Learning Loss for Active Learning

Donggeun Yoo[1,2] and In So Kweon[2]

[1]Lunit Inc., Seoul, South Korea.
[2]KAIST, Daejeon, South Korea.
dgyoo@lunit.io iskweon77@kaist.ac.kr

背景

深度神经网络的性能随着更多标注数据的增加而提高。问题是标注的预算是有限的。解决这个问题的一个方法是主动学习，模型要求人类对它认为不确定的数据进行标注。

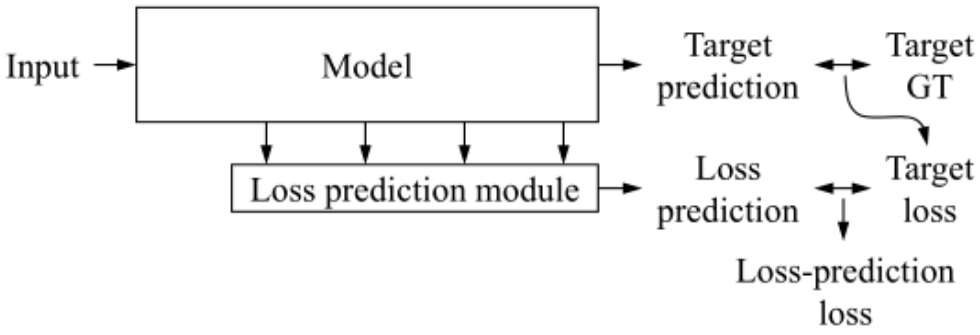
近年来，人们提出了多种方法来将主动学习应用于深层网络。比如基于不确定性的有：基于后验概率的熵、委员会算法以及基于后验概率的 margin;基于数据多样性以及代表性的有:子集选择，即选择代表未标记池的整个分布的不同数据点。

但大多数方法要么是针对目标任务而设计的，要么是针对分类任务。因此在本文中提出了一种新的主动学习方法：在神经网络上附加“损失预测层”，通过它来预测未标记输入的 loss。这种方法不受任务限制(分类、回归或者混合均可)，并且能有效地在各种神经网络中工作，与各种神经网络结合(只要是计算 loss 的神经网络都可以)。

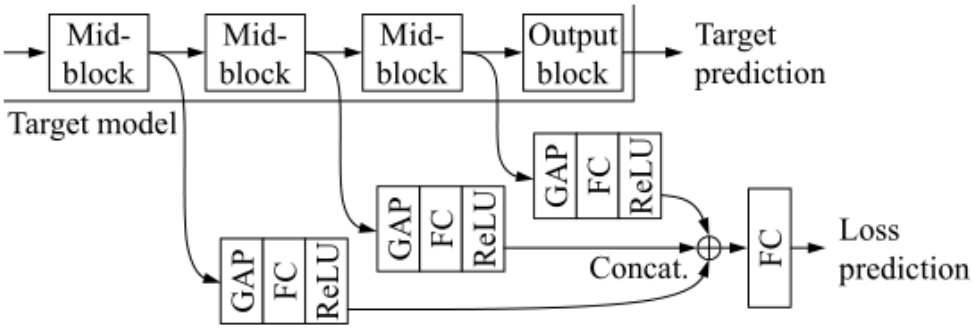
思想

因此本文提出的方法是在神经网络中加一层 loss 预测网络，即输入一个未标注数据,预测它经过模型分类后的 loss 值(因为是无标注数据，因此无法计算真实 loss)，最终选择预测的 loss 值高的数据返回去进行人工标注！

模型



以使用 CNN 进行图分类为例：



训练流程(损失预测层网络可随着神经网络进行学习)：

1.对于标签预测层网络(target prediction)来讲，和传统网络一样，数据从输入层输入后，经过隐层、输出层并最终得到一个输出 y，以及预测标签 y 与真实标签 Y 的真实 LOSS 值。

2.对于损失预测层网络(loss prediction)来讲,输入是几个隐层提取到的特征数据，并最终将他们串联(非相加)起来，输入到 FC 进行输出，最终得到一个预测 loss 值!

3.对于标签预测层网络来讲，他的损失是计算 y 与 Y 的；对于损失预测层网络来讲，他的损失是计算 LOSS 与 loss 的；又因为最终只需要一个损失值进行反向传播，因此

损失值 = L(y,Y) +λ*L(LOSS,loss)； λ为调节参数，L是均方误差MSE，L(y,Y)也就是LOSS

但是在实验中发现，随着模型的训练LOSS在减小，导致此处使用MSE计算LOSS与loss最终得到的效果并不好。因此作者认为需要去掉LOSS变化的影响因素，提出了新的LOSS与loss计算方法：

在一个batch的数据中(batch_size为B)，生成B/2个数据对，然后通过一对数据的损失预测的差异来学习损失预测模块，使得损失预测模块完全抛弃了整体规模的变化。为此，损失预测模块的损耗函数定义为:

L(LOSS,loss)=max(0,-ll(LOSS_i,LOSS_j)*(loss_i-loss_j)+ ξ)

st: if LOSS _i > LOSS _j :

ll(LOSS _i, LOSS _j)=+1

else:

ll(LOSS _i, LOSS _j)=-1

其中，一个数据对是(x_i,x_j)；ξ是调节参数

那么就得到了模型最终的损失值：

$$\frac{1}{B}\sum_{(x,y)\in\mathcal{B}^s} L_{\text{target}}(\hat{y},y)+\lambda\frac{2}{B}\cdot\sum_{(x^p,y^p)\in\mathcal{B}^s} L_{\text{loss}}(\hat{l^p},l^p)$$

$$\text{s.t.}\quad \begin{aligned}&\hat{y}=\Theta_{\text{target}}(x)\\&\hat{l^p}=\Theta_{\text{loss}}(h^p)\\&l^p=L_{\text{target}}(\hat{y^p},y^p).\end{aligned}\qquad(3)$$

AL挑选数据流程:

在损失预测层网络(loss prediction)的最后一个全连接层FC输出时，会得到预测的loss值，我们挑选loss值最高的K个数据进行人工标注!

实验

图像分类任务：（λ=0.1，ξ=1）

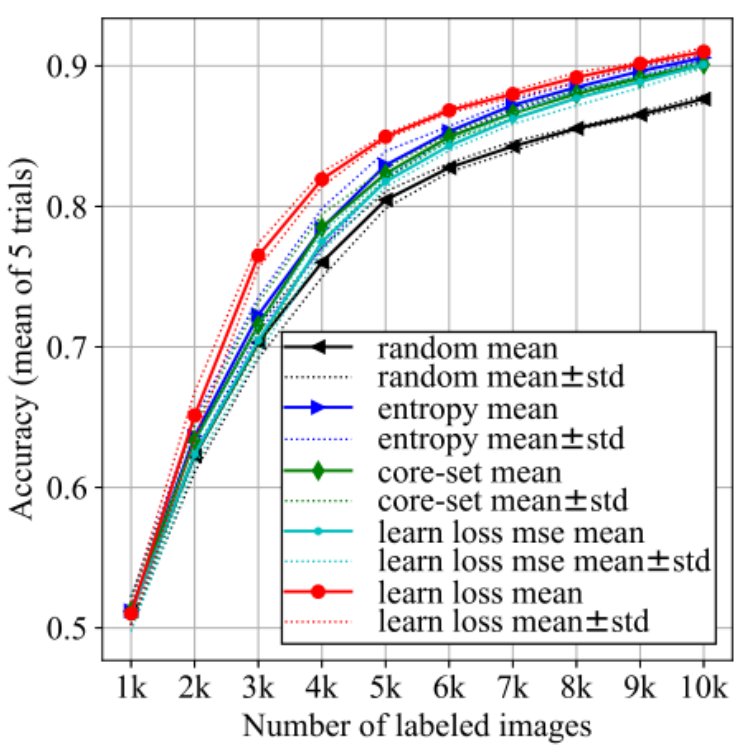


Figure 4. Active learning results of image classification over CIFAR-10.

此图比较了基于随机采样、基于交叉熵、基于核心集抽样、基于 mse 的损失网络学习以及损失网络学习五种方法的模型训练效果！可以发现基于随机采样的效果最差，说明主动学习确实有用！在图中，损失网络学习也理所当然的效果最好！

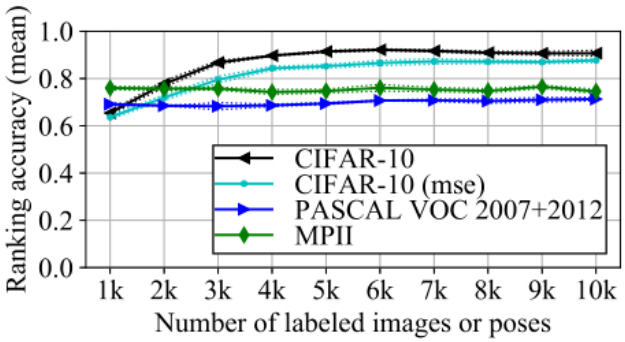


Figure 5. Loss-prediction accuracy of the loss prediction module.

此图展示了在计算 LOSS 与 loss 的损失值时，不同测试集下的模型效果！

接下来在目标检测作为分类和回归的混合任务，以及作为典型回归问题的人体姿势估计任务中，该方法也同样取得了最好的成绩！