**Abstract:**

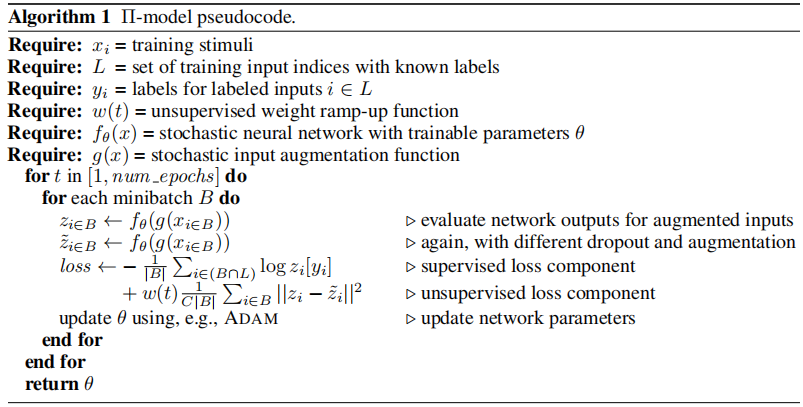
在本文中提出了一种简单有效的训练深度神经网络的半监督方法（只需要小部分的labeled data）。我们介绍了自集成方法，我们建立了一个对无标签数据的一致性预测通过使用不同轮次网络的不同输出（而且在不同规范化和数据增强条件下）。这种集成的预测能够对unlabelled data做出更好的预测。

**Introduction：**

观点——混合网络结构的集成效果要比单独网络的集成效果好，训练总是关注于网络的特定子集，所以这种网络可以看作是这种训练的子网络的隐式网络。我们发展了这个idea——在训练过程中建立集成预测，通过单个网络在不同训练轮次，不同规范化，不同数据增强情况中。

提出了双模型和时序模型——双模型和ladder network有关，时序模型和bootstraping method有关。

**双模型：**



为了控制监督和无监督损失，引入了具有时间依赖的权重参数w(t)，我们比较的是整个输出向量和，这样可以有效的要求他们的dark knowledge(Hinton2015提出)更接近。

因为dropout，相同权重相同输入会产生不同的结果，高斯噪音和增强导致了额外的变化，所以给定相同的原始输入，最小化这种变化就成为了一种合理的目标。

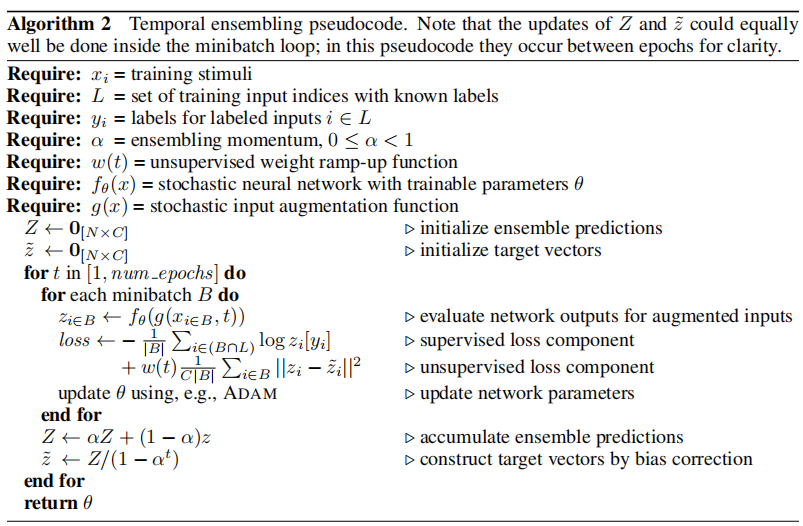
控制loss的权重参数从0开始，前80轮引入高斯曲线，要控制参数足够小，不然网络会因为没有有意义的分类数据获得从而陷入退化问题。

与套模型不同，①在softmax函数后，没有辅助映射在两个分支之间，例如ladder network中的学习过的去噪函数。②套模型中是一个正常分支，一个”腐蚀”分支，而双模型中增强和噪声是同时作用于两个分支。

**时序模型：**

分析双模型我们能够把他拆分成两个步骤——首先分类一次训练集(不更新参数)，然后在相同输入，不同augmentation和dropout条件下再训练一次，将刚刚得到的预测作为target来服务无监督损失部分，因为target只经过一次传播得到，可能会受噪声影响，时序模型通过聚集多次的网络传播预测来减轻这种影响，而且速度是双模型的两倍。

模型通过参数控制ensemble多久以前的training epoch，并且通过时间参数控制不同轮次对产生的target比重的影响。



与双模型对比，优点：

1. 训练速度提升，因为网络只需要评估一次在一轮中。
2. Training targets 收到的噪声更小。

缺点：每一轮需要存储更多的辅助数据。

另一种有趣的可能是时序集成在收集网络预测除了均值以外的统计量。例如，通过追踪网络输出的第二时刻，我们能估计每个输出部分的方差。这使得我们有可能以一种有原则性的方式对网络进行推理。基于这个信息，我们能够在确定性预测中加入更多的权重，而不是不确定性上。

令人惊讶的是，我们的方法在所有标签在使用的时候仍然减少了错误率， 我们认为，这表明一致性要求增加了对在许多分类任务中相当常见的模糊标签的一定程度的抵制，并且它鼓励特征对随机抽样更不变。

我们的方法通过在训练前分配一定比例的随机标签来增加对错误标签的容忍性。因为无监督部分鼓励映射函数实现通过使网络在所有的输入部分附近边平坦，尽管有监督部分鼓励映射函数在有标签输出附近拥有特殊的值。这表明尽管错误的有标签输入确实对构建映射函数有一定的帮助，但是无监督部分使得映射函数和决策边界更加光滑，有效的将输入融合在了连续的集群，尽管额外的正确标签通过有监督部分将集群锁定到正确的输入向量。与常规正则化不同的地方是我们只在可能的输入诱导平滑而不是在整个输入域，未来的研究方向就是映射函数梯度的重要性。