

## 任务 5

我们组尝试了一下用 smooth l1 来计算 loss。

模型沿用任务 2 中的模型，网络是 stage1 的默认网络。

### ----- Smooth L1 基本概念和公式 -----

Smooth L1 Loss 取了 MSE 和 MAE 的优点，当模型计算出的 landmarks 的坐标值和 ground truth 的坐标值差别过大时，使用 MAE 损失，这样对异常点不敏感，解决了梯度爆炸的问题。当 landmarks 的坐标值和 ground truth 的坐标值差别不大时，使用 MSE 损失，loss 收敛更快。

公式

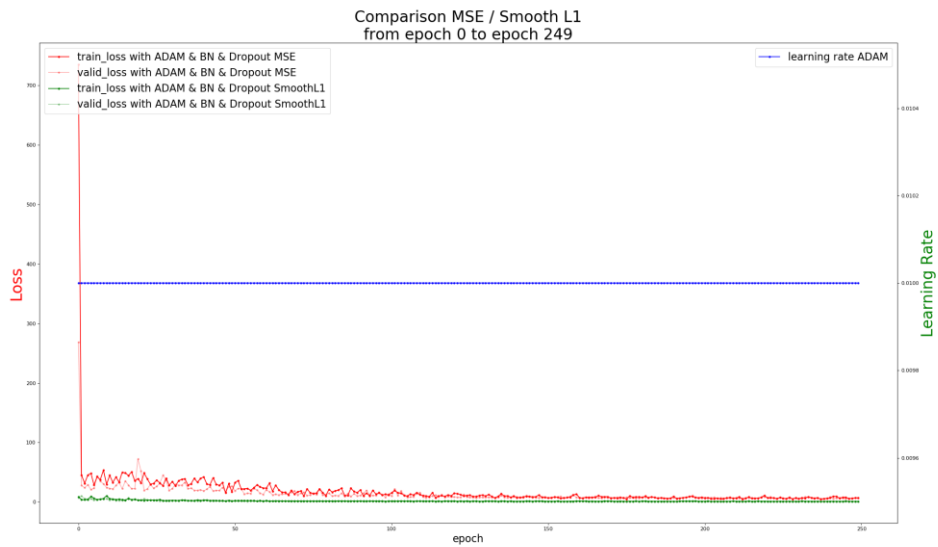
$$\text{Smooth}L_1(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

我们组分别用了 MSE 和 SmoothL1 进行了训练然后做了对比，下面是模型参数

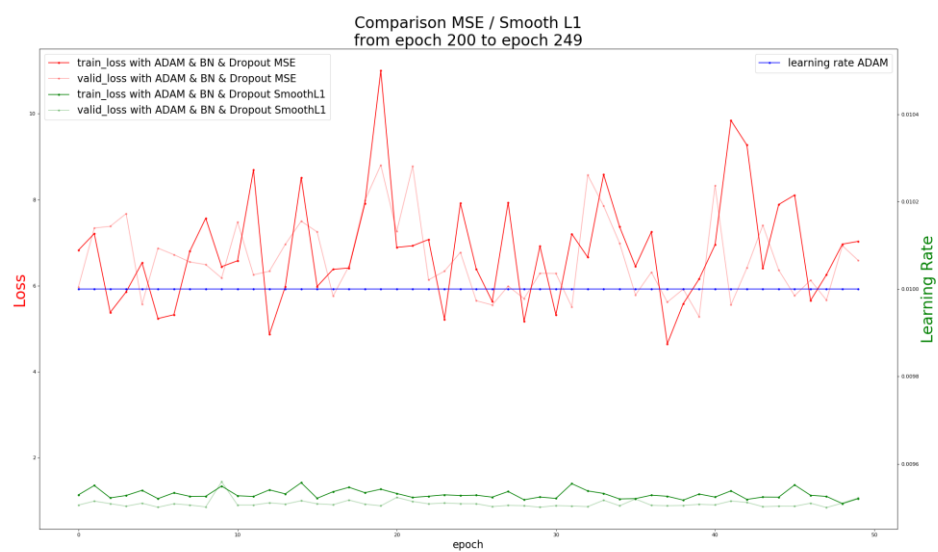
Optimizer: ADAM   Learning rate: 0.01   beta:0.9, 0.999 Batch\_Size:64 Epoch: 250

criterion: MSE / Smooth L1 BN: 分别在 conv1\_1,conv2\_2 和 conv3\_3 后面加了 BN 层

Dropout:在 flatten 后面加了 dropout 层，随机激活 70%的神经元



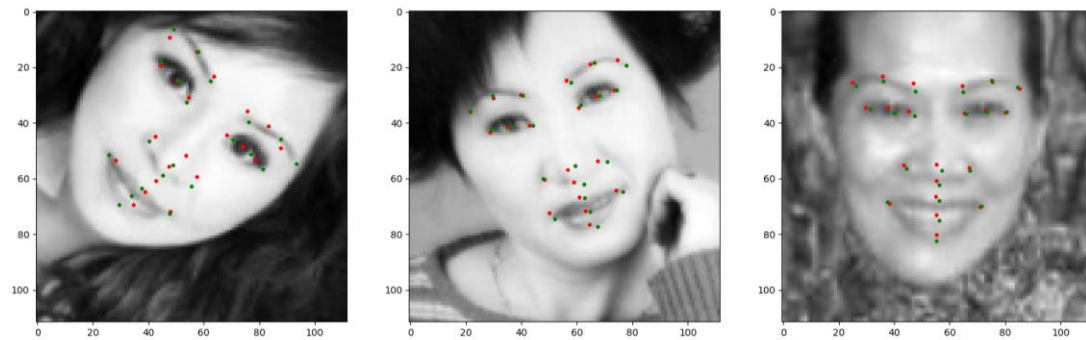
→用 MSE 的话初始 loss 很高，而且在训练初期(前 50 轮 epoch)会有震荡，而用 SmoothL1 在训练初期就能保持非常小的 loss 值。



→最终 MSE 的 loss 能达到 6 左右，而 Smooth L1 的训练 loss 和验证 loss 都能达到 1 以下

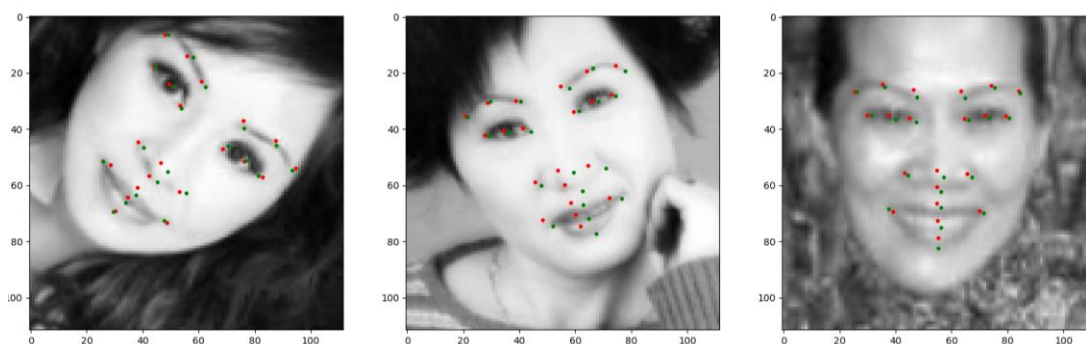
以下为 Test 结果(红点为测试坐标, 绿色点为 ground truth)

## MSE



→ 测试结果比较准确, 个别点仍有误差

## Smooth L1



→ 测试结果比较准确, 个别点仍有误差, 虽然 SmoothL1 的 loss 更小, 但是精度相比 MSE 并没有明显提高!