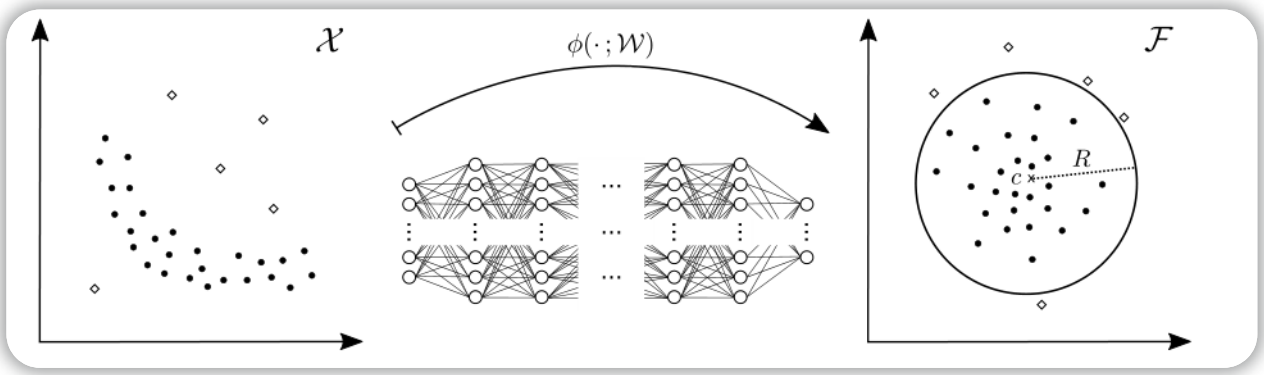
暑期组会 **8.11**

0. **Deep One-Class Classification**方法解析：



- SVDD——支持向量数据描述：

$$\min_{R,a} R^2 + C\sum_i \xi_i$$
$$s.t. \|\mathbf{x}_i - \mathbf{a}\|^2 \le R^2 + \xi_i, \quad \xi_i \ge 0 \quad \forall i$$

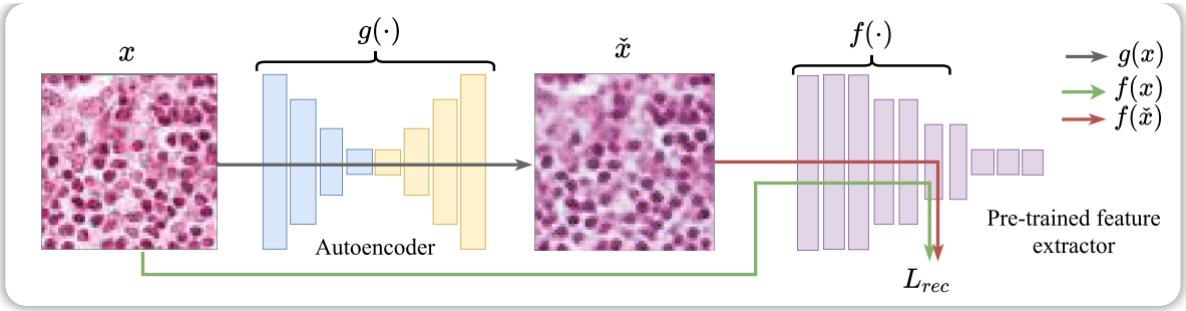
> 1. 模型给出了一个包围全部数据的封闭边界：超球体，球体的特征为球心a和半径R，通过最小化 R^2 来最小化球体的体积，C作为超参数用于控制误差的容忍度，需要满足的条件是：xi（训练样本点）与球心a的距离小于$R^2+\xi$(松弛因子)。
> 2. 而训练样本点为正常样本，如果一个超球体，其包括所有的正常样本点并且体积达到最小，那么异常样本点应该落在超球体之外。

- SVDD的问题——计算可扩展性差，在高维场景下不适用，因此需要深度学习提供特征提取器
- Deep SVDD 通过训练神经网络以将网络输出拟合到最小体积的超球体中来学习提取数据分布变化的共同因素（以此获得一个特征提取器），审神经网络的优化目标为：将输入图片提取为特征向量，再将特征向量映射到超平面中，并使其分布于超球体内，最小化此超球体半径。

$$\min_{\mathcal{W}} \quad \frac{1}{n}\sum_{i=1}^{n}\|\phi(\boldsymbol{x}_i;\mathcal{W})-\boldsymbol{c}\|^2 + \frac{\lambda}{2}\sum_{\ell=1}^{L}\|\boldsymbol{W}^\ell\|_F^2.$$
$$s(\boldsymbol{x}) = \|\phi(\boldsymbol{x};\mathcal{W}^*)-\boldsymbol{c}\|^2,$$

> 1. 目标函数与SVDD类似，将xi替换为φ（xi，W）（W为神经网络参数），优化目标使得降维后的数据点紧密映射到球心
> 2. 异常分数简单地表示为图像特征向量与球心的间距

- 相比之下，其他基于深度学习（自动编码器）的方法——深度自动编码器，其优化目标都是最小化重建误差

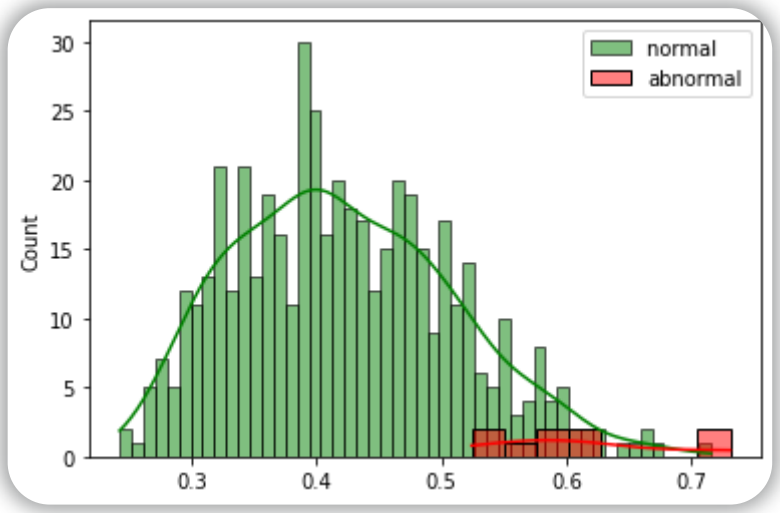


> 如上图，其优化目标为：重建图像x'与输入图像之间的差异作为目标函数

- 自编码器的目标是降维，不直接针对 异常检测 。自动编码器用于异常检测的主要困难在于选择正确的压缩程度，即如何正确将图片降维得到中间特征向量，这个特征向量所含信息的“紧凑型”难以确定
- 相比之下，通过最小化包含数据的超球体的体积将表示的紧凑性纳入 模型训练 目标，从而直接针对 异常检测。

1. 腐腻分类

- Deep One-Class Classification via Interpolated Gaussian Descriptor，改超参数以及随机数种子多次训练。

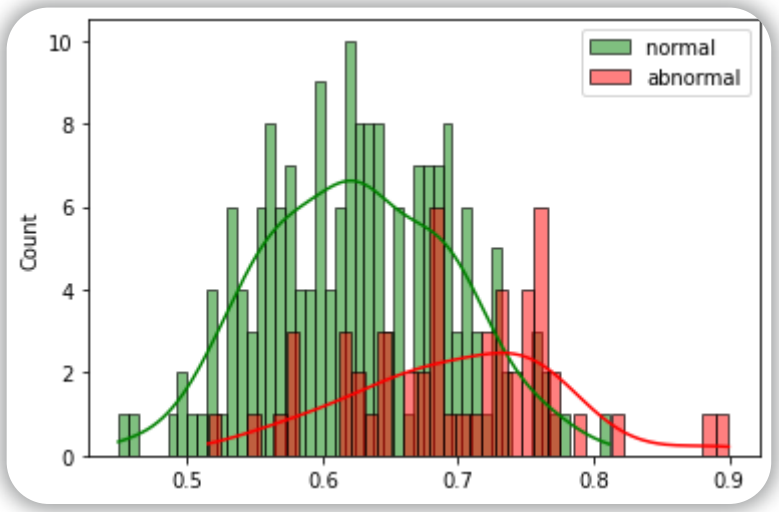


> AUC = 0.9541
>
> 阈值：0.524
>
> 腐：100%
>
> 非腐：87.4%
>
> 整体准确率：87.6%

阈值设定为0.524可以将此测试集的腐都检出不能保证实际使用时可以检出所有腐，因此这一测试结果的泛化能力不高。但是可以对比不同模型的检测能力，统一基准——对比：设定阈值可以识别出所有腐时，非腐的召回率

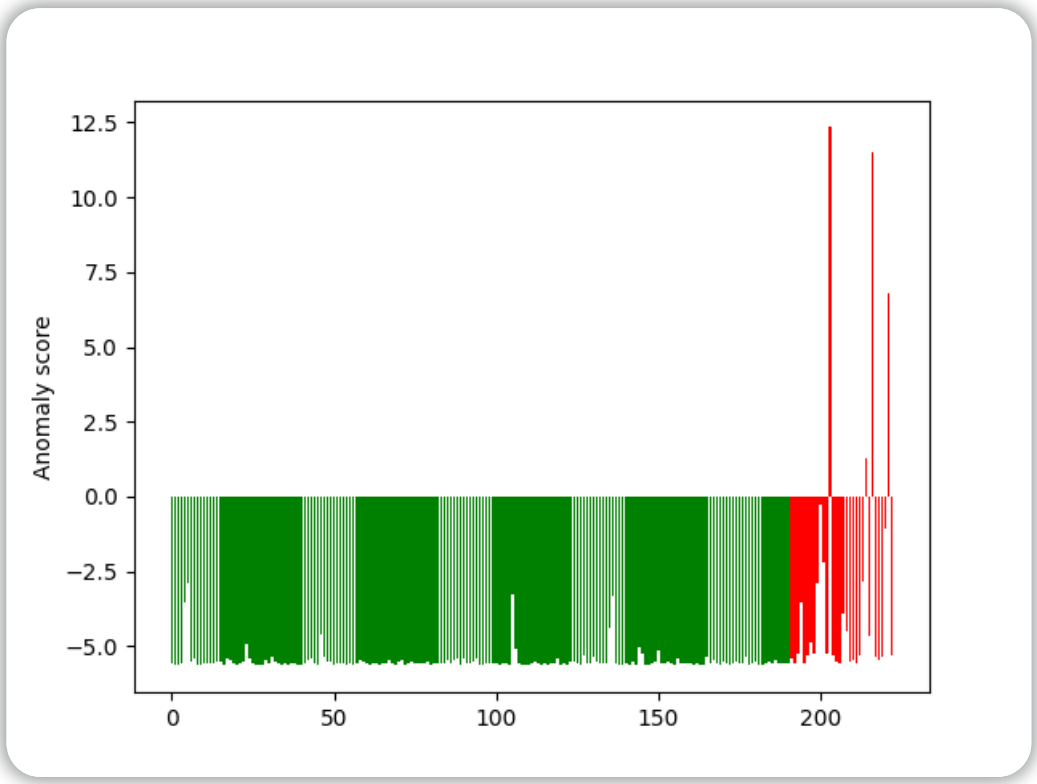
2. 使用异常检测模型将绛红舌/青紫舌于其他舌头分离

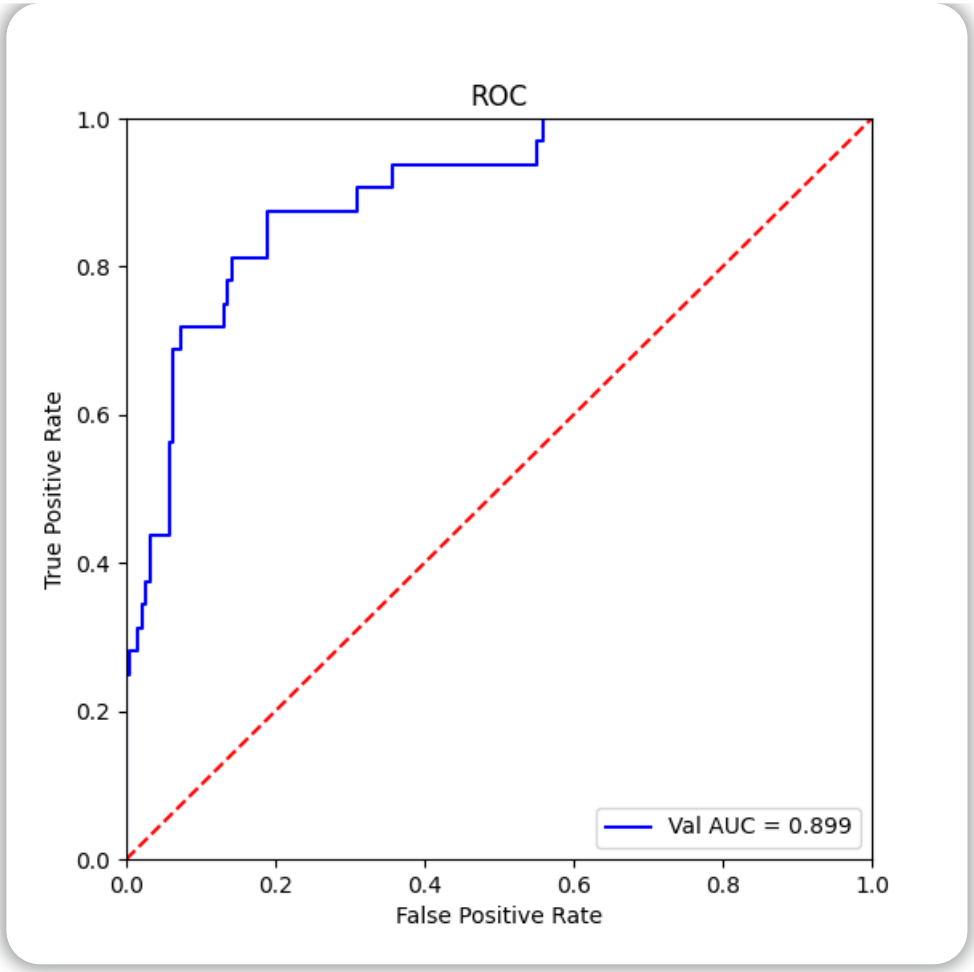
- 无监督方法：
  - Deep One-Class Classification via Interpolated Gaussian Descriptor



> AUC = 0.8501

- 半监督方法：
  - Open-set Supervised Anomaly Detection（异常示例十张）

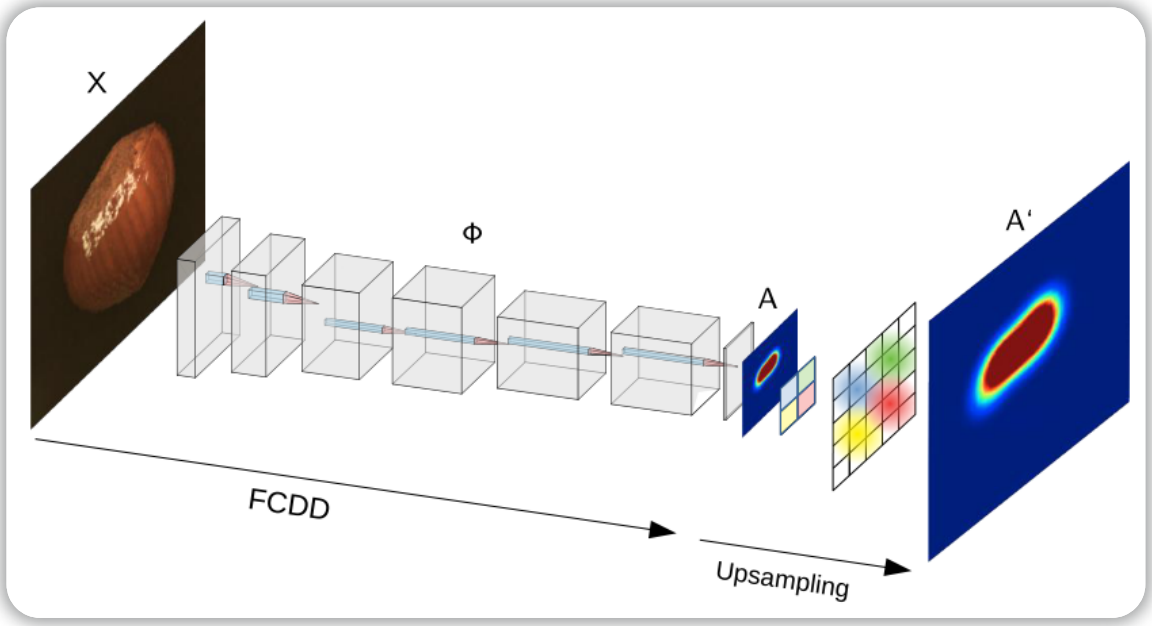




AUC = 0.899

- EXPLAINABLE DEEP ONE-CLASS CLASSIFICATION（ICLR 2021）——可解释的一分类模型

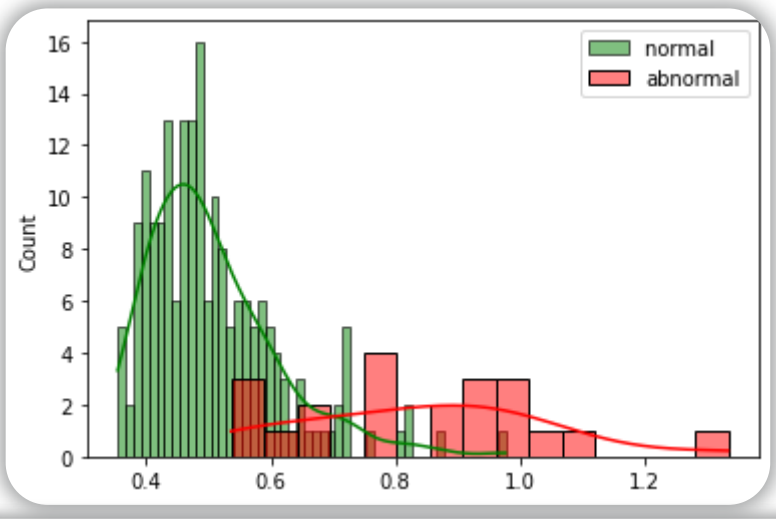
  （异常示例14+7张，测试示例14+6张）

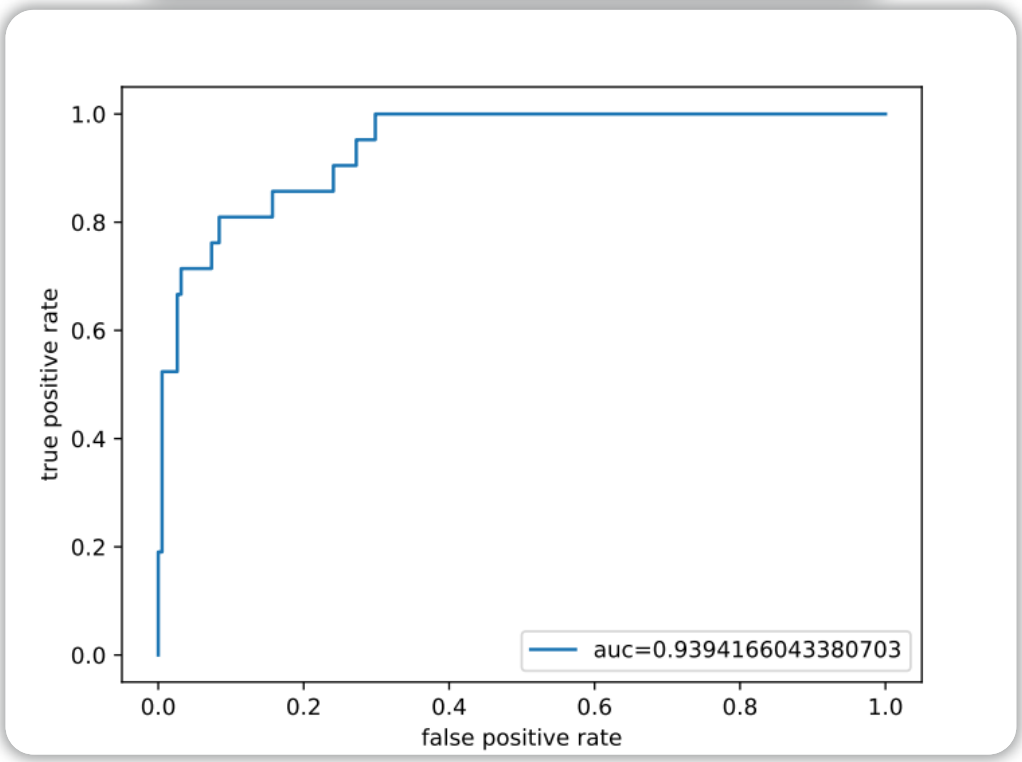


1. 其为DSVDD的半监督版本，训练集中除了正常样本外可以加入少量异常样本

$$\min_{\mathcal{W},\mathbf{c}} \quad \frac{1}{n}\sum_{i=1}^{n}(1-y_i)h(\phi(X_i;\mathcal{W})-\mathbf{c}) - y_i \log\left(1-\exp\left(-h(\phi(X_i;\mathcal{W})-\mathbf{c})\right)\right),$$

2. 其可以通过上采样网络输出异常分布热图





AUC = 0.9394

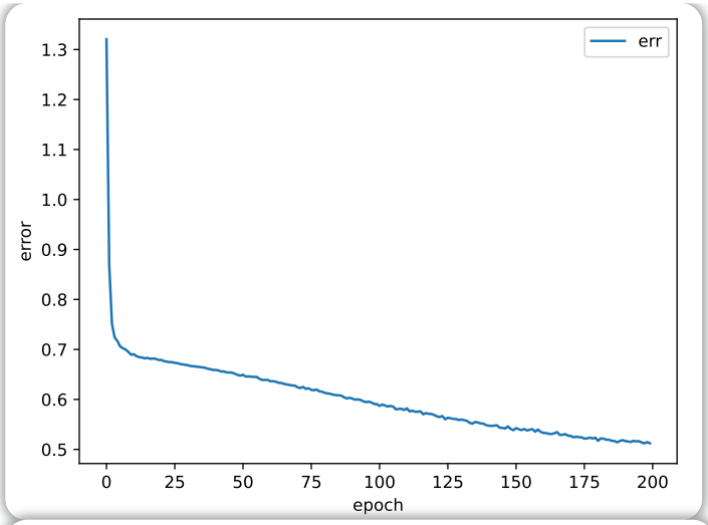
阈值：0.60

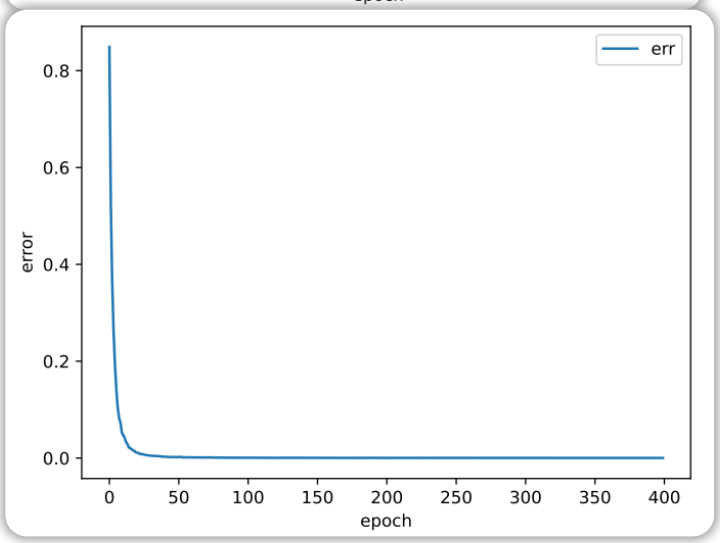
异常召回率：17/21 = 81.0%

正常召回率：84.3%

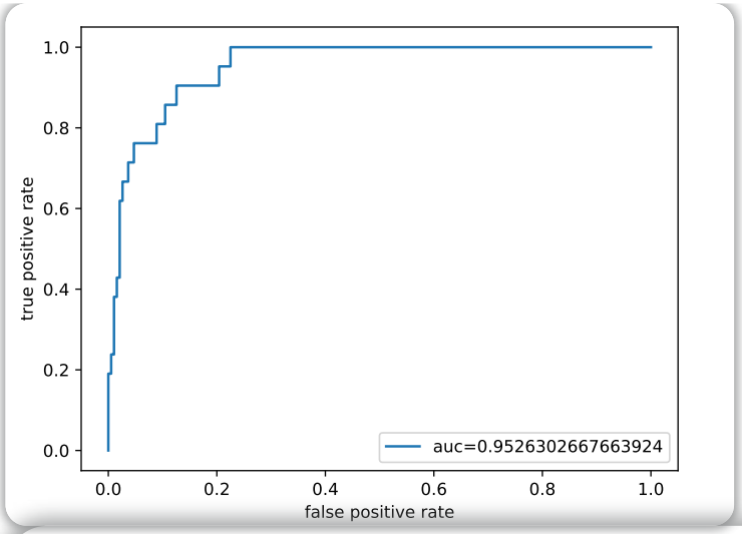
总体准确率：84.0%

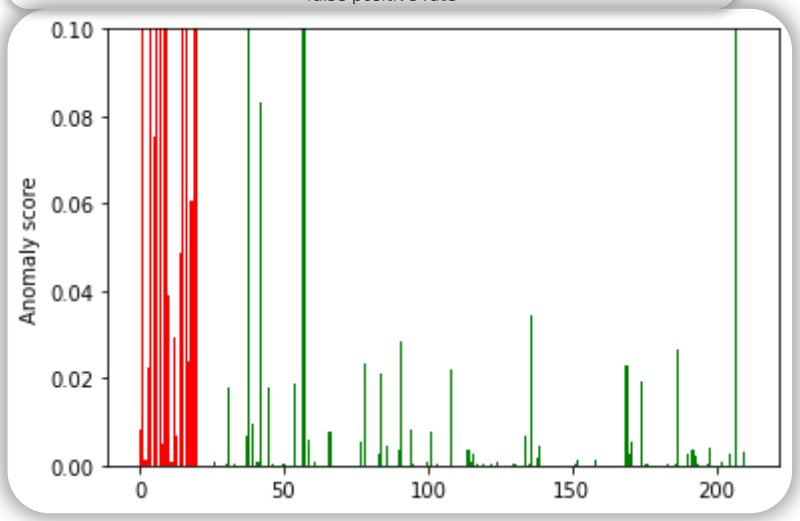
- 上述模型，优化器改为Adam，epoch改为400


err
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
error
0
25
50
75
100
125
150
175
200
epoch


err
0.8
0.6
0.4
0.2
0.0
error
0
50
100
150
200
250
300
350
400
epoch

损失值下降速度对比

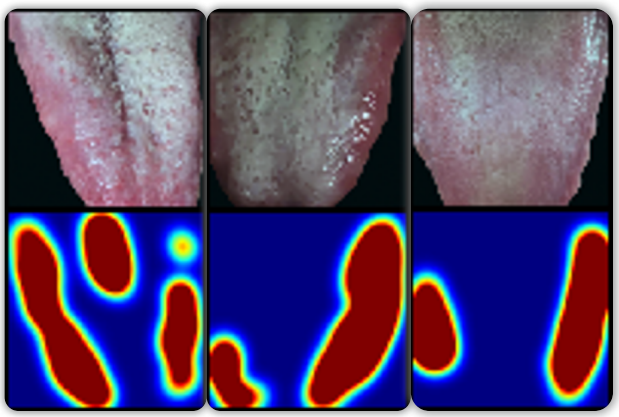




> AUC = 0.9526
>
> 阈值：0.0067
>
> 异常召回率：18/21 = 85.7%
>
> 正常召回率：89.5%
>
> 总体准确率：89.2%

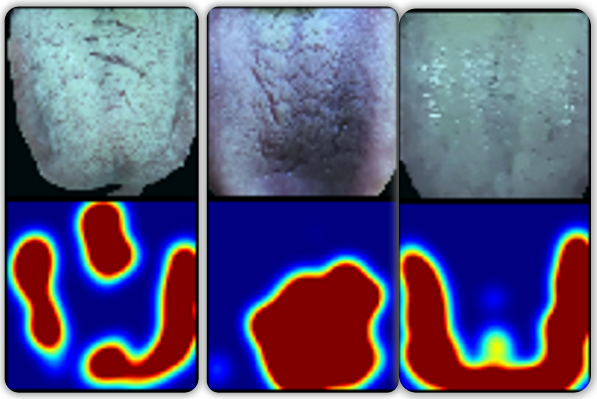
- 模型可解释性：

  1. 对于绛红舌：



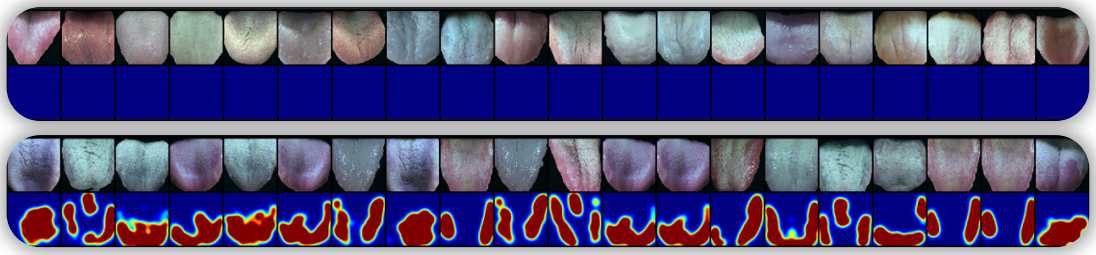
> 可以看出，热度图中，红色部分为舌质，说明异常为舌质偏红处

2. 对于青紫舌

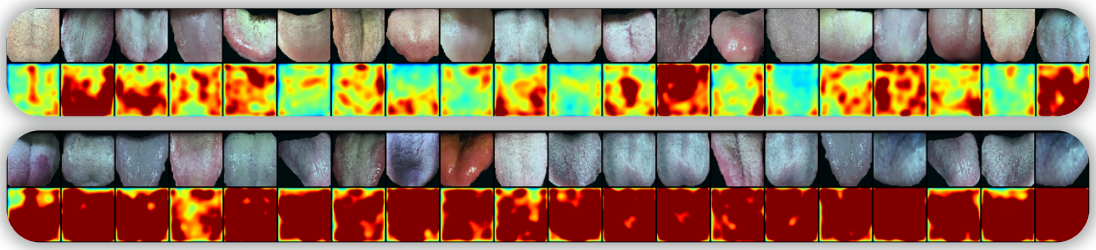


> 和绛红舌的结果类似，异常部分集中在舌质处，说明模型识别异常主要依靠舌质颜色，而非舌苔或者其他特征

3. 热度图总览（上：正常图像，下：异常图像）



与使用sgd优化器时的热图对比：



> 无法给出异常区域，说明训练结果不理想

4. 参考文献：

总结——[2]与[3]都是基于[1]的改进网络：[2]中引入了半监督方法，可以使用部分异常作为示例，并提供上采样网络输出异常分布热图以此提供网络的可解释性；[3]中引入高斯插值方法解决DSVDD的过拟合问题，当数据集数量较少，训练样本中存在异常样本污染时效果较好。

> [1] Ruff, Lukas, et al. "Deep one-class classification." *International conference on machine learning*. PMLR, 2018.
>
> [2] Liznerski, Philipp, et al. "Explainable deep one-class classification." *arXiv preprint arXiv:2007.01760* (2020).
>
> [3] Chen, Yuanhong, et al. "Deep one-class classification via interpolated gaussian descriptor." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. No. 1. 2022.