Unsupervised Representation Learning By Pre-Dicting Image Rotations

1. 摘要

在我们的工作中，我们建议通过训练ConvNet来学习图像特征，以识别应用于它作为输入的图像的2d旋转。

1. 问题重述

因此，最近人们对以无监督的方式学习高水平的基于ConvNet的表征越来越感兴趣，从而避免了对视觉数据的人工标注。其中，一个突出的范式是所谓的自监督学习，它定义了一个无注释的前提任务，只使用图像或视频上的视觉信息，以便为特征学习提供一个替代的监督信号。无监督特征学习的其他成功案例包括基于聚类的方法、基于重建的方法，以及涉及学习生成概率模型的方法。

Unsupervised feature learning：

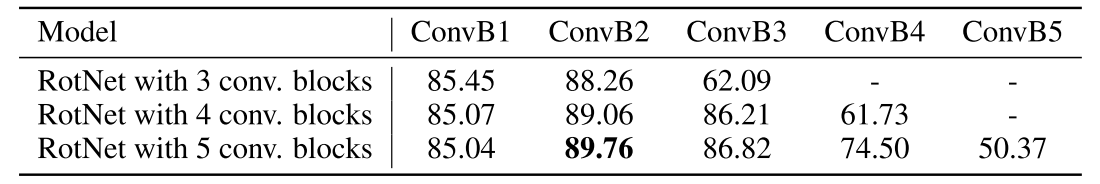
1. Self-supervised learning
2. Clustering based methods
3. Reconstruction based methods
4. Methods that involve learning generative probabilistic models

我们的工作遵循自监督的范式，建议通过训练ConvNets来学习图像表征，以识别应用于其输入的图像的几何变换。更具体地说，我们首先定义一小群离散的几何变换，然后将这些几何变换应用于数据集上的每一幅图像，并将产生的变换图像输入到ConvNet模型中，该模型被训练来识别每一幅图像的变换。在这种表述中，实际上是几何变换的集合定义了ConvNet模型必须学习的分类借口任务。因此，为了实现无监督的语义特征学习，正确选择这些几何变换是至关重要的。我们建议将几何变换定义为图像的0、90、180和270度旋转。因此，ConvNet模型是在识别四种图像旋转中的一种的四向图像分类任务上训练的（见图2）。

1. 详细阐述

没啥可讲的，就是设计了一个新的自监督辅助任务。

1. 实验



我们在表1中报告了CIFAR-10测试集的准确性结果。我们观察到，在所有情况下，由第2个卷积层块（就截至该点的卷积层总数而言，该块实际上有6个深度）生成的特征图达到了最高的准确率，即88.26%到89.06%。第二层之后的信念块的特征逐渐降低了物体识别的准确性，我们认为这是因为它们在旋转预测的自我监督任务中开始变得越来越具体。另外，我们观察到，增加RotNet模型的总深度会导致由早期层（以及第1个conv.块之后）产生的特征图的物体识别性能提高。我们认为这是因为增加模型的深度，从而增加其头部（即顶部ConvNet层）的复杂性，使得早期层的特征对旋转预测任务的具体影响较小。

我们观察到，确实对于4个离散的旋转（如我们所建议的），我们取得了比8个或2个情况更好的物体识别性能。我们认为这是因为2个方向的情况下提供的识别类别太少（即提供的监督信息较少），而在8个方向的情况下，几何变换的区分度不够，此外，引入的4个额外旋转可能导致旋转后的图像出现视觉伪影。此外，我们观察到，在用2个离散旋转训练的RotNet模型中，用90◦和270◦旋转训练的RotNet模型比用0◦和180◦旋转训练的模型取得了更差的物体识别性能，这可能是由于前一个模型在无监督阶段没有 "看到 "通常在物体识别训练阶段使用的0◦旋转这一事实。

1. 评析