

论文 VICREG 的复现

董火荣

摘要

VICREG 是一种图像编码器的自监督方法，具有规则简单、准确率高的优点。VICREG 论文的贡献主要是：（1）提出了一种明确避免崩溃问题的方法，（2）只使用简单的三个规则，与现有的主流技术具有相当的水平，（3）Vicreg 的正则化可以稳定其他方法的训练。基于 VICREG 论文，本文主要复现的内容是在 ImageNet 数据集上训练 vicreg 预训练模型、冻结 vicreg 模型参数，应用到分类任务、对比不同数据集上 vicreg 模型的表现和将 vicReg 应用到下游任务。最后将自己的实验结果与作者的实验结果作对比，分析结果表现。

关键词：自监督；编码器；崩溃；正则化条款

1 引言

最近用于图像表征学习的自监督方法提高了由同一图像的不同视图的编码器产生的嵌入向量之间的一致性。现在主要的挑战是如何防止编码器产生恒定或无信息的向量的崩溃。文中引入了 VICReg（方差-不变-协方差正则化）^[1]，这是一种明确避免崩溃问题的方法，有两个正则化条款分别适用于两个嵌入：（1）将每个嵌入维度的方差保持在一个阈值之上，（2）将每对变量进行去相关操作。与同一问题的大多数其他方法不同，VICReg 不需要诸如以下技术：分支之间的权重共享、批量归一化、特征归一化、输出量化、停止梯度、存储库等，并且在几个下游任务上取得了与技术现状相当的结果。此外，文中的方差正则化项稳定了其他方法的训练，并导致了性能的提高

2 相关工作

图像的自监督编码器主要由以下 4 种。

2.1 对比性学习

在应用于联合嵌入架构的对比 SSL 方法中，一个样本的输出嵌入及其失真版本彼此接近，而其他样本及其失真被推开。该方法最常用于孪生架构，其中两个分支具有相同的架构并共享权重。许多作者使用 InfoNCE 损失，其中排斥力对于更接近参考的对比样本更大。虽然这些方法产生了良好的性能，但它们需要大量的对比对才能正常工作。这些对比对可以从内存库中进行采样，如 SimCLR 中所示，或由当前批数据提供，具有显著的内存占用。对比方法的缺点促使人们寻找替代方法

2.2 聚类方法

基于聚类的方法不是将每个样本视为其自己的类，而是基于某种相似性度量将它们分组为集群。DeepCluster 使用先前迭代中表示的 k-means 赋值作为新表示的伪标签，这需要一个昂贵的异步聚类阶段，并使该方法难以扩展。SwAV 通过在线学习集群来缓解这个问题，同时通过 Sinkhorn Knopp 变换保持分配的平衡分区。这些聚类方法可以被视为聚类级别的对比学习，这仍然需要大量的负面比较才能正常工作

2.3 蒸馏方法

最近的提议，如 BYOL、SimSiam、OBoW 和变体都表明，通过使用受知识蒸馏启发的架构技巧，可以避免崩溃。这些方法训练学生网络以预测教师网络的表示，对于教师网络，权重是学生网络权重的运行平均值，或者与学生网络共享，但是没有梯度通过教师反向传播。这些方法是有效的，但对于它们为什么以及如何避免崩溃还没有明确的理解。或者，图像可以表示为视觉特征字典上的单词包，这有效地防止了崩溃。在 OBoW 中，字典通过离线或在线聚类获得。相比之下，我们的方法明确地防止了两个分支中的独立崩溃，这消除了对共享权重和相同架构的要求，为将联合嵌入 SSL 应用于多模态信号打开了大门。

2.4 最大化信息方法

防止崩溃的一个原则是最大化嵌入的信息内容。最近提出了两种这样的方法：W-MSE 和 Barlow Twins。在 W-MSE 中，一个额外的模块将嵌入变换到其协方差矩阵的本征空间（白化或 Karhunen-Loeve 变换），并迫使由此获得的向量均匀分布在单位球面上。在 Barlow Twins 中，损失项试图使来自两个分支的嵌入向量的归一化互相关矩阵接近同一性。这两种方法都试图产生相互去相关的嵌入变量，从而防止变量携带冗余信息的信息崩溃。因为所有变量在一批中都是标准化的，所以它们既不收缩也不膨胀。这似乎足以防止崩溃。我们的方法借鉴了大麦双胞胎的去相关机制。但它包含两个嵌入的每个变量的显式方差保持项，因此不需要任何规范化。

3 本文方法

3.1 本文方法概述

VICReg 是基于联合嵌入架构的。该架构可能是完全对称的或完全不对称的，两个分支之间没有共享的结构或参数。在文章的大多数实验中，使用的是连体网架构，其中两个分支是相同的，并且共享权重。每个分支包括一个编码器，它输出表征（用于下游任务），然后是一个扩展器，它将表征映射到嵌入空间，损失函数将在嵌入空间里计算。扩展器的作用有两个方面：（1）消除这两种表征形式不同的信息。（2）以非线性的方式扩展维度，以便对嵌入变量进行去相关操作，从而减少表征向量之间的依赖关系。损失函数使用一个距离项 s 让 Z 和 Z' 尽可能接近，用一个方差项 v 防止崩溃，和用一个协方差项 c 对向量的不同维度进行去相关操作。在预训练之后，扩展器被丢弃，编码器的表示被用于下游任务。vicReg 方法的核心结构如图 1 所示：

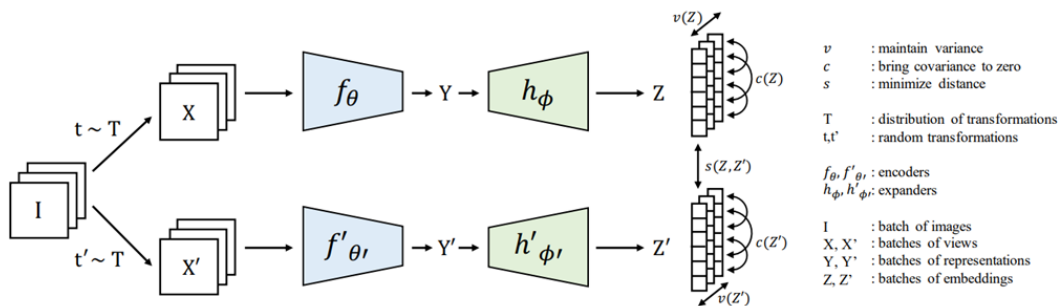


图 1: VICREG 核心结构

3.2 Variance 项

不同图像的方差要高于一个阈值，可以保证对不同图像的输出具有差异性，避免崩溃。用方差来表示图像之间的差异性

$$v(Z) = \frac{1}{d} \sum_{j=1}^d \max(0, \gamma - S(z^j, \epsilon)),$$

$$S(x, \epsilon) = \sqrt{\text{Var}(x) + \epsilon},$$

3.3 Invariance 项

让同一张图像的距离尽可能地近，这里用的是欧式距离。保证特征提取器的正确性。

$$s(Z, Z') = \frac{1}{n} \sum_i \|z_i - z'_i\|_2^2.$$

3.4 Covariance 项

让协方差矩阵尽可能地趋近于 0，降低不同图像的表征之间的相关性。

$$C(Z) = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})(z_i - \bar{z})^T, \text{ where } \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i.$$

$$c(Z) = \frac{1}{d} \sum_{i \neq j} [C(Z)]_{i,j}^2.$$

4 复现细节

4.1 作者做的主要实验

(1) imageNet 上对 VicReg 模型进行预训练。(2) 基于 vicReg 预训练模型，在 ImageNet 上冻结表征进行线性分类任务。(3) 基于 vicReg 预训练模型，在 0.01 和 0.1 比例 ImageNet 样本的微调表征之上进行半监督分类任务。(4) 基于 vicReg 预训练模型，在冻结表征的基础上进行线性分类任务，在 Places205 和 iNat18 数据集上计算 Top-1 准确率，在 VOC07 数据集上计算 mAP。(5) 基于 vicReg 预训练模型，在 VOC07+12 数据集上，利用 AP50 指标评估使用 C4 骨架的 Faster R-CNN 的具有微调功能的目标检测模型。(6) 基于 vicReg 预训练模型，在 COCO 数据集上，使用带有 FPN 主干网络的 Mask R-CNN 完成目标检测和实例分割任务，并用 AP 指标进行模型评估。

4.2 作者提供的代码

作者仅提供部分代码，代码内容涵盖 6.2 的 (1) 到 (3) 的实验。其余实验均没有代码。

4.3 本人完成的工作

根据作者给出的代码，结合自己已有设备特殊性对代码进行微调，保证模型参数不变的情况下顺利地让代码跑通，并完成了 (1) 到 (3) 的实验。其中实验 (1) 用时 3 天，实验 (2)，实验 (3) 用时

几乎都是一天。实验（4）作者没有给出具体代码，本人根据文章要求，基于 vicReg 预训练模型，冻结表征，实现了在 iNat18 数据集上计算 Top-1 准确率的任务。同时我还计算了 VOC07 的 top1 准确率，由于 Places205 数据集目前处于维护状态，故暂时没有涉及。本人做的实验和作者做的实验差异如下：

表格 1 作者做的实验

实验	任务	Top1	Top5	epoch	数据集
(1)	线性分类	73.2	91.1	100	ImageNet
(2)	1%半监督	54.8	79.4	20	ImageNet
(3)	10%半监督	69.5	89.5	20	ImageNet
(4)	线性分类	47.0	----	---	iNat
(5)	线性分类	86.6 (mAp)			

表格 2 本人做的实验

实验	任务	Top1	Top5	epoch	数据集
(1)	线性分类	63.95	85.68	100	ImageNet
(2)	1%半监督	35.43	63.75	20	ImageNet
(3)	10%半监督	63.53	86.02	20	ImageNet
(4)	线性分类	52.45	75.69	100	iNat
(5)	线性分类	41.27	94.27	100	VOC07

表格 3 作者的实验

实验	任务	Top1	Top5	epoch	数据集
(1)	线性分类	47.0	----	84	iNat
(2)	线性分类	86.6(mAp)	---	28	VOC07
(3)	目标检测	82.4(Ap50)	---	24K iteration	VOC07+12
(4)	目标检测	39.4(Ap)	---	90K iteration	coco

表格 4 本人的实验

实验	任务	Top1	Top5	epoch	数据集
(1)	线性分类	52.45	75.69	100	iNat
(2)	线性分类	41.27	94.27	100	VOC07
(3)	目标检测	0.01(Ap50)	---	20	VOC07+12
(4)	目标检测	Loss NaN	---	1	coco

5 实验结果分析

由于作者的模型过大，本人的硬件资源有限，所以迭代次数相比作者训练的要少，故实验精度有所下降。表格 3 和 3 是作者没有给出代码的部分，需要自己实现代码。在做这部分实验做到一半的时候，实验室服务器坏掉，用了我自己的笔记本跑，会对实验效果造成一定影响。

6 总结与展望

本人会关注 VICREG 模型的进展，关注众人对其的改进，并完善自己的实验

参考文献

- [1] ADRIEN BARDES Y L, Jean Ponce. VICREG: VARIANCE-INVARIANCE-COVARIANCE REGULARIZATION FOR SELF-SUPERVISED LEARNING[J]. ICLR, 2022.