# Refine Myself by Teaching Myself : Feature Refinement via Self-Knowledge Distillation

作者：马小莉

代码：

文章(cvpr2021)：

**一句话总结**： 该篇论文提出了一种有效的自蒸馏的方法，无需训练教师模型

**前提知识补充：**

**1.KL离散度：**

相对熵，计算公式如下：

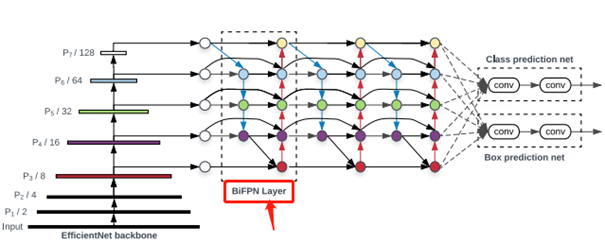


其中P(x)是真实的分布，是目标;Q(x)是拟合分布，是想要改变的分布。KL散度值越小，分布越接近。KL散度越大，表达效果越差

**2.BIFPN：**

出自论文（EfficientDet---双尺度特征融合BiFPN-目标检测）

<https://arxiv.org/abs/1911.09070>



**1.本文希望解决的问题：**

当前蒸馏存在的问题：

a.需要先预训练教师网络，费劲费资源

b.基于数据增强的自蒸馏丢失了局部信息，且其他基于辅助网略的方法和基于数据增强的自蒸馏方法一样， 都没有办法得到更加精细的feature map

**2.本文使用的trick：**

提出了一种FRSKD方法，无需训练教师网络，提出的辅助self-teacher网络结构是在目标检测领域使用的特征网络的基础上发展起来 的。为classifier network自身提供一个精细化的feature map 及其soft label，后通过同时利用soft label 和feature map的内容来进行自知识提取，将精炼后的知识转移到classifier network中

**3.本文效果如何：**

在分类（cifar100&tiny imagenet）和分割任务上普遍涨点

