模式识别报告

XiaoX

Demon Killer

January 2, 2023

摘要： 论文选自 CVPR2022 的一篇，题目为《HCSC: Hierarchical Contrastive Selective Coding》。来自上海交通大学、Mila 魁北克人工智能研究所以及字节跳动的研究者提出了一种具有层级语义结构的自监督表征学习框架，在 ImageNet 数据集上预训练的模型在多个下游任务中取得了 SOTA 性能。
本报告将从文章概要，研究背景，知识联系，个人收获几方面做简要汇报。

关键词： 关键词 1；关键词 2；关键词 3

目录

1 文章概要

自然界中广泛存在如图 1 这种分层的语义概念，比如哺乳动物–狗–拉布拉多，相似图片可能在某个层次上属于同一分类。这种现象在 image 的数据集上是广泛存在的，如果能引入这样的信息会有助于预训练模型迁移到下游任务上。

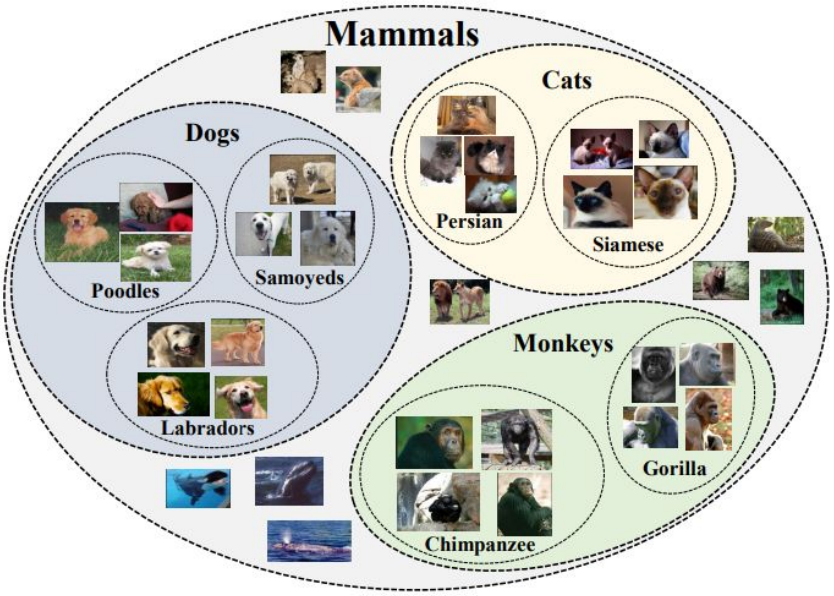


Figure 1. An image dataset always contains multiple semantic hierarchies, *e.g.* "mammals → dogs → Labradors" in the order from coarse-grained semantics to fine-grained semantics.

现有的对比学习模型，作者认为存在以下的问题：

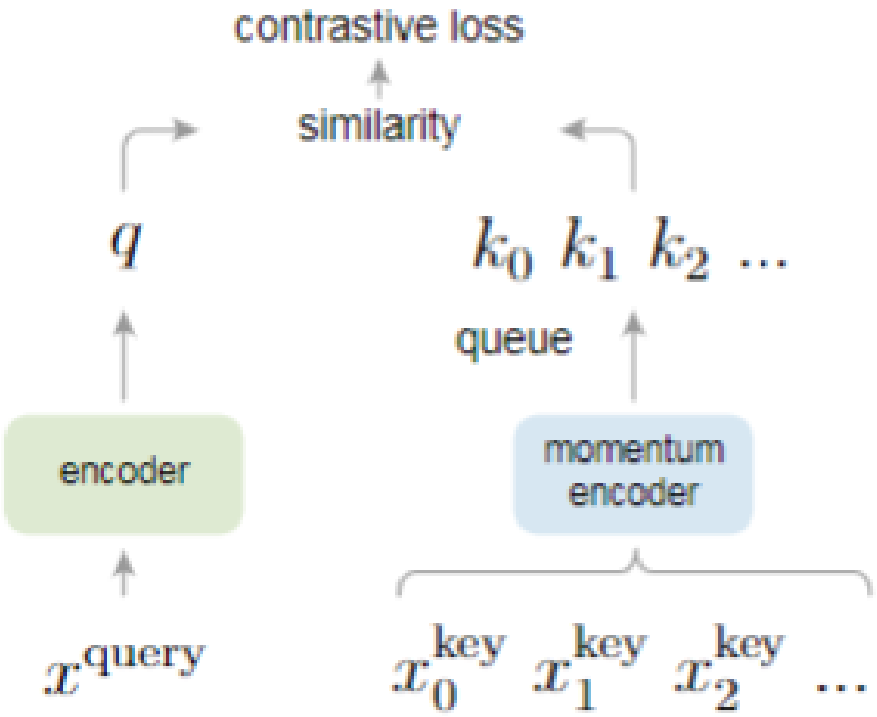
1. 缺乏对上述描述的分层语义结构的建模
2. 对比学习中的负样本不能保证语义是不同的，妨碍了模型的学习

因此本文作者提出了 Hierarchical Contrastive Selective Coding (HCSC) 模型，模型为了解决现有存在的问题，建立了一组动态更新的层次原型，以表示潜在空间中数据的层次语义结构，在训练过程中可以选择更合适的对比样例进行训练。

1.1 研究背景

1.1.1 MoCo

对比学习（*Contrastive Learning*），顾名思义就是对比着进行学习。不需要知道标签的具体信息，因此这是一种无监督学习方法。目的是将类似的图片的特征在特征空间中尽量的靠近，即缩小与正样本间的距离，扩大与负样本间的距离。Moco 将一系列的对比学习方法归纳为一个字典查询的问题。从动态字典的角度看对比学习，相当于在字典中查询，相似的 query 和 key 应尽量靠近。要想获得较好的训练结果，字典应该满足大且保持一致性两个特征。



Moco 在不断更新 encoder 的过程中提出了两点优化策略：1. 使用队列结构存储字典中的键值（负样本），由于训练中每个 batch 在不断 $push\_back$ 和 pop_front, 总体来看这个字典就能变得非常大。2. 引入动量（*Momentum*）参数保证 momentum encoder 的缓慢变化，从而保证 key 的一致性。

$$\theta_k = m\theta_{k-1} + (1-m)\theta_q$$

1.1.2 PCL

PCL 是对 MoCo 的改进，他认为 MoCo 只把自己当作正样本，其他样本均当作负样本的操作太过暴力，毕竟其他样本中有很多都是相似样本。所以作者提出先将样本聚类，自

己的聚类中心（也就是所谓原型）当作正原型，其他原型当作负原型，拉近正原型，推开负原型。聚类采用 k-means，作者采用了不同的 k 值求 loss 取平均。框架是 EM 框架，E 步：根据上一轮动量编码的输出结果计算距离进行 k-means 聚类，计算出原型 c 和浓度 ϕ。M 步：将计算出的原型 c 和浓度 ϕ 带入到下一轮 loss 计算中并进行更新。

这种方法站在聚类后的类层次上的，获得聚集在相应的集群中心周围的紧凑的图像表示，从而捕获一些可以由单一集群层次表示的基本语义结构。PCL 不仅学习低级特征来完成实例识别任务，更重要的是将聚类发现的语义结构编码到学习的嵌入空间中。

1.1.3 HCSC

HCSC 是在 PCL 上的进一步改进工作。他认为 PCL 直接将多粒度原型对比学习的 loss 取平均太过暴力。不同粒度间是有关联的，就比如自然界，哺乳动物–狗–拉布拉多，就是一个粒度的关联。所以作者进行如下改进，对于 PCL 采用不同 k 值的 k-means 的并行操作，作者改成了串行操作，即对聚类产生的原型，将原型进行再次聚类，（框架仍然是 EM 框架），这样循环聚类 L 次产生树结构。在采样负样本的时候，根据树结构计算每层的样本距离，根据距离来采样负样本和负原型，加入计算中。

1.2 知识联系

总体来说这一系列研究都与 Feature Selection 和 Feature Extraction 有关，是为后期应用于分类，检测，分割等下游任务的 pre-train 工作，来提取出更加有效的特征。

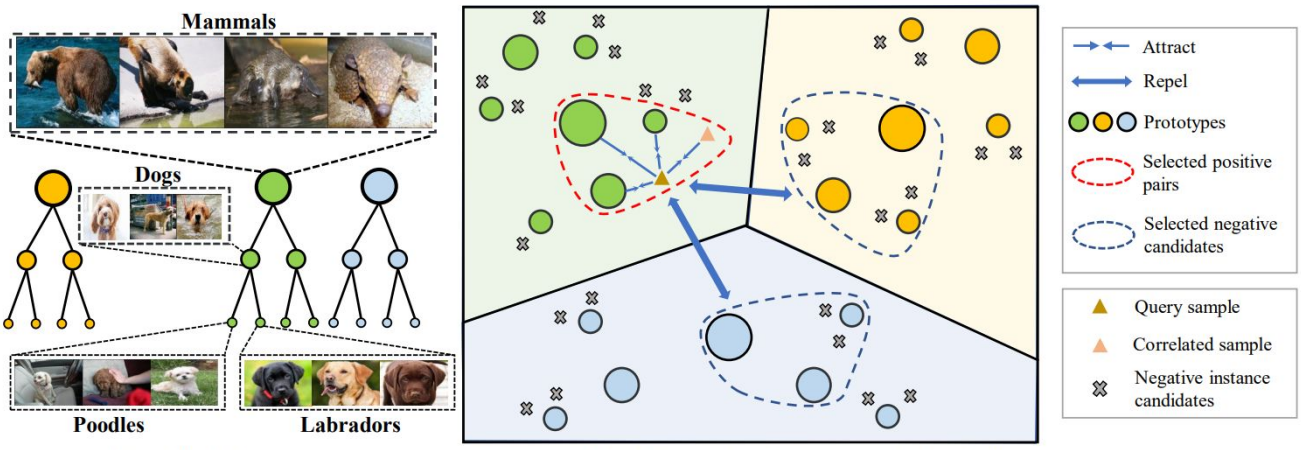
2 文章内容

2.1 实验方法

该工作的方法论框架包含两个重要的模块: 一个是层级语义结构的构建与维护, 另一个是基于层级语义结构的选择性对比学习。

2.1.1 层级语义表征

层级语义结构天然可以通过树状结构来描述：如果将树中的某个节点认为是一个语义类别, 则父节点可以认为是它的上层类别。如下图 (a) 中，「拉布拉多犬」的父节点是「犬类」，而其兄弟节点可以包括「贵宾犬」等。这样的树状结构显然具备一个性质：同一父节点的两个子节点必然也共享更上层的祖先节点，例如「贵宾犬」与「拉布拉多犬」同为犬类, 它们也同为哺乳动物。



(a) Hierarchical Prototypes (b) Instance-wise & Prototypical Contrastive Selective Coding

可以通过对图像特征聚类的方式获得图像的潜在语义类别。聚类中心则可以被认为是代表着某种语义类别的“原型”，基于自底向上的层级聚类思想, 在这些聚类中心的基础上进一步进行聚类则可以得到更高层级的潜在语义类别。在这一过程中, 语义类别的树状结构自然地得以维护。

下图为论文中采用的层次 K-means 聚类算法的流程图，对聚类产生的原型，将原型进行再次聚类，一共 L 次，最终形成一棵树的结构。

Algorithm 1 Hierarchical K-means.

Input: Image representations Z, # semantic hierarchies L, # prototypes at the l-th hierarchy M_l.
Output: Hierarchical prototypes $C = \{\{c_i^l\}_{i=1}^{M_l}\}_{l=1}^{L}$, the undirected edges E between different prototypes.
$\{c_i^1\}_{i=1}^{M_1} \leftarrow$ K-means(Z).
for $l = 2$ **to** L **do**
  $\{c_i^l\}_{i=1}^{M_l} \leftarrow$ K-means$\left(\{c_i^{l-1}\}_{i=1}^{M_{l-1}}\right)$.
  for $i = 1$ **to** M_{l-1} **do**
    $E \leftarrow E \cup \left\{\left(c_i^{l-1}, \text{Parent}(c_i^{l-1})\right)\right\}$.
  end for
end for

2.1.2 选择性对比学习

有了层级信息，则可以在实例（Instance-wise）对比学习中指导负样本的选取，通过每次聚类形成的标签，将同类型（属于相同聚类中心）的负样本剔除。下图为训练过程中，对于不同的 query，删除的同类型负样本

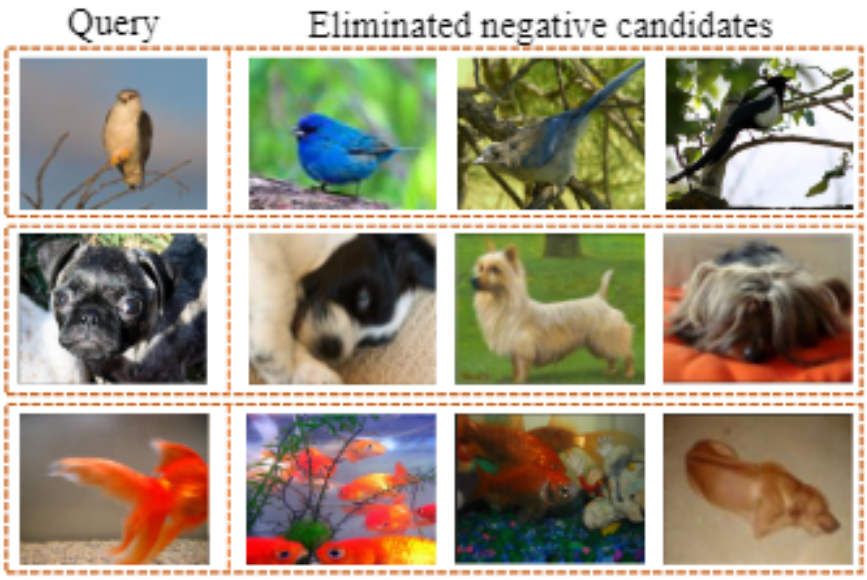


Figure 4. Visualize the query sample and the negative candidates eliminated by our pair selection approach.

类似地，层级原型还可用于辅助原型（Prototypical）对比学习，原型对比学习是图像表征与聚类中心之间的交互，在选取负原型时，可以尽量避免选取与当前聚类中心语义相近的原型，表现在树形结构上就是尽量避免选取当前节点的兄弟节点。具体来说，可以计算候选原型 c_j 与目标原型 $c^l(z)$ 父节点的相似程度，相似度大的则减小其选中的概率。

$$p^l_{Select}(c_j; c^l(z)) = 1 - \frac{\exp[s(c_j, Parent(c^l(z)))]}{\sum_{i=1}^{M_{l+1}} \exp[s(c_j, c_i^{l+1})]}$$

将两种改进后的对比学习损失进行组合得到最终的优化目标

$$\mathcal{L} = \mathcal{L}_{ICSC} + \mathcal{L}_{PCSC}$$

2.2 实验结论

对各种下游任务的大量实验验证了 HCSC 方法的优越性。该论文还给出了直观的可视化结果。在论文中结论部分展示了 HCSC 在 ImageNet 上的聚类结果，在下图中可以明显地看出存在层级结构：叼着鱼的灰熊 => 在水上的熊或者狗 => 在水上的动物。

3 个人体会

模式识别作为人工智能的一个分支领域，上世纪 50 年代以来逐渐形成了理论方法体系并快速发展，提出了统计模式识别、句法结构模式识别、神经网络、深度学习等多种有效的理论与方法，应用上也取得了巨大进展，尤其是视觉模式识别。然而，面向复杂开放环境、小样本、动态场景等挑战，当前依赖大数据学习的主流方法在泛化性、自适应性等方面存在明显不足。这篇论文在自监督学习方面做出了一定贡献，从分类学中不同层级的角度出发，将原型之间以树形结构联系起来，同时结合前人的研究成果，将 PCL 和 MoCo 的针对样本和原型的对比学习方法结合起来，取得了较好的成果，开阔了我的思维和眼界。

参考文献

[1] Kaiming He et al. “Momentum Contrast for Unsupervised Visual Representation Learning” arXiv: Computer Vision and Pattern Recognition (2019): n. pag.

[2] Junnan Li et al. “Prototypical Contrastive Learning of Unsupervised Representations” Learning (2020): n. pag.

[3] Y. Guo et al., ”HCSC: Hierarchical Contrastive Selective Coding,” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 9696-9705, doi: 10.1109/CVPR52688.2022.00948.