

# SimCLR:简单的使用对比学习进行表征学习的框架

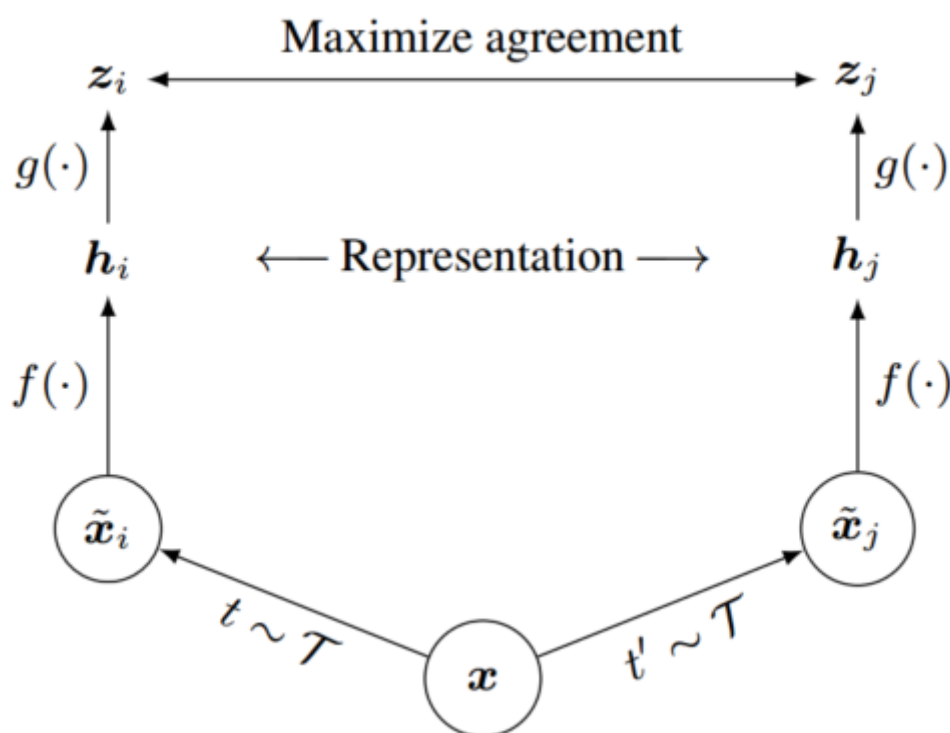
## 写在前面:

目前对于这个框架（模型）的理解是该模型输入图片，输出的是该图像的表征（编码），相当于给图像编码的一个过程，处理完成的编码可以放到下游任务中做进一步的处理。

## 主要贡献:

1. 提出SimCLR，该框架由三部分构成：数据（图像）增广，神经网络编码器（ResNet等能将图像映射为张量的就可以），一个全连接神经网络用于将神经编码器得到的张量映射到潜在空间
2. 基于SimCLR框架提出：相较于监督学习，对比学习更需要数据增强
3. 在使用神经网络编码器得到图像表征 $h$ 后，再连接一个**非线性投射层**（实际上就是多层感知机），效果会更好

## 框架结构:



## 组成部分:

1. 数据增广
2.  $f(\cdot)$ ，将图像映射为表征（其实我的理解就是映射成一个张量），论文中所采用的 $f$ 是ResNet，同时论文强调该框架是灵活的 $f$ 可以换成其他网络架构
3.  $h(\cdot)$ ，一个非线性投射，其实就是带有激活函数的多层感知机，用于将 $f$ 得到的表征进一步映射到隐藏空间中

运行流程：

---

**Algorithm 1** SimCLR's main learning algorithm.

---

**input:** batch size  $N$ , constant  $\tau$ , structure of  $f, g, \mathcal{T}$ .  
**for** sampled minibatch  $\{\mathbf{x}_k\}_{k=1}^N$  **do**  
    **for all**  $k \in \{1, \dots, N\}$  **do**  
        draw two augmentation functions  $t \sim \mathcal{T}, t' \sim \mathcal{T}$   
        # the first augmentation  
         $\tilde{\mathbf{x}}_{2k-1} = t(\mathbf{x}_k)$   
         $\mathbf{h}_{2k-1} = f(\tilde{\mathbf{x}}_{2k-1})$  # representation  
         $\mathbf{z}_{2k-1} = g(\mathbf{h}_{2k-1})$  # projection  
        # the second augmentation  
         $\tilde{\mathbf{x}}_{2k} = t'(\mathbf{x}_k)$   
         $\mathbf{h}_{2k} = f(\tilde{\mathbf{x}}_{2k})$  # representation  
         $\mathbf{z}_{2k} = g(\mathbf{h}_{2k})$  # projection  
    **end for**  
    **for all**  $i \in \{1, \dots, 2N\}$  and  $j \in \{1, \dots, 2N\}$  **do**  
         $s_{i,j} = \mathbf{z}_i^\top \mathbf{z}_j / (\|\mathbf{z}_i\| \|\mathbf{z}_j\|)$  # pairwise similarity  
    **end for**  
    **define**  $\ell(i, j)$  **as**  $\ell(i, j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$   
     $\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$   
    update networks  $f$  and  $g$  to minimize  $\mathcal{L}$   
**end for**  
**return** encoder network  $f(\cdot)$ , and throw away  $g(\cdot)$

---

1. 输入：batch\_size大小 $N$ ，损失函数（归一化温度尺度交叉熵损失）的温度参数  $\tau$ ，图像增广的方式  $\mathcal{T}$ ，将图像映射成表征的网络  $f$ ，以及将表征投影到隐藏空间的网络  $g$
2. 对batch\_size中的每个样本 $k$ ，应用图像增广方式每个样本增广为两个样本 $(x_{2k-1}, x_{2k})$ ，这就是一个正样本对，那其他 $2N-2$ 个与其就构成了负样本对
3. 对于图像增广的每个样本，按照算法1中的顺序投影到隐藏空间中并计算每个正样本对之间的相似度，具体计算方式为归一化的温度尺度交叉熵损失即 $\ell(i, j)$ 。对于框架来言，每个正样本对之间的相似度越大越好，但由于一般优化是按照损失函数最小的方向优化的，为达成一致，采取了取负对数的方式，也就是公式中 $-\log$ 的由来
4. 总体的损失函数定义为：每个样本产生的正样本对间损失和的平均，对其求导更新以 $f, g$ 网络的参数

## 图像增广说明：

在论文中对图像增广实验产生如下结论：

1. 只使用一种单一的变换，没有一种变换可以学到好的表征。在采用组合变换增广后，对比学习的任务变难了但学习到的表征质量提高了。
2. 在组合变换增广中，随机裁剪和随机颜色失真的组合的效果较好。分析只使用随机裁剪效果一般的原因可能是：每张图像存在着局部和整体颜色分布相似的特点，也就是说在没有颜色失真的情况下，只使用颜色分布直方图也能区分图像，神经网络在训练的时候很可能学习到了这种捷径导致效果一般。基于此得到结论：**为了学习到可推广的特征，裁剪和颜色失真的组合是必要的。**
3. 和有监督学习相比，无监督学习对于更强的（颜色）增广受益更大。在有监督学习中甚至会出现强的颜色增广导致效果下降的现象。

## 非线性投影层的说明：

论文经过实验发现：

在  $f$  网络处理完后，在其后面接上一个投影层  $g$ （即使投影层是线性的），也会大大提升  $f$  网络学习到的表征的质量。

经实验分析出的可能原因是经过  $g$  网络的处理，表征会丢失图像增广的部分信息，也就意味着学习到的表征质量提升了（暂时先这样理解）。

## 对比学习受益与更大batch\_size以及更多训练轮数的说明：

更大的batch\_size以及更多的训练轮数都会提供更多的负样本从而提升对比学习的质量。