P2:

Motivation：

但是系统地比较，尤其是跨不同网络的功能，仍然很困难，并且通常是逐层进行的。

对于不同层（一般是不同维数），很难衡量相关性。

希望能找到度量维度不相等、具有非线性相关的两个变量的相似度的方法。

Contribution：

PDC可以评估不同维数特征空间之间的相关性，以及复杂的非线性关系。

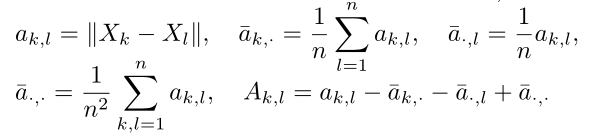
有助于学习相互独立的特征，对对抗性攻击具有更好的鲁棒性。

使潜在变量中的因子相互独立

DC：



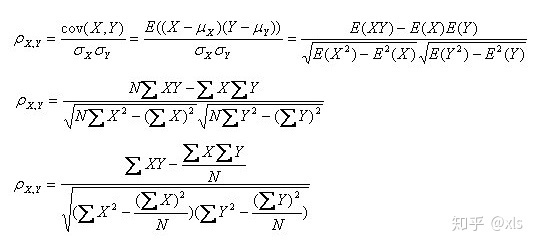
X与Y是不同维数的随机向量，经过这样的方式进行了降维。从各自的p维和q维降至1维研究相关性。



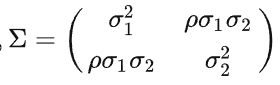
先求向量的2范数，可以看成再求该行的平均、该列的平均，总体平均。这里列平均应该求和号打漏了。

这个函数是自定义的一种方差，相当于是A，B中每个元素对应乘积之和再取算术平均。并且当分母为0的时候额外定义相关系数为0.

Pearson系数



最明显的是第1张图和第4张图，第一张明显是非线性关系，可以看出pearson系数很接近0，（pearson系数为0也无法表示变量间相互独立，只是不存在线性相关性，但是DC为0可以表示变量相互独立），对于非线性关系刻画能力很差，而DC明显好于pearson系数。



第3第4张图是二维正态分布，第4张图的副对角线上都为0，表示x与y线性无关，而又因为是正态分布，可以充要推出x与y相互独立，DC非常接近于0，而pearson系数倾向于表示其中线性相关的部分。

PDC：

在神经网络当中，一个变量受到其他变量的影响，我们要比较两个特征或是变量的相关性，单独用DC会包含进其他变量的影响，这时我们需要比较的是在某个条件下的两个变量的相关性，即排除其他因素的影响。

本文只是举例在线性回归中如何实现。

分别取X和Y正交于Z的部分进行距离相关的计算。

**通过将原始的以u为中心的矩阵A投射到C上，这之间残差和B之间的相关性将可以度量X学到而Z没有学到的东西。**

Optimizing Distance Correlation in Neural Networks：

我们要最小化矩阵A和B之间的内积，就是对应元素的乘积之和，但是同时要保证最大的A和B内积分别都不超过m，因为为了避免计算开销过大，每次更新迭代目标函数只是在大小为m的minibatch上更新。

每个minibatch，我们都会先更新X网络，再更新Y网络，然后在每个minibatch上求一个平均。

最后，作者证明了这样的迭代是收敛的。

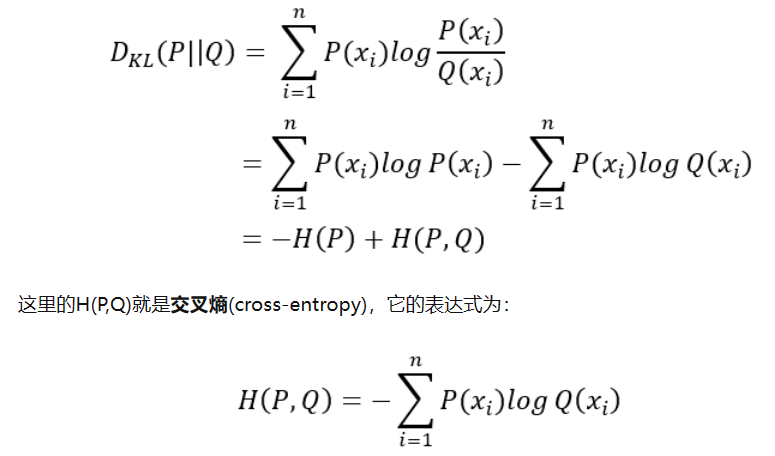
**Experiments：**

**1、Independent Features Help Robustness**

经常观察到，一个训练网络的对抗样本相对容易转移到具有相同架构的另一个网络。

该实验选择同样的架构，在同样地数据集上训练，加入DC正则化项，可以一定程度上让学习到的特征相互独立，防止了特征的迁移。

本实验以分类任务为例。



f1只用交叉熵损失函数训练，f2用交叉熵加上加权的DC损失训练。

g1, g2是特征 （本实验中是用最后一个完全连接层之前的特征）

实验结果中，分别用快速梯度符号法（FGM）和投影梯度下降法（PGD）攻击，可以看到加入DC损失项之后，准确率全都有提升。PGD的最大迭代次数为40。

【像epusilon的东西是对抗性扰动尺度】

为了可视化，嵌入维数为2。

我们可视化了每个维度f1和f2Baseline之间的相关性，以及f1和f2Ours之间的相关性。如果散点图看起来像圆形，我们可以推断这两个模型是独立的。

可以看出，经过DC训练的f2的特征，相关性更小，更加的独立。

**2、Informative Comparisons between Networks**

ViT (Vision Transformer)和ResNet (Residual Network)都是在计算机视觉领域中非常常用的神经网络模型。

ViT是一种完全基于Transformer结构的图像分类模型，它在处理图像时不仅可以捕捉图像中的空间关系，还可以处理图像中的语义信息。ViT是由Google Brain团队于2020年提出的，它将输入的图像划分为固定大小的图块，然后将这些图块转化为向量，输入到Transformer中进行特征提取和分类。ViT相较于传统的卷积神经网络，具有更好的可扩展性和适应性，特别是在处理大尺度图像时表现更加优异。

ResNet是一种残差网络，由何恺明等于2015年提出。ResNet通过引入了残差块极大地增加了网络的深度。相对于传统的卷积神经网络，在训练网络时，会出现梯度消失或梯度爆炸等问题，导致网络无法收敛。而ResNet通过引入残差块，使得不同层之间仅仅相差单层变换，可避免梯度消失的问题，更容易进行网络的训练。ResNet已经被广泛用于图像分类、目标检测、语义分割等计算机视觉领域的任务，并取得了极好的效果。

总的来说，ViT和ResNet都是在计算机视觉领域非常优秀的神经网络模型，它们在深层神经网络方面的特殊处理方式在从感知到任务的处理中取得了巨大的进展。ViT利用自注意机制和transformer框架，具有很强的应用场景适应性，而ResNet通过残差块的机制，可以训练出非常深的神经网络，提高了图像处理的准确率和稳定性。

**A)**使用距离相关的ViT和ResNets不同层之间的相似性

**B)**然后，我们研究在从ViT中去除Resnets的影响(反之亦然)后，网络中保留的残差学习到的东西是什么。

**A)：**

网络之间的相似性, 在这里，我们想证明距离相关是CKA的合理替代方案。

首先，我们表明单个神经网络内各层之间的相似性可以使用距离相关来评估(见图3(a))。我们选择带有补丁16的ViT Base和三个常用的Resnets。所有网络都在ImageNet上进行预训练。对于ViT，在每个块内，我们选择嵌入层和所有的归一化层、注意力层和完全连接层。层数为63层。对于Resnet，我们使用所有卷积层和最后一个完全连接层，这是层数来构建Resnet模型。

左边4幅图，是单个模型中各层之间的相似性。ViT可以分割成小块，浅层与深层的相似度更高。而大多数Resnet模型在网络中显示几部分大块，并且最后几层与浅层具有最小的相似性。

右边3幅图， ViT和Resnets层之间的相似性。在最初的1/6层(以绿色突出显示)中，两个网络具有很高的相似性。最后几层的相似性最小。

利用距离相关性计算相似矩阵的热图，是定性地描述不同网络中不同层的相关性。

接下来是讨论定量地显示不同网络间的相似度。

**B):**

判断两个网络间的关系还未能解决

但可以测量网络X的特征与地面真值标签之间的相似性，然后用部分距离相关将这个想法扩展到多个网络，使我们能够解决判断两个网络间的关系的问题。

【如果相似度越高，我们可以说X的特征空间包含了更多关于真实标签的信息。距离相关可以实现这一点。】

这里，我们选择最后一个完全连接层之前的最后一层作为特征层。

第一次尝试是直接将距离相关测量应用于特征X和one-hot的真值。但是，one-hot的真值嵌入包含的信息非常少，例如，它没有显示“猫”与“狗”、“猫”与“飞机”之间的区别。

因此，使用预训练的BERT在语言上将类标签嵌入隐藏空间。然后，我们测量X的特征空间与预训练的隐藏空间真值之间的DC。



为一个minibatch的特征，为对应标签的嵌入向量。

为了进一步扩展这个度量来度量“剩余”或残差信息，我们X以Y为条件，用部分距离相关计算，从X中移除 Y的影响。



实验结果，值越大表示信息越多。

我们首先展示了两个网络之间的信息比较。注意到，由于ViT达到了最好的测试精度，它也包含了最多的信息。

虽然更好的测试准确性通常与更多的信息相一致，但这并不总是正确的。Resnet 50比Resnet 152包含更多的语义信息，这可能是一种补偿机制。对于Resnet 152，网络足够深，可以专注于语义信息的局部结构(或者是不必要的信息)。

所以，接下来，提出了一种要求网络ΘX学习超越网络ΘY所学的概念的实验

在ImageNet上预训练网络

接下来替换损失函数为最大化下面这个目标函数



实验结果：

我们用部分距离相关代替了Grad-CAM的损失项。结果如图4所示。我们看到，预训练的ViT在不同的位置看到整个图像，而Resnet (VGG)倾向于只关注图像的一个区域。

训练后，ViT(以Resnet为条件)更加关注对象，特别是Resnet焦点之外的位置。这样的实验有助于理解ViT如何超越Resnets (cnn)进行学习。

**3、Disentanglement**

解耦隐变量学习的关键是使隐变量中的因素相互独立。用DC可以衡量不同维度的隐变量的相似性。

重点是半监督解耦来生成高分辨率图像。人们将隐变量分为两类:

(a)interests-一组语义和可解释的属性，例如头发颜色和年龄;

(b) residual——剩余信息。

为了保证有关感兴趣属性的信息不会泄漏到残差表示中，使用该函数限制残差信息。



但是这不是最优的次优，因为可能存在ri不为0但仍然独立于interests的因素

的情况。因此，我们使用距离相关来代替这种损失:



生成器架构采用StyleGAN2。数据集为人脸数据集FFHQ，

属性为:年龄、性别等。

我们使用CLIP[54]对属性进行部分标记，生成用于训练的半监督数据集。除Lres被DC取代外，使用了[21]中的所有损失。

结果(如图5所示)：优化后的模型显示了在不影响剩余特征(如姿势)的情况下改变特定属性的能力。