Abstract:

时序模型在大数据集中很笨拙，因为他在一次迭代中target只会改变一次，MT选择平均模型参数而不是预测结果。

Introduction:

在半监督中对无标注数据加入正则化方法来减少过拟合。

由于模型自己产生targets，他们经常是不正确的。如果给予target太多的权重，会导致不一致性的cost超过错误分类的cost，导致模型更难学习新的东西。

有两种方法提升target quality，①谨慎选择扰动，而不是随意使用加法和乘法噪声。（随机扰动的方法会让模型对某些特定方向的微小扰动表现得很敏感、脆弱，这些方向被称作adversarial direction。）

②谨慎的选择teacher model，而不是随意的拷贝student model。

并非与student model共享参数，而是用student model的EMA参数的算术平均值。现在在每一个step而不是epoch更新参数。优点:①两个模型之间的feedback更快②可以适用于大的学习。

Experiments:

Baseline:13层Convnet，三种扰动：随机变换，水平反转作用于输入图像，高斯噪声作用于输入层，dropout作用于网络。

VAT和MT是互补的方法，他们完全可以结合在一起产生更好的效果。

MT在标签缺少时训练速度和效果都更好。

**消融实验**：

**除去噪声**：输入噪声在有图像增强时是游泳的，Dropout在teacher端只提供了很微小的帮助对比于只加在student端在拥有输入增强时。

**EMA权重和一致性系数**：在一定的范围内保持一定的幅度，但是超过这个范围就退化的非常快。

**解耦分类和一致性**：强耦合版本表现比弱耦合版本表现好，适度的解耦看起来能够对一致性上升有好处？

**改变一致性损失函数**：MSE效果比KL散度好。