域分离网络

康斯坦丁诺斯·布斯迈尔 谷歌大脑 加利福尼亚，山景城

乔治·特里格尔吉斯 帝国理工学院 英国，伦敦

内森·西尔伯曼 谷歌研究 美国，纽约

迪利普·克里希南 谷歌研究 马萨诸塞州，剑桥市

杜米特鲁·埃尔汉 谷歌大脑 加利福尼亚，山景城

摘要

大规模数据的收集和注释成本常常使得机器学习算法在新的任务或数据集上的应用变得代价昂贵，一种规避这种昂贵成本的方法是在可以进行自动数据标注的合成数据集上完成模型训练。尽管这个方法很有吸引力，但这样的模型常常不能从合成图像推广至真正的图像，因此有必要使用域分离算法在这些模型能成功应用之前操纵他们。现有方法或者关注从一个域到另一个域的映射表示，或者关注于学习提取与提取域不变的特性。然而，仅仅只关注了在两个域间创建映射或共享展示，他们忽略了每个域的单个特征。我们假设，对每个域的特有特征进行显式建模可以提高模型提取域不变特征的能力。受私有共享组件分析工作的启发，我们明确地学习了如何在被划分为两个子空间的图像集中进行图像表示的提取：一个组件是每个域的私有组件，另一个组件是跨域共享的。我们的模型被训练成为不止能在源域上执行我们关注的任务，而且使用分区表示来重构来自这两个域的图像。我们新颖的层次结构生成了一个模型，该模型在一系列无监督域适应场景上的性能表现优于最先进的模型，且生成了私有和共享表示的可视化，得以支持对域自适应过程的解释。

1. **介绍**

近日监督学习算法的成功，部分要归功于他们训练的大规模数据集。不幸的是，对这样数据集进行收集、标注和管理工作非常消耗时间，且代价昂贵。另一种选择是在非现实但便宜的设置中创建大型数据集，比如计算机生成的场景。虽然这样的方法提供了针对无限带标签数据的有效承诺，但是在这种设置下训练的模型不能很好地推广至现实领域。受此启发，我们研究了在训练和测试期间数据分布不同的场景中，域不变的学习表示问题。在这种设置下，源数据被标记为一个特定的任务，我们希望将知识从源域迁移到目标域，而我们对该目标域没有基本事实标签。

1. **相关工作**

学习进行无监督域适应是一个开放的理论和实践问题。虽然已经存在了很多现有技术，但我们的文献综述主要关注基于卷积神经网络（CNN）的方法，得益于它们在这个问题[8,18,27,30]上所体现出的经验优势。本-大卫等人提出了目标域的域适应分类器的上界，他们介绍了训练一个二分类器来区分源域和目标域的思想。这个由“域不一致”分类器提供的错误（以及源域特定分类器的错误）结合起来给出了整个分类器的边界。曼苏尔等人将这一理论扩展到了处理多个源域的情况。

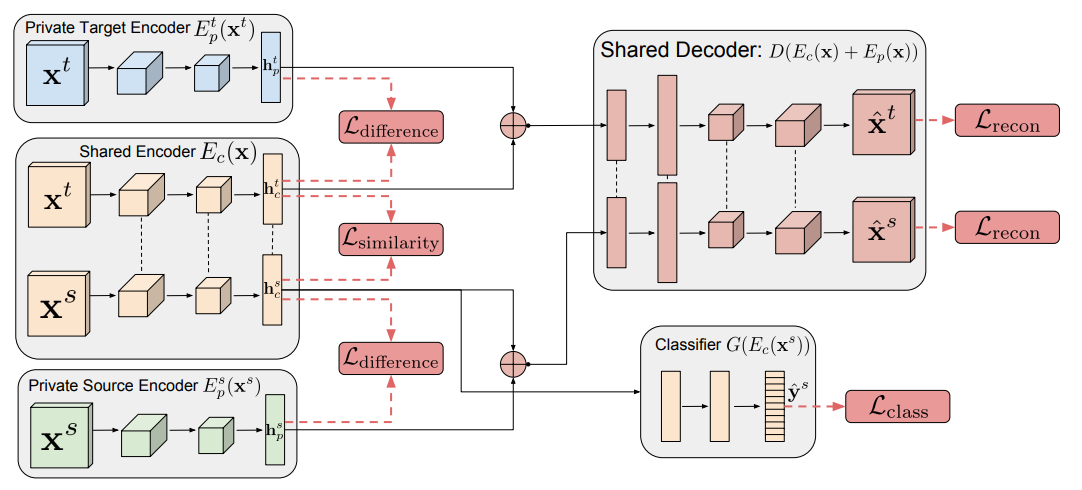
加宁和阿嘉康等人使用对抗性训练来寻找网络中的域不变表示。他们的域-对抗神经网络（DANN）展示了一种结构：其最初的几个特征提取层由两个同时训练的分类器共享，第一个训练用于正确预测源数据上特定于任务的类标签，第二个训练用于预测每个输入的域，DANN使特定于域分类器的域分类损失最小化，同时使得两个分类器共有的参数的域分类损失最大化。通过使用梯度反转层（GRL），这种极小极大优化成为可能。

曾和朗等人提出了该模型的版本，其用最大平均偏差（MMD）度量最小化代替了域分类损失的最大化。MMD度量是通过从每个域的样本集中提取到的特征进行计算的。曾等人提出的深度域混淆网络在CNN体系中的一层有MMD损失，然而朗等人提出了在多层上的MMD损失。

其他相关技术包括学习从一个域到另一个域的转化。在此设置下，在域自适应优化过程中，特征提取管道是固定的。这已经应用于了多种非CNN的方法[9,5,10]，以及最近基于CNN的相关对其算法(CORAL)，该算法利用目标域特征的协方差对源特征进行“重新着色”。

1. **方法**

虽然域分离网络在原则上可以被应用于其他学习任务，但没有失去泛化，我们主要使用图像分类做为跨域任务。给定一个源域上的有标签数据集和一个目标域上的无标签数据集，我们的目标是训练一个分类器用于从源域中归纳出目标域中的数据。类似之前的工作，我们的模型经过训练，使得来自源域的图像表示与来自目标域的图像表示相似。这使得一个基于源域图像训练的分类器可以泛化，因为分类器的输入理论上不受源域的影响。然而，这些表示可能包括与共享表示高度相关的噪声，如萨尔兹曼等人展示的一般。



我们主要的创新之处在于，受最近关于共享空间组件分析工作的启发，DSNs显式地为域表示的私有和共享组件建模，表示的私有组件独立于每个域，且表示的共享组件由两个域共享。为了诱导模型产生这样的分割展示，我们添加了一个损失函数来鼓励这些部分的独立性。最后，为了确保私有展示仍然有效（避免平凡的解决方案）并增加可泛化性，我们还添加了一个重构损失。这些目标的组和是一个模型，它为域和特定于域的私有表示生成类似的共享表示。通过以这种方式划分空间，在共享域表示上训练的分类器，能够更好地进行跨域泛化，因为它的输入不受每个域特有的表示方面的损害。

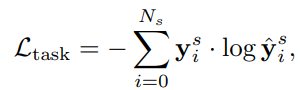
更特别的是，让表示源域的样本的标记数据集，并且让表示目标域的样本的未标记数据集，让是一个函数参数化的图像x，hc代表一个隐藏的表示特性相同或跨域共享。让是一个类似的函数映射图像x隐藏表示hp展示了每个域的私有特征。让是一个解码函数映射一个隐藏表示h，对一张图像x的重构。最后，表示一个特定于任务的函数，由进行参数化，h从隐藏表示映射到特定于任务的预测。得到的域分离网络模型（DSN）如图１所示。

**３.１　学习**

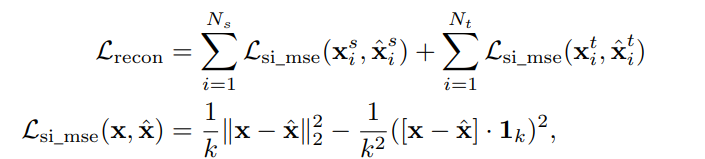
DSN模型的推理由和给出，其中为输入ｘ的重构，且是任务特定预测。训练的目标是对以下参数进行损失最小化：



其中α、β、γ是控制损失项交互的权重。分类损失训练模型来预测我们最感兴趣的输出标签。我们假定目标域是未标记的，这项损失仅运用于源域。我们希望最小化每个源域样本的分类准确性的负对数可能性：

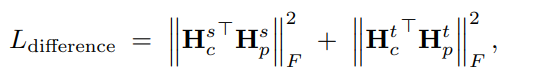


其中是源域输入ｉ的分类标签的一个热编码并且是对模型：的软最大预测。我们使用一个尺度不变的均方误差项对应用于两个域的对损失进行重建：



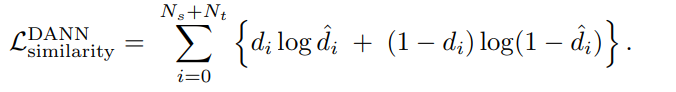
其中ｋ为输入ｘ的像素数量，为长度为ｋ的向量；并且是的平方模。虽然均值平方误差损失传统上用于重构任务，但它会惩罚在缩放项下正确的预测。相反，规模不变的均方误差抵消了像素对之间的差异。这允许模型学习复制被建模对象的整体形状，而不需要在输入的绝对颜色或强度上花费建模能力。在４．３节中通过训练我们最佳DSN模型的一个版本，用传统的均方误差损失代替式３中的尺度不变损失，我们验证了这种重构损失确实是正确的选择。

差异损失也被运用在了两个域，并且鼓励共享和私有编码器去编码输入的不同方面。我们通过每个域的私有和共享表示的软子空间的正交约束定义了损失。设和为矩阵，其行分别为源域数据和目标域数据样本的隐藏共享表示和。相似地，设和为矩阵，其行分别是源域数据和目标域数据样本的隐藏共享表示和。差异损失鼓励了每个域共享和私有表示之间的正交性：



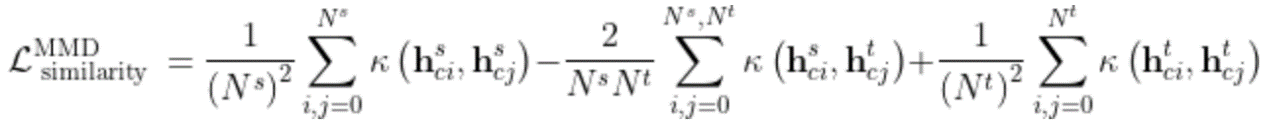
**３.２　相似性损失**

域对抗相似性损失［７，８］用于训练模型产生分类器无法可靠预测编码表示的表示。通过梯度反转层（GRL）和训练用于预测域生成隐藏表示的域分类器可以最大限度的消除这种“混淆”。梯度反转层具有与恒等函数相同的输出，但是反转了梯度方向。形式上，对于某个函数ｆ（ｕ），梯度反转层被定义为：，梯度。域分类器被图共享表示向量，对输入样本标签ｘ的标签的预测通过进行参数化。用梯度反转层进行学习是敌对的，被优化来增加的能力来区分从源域到目标域的图像编码，然而梯度的翻转导致了模型参数对域学习表示的分类准确率降低。最终，我们最大化了有关参数的域预测任务的二项式交叉熵，同时最小化了参数：



其中是对样本ｉ的分类准确性的域标签。

最大平均差损失（MMD）是一对样本之间基于核的距离函数，我们对源样本和目标样本的共享编码之间的平方总体MMD使用有偏统计量：



其中是一个PSD核函数。在我们的实验中，我们使用了一个线性组和多个RBF内核：，其中是标准偏差，是第n个RBF核的权重。其他我们在多-RBF核中包含的核都是可加的，并保证他们的线性合成仍然保持特性。共享的分布特征在学习过程中会发生改变，因此，有一个大范围的核是有益的。并且，多-RBF核的不同组成部分可能会确保我们在不同时间拒绝虚假的零假设。即损失足够高时，分布是不相似的。使用带MMD距离的RBF核的优点是高斯函数的泰勒展开允许我们匹配两个总体的所有矩。需要注意的是，它要求找到最优内核带宽

1. **评价**

我们的动力在于解决问题：在一个干净、合成的数据集上学习模型，而在嘈杂、真实的数据集上进行测试。为了实现这一点，我们评估了之前工作中使用的对象分类数据集，包括MNIST和MNIST-M、德国交通标志识别数据集（GTSRB），还有街景门牌号码数据集（SVHN）。我们也评估了裁剪后的LINEMOD数据集，该数据集是一个用于对象实例识别和三维姿态估计的标准[12,31]。我们测试了以下无监督域适应场景：(a)从MNIST到MNIST-M；(b)从SVHN到MNIST；(c)从交通标志识别到GTSRB的真实场景；(d)从在黑色背景上呈现的合成LINEMOD对象实例到现实世界中的相同对象实例。

我们通过与当前流行的神经网络视觉域自适应技术：相关对齐（CORAL）、域对抗神经网络（DANN）和MMD正则化［３０，１８］。为每一个场景，我们提供两种额外的基线：各自的性能在目标域模型没有域适应和训练。（ａ）在源域（表４中的仅有源）和（ｂ）在目标域（仅有目标），分别做为经验的下界和上界。

我们还没有找到一种普遍适用的方法来优化非监督域自适应的超参数。之前的工作建议使用反向验证。我们实现了这一点（详情见补充材料），但是发现反向验证的准确率和测试的准确率常常不能很好地匹配。理想情况下，我们希望避免使用来自目标域的标签。因为可以认为，如果含有目标域的标签，则应该在训练过程中使用它们。然而，有些应用中带标签的目标域数据集在训练过程中不能进行使用。一个例子是使用ＡｐｒｉｌＴａｇｓ对数据集进行标注。这是一种二维条形码，用于标记对象的姿势，前提是摄像机已经校准且条形码的物理尺寸已知。这些图像不应该在从像素学习特征时使用，因为模型应该能够辨认标记。然而，它们可能是训练期间不可用的测试集的一部分，并且没有标记的等价数据集可以用于非监督域适应，因此我们选择一个目标域的带标记的小的数据集做为我们比较的所有方法的超参数的有效集。所有的方法都使用相同的协议进行评估，因此比较数字是公平和有意义的。在这个有效集上的验证可以作为非监督域自适应的一个令人满意的验证度量的上界，这对我们的知识来说仍然是一个开放性的研究问题，超出了本工作的范围。

**４．１数据集和适应方案**

MNIST到MNIST－M。在这个域适应方案下，我们使用广受欢迎的MNIST手写体数字数据集作为源数据集，且使用MNIST－M，一种提出的MNIST数据集的变体用于非监督域适应。MNIST－M的创造方式是：使用每一个MNIST数字作为一个二元掩码，并用它反转背景图片的颜色。背景图片是从伯克利分割数据集（BSDS５００）中均匀随机采样的。在我们的所有实验中，我们使用它们中的１０００个标签来寻找我们模型的最优超参数。这个方案，像所有三数字适应方案一样，有１０个分类标签。

Synthetic Digits到SVHN。在这个方案中，我们旨在学习一种应用于街景门牌号码的分类器，我们的目标域，来自一个纯合成数字的数据集，我们的源域。合成数据集的创建方法是：用一张图像正中数字的分类准确性中光栅化位图字体序列（一、二、三位数字）。源域数据样本会因尺度的变化而进一步增大，背景颜色，笔划颜色和高斯模糊。我们使用４７９４００个合成数字用在我们的源域训练集。和上述情况类似，我们使用１０００个SVHN标签样本来为我们的模型寻找最优超参数。

SVHN到MNIST。尽管SVHN数据集包含显著的变化（在规模、背景杂乱、模糊、浮雕、倾斜、对比、旋转、序列等等），没有很多实际数字形状的变化。这使得其非常显著地区分于手写数字体，像MNIST，其中形状、厚度变化和数字本身的噪声上有大量的弹性变形。由于两个数据集中的基本真值数字都居中，这是一个适定且相当困难的领域适应场景。综上所述，我们使用１０００个MNIST训练样本用于验证。

Synthetic Signs到GTSRB。我们还使用合成流量数据集进行了实验，从［２０］到真实世界交通标志数据集（GTSRB）［２６］的标志。虽然三位数适应场景有１０个分类标签，这个场景有４３个不同的交通标志。综合标志是通过提取相关的象形图和添加各种类型的变体，包括随机背景，亮度，饱和度，三维旋转，高斯和运动模糊。我们使用９００００个合成标志用于训练，１２８０个任意GTSRB真实－世界标志用于域适应和验证，且和剩下的３７９２９个GTSRB真实标志作为测试集。

Synthetic Objects到ＬｉｎｅＭｏＤ。ＬｉｎｅＭＯＤ数据集由对象的CAD模型组成杂乱的环境和每个物体的高变化的３D姿势。我们使用来自数据集裁剪版本的１１个非对称对象，图像以集中对象进行了裁剪，以完成对象距离识别和３D姿势评估的任务。我们在１６９６２张图片上训练我们的模型，为了能够在黑色背景上渲染而没有附加噪波的这些对象。我们使用一个应用于域适应和验证的包含１０６７３张真实图片的目标域训练集，和一个用于测试包含２６５５张图片的目标域测试集。在这个场景下我们的任务同时包括分类和姿势评估。我们的任务损失因此是：。其中是表示分类准确率的三维姿态的正单元四元数向量，并且是等价预测。第一个是分类损失，类似于实验的剩余部分，第二个是四元数的三维旋转度量的日志，且是姿势损失的权重。四元数是一种方便的角度轴表示，用于三维旋转。在表２中，我们报告了从预测的姿势移动时，需要从分类准确率预测姿势旋转对象的平均角度（在固定的三维轴上）。

**４.２　执行细节**

所有的模型都使用Ｔｅｎｓｏｒｆｌｏｗ进行执行，并且用随机梯度下降加动量进行训练。我们最初的学习速度是每２００００步乘于０.９（小批量）。我们使用了来自每个域的３２个样本的批处理，总共有６４个样本图像以平均值为中心并重新缩放至［－１，１］。为了避免在培训过程的早期阶段分散主要分类任务的精力，在１０００步训练后我们激活了其他适应域的分类损失。对我们的所有实验来说，我们的CNN拓扑基于在文献中使用的方法上，与以前在非监督域适应上的工作相当。我们所有模型的精确结构都在补充资料中进行了展示。

在我们的框架中，CORAL相当于修复我们的共享表示矩阵和，通过两个相关对齐的权重矩阵A使他们常规化然后最小化。对于CORAL实验，我们跟随的建议是，从每个网络的倒数第二层提取源域和目标域的特征。一旦每个域的相关矩阵对齐，我们对目标测试数据上训练的线性支持向量机（SVM）分类器的性能培训数据进行评估。基于对我们每个目标域适应场景验证集，支持向量机惩罚参数得到了优化。针对MMD正则化，我们使用了一个包含１９个RBF核的线性合成。我们在所有模型架构的上允许MMD并且针对参数最小化了式。在多于一层上应用的带有MMD的初步实验没有展示对我们实验和结构的任何性能提升。对DANN正则化，我们应用了梯度反转层和域分类器作为对每个场景的规定。我们优化了式通过最小化参数并且最大化域分类器参数。

对于我们的域分离网络实验，我们的相似性损失总是应用于每个网络的多次卷积和最大池化层后的第一个全连接层。对于每个专有空间编码器网络，我们使用一个简单的卷积和最大池化结构。然后是一个全连接层，其节点数等于等效共享编码器最后一层的节点数。在被反馈到共享解码器D之前，共享和私有编码器的输出已经被添加了。对于后者我们使用了一个反卷积结构，其含有包含300个节点的全连接层，一个10x10x3的调整图层，两个3x3x16的卷积层，一个32x32x16的上采样层，另一个3x3x16的卷积层，其次是重建输出。

* 1. **描述**

带有DANN模型的DSN在我们所有非监督域适应场景下，表现超过了我们实验的所有方法。我们的非监督域分离网络能够同时改善在MMD正则化和DANN上的性能表现。使用DANN作为相似性损失（方程6）要比使用MMD作为相似性损失（方程7）表现地更好，这与仅使用MMD正则化和DANN进行域自适应得到的结果一致。

为了检查软正交约束的影响（），我们使用了我们最好的模型，我们的带有DANN损失的DSN模型，并且通过设置β系数到0移除了这些约束。没有他们，模型在所有场景中都会表现地更差。我们也通过式运行我们的最优模型验证了，我们没有采用更受欢迎的均方误差损失而是采用了尺度不变的均方误差重构损失的选择。通过这一变化，如表3所示，我们得到了表现持续更差的分类结果。

每个域的共享和私有表现被结合用于样本重构。单独解码共享和私有表示给我们提供了重构，作为我们域适应过程的有效描述。在图2中，对于每个可视化我们使用了“MNIST到MNIST-M”和“Synth对象到LINEMOD”场景。在之前的场景中，模型干净地将前景从背景中分离，并产生了一个与源域非常类似的共享空间。因为目标是对源的转换，这一点是被期待的。在后一个场景中，模型能够生成共享表示的可视化，这些表示在源域和目标域之间看起来非常相似，如图2所示，这对于分类和姿势评估是有用的。

1. **结论**

我们在工作中展示了一个深度学习模型，其改进了已经存在的非监督域适应技术。模型通过显式地分离每个域的私有表示以及源域和目标域之间的共享来实现这一点。通过使用现有的域分离技术得以使共享表示相似，软子空间正交技术来使私有和共享表示不同，我们的方法优于所有现有的非监督域适应方法的适应场景，他们关注于从synthetic到现实的情况。

**致谢**

我们想要感谢萨米·本吉奥、凯文·墨菲和文森特·范霍克在这项工作上提出的宝贵意见。我们同时要感谢雅罗斯拉夫·加宁和保罗·沃尔哈特提供了一些我们使用的数据集。