Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning

为损失函数定个框架，码隆CVPR 2019提出图像检索新范式

原创： 思 [机器之心](javascript:void(0);) 1周前



机器之心原创

**作者：思源**

全球计算机视觉顶级会议 CVPR 2019 已于今日在美国长滩落幕。从日前公布的[获奖论文中](http://mp.weixin.qq.com/s?__biz=MzA3MzI4MjgzMw==&mid=2650764315&idx=1&sn=3462076d663fdbd9101ee752b79b1f46&chksm=871ab065b06d3973418c5437bc3fbd6cd7a8d72d11810dcf48ad95102d010a91abf19dd0dd8d&scene=21" \l "wechat_redirect" \t "_blank)，我们能够发现华人声音在该领域愈为响亮，研究成果也备受关注。作为今年大会的黄金赞助商，码隆科技在 CVPR 2019 上不仅发表了优秀的研究工作，同时还与 Google Research 等合办了 FGVC6 Workshop，主办了细粒度商品识别挑战赛。本文将对码隆科技的 CVPR 2019 论文进行解读，介绍码隆是如何为图像搜索任务提出一个通用的损失函数框架，并将该领域近十年的损失函数都统一在该框架下。

对于很多研究者而言，以前我们针对图像搜索任务设计损失函数并没有统一的框架，很多研究者都通过直观理解尝试新的损失函数。但在码隆科技的这篇论文中，研究者探索了图像搜索的核心问题：即如何为损失函数的设计提供一个标准框架，从而通过深度度量学习实现更优质的图像检索。

* 论文：Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning
* 论文地址：https://arxiv.org/pdf/1904.06627.pdf

**统一的损失函数框架：GPW**

General Pair Weighting（GPW）是一种通用样本对加权框架，它希望从底层理解图像检索中的损失函数（深度度量学习）。简单而言，GPW 通过

梯度分析将 深度度量学习 转化为 样本对 的加权问题，

为理解 基于样本对的损失函数提供了统一的视角和有力的工具。

注意这里有两个关键点，即度量学习和基于样本对的损失函数。首先在机器学习中，

度量学习的目的在于学习一种低维空间，

在这个空间内同类样本相距非常近，异类样本距离比较远。

深度度量学习利用卷积网络的强大的特征抽取能力，能学习更好的嵌入空间。

其次，GPW 框架关注的是基于样本对的损失函数，

这类损失函数一般可以表示为嵌入空间中的成对余弦相似性。

例如最直观的对比损失（Contrastive loss），

它希望正样本对（或同类样本对）越近越好，负样本对之间至少要有大于某个给定的距离，具体而言可以表示为：

https://mmbiz.qpic.cn/mmbiz_png/KmXPKA19gWicm0oO9uF8sSnOvNxEY1WicgNBXIClADVchYtUGIZYeQwYN0rHRMprdY1nlhoP9u1NHdPYWYVdffiaA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1

其中 I 表示指示函数，如果 i 和 j 为正样本，则 I\_ij 为 1，否则它就等于 0。我们可以了解到，正样本对的损失就直接是负的相似性 -S\_ij，负样本对只有在相似度高于阈值 λ时损失函数才大于０。从这个简单的案例，我们可以发现这类损失函数的目标，即在嵌入空间中将标签相同的同类样本拉得更近，将标签不同的异类样本推地更远。

既然这类损失函数的基本思想、目标都有这样的共同点，那么我们是不是能从本质上理解它们，并提出一种统一的框架呢？这就正是 GPW 希望做到的。

**从归纳到统一**

为了完成统一，码隆 AI 中心的研究者探索了十多年来比较常用的基于样本对的损失函数，并发现它们的一些本质区别。研究者表示：

「不同损失函数其实是在给予不同的力度来拉近或推远某个样本。例如

对比损失，它的基本思想即平均拉近或拉远所有样本；又例如

三元组损失（triplet loss），它会选取部分样本来实现拉近或拉远。」

既然各种损失函数都在给样本加权以确定「拉或推」的力度，那么使用什么样的函数来加权就非常重要了。在 GPW 框架下，我们可以了解各种损失函数是怎样做这种加权的。因为 GPW 本身相当于一种理论分析，它把各种基于样本对的损失函数都放在一个统一的框架下，所以我们能快速了解各种损失函数如何做加权这一本质。研究者表示，这一本质即是

在学习不同样本对时给它们不同的权重，

如果某个样本包含的信息比较多或比较难学习，那么它就需要比较大的权重。

如果我们再从 GPW 出发解释已有的损失函数，那么就能把它们的加权方式直接写出来，这是过去没有人做到的一点。在原论文中，作者们就从 GPW 框架出发解释了 Contrastive、Triplet、Lifted Structure 等多种基于样本对的损失函数，感兴趣的读者可查阅原论文。

**更强大的 Multi-Similarity 损失函数**

其实 GPW 还有更重要的实践意义，我们可以在它定义的框架下构建性能更强大的损失函数。码隆 AI 中心的研究者就设计一种名为

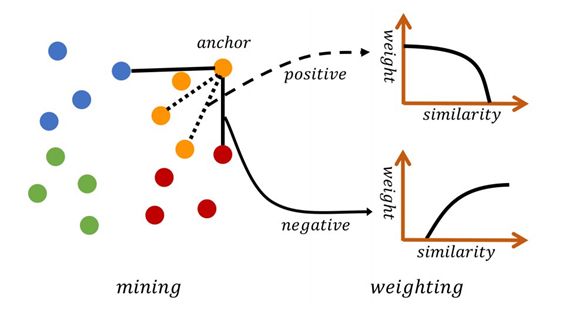
Multi-Similarity 的损失函数（MS Loss），它可以显著提高图像搜索的性能。研究者表示，该损失函数在多个主要的图像检索基准数据库上都获得了当时最好的结果。如下图所示，MS Loss

通过采样和加权两次迭代，实现更加高效的样本训练。

它通过定义自相似性和相对相似性，在训练过程中更加全面地考虑了局部样本分布，从而能更高效精确的

对重要样本对进行采用和加权。这里重要样本对通常是

含有更大的信息量的样本对。



*原论文图 1：MS Loss 的两个阶段，其结合了三种相似性。*

在 MS Loss 的第一阶段采样中，它会通过某种相似性判断哪些样本对学习嵌入空间更重要。例如上图连着黑线的蓝色、红色样本点，它们就比同色的其它样本点重要，信息量更大，因为黄色样本点要是想与这两种样本分离，那就需要把这两个样本推开。

第二阶段的加权则是在采样的基础上进行的，第一阶段采样的样本有正样本（黄色）也有负样本（红色、蓝色），我们需要另外两种相似性来确定它们的重要性。具体而言，

如果正样本相似性越高，那么加的权就越少，因为它已经不太需要进一步拉近距离。

但如果负样本的相似性太高，就表示它离 anchor 非常近，我们需要提供更大的权重以令模型学会把它们推地更远。

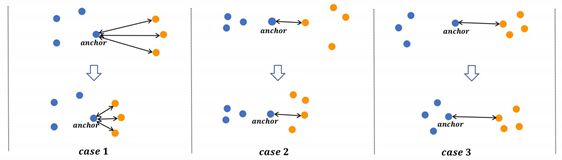
**MS Loss 定义的三种相似性**

前面我们看到 MS Loss 综合考虑了三种相似性，它们能概括目前大多数基于样本对的损失函数。它们可以分为自相似性与相对相似性，正如作者所言，给样本对加权的核心在于判断局部分布 - 即它们之间的相似性，局部样本之间的分布和相互关系并不是仅仅决定于两个样本之间的距离和相似性，还取决于当前样本对与其周围样本对之间的关系。

因此，对于每一个样本对，我们不仅需要考虑样本对本身的自相似性，同时还要考虑它与其它样本对的相对相似性。其中相对相似性又可以分为正相对相似性 (正样本)、负相对相似性（负样本）两种相似性，它们三者共同构建了上图 1 的 MS Loss 两步迭代策略。

* 自相似性（余弦相似性）：随着负样本接近 anchor，样本对相似性增加；
* 负相对相似性（Negative relative similarity）：随着周围负样本聚集在一起，样本对相似性降低；
* 正相对相似性（Positive relative similarity）：随着周围正样本聚集在一起，样本对相似性降低。

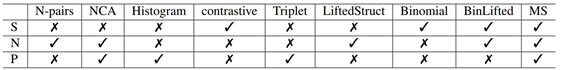
下图展示了上面三种相似性的直观变化，我们希望计算实线样本对间的相似性：



*原论文图 2：负样本对间的三种相似性，从左到右分别为自相似性、负相对相似性、正相对相似性。*

图 2 中 case 1 的相似性是增加的，case 2 和 case 3 的相似性都将降低。因为 case 1 中的自相似性很明显没有考虑与周围样本的关系，所以作者引入了后面两种相对相似性。在 case 2 中，即使 anchor 与负样本的自相似性不变，但我们还需要考虑负样本的近邻。如果负样本间的自相似性增加，那么它与 anchor 间的相对相似就自然降低。case 3 也是同样的道理，正样本间的自相似要是增加了，那么 anchor 与负样本的相对相似性就要降低。

如下表 1 所示，基于样本对的各种损失函数都可以归类到这三种相似性，目前只有该论文提出的 MS Loss 能同时考虑三种相似性。



*原论文表 1：不同损失函数为样本对加权所采用的相似性度量，其中 S、N、P 分别表示自相似性、负相对相似性和正相对相似性。*

作者表示：「其它损失函数之所以没能全部考虑三种相似性，是因为它们并没有特意关注这一本质区别，很多损失函数都是从直觉的角度设计的，因此只考虑到其中一种相似性度量，能考虑到两种的情况都很少。」此外，如果不能将三种相似性都构建进去，那么度量方法总会有一些缺陷，效果也就达不到最优。

**MS Loss 表达式**

前面已经了解到，MS Loss 采用采样和加权交替迭代的训练策略来实现上述三种相似性。研究者表示他们实际上也是站在前人的肩膀上，巧妙地融合表 1 中 Triplet、Lifted Structure 和 Binomial　Deviance 三种损失函数，它们都只采用三种相似性中的一种。

作者说：「我们需要分两步完成目标。因为我们很难构建单个训练和优化步骤来同时实现三种相似性，这样的公式不仅复杂，同时还容易产生冲突。此外，分为两步实现也不会违反 GPW 的基本思想。

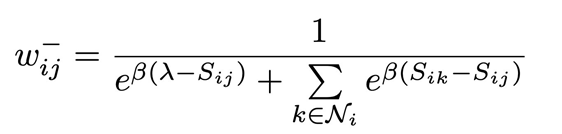
因为第一步的采样就是抛弃一些非常不重要，信息量很少的样本，可以看作将它们的权重设置为 0。这个权重与第二步的权重类似，它们结合起来可以视为一种加权方法。」

总体而言，第一步通过正相对相似性确定哪些是信息量大的样本对，然后第二步通过自相似性和负相对相似性为重要的样本对进一步赋不同的权重。

对于正相对相似性，我们可以度量在相同 anchor 下正样本和负样本对之间的相似性。具体而言，如果 anchor 与负样本的相似性比它与最不相似的正样本对还要大，那么该负样本就是重要的样本。

同理，如果 anchor 与正样本的相似性比它与最相似的负样本还要小，那么该正样本也含有重要信息。这两者加起来就是该 anchor 第一步选出的重要样本。

对于第二步的两个相似性，研究者结合了 binomial deviance 和 lifted structure 两种损失函数，且分别利用了自相似性和负相似性。例如给负样本加权，那么计算式可以表示为：



其中 λ-S\_ij 表示自相似性，S\_ik-S\_ij 表示相对相似性，上式将两种相似性结合起来。如果将分母的这两项分开，那么它就和前面两种损失函数非常像。与上式类似，给正样本加权也由这两部分组成。

最后，作者将采样和加权策略结合起来，从而产生了 MS Loss 这种新型基于成对样本的损失函数。这里省略了很多具体表达式，包括第一步的挑选准则与最后的融合表达式等等。

因为 MS Loss 在大部分图像检索基准数据库上都有很好的性能，且相比最新的方法也有较大的优势，所以想要试试该损失函数的读者不妨查阅原论文细节。

**细粒度商品识别挑战赛**

在 CVPR 2019 中，码隆除了在图像检索方面提出非常优秀的算法，同时在细粒度识别方面还举办 [iMaterialist Challenge on Product Recognition 挑战赛](http://mp.weixin.qq.com/s?__biz=MzA3MzI4MjgzMw==&mid=2650759697&idx=5&sn=48fe55f18575f084fd41820f7a41245a&chksm=871aa66fb06d2f79d294c5034d5f347415e0f62e3f6b24c9d7fb9f5b2247e77decedcb6ae774&scene=21#wechat_redirect)。通过主办此次竞赛，码隆科技希望能够引发学界和产业界对商品识别这一技术领域的更多关注，共同探究提升细粒度物体识别的算法性能，探索出更好、更强的技术路径。

据码隆介绍，该挑战赛共涵盖 2019 类 SKU，超过一百万图像数据，是 CVPR 迄今数据规模最大、种类最多的商品识别竞赛。两个月的赛程中，全球共有 96 支队伍、152 位选手通过 1600 次提交参加了竞赛。

美国时间 6 月 17 日，码隆科技在 FGVC6 Workshop 上公布大赛最终结果，前三名分别为京东 AI 研究院、美团点评视觉图像中心和东信北邮。*https://mmbiz.qpic.cn/mmbiz_png/KmXPKA19gW8Zfpicd40EribGuaFicDBCRH6IOu1Rnc4T3W3J1wE0j6kQ6GorRSgicib0fmNrj3yzlokup2jia9Z0YVeA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1*