CVPR 2019 | 旷视研究院提出ML-GCN：基于图卷积网络的多标签图像识别模型

旷视研究院 [PaperWeekly](javascript:void(0);)

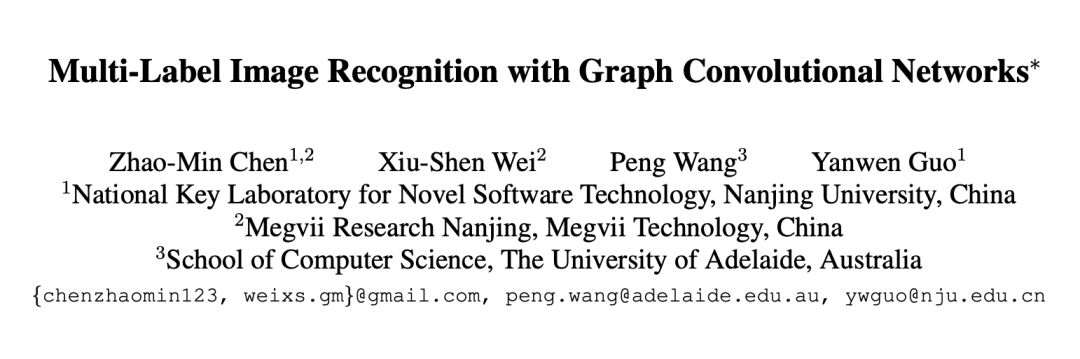
**PaperWeekly**

微信号paperweekly

功能介绍PaperWeekly是一个推荐、解读、讨论和报道人工智能前沿论文成果的学术平台，致力于让国内外优秀科研工作得到更为广泛的传播和认可。社区：http://paperweek.ly | 微博：@PaperWeekly

4月11日

全球计算机视觉三大顶会之一 CVPR 2019 （IEEE Conference on Computer Visionand Pattern Recognition）将于 6 月 16-20在美国洛杉矶如期而至。届时，旷视首席科学家、研究院院长孙剑博士将带领团队远赴盛会，助力计算机视觉技术的交流与落地。在此之前，旷视每周会推出一篇 CVPR'19 接收论文解读文章。**本文是第 5 篇，旷视南京研究院提出了一种基于图卷积网络的多标签识别模型 ML-GCN 以及一种构建相关系数矩阵的二次加权方法。**



**论文名称：**Multi-Label Image Recognition with Graph Convolutional Networks

**论文链接：**https://arxiv.org/abs/1904.03582

* **导语**
* **介绍**
* **方法**
  + 动机
  + 图卷积网络简介
  + 用于多标签识别的GCN
  + 相关系数矩阵
* **实验**
  + MS-COCO实验结果
  + VOC 2007实验结果
  + 分类器可视化
  + 在图像检索上的表现
* **结论**
* **参考文献**
* **往期解读**

**导语**

多标签图像识别（multi-label image recognition）任务的目标是预测一张图像中出现的多个物体标签，其在搜索引擎、推荐系统中应用广泛，长期以来作为计算机视觉和机器学习领域一项基础研究课题备受学界业界关注。

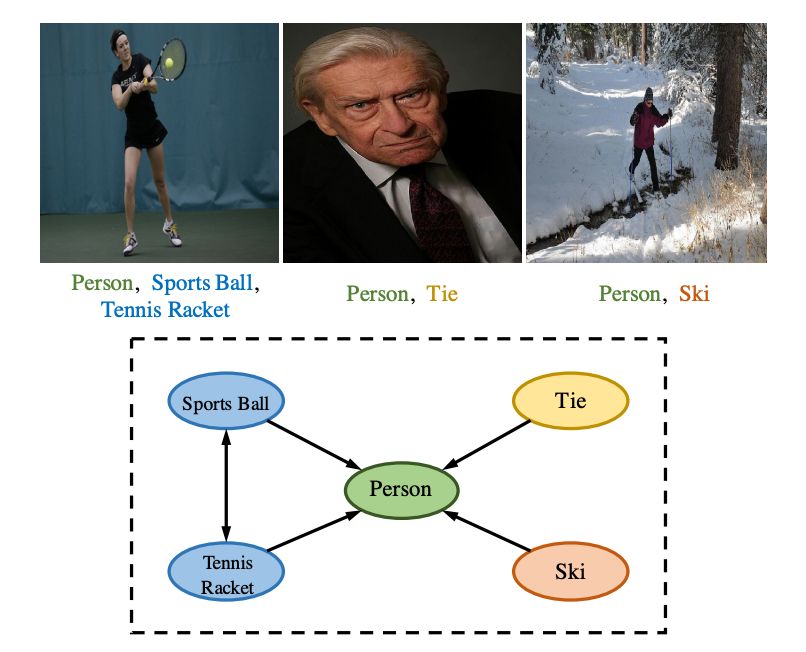
由于多个相关物体通常同时出现在一副图像之中，因此提升识别性能的一个理想方法就是针对多标记识别的核心问题，即“如何有效建模标记间的协同关系”进行探索，对标签之间的依赖性进行有效建模。

为获取和利用这种依赖性，旷视研究院提出一种**基于图卷积网络（Graph Convolutional Network，GCN）的多标签分类模型**。该模型通过 data-driven 方式建立标记间有向图（directed graph）并由 GCN 将类别标记映射（mapping）为对应类别分类器，以此建模类别关系，同时可提升表示学习能力。

此外针对 GCN 中的关键元素 correlation matrix 进行了深入分析和重设计，使其更胜任多标记问题。在两个多标签图像识别权威数据集上的实验结果显示，ML-GCN 明显优于目前所有的最佳方法。另外，可视化结果显示，模型习得的分类器还能保持有意义的语义拓扑结构。

**介绍**

多标签图像识别是计算机视觉领域的一项基本任务，其目标是识别图像中存在的一系列对象。这项技术可应用于医学诊断识别、人类属性识别和零售识别等诸多领域。相比于多类别图像分类，多标签任务的难度更大，因为其输出空间随着类别数目呈指数增大。



**▲**图1：通过构建目标标签的有向图来建模标签依赖性

由于现实世界中很多物体通常会同时出现，因此对标签之间的相关性进行建模就成了多标签图像识别的关键，如图 1 所示。

解决多标签识别问题的一个朴素方法是分离地看待各个目标，通过将多标签问题转换成多组二值分类问题，预测每个目标是否存在。由于深度卷积神经网络在单标签图像分类上取得的巨大成功，二值分类的性能已得到极大提升。

但是这些方法忽视了物体之间复杂的拓扑结构，因此在本质上有局限性。正是这个缺陷促使研究员寻找能够获取并从多个角度探索标签之间相关性的方法。其中的部分方法基于概率图模型或循环神经网络（RNN），可显式地对标签依赖性进行建模。

另一个研究方向是通过注意力机制来对标签相关性进行隐式建模。该方法考虑的是图像中被注意区域之间的关系（可视为局部相关性）。不过即便如此，该方法还是忽略了图像中标签之间的全局相关性（全局相关性需要通过单张图像之外的知识才能推断出来）。

为此，旷视研究院提出基于图卷积网络（GCN）的全新模型，即 ML-GCN（Multi-Label Graph Convolutional Network），用于建立多标签之间的相关性，该方法有其它方法无法具备的扩展性和灵活性。

除了将目标分类器视为一组待学习的独立参数向量外，旷视研究院还提出一种可从标签的先验特征（如词嵌入向量）学习相互依赖的目标分类器方法，它通过一个基于 GCN 的映射函数来实现。随后，生成的分类器再被应用于由另一个子网络生成的图像特征，以实现端到端训练。

由于这些从词嵌入向量到分类器的映射参数是在所有类别（如图像标签）之间共享，因此来自所有分类器的梯度都会影响这个基于 GCN 的分类器生成函数。这可以对标签的相关性进行隐式建模。此外，由于分类器的学习涉及到对标签相关性的建模，因此本文设计了一个有效的标签相关系数矩阵，来引导信息在 GCN 各个节点之间的传递。

**方法**

本节将介绍这一新提出的多标签图像识别模型 ML-GCN。首先是这一方法的动机，接着是一些图卷积网络初步知识，最后是 ML-GCN 模型以及用于相关系数矩阵构建的二次加权方法。

**动机**

如何有效获取目标标签之间的相关性？如何利用这些标签相关性提升分类表现？这是多标签图像识别的两个重要问题。本文使用图（graph）来对标签之间的相互依赖关系进行建模。这种方法能够灵活地获取标签空间中的拓扑结构。

旷视研究员将图中的每个节点（标签）都表示为该标签的词嵌入向量，并提出使用 GCN 直接将词嵌入向量映射到一组互相依赖的分类器上，这些分类器进一步又可直接应用于图像特征以进行分类。**基于 GCN 的模型有两个设计动机：**

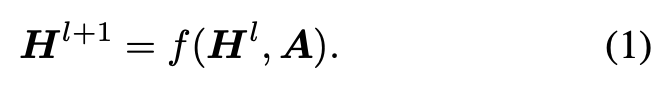
1. 由于从词嵌入向量到分类器的映射参数在所有类别中是共享的，所以习得的分类器能够在词嵌入空间中（语义相关的概念在词嵌入空间中彼此临近）保留较弱的语义结构。与此同时，对于可以对标签依赖性进行隐式建模的分类器函数，所有分类器的梯度都会对它产生影响。

2. 基于标签的共现模式，旷视研究员设计了一个全新的标签相关系数矩阵，可显式地用 GCN 建模标签相关性，让节点的特征在更新时也能从相关联的节点（标签）吸收信息。

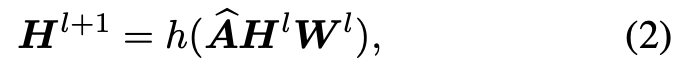
**图卷积网络简介**

图卷积网络可用于进行半监督分类任务，其核心思想是通过节点之间的信息传播来更新节点的表示。

不同于在一张图像局部欧氏结构之上进行操作的标准卷积方法，GCN 的目标是学习一个图 G 的函数 f(., .)。该函数的输入是特征描述https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZMSItA8AZrrt8W5EjDLvHmz8l1YhEVAY6I8djl2d630Z6KHibRYcGqaA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1和相关系数矩阵https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZXkXM275vaFqbiaJ0sSetd5EbuXPqcOW2k9N7T8EQyXiamJjmU5S8LkpQ/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，从而把节点特征更新为https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZcudUjpxQzJAbX0wp7cNxMmkX2YrVvU3ObiaYYclbcY2laSER7BFWPVw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1。每个 GCN 层都可以写成一个非线性函数：



f (., .) 可以表示为：

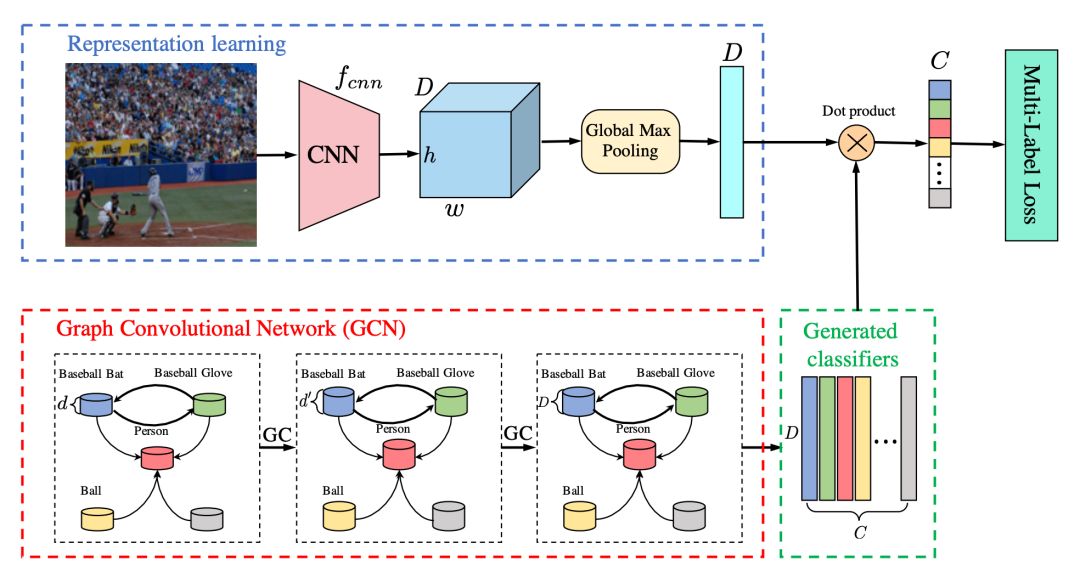


如此一来，便可以通过堆叠多个 GCN 层来对节点之间交织的复杂关系进行建模。

**用于多标签识别的GCN**

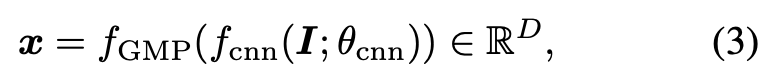
GCN 的设计初衷是半监督分类，其节点层面的输出结果是每个节点的预测分数。不同的是，在 ML-GCN 中，每个 GCN 节点的最终输出都被设计成与标签相关的分类器。此外，不同于其它任务，这里的多标签图像分类任务没有提供预定义的图结构（即相关系数矩阵）。这需要从头构建相关系数矩阵。

图 2 展示了该方法的整体架构，它包含两个主要模块：图像特征学习模块和基于 GCN 的分类器学习模块。



**▲**图2：用于多标签图像识别的ML-GCN模型整体框架

图像特征学习：原则上可使用任意基于 CNN 的模型学习图像特征。本文在实验中使用 ResNet-101 作为实验基础模型；然后应用全局最大池化获取图像层面的特征 x：

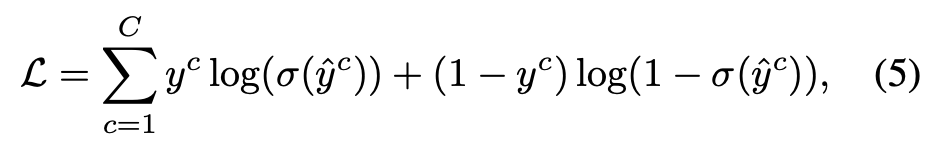


GCN 分类器学习：通过一个基于 GCN 的映射函数从标签特征学习相互依赖的目标分类器https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZETT2HItMt4AOpiasyQuFBaxmUeBERyUMia1N5ltI97VsMYf0N6gCuqow/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1。旷视研究员使用堆叠 GCN，其中每个 GCN 层 I 的输入都取前一层https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZl19XTHaJLoX9HgBnHu9ia9atq50uWQgPI6W1icKzmPwQ4H12raf3VuRA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1的节点特征作为输入，然后输出新的节点特征https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZ7zVy31WU5qwc0NLra96Z7YKiaNuNUA664PCrUficjsZW5IFb5IjD0SfA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1。第一层的输入是词嵌入向量https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZrJWWj1oY0E33k4W4hXVY7yeMZBpB08HribQ8Q9gOuBzGZicz4EuVT5Pw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，矩阵最后一层的输出是分类器https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZ03S00newNzQNy9QwRjRmVuJJ0Z5bKopGL3I7YZGXMBEaURSrRicLIicg/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1。

通过将所学到的分类器应用于图像特征，得到预测分数：



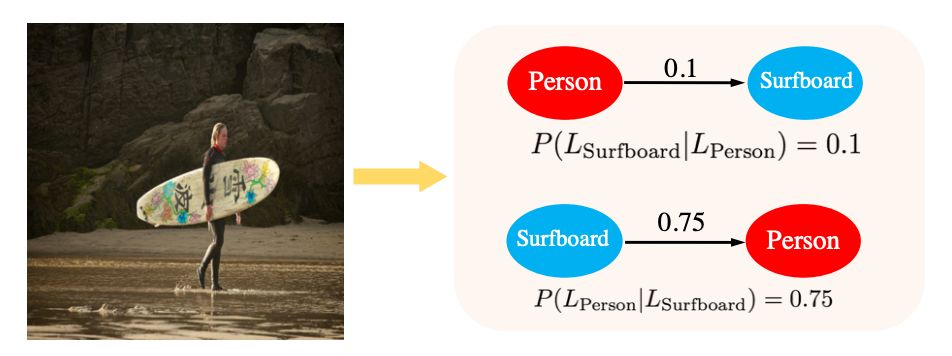
假设一张图像的真实标签是https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZ7m0xibDnE61QyrwUBtF3xGtxSx6hRvUicChiaKVRCjibFuA4P7z7TaicKbg/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，那么整个网络可使用传统多标签分类的损失函数来训练，如下：



**相关系数矩阵**

基于相关系数矩阵，GCN 可在节点之间进行信息传递，因此如何构建相关系数矩阵 A 就成了GCN 模型中一个非常重要的问题。旷视研究员在本文中以数据驱动的方式构建了一个相关系数矩阵，换句话说，相关性可以通过挖掘标签在数据集中的共现模式而来定义。

本文以条件概率的形式（即https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZWm2drNY67dP4OzVibiaMKbnbKN0Ngog2a0jV1DFeoXj0s6q6gp2tFNlw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1）对标签的相关依赖性进行了建模。如图 3 所示， 不等于https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZOBl5IyicPPnKV3ddLg8bQFKA5OOJMQRfJFfFs5JX4lpltFT7ABu62Iw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，因此相关系数矩阵不是对称的。

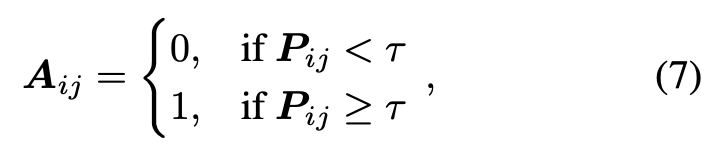


**▲**图3：两个标签之间的条件概率示意

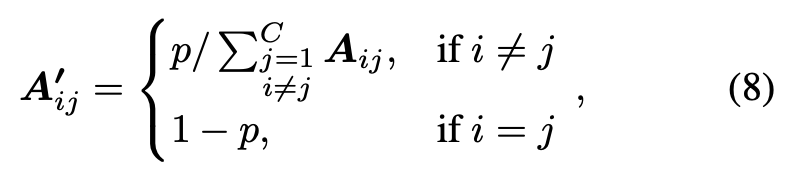
为构建相关系数矩阵，本文首先统计了训练数据集中标签对的出现次数，得到矩阵https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhglv4QwWXVKEeCbfS7OgSLqZ1jPichUccKAu6hCxmgohdHibFywFnfPZHWVIUGq1Vwpy3icjtgtqwWOQg/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，然后使用这个标签共现矩阵得到条件概率矩阵：



然而，上面这种简单的相关性可能有两个缺陷。首先，一个标签和其它标签的共现模式可能表现为长尾分布，其中某些罕见的共现可能是噪声；其次，训练和测试中共现的绝对数可能并不完全一致。因此，本文提出对相关系数矩阵进行二值化处理。具体而言，研究人员将阈值 T 用于过滤噪声边，其中 A 是二值相关系数矩阵：



过度平滑问题：根据 (2) 式可以知道，经过 GCN 后，一个节点的特征是其自身特征和相邻节点特征的加权和。而二值化相关系数矩阵的一个直接问题是其可能导致过度平滑。为了缓解这一问题，本文提出以下二次加权方法：



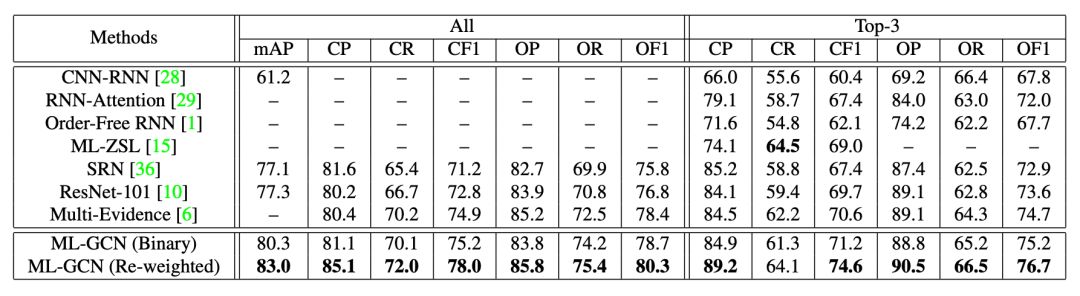
通过这种做法，在更新节点特征时，节点本身的权重是固定的，相关节点的权重则由邻近分布确定。当 p→1 时，不考虑节点本身的特征；当 p→0 时，往往忽略相邻信息。

**实验**

多标签图像识别基准数据集 MS-COCO 和 VOC 2007 上的实验结果表明，本文方法实现当前最优。

**MS-COCO实验结果**

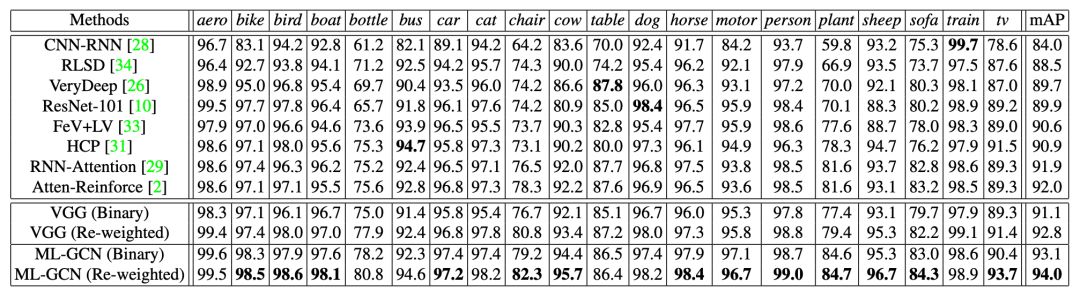
对于 ML-GCN，旷视研究员给出了基于二值相关系数矩阵与基于二次加权相关系数矩阵两个版本的结果，后者的分类表现更好，可以有效缓解上述问题，从而在几乎所有指标上领先其它方法，这证明了新提出的网络与二次加权法的有效性。具体实验结果见表 1。



**▲**表1：在MS-COCO数据集上ML-GCN与之前最佳方法的比较

**VOC 2007实验结果**

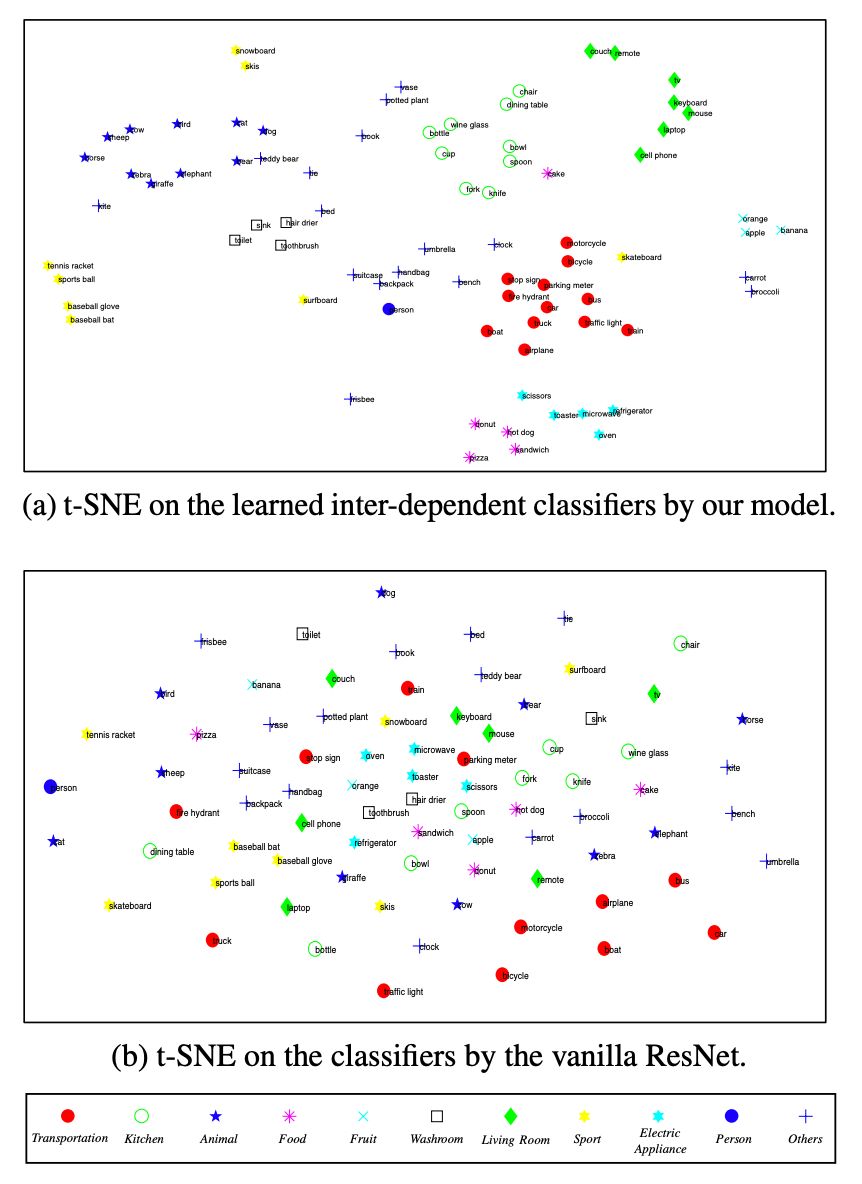
为公平对比，旷视研究员给出了以 VGG 为基本模型的结果。由表 2 可知，使用权重更新方案的 ML-GCN 模型在 mAP 指标上得到 94% 的分数，高出先前最优方法 2%。此外，即使以 VGG 为基础模型，仍然超出先前最佳水平 0.8%。



**▲**表2：在VOC 2007数据集上ML-GCN与之前最佳方法的AP与mAP比较

**分类器可视化**

旷视研究员对采用 ML-GCN 模型习得的分类器与由 vanilla ResNet（即最后一个全连接层的参数）得到的基本分类器进行了可视化分析。结果如图 8 所示。可以清楚看到，由新提出的方法学习到的分类器能够维持语义的拓扑结构。



**▲**图8：在MS-COCO上新模型学习到的互相依赖型分类器与ResNet的基本分类器的可视化

**在图像检索上的表现**

进一步，旷视研究员考察了新模型在图像特征上的学习能力，通过使用 k-NN 算法执行基于内容的图像检索验证由新模型习得的图像特征的鉴别能力，结果表明（见图 7），ML-GCN 不仅能通过有效获取标签依赖性学习更好的分类器，而且还能有助于多标签的图像特征学习。



**▲**图7：使用查询图像得到的排名前五的结果

**结论**

标签相关性建模是多标签图像识别的一大关键问题。为建模和利用这种重要信息，旷视研究院提出基于 GCN 的模型来根据先验的标签特征（比如词嵌入向量）学习互相依赖的目标分类器。

为了对标签相关性进行显式建模，文中设计了一种全新的二次加权方法，可通过平衡节点与其相邻节点来为 GCN 构建一个相关系数矩阵，以更新节点特征，从而有效缓解了妨碍 GCN 性能的两大问题：过拟合与过度平滑。定量和定性实验结果都表明新方法的优越性。

**传送门**

欢迎各位同学加入旷视南京研究院，简历可以投递给南京研究院负责人魏秀参博士。

**邮箱：**weixiushen@megvii.com

**参考文献**

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.

[2] Thomas N. Kipf, and Max Welling. Semi-supervised classiﬁcation with graph convolutional networks. In ICLR, pages 1–10, 2017.

[3] Chung-Wei Lee, Wei Fang, Chih-Kuan Yeh, and YuChiang Frank Wang. Multi-label zero-shot learning with structured knowledge graphs. In CVPR, pages 1576–1585, 2018.

[4] Xiu-Shen. Wei, Quan Cui, Lei Yang, Peng Wang, and Lingqiao Liu, “RPC: A large-scale retail product checkout dataset,” arXiv preprint arXiv:1901.07249, pages. 1–9, 2019.

[5] Zhao-Min Chen, Xiu-Shen Wei, Xin Jin, Yanwen Guo. Multi-label Image Recognition with joint Class-aware Map Disentangling and Label Correlation Embedding. In ICME, pages 1-6, 2019.

**往期解读：**

* [CVPR 2019 | 旷视提出GIF2Video：首个深度学习GIF质量提升方法](https://mp.weixin.qq.com/s?__biz=MzA3NjIzMTk0NA==&mid=2651648935&idx=1&sn=afeedb06b0dd3f838a762ebfde988874&scene=21#wechat_redirect)
* [CVPR 2019 | 旷视Oral论文提出GeoNet：基于测地距离的点云分析深度网络](https://mp.weixin.qq.com/s?__biz=MzA3NjIzMTk0NA==&mid=2651649471&idx=1&sn=8810ee23e1c5d38afa8a1bc1eea98a5e&scene=21#wechat_redirect)
* [CVPR 2019 | 旷视提出超分辨率新方法Meta-SR：单一模型实现任意缩放因子](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495978&idx=1&sn=0905887101d0b5ac21d6905aafccde97&chksm=96ea2eaaa19da7bc1b2cb872b7196d46f81bd801d7c1219c435891090d62f134fb8d476626b4&scene=21#wechat_redirect)
* [CVPR 2019 | 旷视实时语义分割技术DFANet：高清虚化无需双摄](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496126&idx=1&sn=a840b22484fea131b8bf86a2b0964a64&chksm=96ea2e3ea19da728be4732d485af38ea2bcf3a4cd8f0f9b24a6470e05b233b8128592d01fe05&scene=21#wechat_redirect)

https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhgmPEF4lW0pL5weJia5y4xhJbog2pIZZ3ZCgVUDynvus6rCzNKGAAAI6R8jaXTpYPISCMicpFegVdG0g/640?wxfrom=5&wx_lazy=1&wx_co=1

**点击以下标题查看更多往期内容：**

* [CVPR 2019 | 无监督领域特定单图像去模糊](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495696&idx=1&sn=8668a9932f1012d0fd6225b183658fe5&chksm=96ea2f90a19da686dc1ff257877edfdad1fb0257beba8611e85dfaf3f7096f2b2a0b55435cd9&scene=21#wechat_redirect)
* [图神经网络综述：模型与应用](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247493906&idx=1&sn=15c9f18a1ce6baa15dc85ecb52e799f6&chksm=96ea3692a19dbf847c1711e6e194ad60d80d11138daf0938f90489a054d77cfd523bee2dc1d2&scene=21#wechat_redirect)
* [近期值得读的10篇GAN进展论文](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247493987&idx=1&sn=ce1bcdce28e78f4a307743e389f42b10&chksm=96ea36e3a19dbff5cff7f4f1c9d9fc482bb2144d80566319b3d26bce4d9ab80689d38ab2e427&scene=21#wechat_redirect)
* [小样本学习（Few-shot Learning）综述](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496040&idx=1&sn=ebea875571ccc80a70a3166f62b9640b&chksm=96ea2ee8a19da7fe9bf3d9cbf15c3a2fc6e3ca898e48a2d01207a7cdc2cd046c4087eda27637&scene=21#wechat_redirect)
* [万字综述之生成对抗网络（GAN）](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495668&idx=1&sn=e7e959b2bdd7b2763b9207ccb80fa6bc&chksm=96ea3074a19db96208a51d26f7b5b4ef9c3a37a7799ec270becc77203de4294235041ede7206&scene=21#wechat_redirect)
* [可逆ResNet：极致的暴力美学](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495880&idx=1&sn=e3a960836797c7dd4eb2dbb9b9a0b9e7&chksm=96ea2f48a19da65eea9cc157f2110a6856985d322d872c6679a8dd0807c9bb10f85c4d3c3f7d&scene=21#wechat_redirect)
* [小米拍照黑科技：基于NAS的图像超分辨率算法](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495166&idx=1&sn=a158e603651bc4f26836151a9113e856&chksm=96ea327ea19dbb68b8987aca041bb21579a35b1c679e91fd2368c7f2fb7acd58508cd531bdfe&scene=21#wechat_redirect)
* [AAAI 2019 | 基于区域分解集成的目标检测](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495377&idx=1&sn=a86a98d28c58925c417cefb4e4dcb79f&chksm=96ea3151a19db8473483213424ae62c2d782f3f81452379efb7270c55993e0d3aed31f56cc66&scene=21#wechat_redirect)