分类号: 密级:
UDC 学号：20193231016

华南师範大學

South China Normal University

学士学位论文

（学士学位）

面向人机合作分类血细胞图像的拒识别方法研究

学 位 申 请 人： 何海森
专 业 名 称： 信息工程
学 生 学 号： 20193231016
所 在 院 系： 信息光电子科技学院
导师姓名及职称: 马琼雄 讲师

2023 年 4 月 10 日

摘 要

血液病检验通常需要临床医生在显微镜下对血细胞进行分类和计数，这个过程即费时又容易出错。我们希望能够将最近在各项图像处理任务中超越 CNN 的 Transformer 运用在血细胞识别任务中，提高分类的鲁棒性以及稳定性。然而在人体中某些细胞的数量相对较少会导致模型的分类性能下降，因此有必要对其进行研究并对症下药。因此，本文提出了一种具有排斥识别选项的高精度血细胞识别方法——Cross-CCL，该方法包括融合细胞图像多尺度信息的 Crossformer (Cross)和能够提高图像特征区分并排斥低置信度预测结果的类别质心学习 (CCL)。实验结果表明，Cross-CCL 的识别准确率为 94.41%，超过了目前主流的 CNN、Vision transformer、Swin transformer 等方法和最先进的血细胞识别方法。而且 Cross-CCL 可以在保证高识别精度的前提下，减少人类专家的工作量。

关键词：血细胞识别，视觉 Transformer，长尾分布，确定性预测

Abstract

The examination of blood diseases usually requires the clinician to classify and count blood cells under the microscope, which is a time-consuming and error-prone process. We hope to apply Transformer, which has recently surpassed CNN in various image processing tasks, to the blood cell recognition task to improve the robustness and stability of classification. However, the relatively small number of some cells in human body will lead to the degradation of the classification performance of the model, so it is necessary to study and solve the problem. Therefore, this paper proposes a high-accuracy bone marrow cell recognition method with a rejection recognition option--Cross-CCL, which includes Crossformer (Cross) that fuses multi-scale information of cell images and Class Centroid Learning (CCL) that can improve image feature differentiation and reject low-confidence prediction results. Experimental results show that the recognition accuracy of Cross-CCL is 94.41%, which exceeds the current mainstream methods such as CNN, Vision transformer and Swin transformer and the state-of-the-art bone marrow cell recognition method. In addition, Cross-CCL can reduce the workload of human experts while ensuring high accuracy of identification.

Keywords: Blood cell recognition, Vision transformer, Long-tail distribution, Deterministic prediction

目 录

第一章 绪论

1.1 研究背景与意义

血细胞的计数、分类和形态特征在许多血液病的诊断中起着至关重要的作用。目前最常见的方法是由专业医生检查显微镜下的图像生成检验报告。这个过程即费时又容易出错[1]，因此有许多针对某类疾病细胞分类的研究。例如，[2][3]是针对白血病的细胞分类。在他们之前这方面的研究主要是运用传统计算视觉方法[5][6]，这些方法包括但不局限于 PCA、SVM、k 近邻。

通过手工设计特征的方式对血细胞图像进行多分类时通常准确率不高。Acharjee 等人[6]提出了一种用于红细胞半自动计数的方法,它通过霍夫变换检测规定直径的椭圆形和双凹形状的红细胞。Sinha 和 Ramakrishna[1]使用 K-Means clustering 和 EM-algorithm 从背景中分割白细胞，然后人为设计三个特征—形状特征、颜色特征以及纹理特征进行分类。Rezatofghi 和 Soltanian-Zadeh[7]提取细胞的颜色、形态学和纹理特征来识别白血病。这些特征由专家根据细胞特性和经验设计出来。对于一些简单并且类别少的数据集可以获得非常好的结果，但是对于新的数据或复杂的情况它不是一个好的选择。对某个特定数据设计的特征不能直接应用到其他数据中。

最近几年，许多研究人员纷纷将深度学习运用在医学领域，Vgg[8]、Googlenet[9]、Resnet[10]、Densenet[11]等等流行的模型更是取得当时最先进的性能。Zhao J.[12]等人提出了一种基于卷积神经网络的白细胞自动识别与分类系统。而 Luis 等人[44]将卷积神经网络与支持向量机相结合，取得了很好的效果。Andrea A 等人[13]则利用卷积神经网络作为特征提取器，将得到的特征作为 SVM 的输入和仅对卷积神经网络微调进行血细胞分类，他们的结果显示，使用卷积神经网络特征的 SVM 在分类的敏感性、特异性以及精确率都低于端到端的卷积神经网络。CNN 的成功归功于它们逐渐扩大的感受野，可以将结构化图像表示的层次结构作为语义来学习。它通过共享卷积核来提取特征，这样一方面可以极大地降低参数量来避免更多冗余的计算从而提高网络模型计算的效率，另一方面又结合卷积和池化使网络具备一定的平移不变性（shift-invariant）和平移等变性（equivariance）。并且它由浅入深逐步抽象地提取更具备高级语义信息的特

征，而高层的特征表示依赖于底层的特征表示。但研究发现实际感受野是远小于理论感受野[14]，这不利于我们充分的利用上下文信息进行特征的捕获。并且局部信息在经过很多次卷积之后仍然被保留，这说明 CNN 的架构并不是最优的。最近受 Transformer 在 NLP 中取得成功的启发，Dosovitskiy 等人[15]提出了新的视觉识别模型 Vision transformer (ViT)。它的结构与 CNN 不同，主要将图像分成 patchs 序列来做图像识别，在图像分类 、目标检测[16] 以及语义分割[17]上的表现优于 CNN。并且 Vison transformer 拥有比 CNN 更少的图像特异性归纳偏置[18]这使它更具有通用性。另一方面，它使用动态注意力机制[19]，不管输入的序列多长，它都可以捕获序列之间的关系——即它具有全局建模能力。相比 CNN，ViT 的优势就在于利用注意力的方式来捕获全局的上下文信息从而对目标建立起远距离的依赖，从而提取出更强有力的特征，但这也限制了它对局部的细节的关注。而基于局部注意力的 Swin transformer（Swin）[20]则可以捕获细粒度的特征，并且通过可移动窗口（shift window）促使不同区域的信息的流动。

从血细胞的形态上看，一个细胞通常包含许多不同尺度的对象，比如核仁和颗粒就不在同一个尺度。基于全局注意力的 ViT[15]和基于局部注意力的 Swin transformer （Swin）[20]，都不能提取丰富的多尺度信息。Crossformer（Cross）[21]具有提取跨尺度信息的能力，Cross 和 Swin 一样都是使用局部注意力模块，不同的是 Cross 在将图像序列化的时候嵌入了不同尺度的信息。因此从血细胞的形态特点上来看，Cross 更适合用于识别血细胞。

有了好的识别模型，我们还要考虑数据本身的问题。在我们收集的血细胞数据集上，细胞类别的数量呈现出头尾数量严重不平衡的现象，即长尾分布。长尾分布是计算机视觉任务中常见的问题，例如出现异常现象或稀有物种。传统意义上的长尾分布问题是训练集呈现数据不均衡，而测试集均匀分布。这是目前各种针对该问题提出的方法已经非常有效，但仍然无法落地的主要原因。因为实际应用中，类别的分布是不定的，可能呈现长尾分布也可能是均匀分布。不幸的是，我们现有的数据集尾部类的样本极其稀少，主要原因是这些细胞本身在人体中的数量就比较稀少，因此我们的训练集和测试集的分布都是不均衡的。

在现实世界中的长尾数据分布无处不在[22][23]，通常数据规模越大越明显。

特别是与安全和健康相关的应用，比如自动驾驶和医疗诊断，少数类除了数量少还存在难以区分的问题。这对当前深度学习算法的落地应用提出了一个重大的挑战[24][25]，哪怕使用目前最先进的算法也不能够保证网络在生产环境中的对极端少数类仍然有比较好的表现。

现有的解决方法主要有三种类型，分别是类别平衡、特征增强以及模型改进，本文主要讨论类别平衡对血细胞识别的作用。通常类别平衡方法是最容易的解决方法，一种思路是对样本重新抽样，根据不同抽样策略可分为实例平衡采样、类平衡采样、平方根采样以及渐进平衡采样，其中对于分类问题，比较常用的是类平衡采样。在类平衡采样中每个类别被抽取的概率相同，这就会使少数类样本被重复采样，避免网络“遗忘”。不过这种方法也有缺陷，重复采样并不能提供新的信息，即使模型不发生遗忘现象，模型也很容易由于少数类中存在难以区分或复杂的样本而不能做出正确的响应。另一个思路是改进损失函数。带权重的交叉熵函数是在计算损失时对正样本乘以一个系数，该系数等于所有类别数目的中位数除以对应类别的数目，因此尾部类别的权重系数会大于1，而头部类别的权重系数会小于 1。这种方法会降低头部类正样本的梯度，而增大尾部类正样本的梯度，负样本梯度不变。Balanced softmax（BlSoftmax）[26]除了损失值重新加权外，还在训练期间使用标签频率来调整模型预测，以便通过先验知识来缓解类别不平衡的偏差。Focal loss[27]对标准的交叉熵做了巧妙的改进，使之能够相对增大难分样本的梯度，减少易分样本的梯度。一般来说，样本不平衡会导致尾部类训练不充分从而导致识别困难，一个类可以被视为其他类的负样本，导致尾部类接受更多的抑制性梯度，而 Focal loss 从样本区分的难易程度入手解决这个问题，并且难分样本不局限于尾部类，因此 Focal loss 一定程度上能够提升了模型的性能。既然少数类是难以识别的，那么 Center loss[28]通过增加类间距离使类间差异更加明显也是一种可行的方法。Center loss 最早是用于人脸识别任务上，辅助 Softmax loss 进行人脸任务的训练，最小化样本与相应类别中心的距离。还有一些基于这三种损失函数的改进函数，但本质上都是使少数类的差异更加明显。

在想方设法提高模型性能的同时，我们也要思考一个问题，深度神经网络模型拥有一个看似很高的准确率真的可以被放心地进行临床上应用吗？有时不

可靠的预测结果会影响医生对疾病判断。医疗检查是高风险领域，一旦误诊会对病人造成影响，因此在医疗检查领域，人工智能不是取代人类专家而是辅助人类专家更好地完成检查。为了减少检验专家的工作量，又能够提供可靠的结果，必须将模型不能够确定的细胞交给检验专家识别。Guo 等人[29]提出一种带拒识别选项的血细胞的识别方法。他们使用归纳一致性预测器（Inductive Conformal Predictor, ICP）[30]计算校准集来得到不合格分数的分布的无偏估计。ICP 是一致性预测器的[31][32]其中一个变体，它是一种带置信度的域预测机器学习算法。它通过将所有满足置信度的假设类别作为结果输出，从而实现风险的可控。这种方法为人机协作在血细胞识别上的临床应用提供了一个良好的解决方案。它可以使用任意模型作为骨干网络，但 ICP 给出可靠的结果的过程是非常耗时。

理想情况下，该网络应该只在对其能力有高度确定性的情况下进行预测，而将其余无法预测的情况留给医生。因此，为了促进神经网络在血细胞识别中的临床应用，我们提出了一种端到端的细胞形态识别框架，即具有排斥识别选项的类别质心学习(Class Centroid Learning, CCL)。在 CCL 中，通过最小化样本特征向量到正确质心的距离来训练骨干，同时最大化样本特征向量到其他质心的距离，其更新方式类似于指数移动平均[33]。为了使最终的预测更加可靠，我们提出了一种新的训练质心的方法，它可以分离不同的质心，即激励这些质心相互远离。只有同一类别的样本才会聚集在一个质心附近。当骨干输出的特征向量距离除质心 i 外的每一个质心都特别远时，我们可以认为该特征向量属于类别 i，但一旦它在两个质心中间，我们就无法判断它属于哪个类别，这时可以留给专家来识别。

1.2 主要贡献

本篇论文中，我们将多尺度的特征融合提升识别血细胞的能力，并尝试通过修改损失函数来解决数据不平衡的问题，最后我们为神经网络在细胞形态学识别等医疗高风险领域的应用提出了一种新的方式，降低由于网络预测错误带来的风险

本文的主要贡献如下：

（1） 本文提出了一种具有拒识别识别选项的高精度血细胞识别方法——Cross-CCL。它包括两个部分，使用网络结合多尺度学习来识别血细胞和拒绝识别低置信度血细胞。基于我们提出的 CCL 框架，我们可以推进人机协作方法在血细胞形态学检查中的实施。我们的方法可以为专家处理 78.93% 的数据，并保证其错误率小于 2%。

（2） 我们在血细胞识别中使用了结合跨尺度信息的 Transformer。我们证明了包含多尺度信息的结构确实有利于血细胞识别。并且为了检验模型对血细胞数据集的效果，本文提出一种新的评估长尾分布算法有效性的指标，用于检验各类模型及算法对长尾分布数据的有效性。

（3） 此外，为了使不同的类别质心的区分度更加明显，我们根据指数移动平滑(Exponential Moving Average)提出了一种更新质心的方式。它能够让质心相互分离，变得更有区分度。

1.3 章节安排

本文后续的章节安排如下。第二章主要介绍本文使用到的损失函数以及数据集。第三章将介绍能够结合多尺度信息的识别模型和我们提出的质心类别学习，在质心类别学习中，我们将详细描述它的训练机制以及工作原理。第四章我们将展示各种模型的比较结果、我们的类别质心学习跟常用的长尾分布损失函数比较结果，以及类别质心学习与归纳一致性预测器的拒识别能力。最后第五章我们将对全文做一个总结与展望。

第二章 数据集及常用损失函数

2.1 Softmax loss

Softmax loss 是图像分类最常用的损失函数，通过将模型的输出结果输入进 softmax 函数再由用叉熵计算损失，具体实现如下：

$$S_j=\frac{e^{a_j}}{\sum_{k=1}^{N} e^{a_k}} \tag{2-1}$$

$$L=-\log\left(p_y\right) \tag{2-2}$$

公式（2-1）为 softmax 函数，其中 N 为类别数量。公式（2-2）是交叉熵损失，p_y是对应标签 y 的概率值，即其他类别对应的概率值不参与损失计算。

2.2 Focal loss

Focal loss 从样本的辨别难度这个角度来解决样本不平衡的模型训练问题。一般而言少数类由于样本较少，模型对这些类训练不充分，因此也可以视作困难样本。这是 Focal loss 解决长尾分布的思路，相比交叉熵损失，Focal loss 对于分类不正确的样本，损失不变，对于分类正确的样本，损失变小。 整体而言，相当于增加了分类不准确样本在损失函数中的权重。具体形式如下：

$$L=-\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{2-3}$$

其中p_t反映模型预测的结果与真实标签的接近程度，γ为大于 0 的可调因子。

2.3 balanced softmax

一个简单解决样本不平衡的方法就是在损失函数中添加权重因子，由此调节不同类别在模型训练时的梯度。公式如下所示：

$$L=-\log\left(\frac{\pi_y \exp(z_y)}{\sum_j \pi_j \exp(z_j)}\right) \tag{2-4}$$

2.4 Center loss

Center loss 的主要作用是使同类样本更加紧凑。一般情况下，图像分类主要是将模型的输出通过 softmax 归一化后再用交叉熵计算损失。这会使类别之间

可分但类内分布不紧凑，Center loss 通过给每个类别设置一个中心，并将同类别的样本聚集在中心来解决这个问题。损失函数如下：

$$L=L_s+\lambda L_c=-\sum_{i=1}^{m}\log\frac{e^{w_{y_i}^T x_i+b_{y_i}}}{\sum_{j=1}^{n}e^{w_j^T x_i+b_j}}+\frac{\lambda}{2}\sum_{i=1}^{m}\left\|x_i-c_{y_i}\right\|_2^2 \tag{2-5}$$

其中 m 表示一批数据的样本数，n 表示样本类别数目，而λ为比例因子。从公式（2-5）可知，最终的损失函数由 softmax loss 和 Center loss 组成。Softmax loss 使类间可分，而 Center loss 使类内紧凑，从而使少数类也能被识别出来。

2.5 血细胞数据集介绍

我们使用的数据集是由南方医科大学南方医院通过显微镜以及相机收集来的，总计 15574 张图片。该过程使用 10 倍物镜，100 倍目镜来观察细胞形态并由专业的医生完成标注（使用自主开发的标注软件）。我们使用 13 个类别进行识别，包括退化细胞、早幼粒细胞、中晚幼粒细胞、成熟粒细胞、其他粒系细胞、原早幼红细胞、中晚幼红细胞、原幼稚淋巴细胞、淋巴细胞、单核系细胞、浆细胞、巨核细胞、原始粒细胞。它们的数量和示例如表 2-1 和图 2-1 所示。

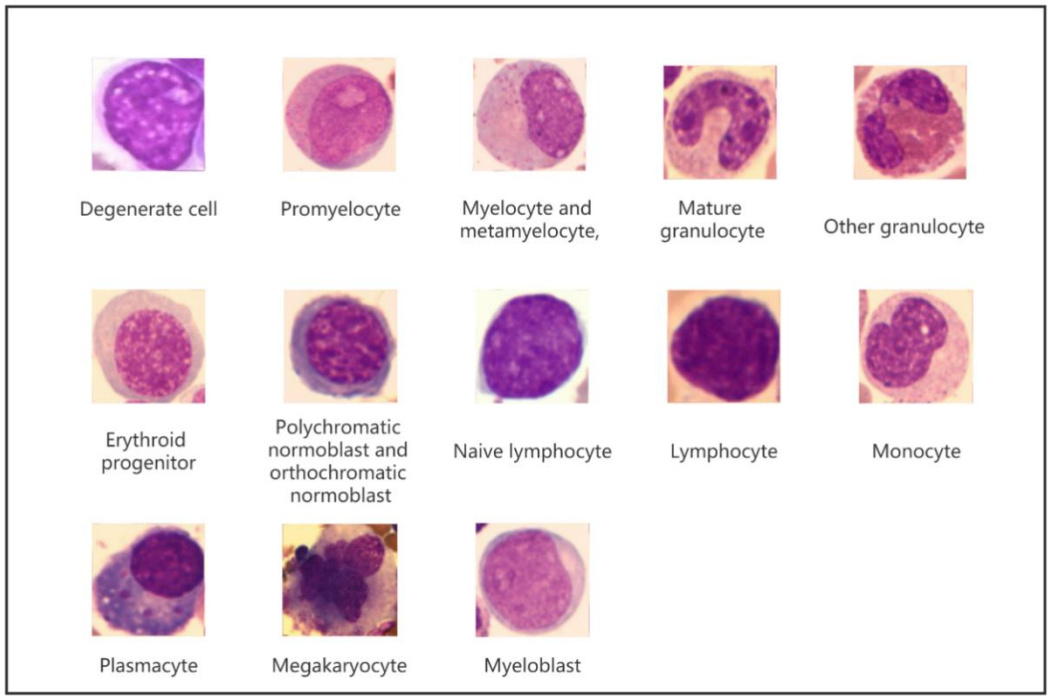

图 2-1 各类血细胞示例

表 2-1 细胞类别及数量

类别	数量
退化细胞	1017
早幼粒细胞	336
中晚幼粒细胞	2417
成熟粒细胞	2774
其他粒系细胞	410
原早幼红细胞	909
中晚幼红细胞	2752
原幼稚淋巴细胞	3207
淋巴细胞	1061
单核系细胞	479
浆细胞	35
巨核细胞	38
原始粒细胞	139
总共	15574

第三章 血细胞识别算法与拒识别机制

在本章中，我们将介绍一种新的血细胞识别方法，该方法由两部分组成。第一部分包括分类网络 Cross 的主要功能模块以及如何提取多尺度信息。第二部分是类质心学习(Class Centroid Learning, CCL)的实现原理，包括如何使用类质心学习训练网络，以及如何拒绝识别不确定单元格。整体流程图如图 3-1 所示。

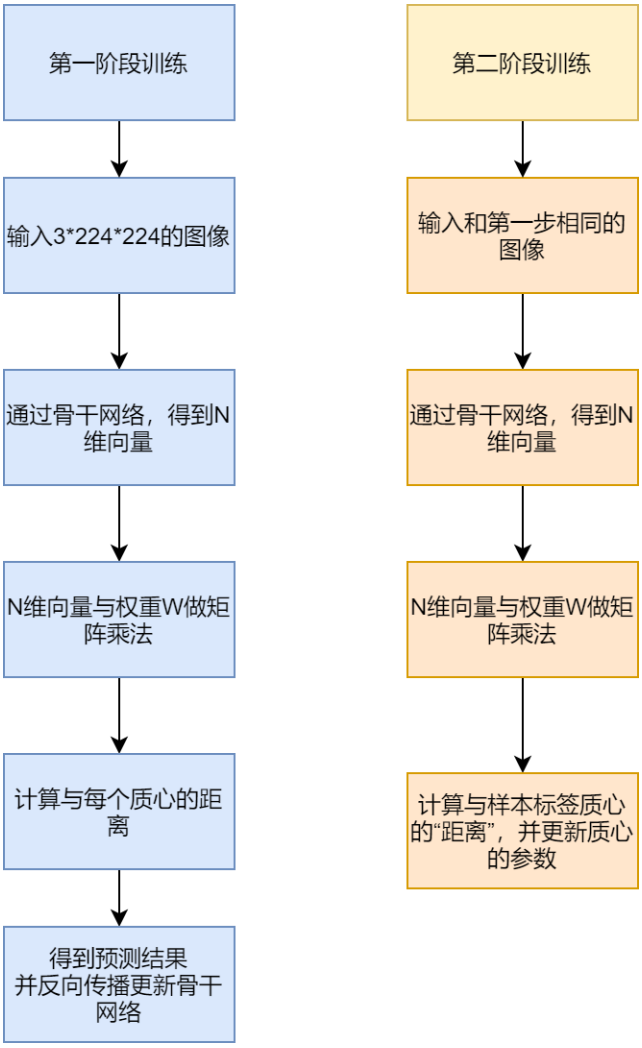

图 3-1 CCL 方法训练流程图

3.1 多尺度信息嵌入

基于全局注意力的 ViT[16] 的每个图像补丁的分辨率一般为 16 或者更大，这将使图像的细节损失。而基于局部注意力的 Swin transformer[20]的图像补丁虽然拥有更小的分辨率，但却让它失去了长距离建模的能力。我们使用

Crossformer[21]的补丁嵌入模块，使每一个 token 都拥有丰富的跨尺度信息。

Crossformer 采用多个不同 kernel 生成图像的补丁（patchs）。为了生成相同数量的补丁还必须使这几个 kernel 的步长保持一样。该模块使用 4*4 的 kernel 进行采样得到细粒度特征，其他大于 4*4 的 kernel 来获取更大尺度的特征。将不同 kernel 生成的嵌入（embedding）合并在一起，即可让 patchs 拥有不同尺度的信息，如图 3-2 所示。实际应用中使用四个卷积来实现这一过程，所有卷积的输入通道都是 3 通道，输出通道一般让小的 kernel 更大一些，比如图 3-2 中 4*4 的 kernel 对应的输出为 48 个通道。经过这四个卷积之后，我们将这四个卷积产生的特征图展开成一维，并在通道（channel）这一维度进行拼接。

3.2 类别质心学习算法

在本节中，我们将介绍 CCL 的整个实施过程。在 3.2.1 节中，我们将介绍如何在训练中计算 Cross 的特征向量到质心的距离和损失函数。在 3.2.2 节中，我们将介绍如何训练和如何更新质心。接下来，在 3.2.3 节中，我们将描述如何计算置信水平以及如何拒绝识别未确定的细胞。

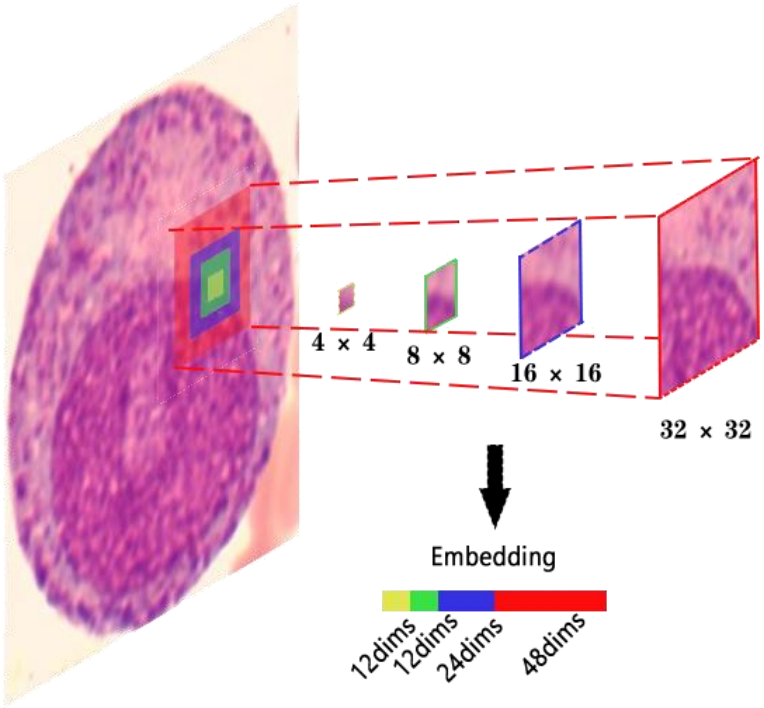

图 3-2 将不同尺度的 embedding 合并为一个 embedding

3.2.1 结构

将图像输入 Cross 后，得到特征向量，然后计算每个类别到质心的距离作为最终的预测结果。也就是说，每张图片的输入对应一个 c 维输出向量，向量的最大值就是预测的类别，向量中的每个值计算如下:

$$D_i(f_\theta(x),CE_i)=\exp\left(-\frac{\frac{1}{M}\|W_i f_\theta(x)-CE_i\|}{2\sigma^2}\right) \tag{3-1}$$

这里$D_i(f_\theta(x),CE_i)$是第 i 个类别的输出。$f_\theta(x)$是 Cross 最后输出的特征向量。另外$f_\theta(x)\in R^N$，N 为向量的长度。CE 是质心的参数，$CE\in R^{M\times C}$，而CE_i 是类别 i 的质心，$CE_i\in R^M$，其中 C 是类别的数量，M 是质心的长度，W_i是类别 i 的权重矩阵，$W_i\in R^{N\times M}$，而$W\in R^{N\times M\times C}$.需要注意的是，我们的这里 Cross 仅仅是作为特征提取器，分类则是通过计算特征向量与不同的质心来完成的。

另外，损失函数定义为：

$$L(x,y) = \sum_i\left[y_i\log(D_i)+\left(1-y_i\right)\log(1-D_i)\right] \tag{3-2}$$

其中y_i是数值为 0 或 1 的真实标签，表示是否属于第 i 个类别。D_i表示样本特征向量与第 i 个类别的距离，离质心越近，D_i越接近 1

3.2.2 类别质心更新

训练 Cross-CCL 的过程分为两个阶段，第一阶段更新 Cross 的参数和 W，第二阶段更新质心的参数。

首先，在第一阶段，我们将一批数据输入到 Cross 中，以获得相应的特征向量。然后利用式(3-1)计算特征向量到每个质心的距离，得到预测结果。然后利用损失函数计算预测结果与实际标签之间的误差，进行反向传播，更新 Cross 的参数和 W。

在第二阶段，我们使用了一种类似于指数移动平均的方法来更新质心，这种方法在 Van den Oord 等人[34]的附录中也有介绍，但我们与他们略有不同。我们的更新方法将根据标签更新类别质心。如果在一个小批中没有特定类别的样

本，那么这一次该类的质心不会更新。为了增加类之间的差异，我们使用了类似于指数移动平均的方法，使类别 i 的质心远离其他类别的质心。

$$CE_i^t=CE_i^{t-1}+\gamma\cdot\{[W_i f_\theta(x)-CE_i^{t-1}]\cdot y_i\} \tag{3-3}$$

$$CE_i^t=CE_i^{t-1}+\beta\cdot\left(CE_i^{t-1}-\frac{1}{C-1}\cdot\sum_j^{j\neq i}CE_j^{t-1}\right) \tag{3-4}$$

其中 y 是经过 one-Hot 编码的标签，y_i表示第 i 个类别的编码值。CE_i^t是第 t 次更新得到的质心，而CE_i^{t-1}则是第 t-1 次更新得到的质心。γ和β通常取 0.0001。

我们通过式(3-3)对样本的特征向量进行更新，使得同一类别的样本最终聚集在相应的质心附近，而式(3-4)则使其他类别的质心远离特定类别的质心，即增加类别之间的差异。我们先用式(3-3)更新 CE，用式(3)更新 CE 后再用式(3-4)。

使用与第一阶段相同的图像通过网络前向传播获得特征向量后，使用式(3-3)和式(3-4)依次更新 CE。需要注意的是，虽然使用与第一阶段相同的图像获得相应的特征向量，但由于网络参数在第一阶段训练时已经更新，所以所有的特征向量都与第一阶段不同。在更新完质心后，接着使用下一批图像重复这两个阶段来训练网络和质心。

3.2.3 拒绝识别机制

在血细胞识别中，我们期望模型所做的预测是可靠和可用的。事实上，由于外部环境等因素，模型做出的判断并不总是可靠的。因此，我们只需要模型给出可靠的预测，即拒绝模型不能区分的样本。

我们定义一个参数置信度阈值c_{th}作为是否拒绝识别的分界线，模型预测的置信度如下所示：

$$c=1-\sum_i D_i+\max(D) \tag{3-5}$$

需要注意的是，输出是由高斯核函数计算的，所以$\sum_i D_i\neq 1$。这里 max(·)是取最大值函数。如果 c 大于等于c_{th}，我们认为结果是可靠的，可以输出预测结果。如果 c 小于c_{th}，那么我们就应该拒绝识别这个样本，并由专家来识别。由于c_{th}受 backbone 的架构以及深度的影响，所以在 4.3 实验中，我们通过使用校准集对c_{th}进行参数搜索，得到一个符合骨干网络的置信水平的c_{th}。

第四章 结果和讨论

我们在血细胞数据集上验证了多尺度信息组合方法的有效性，各种损失函数平衡长尾分布的效果以及 CCL 在拒绝不确定细胞识别方面的有效性。在这个实验中，我们将数据集分为训练集(80%)、验证集(15%)和校准集(5%)。在 ICP 方法中，使用校准集来估计样本的概率分布，即先验概率分布。这类似于 CCL 中的质心，因此骨干的输出只能在校准集的特定类别内被认为是可靠的。在 CCL 中，使用校准集来寻找骨干网络的最佳置信阈值。

我们选择 AdamW[35]作为优化器，初始学习率为 0.0001，β_1为 0.9，β_2为 0.999，权值衰减为 0.01。共训练 100 个 epoch，在第 30 个训练 epoch 和第 60 个训练 epoch 时，学习率衰减，衰减率为 0.1。此外，所有模型都使用预训练的参数。我们使用 PyTorch(一个开源的 Python 机器学习库)来构建模型和创建算法，所有模型都是在使用 Nvidia Titan XP GPU 的 Linux 系统上进行训练的。

4.1 模型比较

我们选择 RegNetY[36]和 ResNeXt50[37]作为 CNN 的代表。我们选择 ResNeXt50 的结果作为基线，是因为 Mateck, C 等人[38]使用 ResNeXt50 作为血细胞细胞形态的分类模型，并证明 ResNeXt50 优于以前的方法，同时对许多具有直接临床相关性的细胞类别实现了极好的准确性。之所以选择 RegNetY，是因为它是近年来最先进的 CNN 模型。对于 Transformer，我们选择了 ViT、Swin 和 Cross。ViT 和 Swin 分别是全局和局部注意力的代表模型。即使是现在，它们仍然具有很强的竞争力。我们选择 RegNetY 中的 4G 和 8G，变压器中的 T 和 S 进行模型比较。RegNetY 的 4G 和 8G 的参数量分别对应 Transformer 中的 T 和 S 的参数。此外，CCL 方法还可以使用其他网络作为骨干网络。

在这里我们使用四种指标来衡量模型及方法的性能，分布是准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1 分数（F1 score），定义如下：

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{4-1}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{4-2}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{4-3}$$

$$\text{F1}=2\cdot\frac{\text{precision}\cdot\text{recall}}{\text{precision+recall}} \tag{4-4}$$

其中 TP 为正类并且也被判定成正类的样本数，FN 为本为正类但判定为负类的样本数，FP 为本为负类但判定为正类的样本数，TN 为一个实例是负类并且也被判定成负类的样本数。

从表 4-1 中我们可以看到，使用 CCL 的卷积神经网络在 ResNeXt50、RegNetY-4G 和 RegNetY-8G 的准确率分别提高了 0.25%、0.22%和 0.44%。用 CCL 训练的 CNN 的召回率略低于常规训练的结果。但精确率有明显提高，特别是 RegNetY-4G 和 ResNeXt50 在使用 CCL 后提高了 1.63%。RegNetY 优于 ResNeXt-50，因为 RegNetY 的模型设置参数是通过参数搜索获得的，并且结构类似 ResNeXt。RegNetY 可以被认为是 ResNeXt 架构的一个很好的解决方案。从这些数据可以看出，CCL 在准确率和精确率方面表现良好。即使使用 CCL 后召回率略有下降，其 F1 分数仍然有一些优势。

ViT 在使用 CCL 之后基本都得到提升，ViT-T-CCL 的准确率提升了 0.25%，精确率提升了 0.29%，召回率提升了 1.02%。而 ViT-S-CCL 的准确率仅提升了 0.09%，精确率提升了 1.09%，召回率下降了 0.33%。CCL 在 Swin 上的表现也和在 ViT 上差不多，Swin-T-CCL 和 Swin-S-CCL 的准确率分别提升了 0.19%和 0.42%，它们的精确率分别提升了 1.53%和 0.99%，只有召回率会略微下降一点。值得注意的是，Cross 是所有骨干网络中表现最好的，其中 Cross-T-CCL 所有指标都优于 Cross-T 以及基线。从表 4-1 分析可知，单独在 ViT 或 Swin 上使用 CCL，得到的提升都不如 Cross。而结合多尺度信息的 Cross 却能够得到明显的提升。说明 CCL 能够帮助模型挖掘出更多有用的语义信息，尤其是在多尺度信息中。另外 Cross 比 Swin 好是因为更大尺度的信息一定程度上弥补了局部注意力机制不能建模长距离依赖关系的能力。

表 4-1 不同模型的比较以及使用 CCL 后的变化

Model	Accuracy	Precision	Recall	F1 score
ResNeXt-50[37]	93.61	89.42	90.48	89.95
ResNeXt-50-CCL	93.86	91.05	89.21	90.12
RegNetY-4G[36]	93.71	89.03	91.01	90.01
RegNetY-4G-CCL	93.93	90.66	89.79	90.22
RegNetY-8G [36]	93.68	91.03	89.54	90.28
RegNetY-8G- CCL	94.12	91.69	89.21	90.43
ViT-T/16[15]	93.87	89.83	90.30	90.06
ViT-T/16-CCL	94.12	90.12	91.32	90.72
ViT-S/16[15]	93.87	89.65	90.38	90.01
ViT-S/16-CCL	93.96	90.74	90.05	90.39
Swin-T[20]	94.09	89.91	90.86	90.38
Swin-T-CCL	94.28	91.44	89.81	90.62
Swin-S[20]	93.77	89.35	90.54	89.94
Swin-S-CCL	94.19	90.34	89.65	89.99
Cross-T[21]	94.19	90.82	92.08	91.45
Cross-T-CCL	**94.44**	91.99	**92.55**	**92.27**
Cross-S[21]	94.41	90.55	92.07	91.30
Cross-S-CCL	94.35	**92.32**	91.85	92.08

我们选择其他网络进行比较的一个原因是，在 CNN 时代，用于大多数血细胞研究的 CNN 很难与最新的最先进的网络相提并论。Mateck, C 等人[38]也证明了他们直接使用 ResNeXt50 优于以往的方法。我们现在比较的方法是在 ResNeXt50 之后发布的最先进的方法。

在我们的不使用 CCL 的分类结果中，Transformer 整体上要优于 CNNs，这一方面也体现 Transformer 这种结构确实能够捕捉到比 CNN 多的有用信息。ViT 能够比 CNNs 优秀，主要是因为它的全局建模能力，也就是它的感受野要比 CNNs 大。与之相反的 Swin，虽然在计算注意力时没法做到建模长距离的关系，但它的可移动窗口（shift windows）让不同的窗口的信息可以流通，因此从整

体上来说，它的感受野也同样比 CNNs 要大。对于架构和 Swin 相似的 Cross 来说，它主要是在每一个 patch 中融合了不同尺度的信息，能够捕获到不同层次的特征。这有助于网络区分不同尺度的物体，比如在图 3-2 中小的卷积核提取到的信息可能与小颗粒相似，这样就很难区分是小颗粒还是细胞核的边界。从分类结果上看，Cross 的准确率是要比其他两个 Transformer 要高的。这也印证了不同尺度的信息有助于血细胞形态识别。

另外，当我们比较同种网络的不同型号时发现，一些大型的网络的准确率反而没有小型号好。按照经验，一般使用一个更大的网络可以提取到更丰富的特征，准确率也因此会更高。然而在这里 RegNetY 和 Swin 出现了模型退化现象。我们分析是这两个网络的浅层网络已经学到足够好的知识，冗余的层数虽然由于残差连接的存在可以接近恒等映射，但不能保证没有损失，这部分损失也许就是造成准确率下降的原因。另一边 Cross 的准确率还能够继续提升，我们认为是多尺度信息发挥的作用，使深层的网络还能够继续提取到丰富的特征。

4.2 长尾分布下的算法对比

在长尾学习中，衡量算法的有效性通常是展示其在所有类的总体表现以及头、中、尾部类中的表现。在不同的任务中所使用的评价指标是不同的。例如，Top-1 准确率或错误率是长尾图像分类中广泛使用的指标，而全类平均精度(mAP)[39]则用于长尾目标检测和实例分割。

通常情况下，图像分类采用 Top-1 准确率衡量算法对长尾分布数据的效果，但这种方法有个缺点，少数类的数目特别少对总体准确率的影响有限，而多数类由于样本较大，准确率也比较容易提升。因此为了能够更好且无偏的衡量 Focal loss、Center loss、BlSoftmax、CCL 的有效性，我们设计了一个新的衡量指标。理想情况下，我们期望输入是均匀分布的，当数据不平衡偏离均匀分布时，少数类对模型性能评估的权重应当上升，公式如下所示：

$$Ld = 1-\sum_i k_i \cdot \left(1-\frac{2R_iP_i}{R_i+P_i}\right) \tag{4-5}$$

其中 k 根据数据集的类别的数量来确定，具体而言第 i 个类别的 k 值为$k_i=\frac{1}{num}-\frac{cls_i}{all}$，num 为类别的数量，$cls_i$为第 i 个类别的样本数量，all 为样本总数。从公式

上看，当多数类的 F1 分数越接近 1 时，Ld 越大，对 Ld 的影响程度取决于偏离平均分布的程度。对长尾分布来说，多数类和少数类的样本数量肯定偏离平均分布，并且偏离程度较大，因此系数 k 也比较大，不同的是多数类的 k 值为负数，而少数类为正数。也就是说我们对多数类的要求比较高，因为多数类样本多更容易训练。而少数类样本少，一般来说我们需要模型对少数类也要拥有不错的识别准确率，因此需要增加少数类在模型评估中的权重。

表 4-2 不同方法对长尾细胞数据集的效果

Model	Recall	Precision	Ld
Swin-T-General	90.86	89.91	95.61
Swin-T-Focal	90.06	91.49	96.72
Swin-T-BlSoftmax	91.35	89.74	96.81
Swin-T-Center	87.15	90.70	94.95
Swin-T-CCL	89.81	91.44	96.52
Swin-T-Focal-CCL	90.88	91.65	96.78

本文提出的 CCL 方法可以使用与 softmax 无关的损失函数，因此我们在对比实验中增加 Swin-T-Focal-CCL 的对比。而 BlSoftmax 依赖 softmax 对输出进行归一化，因此不能与 BlSoftmax 一起使用。Center loss 的输入是模型的输出的特征向量，对特征向量进行的操作反而会影响 CCL 的训练。

从表 4-2 可以看到，BlSoftmax 和正常训练（General）的精确率基本一致，而 BlSoftmax 的召回率比正常训练要高 0.49%。也就是说通过对损失值进行重加权以及通过先验知识来缓解类别不平衡的偏差可以小幅提升模型的召回率。Center loss 的召回率出现了明显的下降，虽然它能扩大类间距离，但在血细胞数据集上相似的类别较多，因此这些相似类别决策域难免会发生重叠。换句话说，Center loss 的原理和度量学习类似，由于它是使用梯度下降更新决策域，相似的类别依然会聚在一起，因此造成召回率下降。Swin-T-Focal-CCL 是我们结合 Focal 和 CCL 两种算法训练的网络，召回率不是最高的，但解决了 CCL 召回率下降的问题，并且精确率和 Ld 相比单独使用 CCL 和 Focal loss 都有提升。值得注意的是，我们在训练时对不同类别进行了加权，放大了少数类的梯度值，其他类不变。Focal 能够在一定程度上能够提升 CCL 的召回率，从原来的下降 1% 回到正常水平。对长尾分布数据分类的能力明显是 BlSoftmax 的方法最好，但 Focal-CCL 与 BlSoftmax 的差距不大。除了 Center loss 其他方法相较正常训练的 Ld 都有明显的提升，并且相差不大。CCL 也能在一定程度上改善长尾分布带来

的影响，综合 CCL 的应用场景来看，实际应用应选择 CCL 或者 Focal-CCL。

4.3 拒绝不确定细胞的能力

首先，我们定义拒识别率 *Rr*，这是被拒绝识别的细胞与总数的比率。接下来，定义接受识别的准确率 *Ra*，这是细胞被接受识别的部分的准确率。

为了测试模型对细胞的识别能力，我们定义了可靠性 *R* 作为度量指标。可靠性 *R* 是接受识别的细胞百分比乘以接受识别的细胞准确率:

$$R=(1-Rr)\cdot Ra \tag{4-6}$$

在实际应用中，我们需要识别准确率尽可能高，而将另外一部分不确定的细胞交给专家来完成。由定义可知可靠性 *R* 越接近 100%，需要专家检查的细胞越少。为了比较 CCL 的优势，我们与 Guo 等人[29]提出的方法进行对比。我们在这里比较模型的 *Ra* 在 99 以上时对应的 *R* 值。需要注意的是，我们是在校准集中让两种方法的 *Ra* 达到 99 以上，也就是期望它在真实环境下的测试也能达到 99 以上。但实际上不管是什么方法或模型，测试结果都会小于估计的。

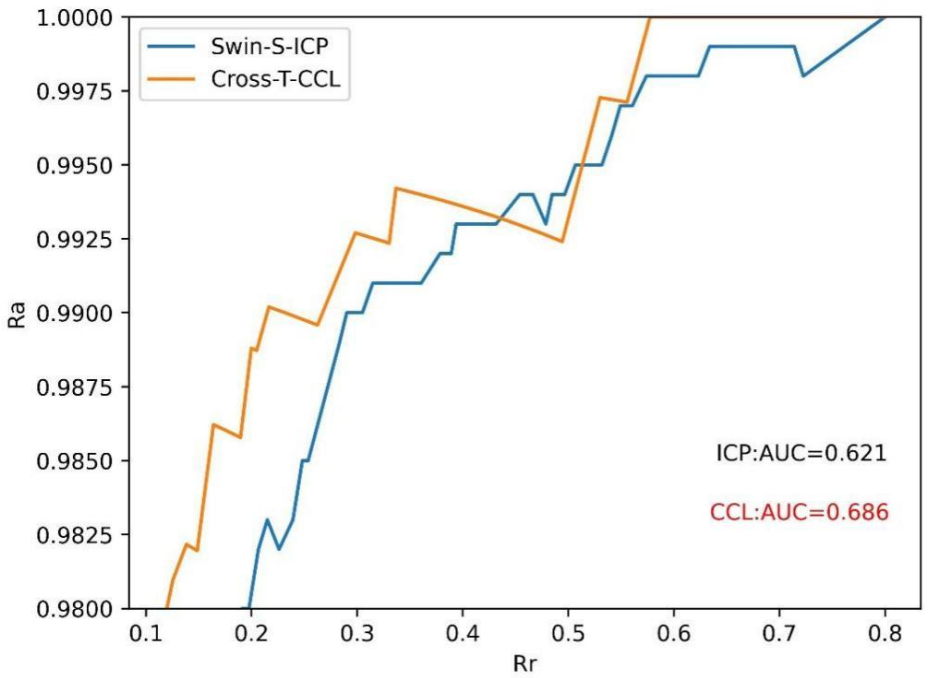

图 4-1 在 0.98-1.0 范围内，不同拒识别率 *Rr* 下对应的接受准确率 *Ra*

从表 4-3 中可以看出，在相同模型上，CCL 明显优于 ICP。对比不同模型的可靠性，可以看出 Transformer 在血细胞这种细粒度的数据集上，依然能够领先

先进的卷积神经网络。在这个实验中， Cross-T-CCL 的 *Rr* 值最低，意味着需要专家进行血细胞形态识别的细胞仅有 21.07%，远低于 Swin-S-ICP 的 27.66%,并且保证剩下的 78.93%是可靠可接受的

图 4-1 展示在 0.98 到 1.0 这个区间中，两个最优组合的比较，这里我们选择了 Swin-S-ICP 和 Cross-T-CCL 来测试 ICP 和 CCL，因为它们在同类中表现最好。在图中，我们希望在保持高准确率的情况下，*Rr* 越低越好。从这两幅图可以看出，在 0.98-1.0 这个区间中，在同样的准确率下 Cross-T-CCL 的拒识别率 *Rr* 明显低于 ICP。另外，比较两者曲线下的面积，我们可以发现在图中 CCL 的 AUC 比 ICP 高 0.065，因此不管是两者数值的比较还是直接观察曲线都可以发现 CCL 的整体性能是要优于 ICP 的。

表 4-3 不同的骨干网络使用 ICP 或 CCL 后的结果

Method	Rr	Ra	R
RegNetY-4G-ICP	33.00	99.00	66.33
RegNetY-4G-CCL	26.25	98.79	72.86
RegNetY-8G-ICP	32.87	99.10	66.53
RegNetY-8G-CCL	26.38	98.96	72.85
ViT-T-ICP	29.01	98.70	70.07
ViT-T-CCL	24.56	98.70	74.46
ViT-S-ICP	28.08	98.90	71.13
ViT-S-CCL	24.75	98.86	74.39
Swin-T-ICP	30.63	98.83	68.56
Swin-T-CCL	25.23	98.86	73.92
Cross-T-ICP	28.00	99.10	71.35
Cross-T-CCL	21.07	98.60	77.82
Swin-S-ICP	27.66	98.88	71.53
Swin-S-CCL	23.79	98.83	75.32
Cross-S-ICP	28.76	98.80	70.39
Cross-S-CCL	21.28	98.54	77.57

另外，神经网络在真实环境下测试，会出现精度下降的情况，我们很难度量这种精度损失。即使不考虑精度下降的情况，准确率的大小也会影响医生做出的诊断结果。此时，将不确定的细胞交给专家识别是最稳妥的，既能保证高

准确率又能减少工作量。从结果上看，我们能保证 78.93%数据的准确率是高于98%的。虽然离 100%准确还有差距，但在神经网络的分类水平已经确定的情况下，要达成更高的准确率必然会造成更多的损失。因此我们没有必要再设置更加苛刻的条件达成更高的准确率，即使是人类专家也不能保证不出错。我们考虑一种情况，一个全自动白细胞识别系统[40]给出的结果只有 95%是正确，并且给出的单核细胞的比例在区间[0.15,0.2]，即检测出来的单核细胞占总数的比例在 0.15~0.2 之间。根据 FAB 标准，急性髓单核细胞白血病的标准之一是单核细胞大于 20%[41]，此时，我们没法确定真正的比例是否超过了 20%，因为模型只有 95%的准确率，因此需要再检查一遍。如果网络能够拒识别不能够确定的细胞，就能降低出现这种情况的可能性。比如使用 CCL 之后，单核细胞的比例在[0.18,0.2]中我们才需要重复检查。从统计学的角度来说落在这个区间的概率要比前一个区间([0.15,0.2])小很多。从这个特例，我们也能看到拒绝识别不确定细胞这一方法的优势。此外，它也可以用在其他需要高精度的医学检测，比如皮肤病变分类[42]和彩色眼底图像分类[18]。

4.4 t-SNE 嵌入

我们使用 t-SNE[43]嵌入分析了网络(Swin-S, Cross-S, ResNeXt50)提取的特征，以证明我们提出的方法的先进性。每个点代表一个样本，不同颜色和形状的组合代表一个类别。

图 4-2 ResNeXt50, ResNeXt50-CCL, Swin-T, Swin-T-CCL, Cross-T, Cross-T-CCL 的二维嵌入结果

从图 4-2 中我们可以看到，无论使用哪种方法，相似的类别都映射到相邻的位置，例如原早幼红细胞和中晚幼红细胞是不同生长阶段的红细胞，它们的特征会非常相似。CCL 可以显著地使同一类别的样本更加紧凑，增加类别之间的

距离。这对于细粒度分类非常有效，因为细粒度数据集在类别之间往往没有什么区别，这使得模型很难正确分类。ResneXt50-CCL 虽然具有较好的识别能力，但效果明显低于 Swin-T-CCL 和 Cross-T-CCL。此外，Cross 还在一定程度上增加了类间距离，说明它可以有效提取不同尺度的特征，从而提高了不同类别之间的区分能力。CCL 对 Cross 的影响最为明显，这表明结合不同尺度的信息可以有效提高模型识别细胞的能力。

4.5 时间复杂度

如前所述，ICP 的推断过程非常耗时，我们在表 4-4 中显示了两种方法的时间复杂度，表中 ICP 时间复杂度 K 为类别数与校准集数据量的乘积。两种方法的时间复杂度都是线性的，但 ICP 比 CCL 要高。因为 K 是一个大于或等于 1 的系数，所以通常不可能看起来是 1。一般情况下，类别和校准集的样本量越大，K 越大，而为了尽可能接近原始数据分布，校准集会很广泛，这将导致较长的推理时间。

我们使用 Nvidia Titan XP 测试 Cross-T 在两种算法上的推理时间，如表 4-5 所示。此外，校准集中的图片数量为 781 张。ICP 处理 2354 张图像的时间为 26.14 秒，是 CCL 时间的 3 倍多。需要注意的是，这里的时间还包括网络的推断时间。总之，我们的方法也比 ICP 方法消耗更少的时间，并且不受类别数量和其他数据的影响。

表 4-4 ICP 和 CCL 的时间复杂度

Method	Time complexity
ICP	O(Kn)
CCL	O(n)

表 4-5 Cross-T 在两种方法上识别验证集的 2354 张图像

Method	Mean time
Cross-T-ICP	26.14s
Cross-T-CCL	8.39s

第五章 结论与展望

本文我们主要讨论在医学领域中，如何提升使用深度学习模型识别血细胞的能力。传统的分类方法往往需要手动设计特征，并且这些特征往往是基于专家经验的，对于新的数据或复杂的情况难以应用。而深度学习可以对输入数据进行端到端的学习和特征提取，并且在医学领域中已经取得了很好的效果。其中，卷积神经网络作为一种流行的深度学习模型，通过共享卷积核来提取特征，并从浅到深逐步抽象地提取更具有高级语义信息的特征。然而，CNN 在处理局部信息和利用上下文信息方面存在一些不足，这启发了研究人员开发新的模型。目前，Vision transformer 和 Swin transformer 是代表 Transformer 的两种模型，它们可以有效地捕获上下文信息并提取更强有力的特征。尤其是 Swin transformer，它通过可移动窗口的方式，可以更好地捕获细节和局部信息。在实际表现中，CNN 最优结果为 ResNetY-4G 的 93.71%，ViT 以 93.87%略高于 CNN，而 Swin 和 Cross 分别为 94.09%和 94.41%，其中 Cross 的最优结果比 CNN 高了 0.7%。从基础模型的比较中，我们可以看出基于注意力机制的 Transformer 完全可以替代 CNN，并且对于血细胞这种包含细粒度特征的数据来说，更适合用基于局部注意力的 Transformer。

此外，从血细胞的形态特点上来看，一个细胞通常包含许多不同尺度的对象，如核仁和颗粒不在同一个尺度。然而，基于全局注意力的 ViT 和基于局部注意力的 Swin transformer 均无法提取丰富的多尺度信息。这启发了研究人员提出了具有提取跨尺度信息能力的 Crossformer，它与 Swin transformer 一样使用局部注意力模块，并通过将图像序列化时嵌入不同尺度的信息来提取跨尺度信息。因此，从血细胞的形态特点看，Crossformer 更适合用于识别血细胞。 另外，在实际使用中，我们发现我们的血细胞数据集中细胞类别的数量呈现出头尾数量严重不平衡的长尾分布现象。针对这种情况，现有的解决方法主要有三种类型，分别是类别平衡、特征增强以及模型改进。本文主要使用损失函数来解决长尾分布问题。从实验结果来看，我们的类别质心学习与 Focal loss 一起使用可以达到非常好的效果，召回率 90.88%仅次于 Blsoftmax 的 91.35%，精确率 91.65%最高并且比 CCL 高 0.21%，在我们提出的指标 Ld 上达到 96.78%略低于最高值 Blsoftmax 的 96.81%。这些方法在解决长尾分布问题方面已经有了很好

的效果，但在解决真实环境下的样本不平衡的问题上，仍需要更多的探索和发展。

同时为了确保模型的可靠性，需要让人工智能和专家共同协作，而不是完全取代专家。我们提出了一种类别质心学习方法来训练骨干网络，以提高血细胞识别的准确度，并将不能确定的图像交给专家识别。在同类型的模型及相同接受识别的准确率中，我们的 CCL 的可靠性 *R* 比 ICP 要高 3.26~7.18%，并且有更低的拒识别率。

总而言之，我们提出的方法不仅在分类上是最优的，在一定程度上缓解长尾分布带来的模型性能下降，而且在保证较高精度的情况下，拒绝识别低置信度细胞的任务中也是最优的。我们的 Cross-CCL 可以分担近 80%的识别工作量，同时保持 98%以上的高精度。我们提出的方法也适用于其他需要较高精度的领域，如彩色眼底图像的分类。未来，我们将继续优化针对血细胞形态特征的识别网络，继续提高 CCL 的性能，希望其拒识别率能够接近分类错误率，同时保持大于 99%的准确率。

参考文献

[1] Sinha N and Ramakrishnan A G. Automation of differential blood count[J]. TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, 2003, 2: 547-551.

[2] Lee L H, Mansoor A, Wood B, et al. Performance of cellavision DM96 in leukocyte classifification[J]. Journal of Pathology Informatics, 2013, 4(1): 14-14.

[3] Young I T. The classifification of white blood cells[J]. IEEE Transactions on Biomedical Engineering, 1972, BME-19(4): 291–298.

[4] P rakisya N P T, Liantoni, F, Hatta, P, Aristyagama, Y H and Setiawan, A. Utilization of K-nearest neighbor algorithm for classification of white blood cells in AML M4, M5, and M7[J]. Open Engineering, 2021, 11(1): 662-668.

[5] Joshi MD, Karode AH, Suralkar S. White blood cells segmentation and classification to detect acute leukemia[J]. Int J Emerging Trends Technol Computer Sci (IJETICS), 2013;2(3):147-151.

[6] Acharjee S, Chakrabartty S, Alam M I, Dey N, Santhi V and Ashour A S. A semiautomated approach using GUI for the detection of red blood cells[J]. 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016: 525-529.

[7] Rezatofighi S H, Khaksari K, Soltanian-Zadeh H. Automatic recognition of five types of white blood cells in peripheral blood[J]. Image Analysis and Recognition, 7th International Conference on Image Analysis and Recognition, 2010, 6112:161-172.

[8] Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. In International Conference on Learning Representations, 2015.

[9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.

[10] He K, Zhang X, Ren S, and Sun J. Deep Residual Learning for Image Recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[11] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. Densely Connected Convolutional Networks[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261-2269.

[12] Zhao J, Zhang M, Zhou Z, et al. Automatic detection and classification of leukocytes using convolutional neural networks[J]. Med Biol Eng Comput, 2017, 55(8): 1287–1301.

[13] Acevedo A, Alférez S, Merino A, Puigví L and Rodellar J. Recognition of peripheral blood cell images using convolutional neural networks[J].Computer Methods and Programs in Biomedicine, 2019, 180: 105020.

[14] Luo W, Li Y, Urtasun R and Zemel R. Understanding the effective receptive field in deep convolutional neural networks[J]. In Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 4905–4913.

[15] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. In International Conference on Learning Representations, 2021.

[16] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. End-to-end object detection with transformers[J]. European Conference on Computer Vision, 2020, 12346: 213–229.

[17] Ranftl R, Bochkovskiy A and Koltun V. Vision transformers for dense prediction[J]. arXiv preprint arXiv:2103.13413, 2021.

[18] Gómez-Valverde J J, Antón A, Fatti G, et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning[J]. Biomed. Opt. Express, 2019, 10(2): 892-913.

[19] Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need[J]. In Proceedings of NeurIPS, 2017: 5998–6008.

[20] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[J]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9992-10002.

[21] Wang W, Yao L, Chen L, et al. CrossFormer: A Versatile Vision Transformer

Based on Cross-scale Attention[J]. International Conference on Learning Representations, 2021.

[22] Liu Z, Miao Z, Zhan X, et al, Large-scale long-tailed recognition in an open world[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[23] Cui Y, Jia M, Lin T-Y, et al. Classbalanced loss based on effective number of samples[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[24] Cao K, Wei C, Gaidon A, Arechiga N and Ma T. Learning imbalanced datasets with label-distribution-aware margin loss[J]. Advances in neural information processing systems, 2019, 32: 1567-1578.

[25] Tan J, Wang C, Li B, et al. Equalization loss for long-tailed object recognition[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11659-11668.

[26] Ren J, Yu C, Ma X, et al. Balanced meta-softmax for long-tailed visual recognition[J]. Advances in Neural Information Processing Systems, 2020, 33: 4175–4186.

[27] Lin T-Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE International Conference on Computer Vision (ICCV), 2017: 2999-3007.

[28] Wen Y, Zhang K, Li Z, Qiao Y. A Discriminative Feature Learning Approach for Deep Face Recognition[J]. European Conference on Computer Vision, 2016, 9911.

[29] Guo L, Huang P, He H, et al. A method to classify bone marrow cells with rejected option[J]. Biomedical Engineering/Biomedizinische Technik, 2022, 67(3): 227–236.

[30] Papadopoulos H, Proedrou K, Vovk V, et al. Inductive Confidence Machines for Regression[J]. 13th European Conference on Machine Learning Springer-Verlag, 2002.

[31] Vovk V, Gammerman A, Shafer G. Algorithmic Learning in a Random World[M]. New York:Springer, 2005: 191–197.

[32] Gammerman A, Vovk V. Hedging predictions in machine learning[J]. The Computer Journal, 2007, 50(2): 151–163.

[33] Klinker F. Exponential moving average versus moving exponential average[J]. Mathematische Semesterberichte, 2011, 58: 97–107.

[34] van den Oord A, Vinyals O, Kavukcuoglu K, et al. Neural discrete representation learning[J]. Neural Information Processing Systems, 2017: 6306–6315.

[35] Loshchilov I and Hutter F. Fixing Weight Decay Regularization in Adam[J]. ArXiv, 2017, abs/1711.05101.

[36] Radosavovic I, Kosaraju R P, Girshick R, et al. Designing network design spaces[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10428–10436.

[37] Xie S, Girshick R, Dollár P, Tu Z and He K. Aggregated residual transformations for deep neural networks[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5987–5995.

[38] Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks[J]. Nat Mach Intell, 2019, 1(11): 538- 544.

[39] Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I and Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[40] Shahin A I, Guo Y, Amin K M, et al. White blood cells identification system based on convolutional deep neural learning networks[J]. Computer Methods and Programs in Biomedicine, 2019, 168: 69-80.

[41] Theml H, Diem H, Haferlach T. Color Atlas of Hematology[M]. New York:Thieme, 2004: 92-93.

[42] Hsu B W-Y, Tseng V S. Hierarchy-aware contrastive learning with late fusion for skin lesion classification[J]. Computer Methods and Programs in Biomedicine, 2022, 216.

[43] Der Maaten L V, Hinton G E. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008: 2579-2605.

[44] Vogado L H S, Veras R M S, Araujo F H D, et al. Leukemia diagnosis in blood

slides using transfer learning in CNNs and SVM for classification[J]. Engineering Applications of Artificial Intelligence, 2018, 72: 415-422.

附录

类别质心学习的伪代码如下：

```
def forward(x):
    Feature = backbone(x)  # 将输入通过骨干网络
    Output = classifier(feature)  # 将网络输出的特征作为分类器的输入
    Return output


def classifier(x):
    x = x * w # x 的大小为 BN w 为 NMC，矩阵乘积之后为 BMC
    Dist = mean((x - center)^2, dim=1)  #dist 大小为 BC
    Dist = exp(-dist/sigma**2)  #sigma 通常取 1 效果比较好
    Return dist
def update_centroid(x, y):
    feature = backbone(x)
    # 更新质心
    Dist = feature * w - centroid.view(1,M,C) # 这里 view 的操作为增加维度，与被减向量的维度对齐，增加的维度的大小将从 1 自动对齐
    Dist = sum(dist.T · y.view(B, C, 1), dim=0).T  # dist 后两个维度转置与 y 的扩展做哈达玛积
    Grad = dist / sum(y, dim=0).view(1, C)
    Centroid =  centroid + gamma * grad
    #使不同质心彼此分离
    Mask = 1 - diag(C, C)
    Depart = centroid.view(M, 1, C) · Mask.view(1, C, C) # 哈达玛积，depart 大小为 MCC
    Centroid = centroid + gamma*(centroid - mean(depart, dim=2))
```

致谢

转眼间，本科学习生活即将结束。在此篇论文完成之际，我要对曾经关心过我、帮助过我的人致以由衷的感谢。以前每次出远门读书，心里总会有种害怕的感觉，害怕面对新事物，现在已经没有这种感觉，我想这也是我的成长吧。

首先衷心感谢我的导师马琼雄，真心感谢尊敬的导师给予我的教诲。在马老师的指导下，我得以接受科研训练。马老师为人严谨，知识渊博，在我论文和专利的写作过程中，马老师用他丰富的经验，指导我写作与修改，让我少走了许多的弯路。

然后感谢同窗的各位同学，尤其是我的室友们。你们在我的学习和生活中提供了大量的无私帮助，这份同窗之情将是我最值得留恋的回忆。

最后再次感谢我的父母、家人、朋友给我学业和生活的帮助与关心，感谢你们的支持，让我得以顺利完成学业

何海森

2023 年 4 月 1 日