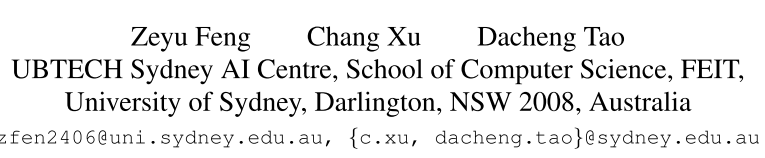
Self-supervised representation learning by rotation feature decoupling



1. 摘要

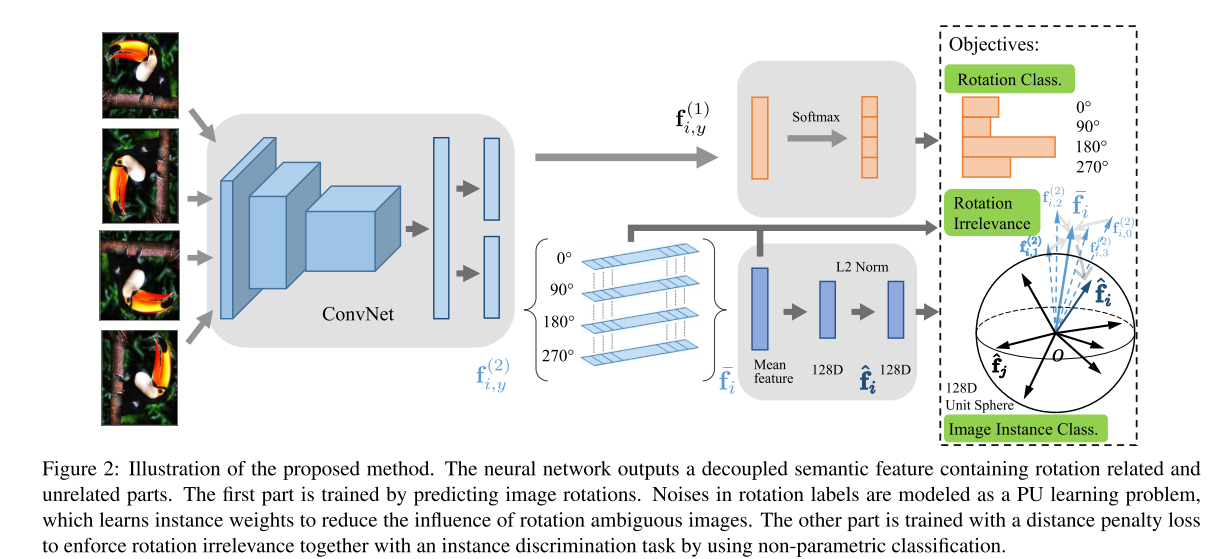
我们介绍了一种自监督的学习方法，该方法专注于表征的有益特性以及它们对现实世界任务的概括能力。该方法将旋转不变性纳入特征学习框架，这是视觉表征的许多良好和研究的属性之一，以前基于深度卷积神经网络工作的自监督表征学习方法很少欣赏或利用。具体来说，我们的模型学习了一个包含旋转相关和不相关部分的分离表征。我们通过联合预测图像旋转和辨别单个实例来训练神经网络。特别是，我们的模型将旋转判别与实例判别相分离，这使我们能够通过减轻旋转标签噪声的影响来改善旋转预测，并在不考虑图像旋转的情况下判别实例。由此产生的特征对更多不同的任务具有更好的泛化能力。

1. 问题重述

现有的工作大多集中在设计各种辅助任务，而很少关心学习到的表征拥有哪些属性，以及它们是否真的有利于现实世界任务中的概括。例如，高层次的表征应该传达一个清晰的解释或某些变量因素的依赖性[5]。最近的一个尝试是预测图像的旋转[17]。在这种方法中学习到的特征可以在各种任务中得到很好的归纳，并取得最先进的性能。然而，这些特征对旋转变换是有辨别力的，因此不能使支持旋转不变性的视觉任务受益。此外，值得注意的是，在实践中并非所有的例子都是可以旋转确定的。图像的方向不仅对圆形物体是模糊的，而且对图像中的许多其他物体也是方向不可知的，例如，从上面看的一些物体或对称的形状，如图1所示。旋转这些物体不会对我们的描述或理解产生重大影响。（作者觉得别人的工作都是在做辅助特征设计，但是对学习到的表征到底怎么样并不关心。他认为表征的旋转特性很重要，但是不是所有的物体都适合旋转特征，有些不具有旋转特性。所以他要解耦旋转特性。实际上，他也没有办法完全解释CNN学习到的表征拥有哪些属性，他只是觉得旋转属性很重要，并且做了实验来验证。）

在本文中，我们提出了一种新的自我监督学习算法，该算法通过旋转预测任务和实例判别任务对表征进行解耦。学习到的实例特征由两个成分组成，分别是旋转判别性和旋转无关性。旋转判别性特征可以通过预测图像的旋转来发现，这种方法简单而有效，在一些基准上取得了最先进的结果[17]。关于数据集中那些方向不可知的图像，自动分配的旋转标签通常包含噪声，这自然导致了一个positive的无标签学习的概率困难。默认方向的原始图像是积极的实例，而旋转后的副本是无标签的实例，可以是积极的也可以是消极的。如果旋转后的副本的变换不能被明确识别，我们就把它当作未标记集中默认方向的阳性实例（见补充材料中的图1）。另一方面，我们通过惩罚同一图像在不同旋转下的特征之间的距离差来学习旋转无关的特征。非参数方法被用来区分基于这些旋转无关的特征的不同实例。因此，这些特征在实例层面上具有判别能力。

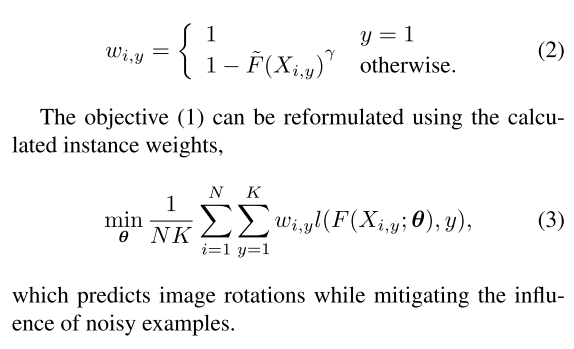
1. 详细阐述



* 1. Noisy rotated images

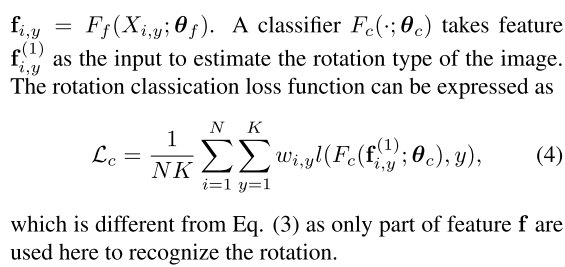
在旋转预测中引入的先决条件是模型可以满足大多数自然图像的要求，这些图像中的物体一般都处于正面姿态。这类图像通常有一个默认的方向。图像的任何旋转都会导致一个不寻常的物体方向，这一点可以毫无疑问地由人眼来指定。ImageNet等数据集中的许多实例都有这样的观察结果，适合于旋转预测的任务。尽管它简单而有效，但这一前提对于图像中许多方向不确定的物体来说是失败的，例如一些从顶部看的物体或对称的形状（见图1）。识别这些图像的精确旋转变换在实践中是没有意义的，而且在任何情况下不加思考地应用ConvNets只会给模型训练带来混杂因素。此外，在RotNet中学习的特征是对旋转角度的辨别。它们在旋转不可知的图像数据集中并不受欢迎，如浮游生物[8]和ISBI 2012电子显微镜分割挑战[3]。我们首先描述了如何减少旋转噪声（即旋转不变的图像）标签的影响，并在下一小节介绍了学习旋转的不相关因素。

我们将数据集中的原始图像视为默认方向的图像，并将其标记为正面例子。没有标签的例子包括所有旋转过的副本，其中一些在旋转后仍处于默认方向。因此，这些图像的自动分配的旋转标签对RotNet来说是有噪声的。因此，如果把所有未标记的数据都当作负面的例子，预测输入图像是否被旋转是一个二元分类问题[4]。在PU学习中，有研究表明，估计的条件概率与噪声率和一个例子是否干净的信心有关[40, 19]。我们建议使用估计的概率对每个旋转的图像进行加权，并减少旋转模糊图像的相对损失。首先，训练一个ConvNet模型来进行二元分类。我们用F˜(Xi,y)表示从这个预训练的模型中估计的图像为正的概率。我们为每个实例添加一个可调整参数γ的交叉熵损失的权重。



仅仅与图像旋转有关的图像特征对于涉及旋转无关的图像的下游任务来说是不实用的。另一个解决方案是用与图像旋转无关的额外特征来补充旋转相关的特征。我们通过开发一种特征解耦算法来实现这一目标，该算法学习一种语义特征，该特征部分与图像旋转有关，部分与之无关，具有辨别能力。该特征的第一部分享有从估计图像旋转的任务中继承的好处。由于与图像旋转无关，另一部分则适用于一些与方向无关的任务。

**旋转分类**。我们假设图像X的高级特征可以表示为f=[f(1)T, f(2)T] T 其中f(1)明确地与图像旋转有关，而f(2)负责与旋转变换无关的信息。我们用ConvNet来表示，它将输入的旋转图像Xi,y映射到一个基于固定大小向量的特征提取器Ff(-; θf)，参数为θf。

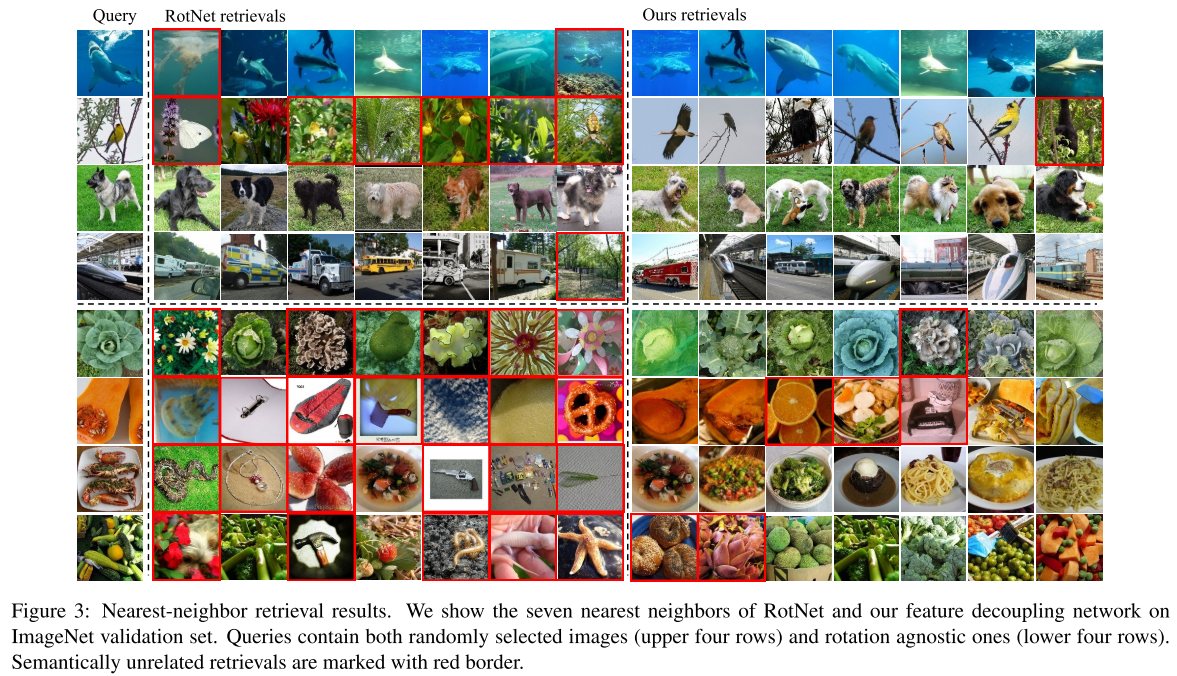


**旋转不相关**。为了实现旋转无关的目标 相关的特征，我们强制要求不同旋转角度的同一图像的特征之间具有相似性。

1. 实验

我们首先在ImageNet ILSVRC 2012验证集上进行最近邻检索，以测试所学特征在捕捉语义方面的能力。我们与RotNet的基线进行比较，以了解特征去耦的效果。对于我们的模型，我们从fc7层提取的特征提取器网输出的4,096维向量中获取特征。我们使用余弦相似性来工作Ff(-; θf)。相应地，对于RotNet的特征是计算特征之间的距离。一些例子的检索结果在图中从左到右按距离增加的顺序排列。

图3中，一些例子的检索结果按照距离增加的顺序从左到右排列。RotNet和提议的模型都能够捕捉到某些类别的图像的语义。随机选择的图像，其中包含突出的物体和旋转不明确的图像，其结果对RotNet和我们的模型都是令人满意的。我们的模型有时可以捕捉到更细微的相似性。例如，在第二行，RotNet检索到一些类似的背景植物，而不是前景物体鸟。对于子弹列车，我们的模型成功地找到了相同类别的图像，而不仅仅是一般的车辆。此外，对于一些旋转不可知的图像查询，RotNet不能提取图像中物体的潜在信息。RotNet检索到的许多图像与查询完全无关（用红边标注）。这很可能是因为RotNet更关注物体的形状，对不同的实例辨别能力较差。相反，我们的模型可以为这些查询返回更多语义相似的图像，这证实了我们的模型在实例层面的辨别能力。



1. 评析