

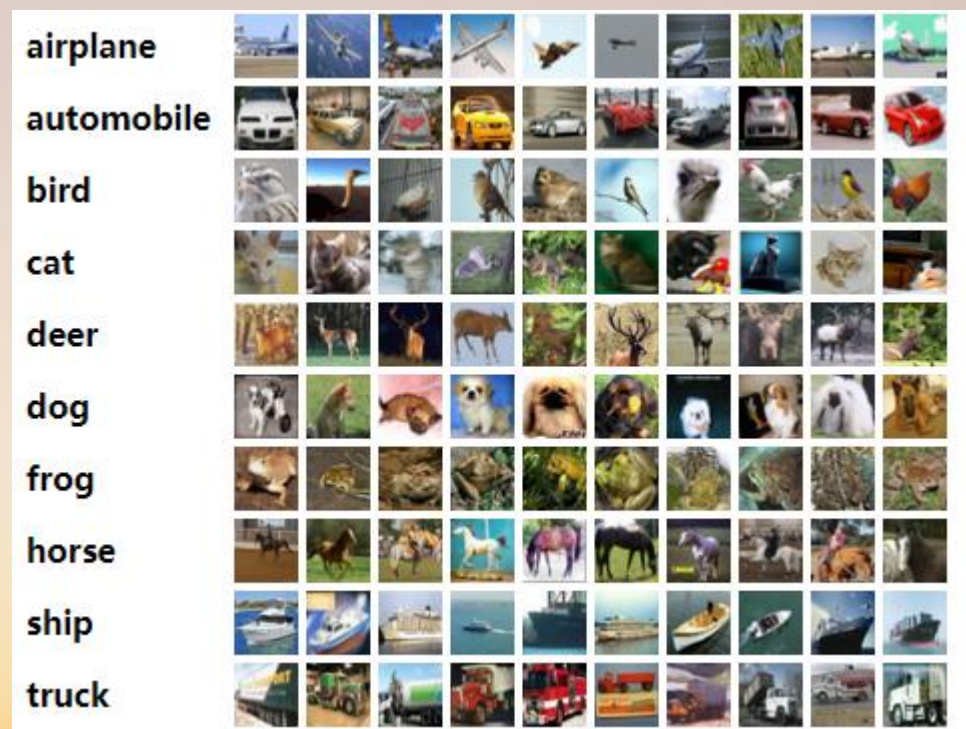
论文题目: A Simple Framework for Contrastive Learning of Visual Representations

本文提出了一个简单的视觉表示对比学习框架SimCLR。简化了最近提出的对比自监督学习算法，而不需要专门的架构或memory bank。为了理解是什么使对比预测任务能够学习到有用的表示，我们系统地研究了框架的主要组成部分。我们发现：

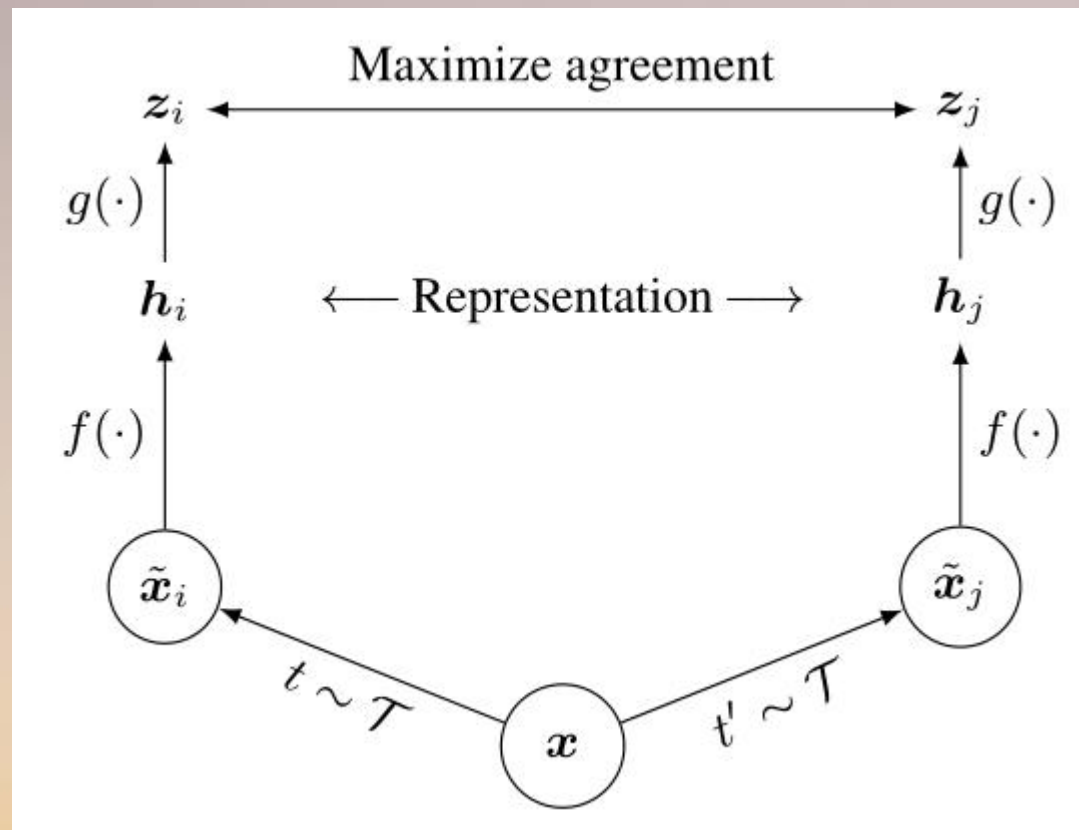
- (1)数据扩增的组合在定义有效预测任务中起着关键作用；
- (2)在表示和对比损失之间引入一个可学习的非线性变换，极大地提高了表示学习的质量；
- (3)与监督学习相比，对比学习受益于更大的批量和更多的训练步。

数据集介绍:

CIFAR-10 数据集包含 60,000 张 32x32 彩色图片，分为 10 个类别，分别为飞机、汽车、小鸟、猫、小鹿、狗、青蛙、马、船、卡车。每个类别 6,000 张。分为训练图片 50,000 张，测试图片 10,000 张。



基本框架:



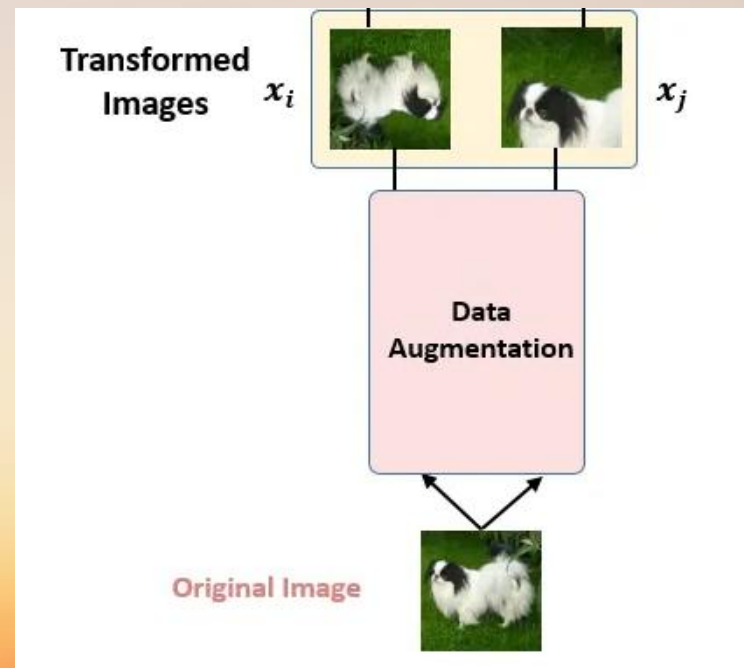
数据增强:

假设现在有1张任意的图片 x ，叫做Original Image，先对它做数据增强，得到2张增强以后的图片 x_i 和 x_j 。数据增强的方式有以下3种：

随机裁剪之后再resize成原来的大小 (Random cropping followed by resize back to the original size)。

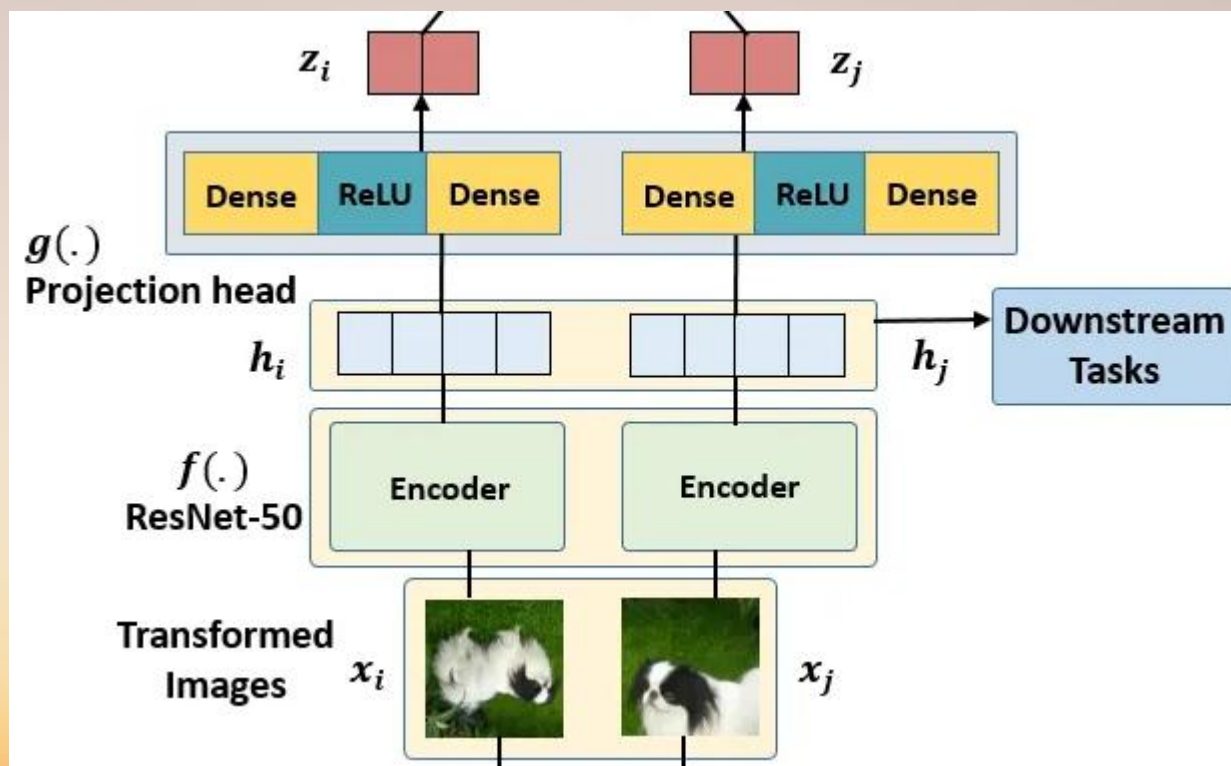
随机色彩失真 (Random color distortions)。

随机高斯模糊 (Random Gaussian Deblur)。



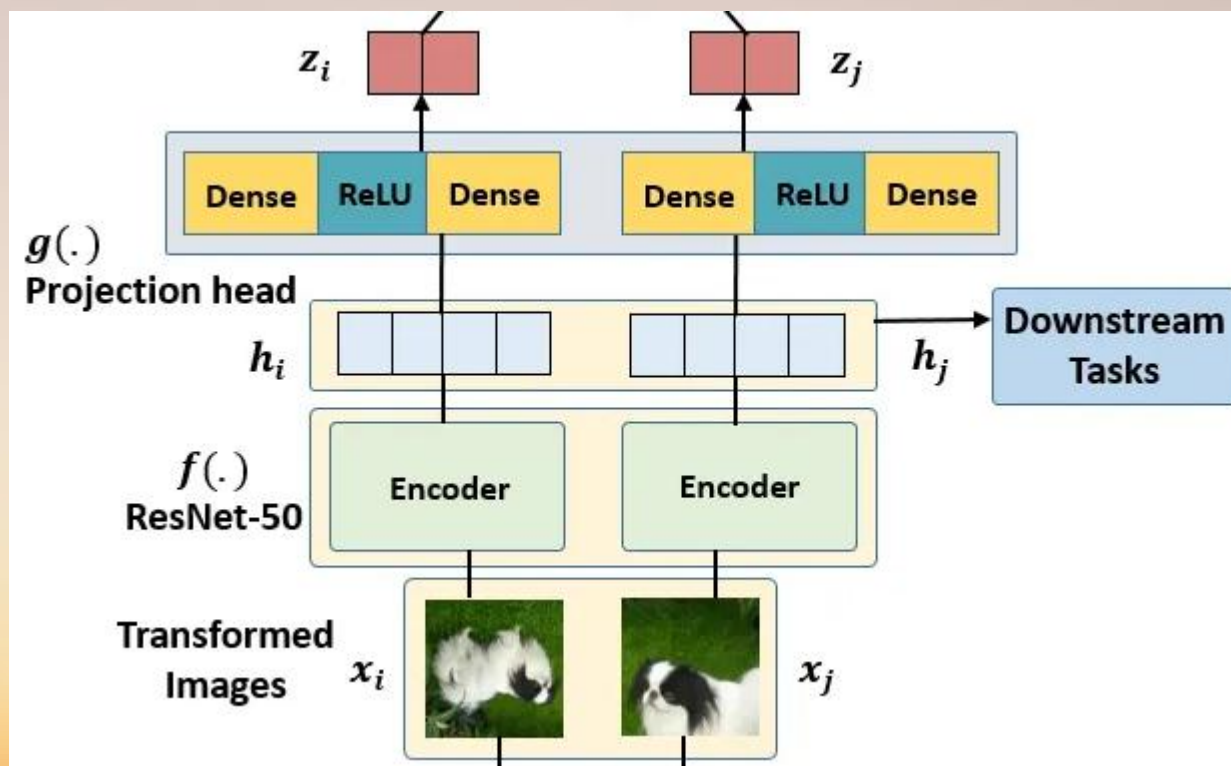
Encoder and Projection head:

Encoder选用Resnet50, 将 x_i , x_j 输入到Encoder得到representation h_i , h_j 。



Encoder and Projection head:

Encoder得到的representation h_i , h_j , 再通过Projection head进一步提取特征, 预测头是一个两层的MLP, 得到新的representation z_i , z_j 。利用 z_i , z_j 求 loss 完成训练, 训练后丢弃预测头, 保留 Encoder 用于获取 visual representation



余弦相似度：

把计算两张augmented images x_i, x_j 的相似度转换成计算两个projected Representation z_i, z_j 的相似度。

$$s_{i,j} = \frac{z_i^T z_j}{\tau ||z_i|| ||z_j||}$$

NT-Xent loss:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)}$$

对下游任务Fine-tune:

通过对比学习，在没有任何标签的情况下训练好模型，使其Encoder的输出可以像正常有监督训练的模型一样表示图像的representation信息。

对下游任务Fine-tune:

通过对比学习，在没有任何标签的情况下训练好模型，使其Encoder的输出可以像正常有监督训练的模型一样表示图像的representation信息。

测试结果:



恳请老师和同学批评指正，谢谢！