**SRNets：基于度量学习与深度卷积神经 网络的笔迹检验模型**

杭州学军中学海创园学校高三（2）班 王子杭

**摘 要**

笔迹检验是一项通过分析手写笔迹特征来确定文字是否为某人书写的技术。笔迹检验的常规检验过程需要大量的人工特征分析，以达到辨别检材是否是当事人所书写。作为一项具有较大难度，且受到很多主观因素影响的检验技术，传统人工检验方式难度过大，需要借助计算机的辅助。我提出了一项基于度量学习与深度卷积神经网络笔迹检验模型——Signature Recognition Network (SRNets，包含SiameseSRNet, FCSRNet, ArcSRNet)，旨在借助机器学习的力量辅助笔迹检验。

实验代码已在GitHub上开源——

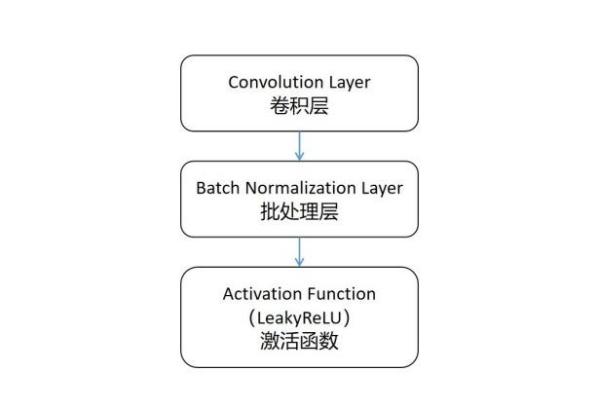
<https://github.com/EclipseR33/Signature-Recognition-Network>

1. **介绍**

传统笔迹检验过程有较多的难点：1.笔迹检验样本少，文字特征难辨别2.样本采集比较随意，导致可比对检验的样本少3.对每个字进行深度的特征分析，对检验人员业务能力要求十分高，能从事人员少。

基于度量学习进行推理的机器学

习模型在面对笔迹检验任务时会有以下几点困难：1. 由于笔迹检验模型训

图片1. 卷积模块

练时需要考虑多数人的手写习惯，要实现准确，训练时期的类别要多(一个人作为一个类别)；2. 数据采集过程中存在固有局限因素使得可参考样本少；3. 由于需要将样本扫描再使用计算机进行处理，所以在此过程中会损失运笔信息、力度信息；4. 纸张背景有可能是不统一的，且拍摄设备、光线强度与拍摄成品的亮暗部分都可能是随机的。采集的信息会有所缺失。

在考虑上述众多难点后，我设计了SRNets的模型结构。

SRNets主体采用了深度卷积神经网络[1]。SRNets主体部分的功能为提取特征。其采用了统一的残差模块——由两个卷积模块组成【图片1】、包含捷径连接(Shortcut connections[2])且具有的残差学习(Residual Learning[2])能力的模块。每一个卷积模块都由3×3卷积层、批归一化层、激活函数层组成。在主体网络中并没有出现池化层，所有类池化操作都由残差模块前的一个卷积层步距为2的卷积模块代替。

在主体部分后有两种结构1.加性角度间隔层，实现了加性角度间隔损失函数(Additive Angular Margin Loss[3])。2. 用于计算两个输入特征图相似度的全连接层。

1. **任务阐述**

在考虑任务的可行性与数据采集的便捷性等其他相关因素下，我模拟了笔迹检验的任务，并且在这个具体任务要求下采集数据、训练模型、测试模型。

2.1 具体任务阐述：

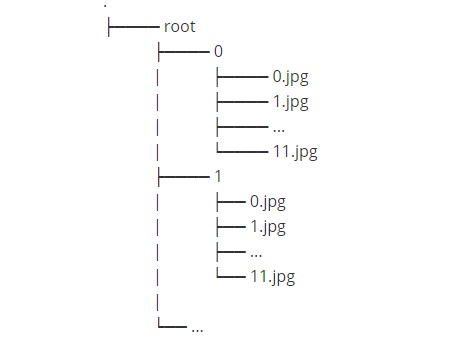
2.1.1 数据准备

一份样本在这里指一份具有类别标注的数字化图像。一份图像中包含有特定文字，这份图像可以是任意摄像设备在任意光线条件下排出的一个字、字母、数字，并且可以通过人眼辨认出笔迹与纸张背景。文件结构与标注方式展示在 【图片2】中。在本次实验中由于条件限制，我选择了12个汉字。每一个志愿者写下的这些 汉字共同组成了笔迹检验数据集。【图片3】中展示了一位志愿者的手写汉字在二值化后的效果。被输入模型的图片都会被二值化处理至如【图片3】所

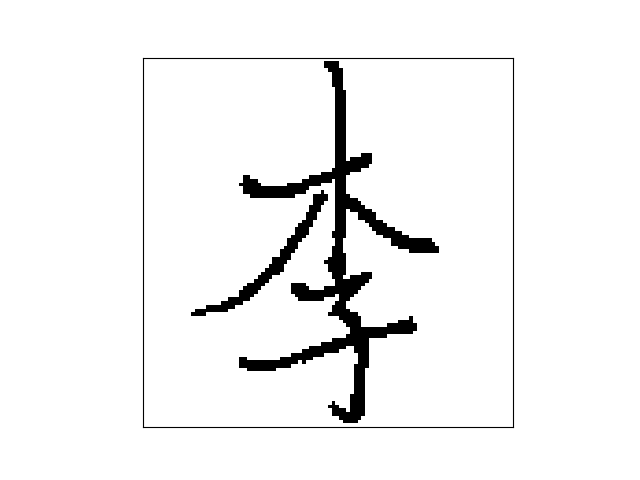
示。本次实验总共收集了199位志愿者的笔迹数据。

2.1.2 期望输出

由于笔迹检验的目的在于检验判



图片2. 文件夹图示。root为数据集根目录，root下的每一个文件夹各代表一个类别，也就是root下每一个文件夹中所有样本为一个志愿者所写。例如root/0/下所有文件为0号志愿者所写的样本，其中包含12张图片。(开源的实验代码中./dataset/indexer.py实现了对上述根目录下所有文件的重命名功能)

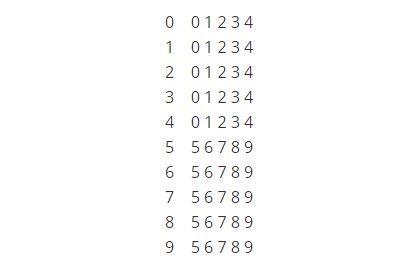


图片3. 二值化处理后的样本

别，所以模型输出应包含分类信息。

模型的预期输出可以根据输入类型的不同有以下两种模式：

1. 当输入模型的数据为两份样本A、B时，模型需要输出这两份样本的相似度，并且判别这两份样本是否为同一个人所书写。这类期望输出所对应的



图片4. 匹配列表。

数据集称作A类数据集。

（2）当输入模型的数据为多个样本（样本数多于2，同时样本可以由2个及以上的人书写）时，模型需要输出一个匹配列表，如【图片4】所示。匹配列表中左侧数字为列表的索引，每一个输入样本都以此赋予自然数编号，左侧索引代表了每个样本。右侧即为列表内容，描述了这个样本的匹配对象，包含了样本空间中每一个与该样本书写者相同的样本编号（包括自身）。例如0号样本与0，1，2，3，4号样本均为同一个书写者所书写，所以在0号样本中与之匹配的样本包含2，也就有图示中第一列“0：0，1，2，3，4”的匹配数据。但5，6，7，8，9几个样本书写者与0号书写者不同，所以在0号的匹配数据中不包含这几个样本。这类期望输出所对应的数据集称作B类数据集。

以上两种输出模式由SRNets的两种不同分类器、后处理结构实现。

2.2 难点说明

在模型训练过程中：

1. 不同场景下不同设备的拍摄效果不同。SRNets是一个自主提取特征的机器学习模型，这要求在预处理过程中需要减小光照条件导致明暗部分不同被视作特征
2. 实验中书写者、也就是预测类别多（199），但是 每个类别中的样本量少（12）。对比常规的数据集：CIFAR-10(类别:每个类别中样本数=10:6000), CIFAR-10(100:600)。

在模型推理过程中：

1. 不确定的输入方式需要多种模型结构实现；
2. （2）训练样本不足会导致数据集本身表达出的信息不足以支撑模型学习，进而产生欠拟合的现象，同时泛化性能不足。
3. **模型结构与训练推理**
   1. 特征提取部分

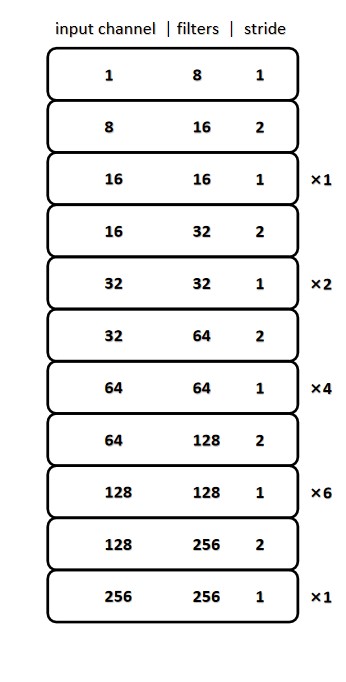
3.1.1 卷积神经网络(CNN)

卷积神经网络最早在[4]中以LeNet-5模型被提出。卷积神经网络为提供了自主学习特征的结构基础，因此在此模型进行笔迹检验任务实现过程中无需设计者对于每一个字的特征进行专门提取。通过卷积层、池化层、激活函数交替连接组成模型主体，这个的主体结构作为特征提取部分便于提取图像特征。其中卷积层中的卷积核以一种不同于全连接层的独特的滤波器式的权重共享机值极大提高了特征提取能力，同时也利于模型训练。在特征提取部分之后LeNet-5使用常规全连接层直接连接到预测结果。通过特征图而不是原图像进行分类也减少了需要学习的参数量。卷积神经网络是SRNets的最基本要素。

近期关于特征提取部分也出现了卷积神经网络以外的解决方案，例如Vision Transformer[5]。由于笔迹检验任务中笔迹信息只用很小尺寸的图片(32×32至128×128)就能表达出，而且我所采集的数据集的样本数量不够多。从目前Vision Transformer的应用实例来看这项技术在这个场景中难以有比卷积神经网络更好的表现。

3.1.2 残差学习

随着卷积神经网络层数加深，退化问题(degradation problem)十分显著，即训练正确率不再随着卷积神经网络层数加深而提高。难以实现恒等变换这类线性变换被视作是退化问题的根本原因。因此残差学习中所加入的捷径连接(Shortcut connections)通过采集残差模块的输入与模块输出特征层相加形成的线性变换与原本卷积层的非线性变换使模型在层数较多的情况下也能有良好的表现。ResNet系列也得以达到50、101、152层。[2]中更有对1202层ResNet模型在CIFAR-10数据集下的训练实验。相同训练次数下1202层ResNet模型与110层ResNet模型表现出相近的训练损失。这也证明了捷径连接的有效性。况且在ResNet之前的VGG模型最多带权重卷积层与全连接层之和仅有19层。



图片5. SRNet-5-[1, 2, 4, 6, 1]的特征提取部分，其中右侧带有×的是卷积区块，其它为普通残差结构。

3.1.3 SRNets中的特征提取部分

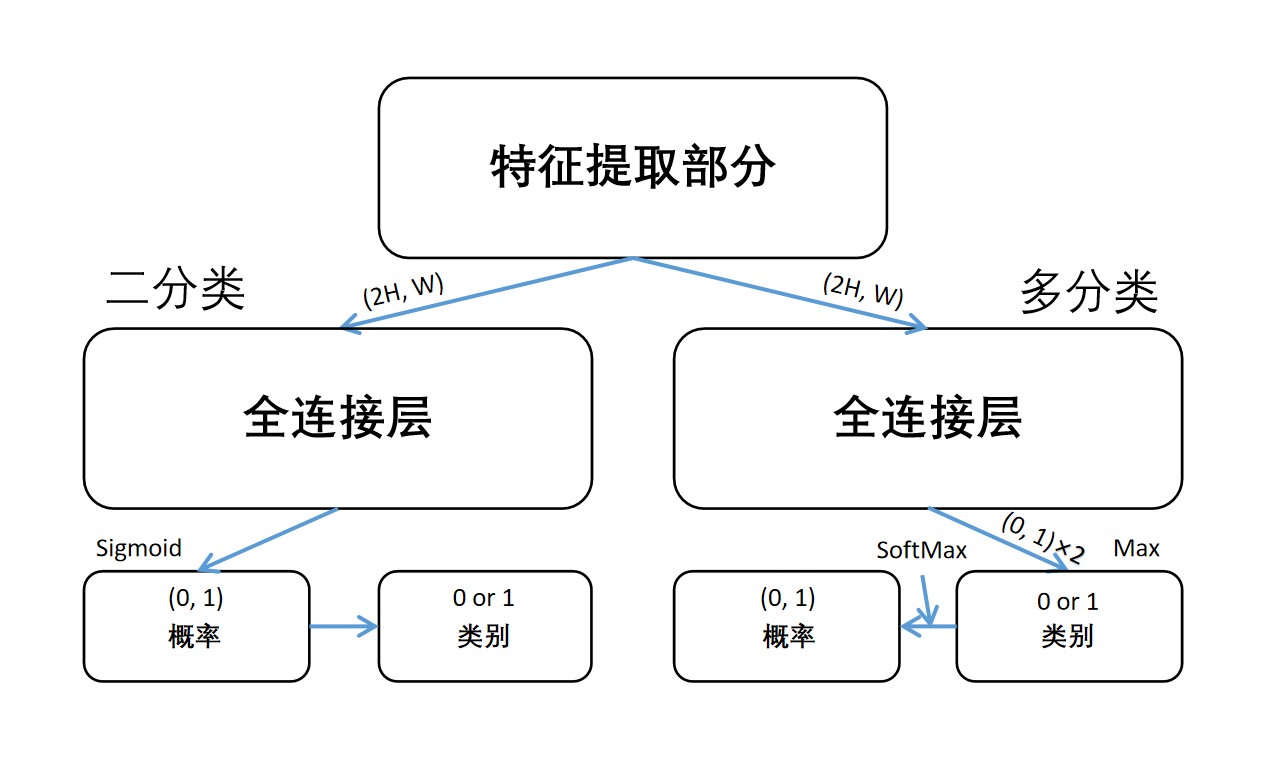
SRNets特征提取部分由多个卷积区块组成。卷积区块是多个残差结构

的集合，其中的残差结构依次排列。具体结构如【图片5】所示。

* 1. 分类器、后处理部分

3.2.1 适用于A类数据集——Siamese SRNet

对于这类数据集，我采用孪生网络的思想处理[6]。训练过程中，A类数据集会给出两个样本，标签标注这两个样本是否是同一类别。特征提取部



图片6. A类数据集解决方案图示

分分别接收两个样本，分别计算出两张特征图，其大小均为(*H*, *W*)，其中*H*代表了特征图高度，*W*代表特征图宽度。由于模型权重完全相同所以这两张特征图的维度与尺寸完全相同。

分类器部分首先在高度上拼接这两张特征图得到一张大小为(2*H*, *W*)的特征图。通过一个适应性平均池化层(nn.AdaptiveAvgPool2d，PyTorch)再将展平后的这张特征图输入全连接层。全连接层也因此可以定义为(2*H*,1)或(2*H*,2)，其中第一个参数代表全连接层的输入维度，第二个参数即为输出维度。

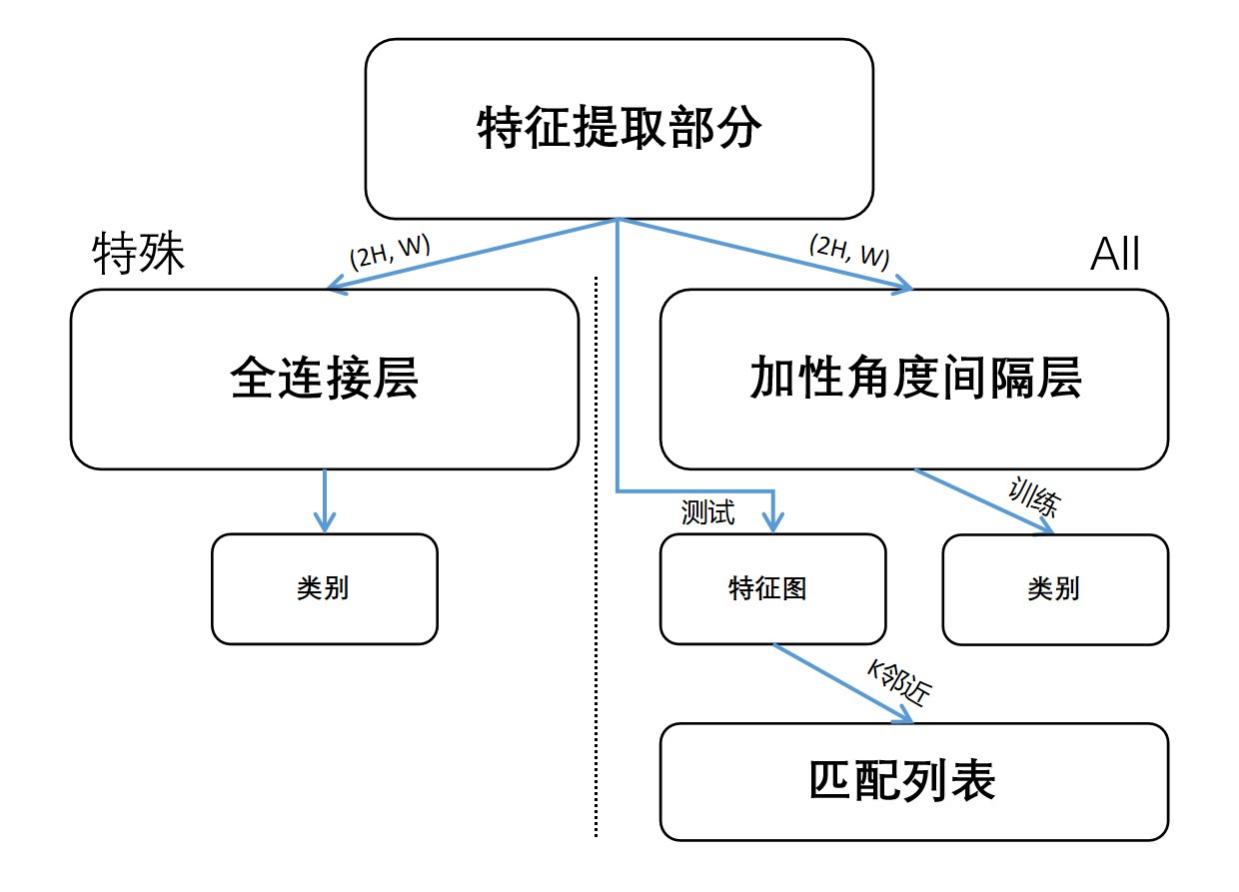
第一种定义是将这个任务视作一个二分类任务，因此输出只用通过Sigmoid函数就可以得到(0, 1)的输出，这个输出可以直接作为相似度的预测值，再设置阈值将其转换为0、1就可以实现判别；

第二种定义是将这个任务视作一个多分类任务，用两个二分类器（第一种定义中使用的）来分别得到两个(0, 1)的输出，再比较这两个输出的大小，计算出值较大的分类器所代表的编号就是模型的输出。相似度预测值就只需要再通过一个SoftMax层就可以得到了。

3.2.2 适用于特殊B类数据集——FCSRNet

特殊B类数据集指的是测试过程中所有的书写（类别）者都是训练过程中出现过的。并且整个数据集的书写者较少（10人以下）。

针对这种数据集模型单独处理每一个样本。得到一个样本的特征图后将特征图依次送入适应性平均池化层（随后展平）、全连接层。这个全连接层参数是(*H*, *N*)，其中*H*代表特征图



图片7. B类数据集解决方案图示

高度，N代表书写者数（即类别数）。在模型训练中使用了交叉熵损失函数(Cross Entropy Loss, (1))

(1)

由于测试过程中的所有书写者都在训练过程中被模型学习过，测试过程就可以直接依据全连接层输出指向的类别生成匹配列表。在衡量这类数据集中模型的表现时，我选择直接使用正确率进行判断。【图片7】

3.2.3 适用于所有B类数据集 ——ArcSRNet

普通B类数据集中除了包含（2）中展示的特殊情况，训练集中的书写者数可以更大、测试集中可以包含训练集之外的书写者。

但在实验中发现，由于笔迹检验中不同书写者的书写差别很大，测试集中新增的书写者写出的样本中可能包含很多训练集中不存在或者不重要的特征。这也使得测试集中添加更多的书写者来进行测试有不合理之处。从传统的笔迹检验过程中在面对新的书写者时也有人为的学习行为，所以在这里我不会在测试集中添加更多的书写者。

SRNets中的ArcSRNet专门为测试集书写者与训练集书写者不同的数据集提供了解决方案。换言之，ArcSRNet对类别没有限制，而FCSRNet要求类别严格对应。

在ArcSRNet的特征处理部分之后是加性角度间隔层。其基础是一个全连接层，参数为(*H*, *N*)，其中*H*代表特征图高度，N代表书写者数。ArcSRNet的训练过程与（2）类似，都是学习一个样本对应到自身书写者这一对应关系。

ArcSRNet的加性角度间隔层在训练过程中与交叉熵损失函数共同实现加性角度间隔损失函数(Additive Angular Margin Loss,(2))。

(2)

(3)

(4)

基于余弦相似度的加性角度间隔损失函数，在角度值上添加了margin参数(3)。对角度更加直接的处理使得加性角度间隔损失函数在减少类内距、增加类间距上比Large Margin Cosine Loss[7] (4) 更加有效。类内距小，类间距大这个特性也使得ArcSRNet在推理产生匹配列表过程中比仅使用交叉熵损失函数训练的模型更容易区分不同类别的样本。

推理过程中需要记录ArcSRNet对每一个测试样本的处理后产生的特征图。通过邻近算法（这里采用K-邻近）对每一张特征图都分别计算其与所有特征图（包括自身）的距离（闵氏距离或余弦距离）。 为了计算出B类数据集中特有的匹配列表，需要人为选取一个阈值。这个阈值用来判别两张特征图是否被视作是同一个书写者书写。若两特征图距离小于阈值则会被视作由同一个人书写，这两个样本也会分别出现在对方的匹配列表中。

阈值的确定方式是在一个指定的阈值区间（包含步距）内寻找训练集F1 score最优的阈值，将这个阈值作为最终阈值。其中F1 score是一个用于衡量二分类精度的指标。F1 score的计算公式如公式（3）。

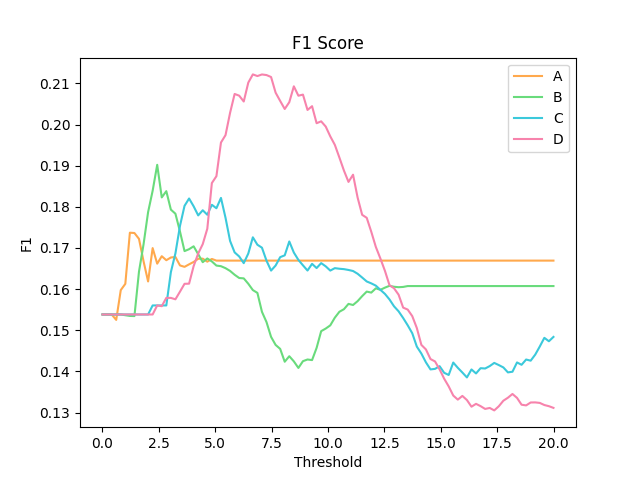
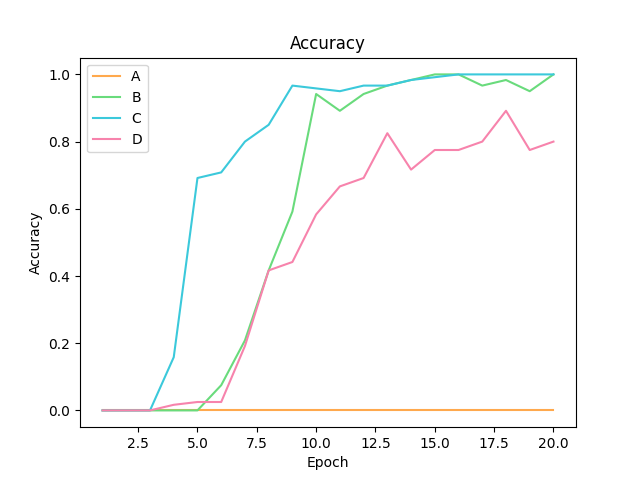
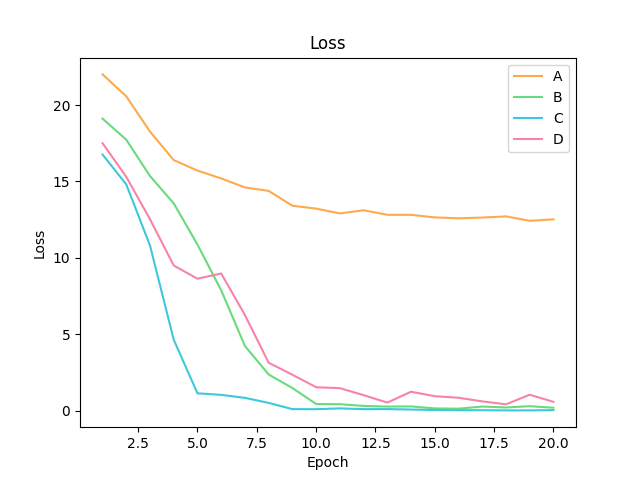
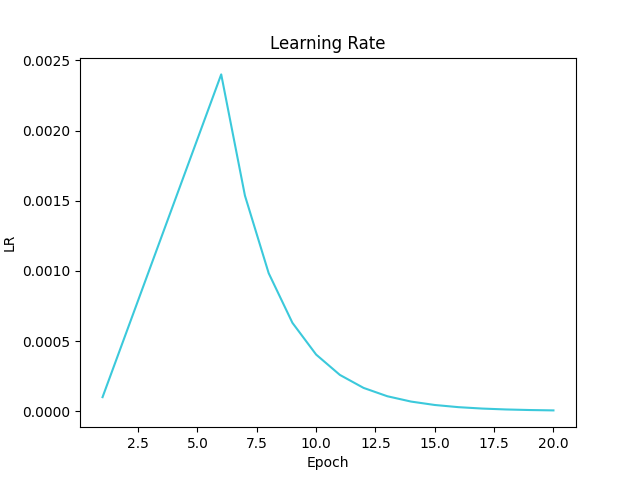
(3)

在得到阈值后，可以在测试集中进行推理得到匹配列表。【图片7】

1. **实验**

在实验中，我选择了四种特征提取网络替换到ArcSRNet上在B类数据集中进行实验，最终得到了损失值（Loss），正确率（Accuracy），F1 score的图像【图片8】。其中损失值、正确率、学习率（Learning Rate）均为训练过程中获取，横坐标为训练时的Epoch。F1 score则是在训练完成后在训练集上尝试不同阈值后得到的，该图像在绘制前经过三次样条插值处理（Cubic Spline Interpolation）。

最终得到的F1 score没有达到较高水平可能是由于数据集较小、数据集样本没有表达出足够的特征、模型没有经过预训练。

****

图片8.

A：ArcSRNet-3-[1, 1, 1];B：ArcSRNet-4-[1, 1, 1, 1];

C：ArcSRNet-5-[1, 1, 1, 1, 1]; D：ArcSRNet-5-[1, 2, 4, 6, 1]

注：ArcSRNet-*N*-*Blocks*，其中*N*代表模型卷积区块数量，每一个区块的输入通道数与输出通道数都是固定的分别为(16, 16), (32, 32), (64, 64), (128, 128), (256, 256)；*Blocks*是一个数组，代表每个区块中残差结构的数量。

除了笔迹检验，SRNets适用于一般的匹配任务，可以解决商品匹配（宣传图）、物品分类、医疗影像或单细胞生物学上的度量学习任务。

**参考文献**

1. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
2. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.
3. J. Deng, J. Guo, N. Xue and S. Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In CVPR, 2019.
4. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In IEEE, 1998.
5. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In ICLR, 2021.
6. J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Sackinger, R. Shah. Signature Verification using a "Siamese" Time Delay Neural Network.In INT J PATTERN RECOGN, 1993.
7. H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu. CosFace: Large Margin Cosine Loss for Deep Face Recognition.In IEEE, 2018.