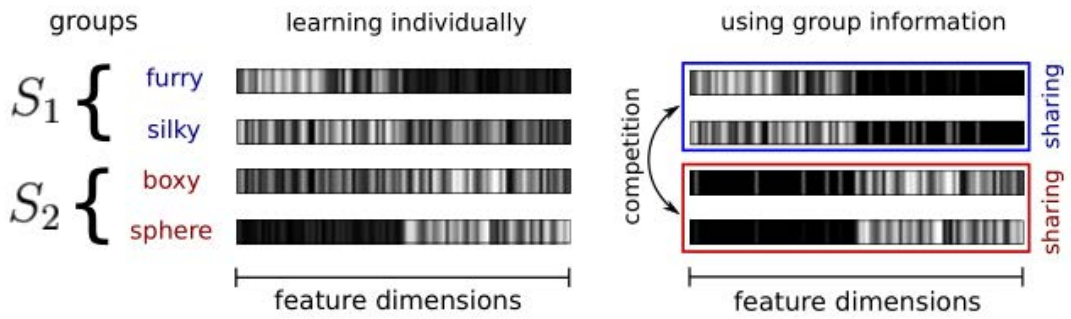

图 2：我们的想法的轮廓。我们展示了每个属性的权重向量（绝对值），这些属性是通过标准（左）的方法和已经提出的方法（右）学习出来的。分配给一个特征维度的的权重越高（较浅的颜色），就会有越多的属性依赖于这个特征。在这种情况下，我们的方法将有助于解决“柔滑”和“四四方方”的问题，这在训练集中是高度相关的两种属性，所以标准学习方法会将其混为一谈。

3.3. 使用属性共享和竞争的联合属性学习

我们学习方案的输入是：（1） N 个训练图像的描述，每一个描述都被表示为 D 维的向量 X_n，（2）所有属性中的相关属性的标记（二进制），下标为 $a=1,\ldots,M$，还有（3）语义属性组 $S_1,\ldots,S_L$。定义 $X_{N\times D}$ 矩阵，这个矩阵有训练图像描述组成。我们将矩阵 X 的第 n 行视为行向量 X_n，X 的第 d 列为列向量 X_d。X_n^d 定义为 X 的 $(n,d)^{th}$ 的值。相似地，训练属性标记可以被表示为 $Y_{N\times M}$，行为 Y_n 列为 Y_m。

因为我们希望强加于属性模型之间的关系的约束，我们同时在多任务环境下学习属性，其中每个“任务”对应一个属性。学习方法输出参数矩阵 $W_{D\times M}$，矩阵的列对应于 M 个属性的分类。我们使用 logistic 回归分类器，以及损失函数：

$$L(X,Y;W)=\sum_{m,n}\log(1+\exp((1-2y_n^m)X_n^TW^m))$$

每个分类器具有对应于“权重”的每个特征的维度，用于检测该属性。需要注意的是矩阵 W 的行向量 W_d 表示所有属性特征维度 d 中的使用情况；在 W_d^m 中的 0 意味着特征 d 在属性 m 中没有被使用。

公式 我们的方法有一个前提，就是语义相关的属性（有一些）更易于在同样的的图像特征中被确定，并且语义不相关的属性依赖于（至少有些是这样的）不相关的特征。以这种方式下，在特征空间中对一个属性的支持，也就是说在一系列维度中都不为 0 权重的属性，强烈地和其语义联系程度相关。我们的目标是通过与有效利用提供的语义分组，包括（1）组内共享特征（2）组间竞争特征，有效地利用所提供的语义分组。我们这其转化为结构稀疏性的问题，在输出属性空间中结构是通过分组来表示的。图 2 显示了我们的方法预期的效果。

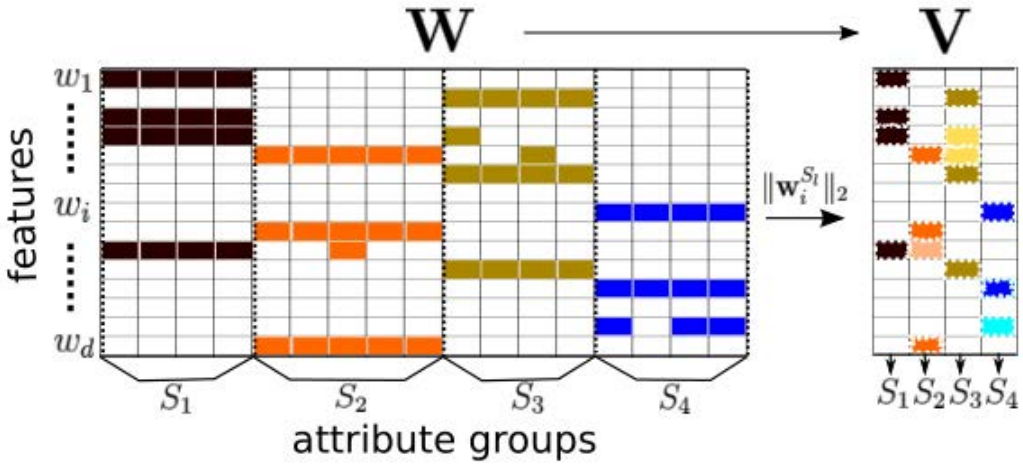

图 3：在应用 “带有惩罚因子的 lasso 回归”之前特征选择矩阵W 中分组过的列的“崩溃”。在W 矩阵和V 矩阵中非零样本被涂黑了。在V 矩阵中被涂黑的程度代表了在那个组中有多少个属性选择了那个特征。

为了给我们的方法搭建平台，我们接下来讨论现有的两种特征选择方法，这两种方法都会被用到在第 4 节中。第一个是一个简单的自适应单任务 lasso 方法[26]。最初 lasso 正则化被用来学习单一属性m，在我们的设定中，这将会定义为$\|W^m\|_1$。众所周知，这样的凸正则化会产生一个对于稀疏的好的近似，这会通过非零样本的数量生成，$\|W^m\|_0$。

总结以上，我们可以扩展单任务 lasso 方法为多任务方法，来产生一个“全局竞争”的最小化 lasso 目标：

$$W^* = \arg\min_W L(X, Y; W) + \lambda \sum_m \|W^m\|_1$$

其中$\lambda \in \mathrm{R}$是一个归一化的参数用来平衡稀疏性和分类损失。注意归一化的第二项可以写为：$\sum_m \|W^m\|_1 = \sum_d \|W_d\|_1 = \|W\|_1$。这表明了归一化是如何与$W$矩阵的行列数匹配的，而且可以想见，就像（1）在每一个任务列W^m中鼓励稀疏，以及（2）在每一个属性行W_d中增强稀疏。后者有效地为特征维度d制造了所有任务之间的竞争。

相反的，针对联合特征选择的“全共享”多任务 lasso 方法提升了所有任务中的共享程度，这是通过最小化下述目标函数实现的：

$$W^* = \arg\min_W L(X, Y; W) + \lambda \sum_d \|W_d\|_2 \text{。}$$

为了验证这将会在所有属性间鼓励特征共享，注意归一化的式子可能会被写为ℓ_1形式$\|V\|_1 = \sum_d \|W_d\|_2$，单列矩阵$V$由$W$的列用$\ell_2$方法组成，例如$d^{th}$变为$v_d = \|W_d\|_2$。矩阵$V$的$\ell_1$标准形式稀疏矩阵$V$会更适合于这个方法。这也就意味着每一个分类器必须只选择对于其他分类器也是有帮助的特征。也就是说，W应该要么一行全为零或者全不为零。

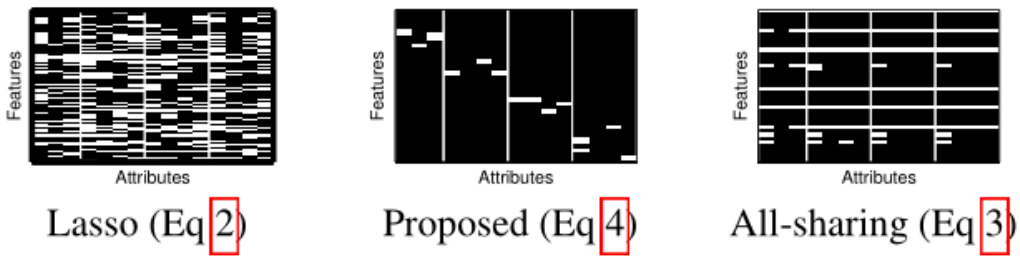

图 4：这是在 CUB 数据上通过不同结构稀疏方法学习得到的W矩阵的一部分（阀值，绝对值）。细的垂直线分散在各个属性组。

我们现在定义我们的目标，一个语义信息中级方法，位于在等式 2 和等式 3（图 4 中）的极端情况中间的。我们的最小化目标保留了竞争——包括ℓ_1形式的常规 lasso 横跨分组，以及在语义组内应用了ℓ_{21}类型的共享：

$$W^* = \arg\min_W L(X, Y; W) + \lambda \sum_{d=1}^{D} \sum_{l=1}^{L} \left\| W_d^{S_l} \right\|_2 \quad (4)$$

$W_d^{S_l}$ 是一个行向量，包含着 W_d 矩阵中的一个样本子集，也就是说，那些在语义组 S_l 中确定的样本子集。这样的归一化会在语义组中限制列的“崩溃”影响，所以 V 不再是一个简单的列向量，而是一个 l 列的矩阵，每一列对应与每个组。图 3 将这个想法可视化了。注意稀疏性是如何将特征竞争以及不相关属性这两者联系起来的，以及在语义组内共享属性。

我们的模型融合了以前的提法并且提出了一中“位于中点”的方法。只有一个组 $S_1 = \{1, 2, ..., M\}$ 包含所有属性，公式 4 可简化为公式 3。同样地，为每个属性设置属于自己的组 $S_m = \{m\}$ 将会产生等式 2 的 lasso 公式。图 4 说明了在结构稀疏中他们之间各自的区别。标准的 lasso 方法的目的是在所有任务中丢弃越多特征越好，标准的“全分享”方法的目的是只使用可以由多任务共享的特征。与此相反，我们提出的方法在属性中寻找可共享的特征，而抵抗在不相关属性中的特征共享。

就像我们将在实验结果中表明的那样，这减轻了偶然的互相关属性的影响。使属性组之间彼此远离，这有助于在词汇表内消除无关属性的相关性。即使“棕色”和“毛茸茸”总是在训练的时候同时出现，他们的分类器也不会把他们搞混。同时，组内特征共享将组内标签聚集在一起用于特征选择，减轻偶然的相关性——不仅在词汇表内，也在可以被捕获的视觉属性中（可命名的或其他）。例如，假设“钩状喙”和“棕色肚子”是指经常共同出现的属性;如果“棕色肚子”与更易于学习的“黄肚” 共享一组，那么压力就会落在棕色和黄色肚子都可以共享的特征维度上，这会间接导致“钩状喙”不联结特征。

然而，我们强调，分组只是一个先决条件。我们的方法更易于语义相关的属性分享，但它不是一个硬约束，而且错误分类造成的损失也在决定哪些特征是相关的这个问题上扮演着一个重要的角色。

优化 方程 4 的形式的混合形式正则化在优化时是非光滑和非平凡。这种形式频频出现在结构学习文献[32, 2, 1, 11]。如在文献[11]中，在使用梯度下降方法来优化

一个目标结果的平滑近似之前，我们通过将归一化 2 范式转化为双重形式而重新制定了目标。见附录。

4. 实验和结果

数据集 我们使用三个数据集的 422 个属性：（1）CUB-200-2011 Birds (“CUB”)[28]，（2）动物属性（“AwA”）[17]，（3）aPascal/aYahoo (“aPY”)[5]。数据统计汇总于表 5。和常见的做法一样，我们将数据划分成“见过”和“未见过”类别。这个想法是为了在已经见过的物体上学习属性，并将其应用到新的没有见过的物体。这种压力测试的泛化能力，自相关模式自然会偏向新的对象。AwA 和 aPY 上见过和未见过分类是实现指定的。对于 CUB，我们随机选择 200 个中的 100 个作为“见过的”。

表 5：数据集统计总结

	类别		属性		特征	
数据集	见过	未见过	数量(m)	组(l)	成功	D
CUB	100	100	312	28	15	375
AwA	40	10	85	9	1，21	290
aPY-25	20	12	25	3	7	105

特征 3.2 节定义了基本的特征提取过程。在 AwA 上，我们使用数据集提供的特征（4 通道的全局描述，3 层的金字塔结构，在两个通道上有 4×4+2×2+1=21 个窗口）。对于 CUB 和 aPY，我们使用文献[5]的作者的代码计算特征。在 aPY 上，我们使用一个一层金字塔结构，每个窗口有 3×2+1=7 个窗口，和文献[5]一样。在 CUB 上，我们在已有的标注过的部分来提取特征。为了避免很少出现的小部分样本，我们将数据集的样本限制在最平常可见的范围上。更多细节见附录。

语义组 为了定义语义组，我们很大程度上依赖现有的数据。CUB 指定了 28 个属性组[28]（头部颜色，背部模式等等）。对于 AwA，其作者建议 9 个分组在文献[16]中（颜色，纹理，形状等等）。对于 aPY，它并没有预先指定的属性分组，我们依照文献[16](“aPY-25”)的建议分了 25 个属性组（总共 64 个），分成了形状，材质和面部属性组。所有分组细节见附录。

正如在 3.2 节所讨论的，我们的方法需要属性组和图像的描述是相互兼容的。例如，如果语库的描述没有描述空间顺序，那么基于其位置的属性的分组将不会

起作用。然而，我们的研究结果表明，这种兼容性是容易满足。我们的方法成功地利用预先指定的属性组，以及预先指定的特征表示形式。

基线 我们比较了四种方法。两种单任务的学习基线，其中每个属性分别学习：（1）“标准”：ℓ_2 logistic 回归和（2）“聪明分类”：第二节提到的文献[5]中基于特征选择框架的物体类别标记方法（为了统一在最后使用 logistic 回归代替了支持向量机）。另外两个是在第三节中提到的稀疏多任务方法：（3）“lasso”（公式 2），以及（4）“全分享”（公式 3）。所有的方法都产生 logistic 回归分类器，并使用相同的输入特征。所有参数（在所有方法中都用到的 λ，以及文献[5]中的第二个参数）都被未见过的分类数据验证过了。

Tasks	Attribute detection scores (mean average precision)								Zero-shot DAP acc.(%)		
Datasets	CUB			AwA		aPY-25			CUB	AwA	aPY-25
Methods	U	H	S	U	S	U	H	S	[100 cl]	[10 cl]	[12 cl]
lasso	0.1783	0.2552	0.2219	0.5274	0.6175	0.2713	0.2925	0.3184	7.345	25.32	9.88
all-sharing [1]	0.1778	0.2546	0.2217	0.5378	0.6021	0.2601	0.2934	0.2560	7.339	19.40	6.95
classwise [5]	0.1909	0.2756	0.2406	N/A	N/A	0.2729	0.2776	0.3595	9.149	N/A	20.00
standard	0.1836	0.2706	0.2369	0.5366	**0.6687**	0.2727	0.2845	**0.3772**	9.665	26.29	**20.09**
proposed	**0.2114**	**0.2962**	**0.2654**	**0.5497**	0.6480	**0.2989**	**0.3318**	0.3021	**10.696**	**30.64**	19.43

表 6：属性检测分数（左，大概）和无样本识别（右，精确）。分数越高越好。U，H 和 S 分别是指未见的，几乎没见过，全部见过的测试集（4.1 节）。我们的方法通常优于现有的方法，尤其是在属性相关训练和测试数据（即 U，H，和无样本（4.2 节）的情况）。

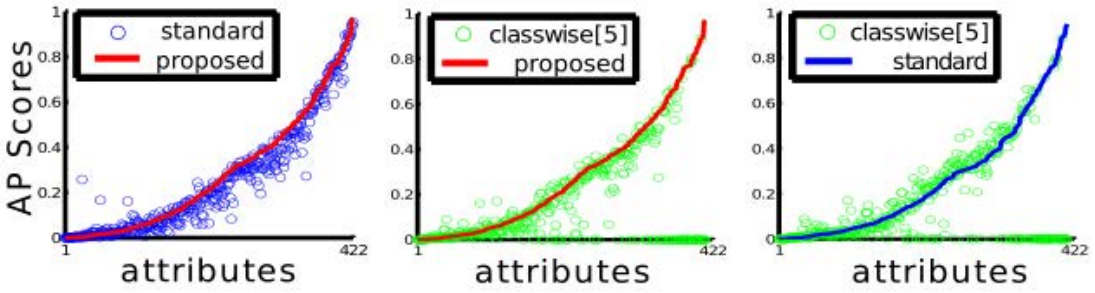

图 7：所有数据集上的属性检测结果

4.1. 属性检测精度

首先，我们测试的基本属性检测精度。为了此任务，每一个测试图像被标记有一个二进制标签在在词汇表的每个属性上。属性模型在一个随机选择 60% “可

见”类别数据集上训练，在三个测试数据上测试：（1）未见过：没见过的样本（2）全部见过：看可见类别的其他样本（3）几乎见过：全部见过的一个子集。要创建几乎见过数据集，我们首先计算出一个二进制类属性的关联矩阵的属性标签，基于每个看到的样本实例的阈值平均值。然后为每个组成违反了类级别标签矩阵中，如该属性的实例属性，大象（灰色），猫耳合住（耳）。

总的结果 表 6 显示了所有的属性的平均 AP 得分，在每数据集上。在所有这三个数据集，我们的方法有显著的泛化能力，在所有的未见类别和几乎未见数据上。

同时文献[5]中“智能类别”技术有助于在一定程度上消除属性相关，比 aPY-25 和 CUB 的“标准”更高，“标准”比所提出的方法更弱。该方法假设同一对象的实例帮助隔离属性；然而，如果两个属性总是在相同的对象中共同变化（例如，如果汽车的车轮总是金属），那么该方法仍然是容易利用相关的特征。此外，对于每一个类别样本的正面样例和反面样例的大量需求可能会是一个实际的负担（并使得它不适用于 AwA）。相比之下，我们的想法，共同学习的属性和特点，它们之间是不容易受到相同的对象关系影响，并且不使这样的标签要求。我们的方法在每个数据集中都优于这种最先进的方法。

两个多任务基线（lasso 和全分享）通常是最弱的，验证该语义在决定何时共享将发挥重要的作用。事实上，我们发现，全共享/全竞争正规化一般会伤害的模型，导致验证过的正规化权重 λ 仍然相当低。

图 7 绘制了从所有数据集中 422 个独立属性的“未见过”的结果。在这里，我们将展示配对的三个最佳方式的比较：我们提出的，“智慧分类”[5]，和标准方法。对于每个图，属性按照一种方法检测能力递增的顺序排列。在几乎所有的 422 个属性，我们的方法均优于标准的学习方法（第一图）和先进的“智慧分类”法（第二图）。

图 8：成功案例：注解显示的是我们的方法的与事实相符的属性预测。 Logistic

回归基线（“标准”）在所有情况下都失败了。

图 9：失败案例：案例在那里我们的预测（如图所示）是不正确的，而“标准”方法是成功的。

图 10：鸟部分的贡献（显示为亮点），以正确检测特定的属性。我们的方法看起来在正确的地方出现的更多，想比于标准的单任务的基线。

“学习正确的东西”的证据 比较全看到和难以看到的情况的结果，我们看到的证据表明，我们的方法的会获益，由于其保留属性语义的能力。在 aPY-25 和 AwA 上，我们的方法在全看到情况下表现不佳，而这提高了对看不见的和难以看到测试集性能。这符合我们的期望，成功地解决了相关的训练数据的问题：它新的测试集的泛化能力在更好，在具有类似的相关性的测试集上花费也会很少（其中一个学习者将从学习相关性获益）。

在图 8 中，基于被标准方法错误标记但是我们的方法正确的标记的情况，我们提出了定性的证据。例如，楔形“熨斗状”建筑物（第 1 行末尾）的形式被正确标记不是“3 维的四四方方”，而在渣土鸟（第 2 行末尾）被正确标记为没有“褐色下半身”，因为它被黑色的污垢沾满了。与此相反，基线预测基于相关线索的属性（例如，城市场景通常四四方方，不是楔形），未能在这些图像上成功。

图 9 显示了一些失败的案例。常见故障的情况下，我们的方法是，当图像模

糊，对象是非常小的，或信息缺少的情况下，从共同出现的方面学习环境会有所帮助。在低分辨率“羽毛”的情况下，例如，认识鸟的其他部分可能有助于正确识别“羽毛”。

我们保留语义的做法会有更多定性的证据，来自于研究影响不同的方法决定的特征。CUB 的基于部分的表达允许我们可视化不同部分所给与的属性贡献程度（见附录）。图 10 显示我们的方法是怎么样着重于鸟的部分与适当的空间区域相吻合，而基线方法将其作为相关的特征。例如，“褐翅”的图像上，当基准方法注重于头部，我们的方法几乎完全突出了翅膀。

4.2. 零样本物体辨识

接下来我们将展示保留属性语义对零样本物体辨识的影响。如果更随文献[17]中的设定，其目标是为了学习文字描述对象的类别，但没有训练图像（例如，“斑马条纹有四条腿”），这使属性的正确性至关重要。我们输入属性的可能性，对于零样本学习，从每个方法的模型到直接的属性预测（DAP）框架[17]（见附录了解详细信息）。表 6 展示出了结果。我们的方法在多级精度上得到大幅提高在两个大型数据集上（CUB 和 AwA）。它是略高于“标准”和“智慧分类”在 aPY-25 数据集上的结果是更糟糕的，不管我们显著更好的属性检测（4.1 节）。我们相信，这可能是由于使用 DAP 识别少量属性是更不可靠，就像在 aPY-25 上（25 个属性）。

4.3. 语义属性的类别发现

最后，我们说明类别发现方面的影响。认知科学家提出，自然类别是一个凸集，在概念上来说，其中心对应于“心理素质方面”[6]。这促使我们使用属性进行类别发现。将语义视觉属性视为视觉分类的概念空间。我们使用 k-means 算法来聚类每一个方法属性存在的可能性，从而来发现凸聚类。我们设定 k 为类别的真实数量。我们比较了每种方法的聚类，在三个数据集上使用真实的“未见”分类。对于 CUB，我们测试的两个 100 种（CUB-S），以及生物分类（CUB-F）。性能通过 NMI 分数来测定，这会在一个给定的聚类和一个真实的类别间测试共享的信息，而不需要硬性分配聚类类别。

表 11 显示了结果。我们的方法比基线方法在所有任务上都显著地表现地更好。如果我们要替换聚类的事实属性签名，我们会感觉到一个上界（最后一行）。这表明，（1）对于发现来说，视觉属性确实构成了一个似是而非的“概念空间”；（2）提高的属性的学习模式可能会使得的高层次的视觉任务获益。

Methods / Datasets	CUB-s	AwA	aPY-25	CUB-f
lasso	0.5485	0.1891	0.1915	0.3503
all-sharing [1]	0.5482	0.1881	0.1717	0.3508
classwise [5]	0.5746	N/A	0.1973	0.3862
standard	0.5697	0.2239	0.1761	0.3719
proposed	**0.5944**	**0.2411**	**0.2476**	**0.4281**
GT annotations	0.6489	1.0000	0.6429	0.4937

表 11：未见类别的发现的 NMI 分数（4.3 节）。越高越好

5. 结论

我们介绍了一种方法使用语义来指导属性的学习。我们在三个数据集中大量的实验支持了我们的两个主要议题：（1）我们的方法克服了误导性的训练数据的相关性，成功地学习语义视觉属性，和（2）在学习到的属性中保留语义是有益的，作为一个高级别任务的中间步骤。在今后的工作中，我们计划研究重叠属性组的影响，并探索方法来自动挖掘语义信息。

致谢：我们要感谢 Sung JuHwang 有益的讨论。这项研究是由美国国家科学基金会支持的，作为 NSF IIS-1065390 和 NSF IIS-1065243 一部分。

参考文献

[1] A. Argyriou， T. Evgeniou， and M. Pontil. Multi-Task Feature Learning. In NIPS， 2007.
[2] F. Bach. Consistency of the group lasso and multiple kernel learning. In JMLR， 2008.
[3] S. Branson， C. Wah， F. Schroff， B. Babenko， P. Welinder， P. Perona， and S. Belongie. Visual recognition with humans in the loop. In ECCV， 2010.
[4] X. Chen， Q. Lin， S. Kim， J. G. Carbonell， and E. Xing. Smoothing proximal gradient method for general structured sparse regression. In AAS， 2012.
[5] A. Farhadi， I. Endres， D. Hoiem， and D. Forsyth. Describing Objects by Their Attributes. In CVPR， 2009.
[6] P. Gardenfors. Conceptual spaces as a framework for knowledge representation. In Mind and Matter， 2004.

[7] S. J. Hwang， K. Grauman， and F. Sha. Learning a Tree of Metrics with Disjoint Visual Features. In NIPS， 2011.
[8] S. J. Hwang， F. Sha， and K. Grauman. Sharing features between objects and their attributes. In CVPR， 2011.
[9] L. Jacob， F. Bach， and J.P. Vert. Clustered Multi-Task Learning: A Convex Formulation. In NIPS，2008.
[10] Z. Kang， K. Grauman， and F. Sha. Learning with whom to share in multi-task feature learning. In ICML.
[11] S. Kim and E. Xing. Tree-guided group lasso for multi-response regression with structured sparsity， with an application to eQTL mapping. In AAS， 2012.
[12] A. Kovashka， D. Parikh， and K. Grauman. WhittleSearch: Image Search with Relative Attribute Feedback. In CVPR， 2012.
[13] A. Kumar and H. Daume III. Learning task grouping and overlap in multi-task learning. In ICML，2012.
[14] N. Kumar， P. N. Belhumeur， and S. K. Nayar. Facetracer: A Search Engine for Large Collections of Images with Faces. In ECCV， 2008.
[15] L. Xiao and D. Zhou and M. Wu. Hierarchical Classification via Orthogonal Transfer. In ICML，2011.
[16] C. Lampert. Semantic Attributes for Object Categorization (slides). http://ist.ac.at/chl/talks/lampert-vrml2011b.pdf， 2011.
[17] C. Lampert， H. Nickisch， and S. Harmeling. Learning to detect un- seen object classes by between-class attribute transfer. In CVPR， 2009.
[18] D. Mahajan， S. Sellamanickam， and V. Nair. A joint learning framework for attribute models and object descriptions. In ICCV， 2011.
[19] S. Parameswaran and K. Weinberger. Large margin multi-task metric learning. In NIPS， 2010.
[20] B. Romera-Paredes， A. Argyriou， N. Bianchi-Berthouze， and M. Pontil. Exploiting unrelated tasks in multi-task learning. In AISTATS， 2012.
[21] B. Saleh， A. Farhadi， and A. Elgammal. Object-Centric Anomaly Detection by Attribute-Based Reasoning. In CVPR， 2013.
[22] W. Scheirer， N. Kumar， P.N. Belhumeur， and T.E. Boult. Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search. In CVPR， 2012.
[23] A. Shrivastava， S. Singh， and A. Gupta. Constrained semi-supervised learning using attributes and comparative attributes. In ECCV， 2012.
[24] B. Siddiquie， R. Feris， and L. Davis. Image Ranking and Retrieval Based on Multi-Attribute Queries. In CVPR， 2011.
[25] F. Song， X. Tan， and S. Chen. Exploiting relationship between attributes for improved face verification. In BMVC， 2011.
[26] R. Tibshirani. Regression shrinkage and selection via the lasso. In RSS Series B， 1996.
[27] A. Torralba， K. Murphy， and W. Freeman. Sharing visual features for multiclass and multiview object detection. In PAMI， 2007.
[28] C. Wah， S. Branson， P. Welinder， P. Perona， and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. 2011.
[29] G. Wang and D. Forsyth. Joint learning of visual attributes， object classes and visual saliency. In CVPR， 2009.

[30] Y. Wang and G. Mori. A discriminative latent model of object classes and attributes. In ECCV，2010.
[31] F. Yu， L. Cao， R. Feris， J. Smith， and S. Chang. Designing category-level attributes for discriminative visual recognition. In CVPR， 2013.
[32] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. In RSS Series B， 2006.
[33] Y. Zhou， R. Jin， and S.C.H. Hoi. Exclusive lasso for multi-task feature selection. In AISTATS，2010.