蘭州大學
LANZHOU UNIVERSITY

认知科学基础第十二周作业

实验报告

题　　目：	第七次作业
上课时间：	2021年11月25日
授课教师：	刘振宇
姓　　名：	周功海
学　　号：	320190903781
日　　期：	2021年11月25日

基于ArcFace论文模型对本节题目的解读与反思

周功海，320190903781

兰州大学信息科学与工程学院

摘　要： ArcFace/InsightFace（弧度）是伦敦帝国理工学院邓建康等在2018.01发表，在SphereFace基础上改进了对特征向量归一化和加性角度间隔，提高了类间可分性同时加强类内紧度和类间差异。我本次检索论文即为本篇，标题为ArcFace: Additive Angular Margin Loss for Deep Face Recognition

关键词： Face recognition model, Arcface,

Interpretation and reflection on the topic of this section based on the ArcFace paper model

Zhou Gonghai

School of Information Science and Engineering, Lanzhou University

Abstract: ArcFace/InsightFace was published by Deng Jiankang and others at Imperial College London in January 2018. Based on SphereFace, it improves the normalization of feature vectors and additive angle intervals, which improves the separability between classes and at the same time strengthens intra-class tightness and inter-class differences. The paper I searched this time is this one, and the title is ArcFace: Additive Angular Margin Loss for Deep Face Recognition

Key Words: Figure; Traversal; Breadth-First-Search; and Depth-First-Search

1　题目
2　人脸识别简介
2.1　人脸识别分类
2.2　人脸识别流程
2.2.1　人脸图像的采集与预处理
2.2.1.1　人脸图像的采集
2.2.1.2　人脸图像的预处理
2.2.2　人脸检测
2.2.2.1　基于肤色模型的检测
2.2.2.2　基于边缘特征的检测
2.2.2.3　基于统计理论方法
2.2.3　人脸特征提取
2.2.4　人脸识别
2.2.5　活体鉴别
2.3　人脸识别主要方法
2.3.1　基于特征脸的方法
2.3.2　基于几何特征的方法
2.3.3　基于深度学习的方法
2.3.4　基于支持向量机的方法
2.3.5　其他综合方法
3　论文阅读：
4　论文介绍

4.1 摘要
4.2 现有方法缺陷
4.3 提出的方法
4.4 SphereFace与CosFace的比较
4.5 部分数据集评估结果
4.6 Conclusions
5 回答题目中的问题
5.1 人脸识别模型:
5.2 实现方式:
5.3 问题二：二者的异同
5.3.1 人工神经网络结构是基于生物学观察
5.3.2 人的视觉系统与机器的差别
5.3.3 人脑会构造事物的 3D 模型
5.3.4 大脑的神经元比计算机的集成电路慢得多。
5.3.5 确定性与非确定性

1 题目

检索一篇3 年内发表的有关人脸识别的SCI 文章，阅读并回答下面问题：

1.文章使用的人脸识别模型是什么样的？它以何种方式抽取人脸的何种特征用以识别人脸？这种模型的局限性何在？

2.请你详细对比人脸识别的计算机模型和本课学的认知神经科学模型的异同。

2 人脸识别简介

解答：我认为在谈人脸识别时，有必要重申下人脸识别的定义，以便之后更好的展开:

2.1 人脸识别分类

人脸识别问题宏观上分为两类：1. 人脸验证（又叫人脸比对）2. 人脸识别。

人脸验证做的是1 比1 的比对，即判断两张图片里的人是否为同一人。最常见的应用场景便是人脸解锁，终端设备（如手机）只需将用户事先注册的照片与临场采集的照片做对比，判断是否为同一人，即可完成身份验证。

人脸识别做的是1 比N 的比对，即判断系统当前见到的人，为事先见过的众多人中的哪一个。比如疑犯追踪，小区门禁，会场签到，以及新零售概念里的客户识别。

这些应用场景的共同特点是：人脸识别系统都事先存储了大量的不同人脸和身份信息，系统运行时需要将见到的人脸与之前存储的大量人脸做比对，找出匹配的人脸。

2.2 人脸识别流程

人脸识别技术原理简单来讲主要是三大步骤：一是建立一个包含大批量人脸图像的数据库，二是通过各种方式来获得当前要进行识别的目标人脸图像，三是将目标人脸图像与数据库中既有的人脸图像进行比对和筛选。

根据人脸识别技术原理具体实施起来的技术流程则主要包含以下四个部分，即人脸图像的采集与预处理、人脸检测、人脸特征提取、人脸识别和活体鉴别。

图1: 人脸识别技术流程

2.2.1 人脸图像的采集与预处理

人脸图像的采集与检测具体可分为人脸图像的采集和人脸图像的检测两部分内容。

2.2.1.1 人脸图像的采集

采集人脸图像通常情况下有两种途径，分别是既有人脸图像的批量导入和人脸图像的实时采集。一些比较先进的人脸识别系统甚至可以支持有条件的过滤掉不符合人脸识别质量要求或者是清晰度质量较低的人脸图像，尽可能的做到清晰精准的采集。

既有人脸图像的批量导入:即将通过各种方式采集好的人脸图像批量导入至人脸识别系统，系统会自动完成逐个人脸图像的采集工作。

人脸图像的实时采集:即调用摄像机或摄像头在设备的可拍摄范围内自动实时抓取人脸图像并完成采集工作。

2.2.1.2 人脸图像的预处理

人脸图像的预处理的目的是在系统对人脸图像的检测基础之上，对人脸图像做出进一步的处理以利于人脸图像的特征提取。

人脸图像的预处理具体而言是指对系统采集到的人脸图像进行光线、旋转、切割、过滤、降噪、放大缩小等一系列的复杂处理过程来使得该人脸图像无论是从光线、角度、距离、大小等任何方面来看均能够符合人脸图像的特征提取的标准要求。

在现实环境下采集图像，由于图像受到光线明暗不同、脸部表情变化、阴影遮挡等众多外在因素的干扰，导致采集图像质量不理想，那就需要先对采集到的图像预处理，如果图像预处理不好，将会严重影响后续的人脸检测与识别。研究介绍三种图像预处理手段，即灰度调整、图像滤波、图像尺寸归一化等。

灰度调整:因为人脸图像处理的最终图像一般都是二值化图像，并且由于地点、设备、光照等方面的差异，造成采集到彩色图像质量不同，因此需要对图像进行统一的灰度处理，来平滑处理这些差异。灰度调整的常用方法有平均值法、直方图变换法、幂次变换法、对数变换法等。

灰度调整:

因为人脸图像处理的最终图像一般都是二值化图像，并且由于地点、设备、光照等方面的差异，造成采集到彩色图像质量不同，因此需要对图像进行统一的灰度处理，来平滑处理这些差异。灰度调整的常用方法有平均值法、直方图变换法、幂次变换法、对数变换法等。

图像滤波:

在实际的人脸图像采集过程中，人脸图像的质量会受到各种噪声的影响，这些噪声来源于多个方面，比如周围环境中充斥大量的电磁信号、数字图像传输受到电磁信号的干扰等影响信道，进而影响人脸图像的质量。为保证图像的质量，减小噪声对后续处理过程的影响，必须对图像进行降噪处理。去除噪声处理的原理和方法很多，常见的有均值滤波，中值滤波等。目前常用中值滤波算法对人脸图像进行预处理。

图像尺寸归一化:

在进行简单的人脸训练时候，遇到人脸库的图像像素大小不一样时，我们需要在上位机人脸比对识别之前对图像做尺寸归一化处理。需要比较常见的尺寸归一化算法有双线性插值算法、最近邻插值算法和立方卷积算法等。

2.2.2 人脸检测

一张包含人脸图像的图片通常情况下可能还会包含其他内容，这时候就需要进行必要的人脸检测。也就是在一张人脸图像之中，系统会精准的定位出人脸的位置和大小，在挑选出有用的图像信息的同时自动剔除掉其他多余的图像信息来进一步的保证人脸图像的精准采集。

人脸检测是人脸识别中的重要组成部分。人脸检测是指应用一定的策略对给出的图片或者视频来进行检索，判断是否存在着人脸，如果存在则定位出每张人脸的位置、大小与姿态的过程。人脸检测是一个具有挑战性的目标检测问题，主要体现在两方面:

人脸目标内在的变化引起:（1）人脸具有相当复杂的细节变化和不同的表情(眼、嘴的开与闭等)，不同的人脸具有不同的外貌，如脸形、肤色等；（2）人脸的遮挡，如眼镜、头发和头部饰物等。

外在条件变化引起：（1）由于成像角度的不同造成人脸的多姿态，如平面内旋转、深度旋转以及上下旋转等，其中深度旋转影响较大；（2）光照的影响，如图像中的亮度、对比度的变化和阴影等；（3）图像的成像条件，如摄像设备的焦距、成像距离等。

人脸检测的作用，便是在一张人脸图像之中，系统会精准的定位出人脸的位置和大小，在挑选出有用的图像信息的同时自动剔除掉其他多余的图像信息来进一步的保证人脸图像的精准采集。人脸检测重点关注以下指标：

检测率: 识别正确的人脸/图中所有的人脸。检测率越高，检测模型效果越好；
误检率: 识别错误的人脸/识别出来的人脸。误检率越低，检测模型效果越好；
漏检率: 未识别出来的人脸/图中所有的人脸。漏检率越低，检测模型效果越好；
速度: 从采集图像完成到人脸检测完成的时间。时间越短，检测模型效果越好。

目前的人脸检测方法可分为三类，分别是基于肤色模型的检测、基于边缘特征的检测、基于统计理论方法1，下面将对其进行简单的介绍：

2.2.2.1 基于肤色模型的检测

肤色用于人脸检测时，可采用不同的建模方法，主要有高斯模型、高斯混合模型，以及非参数估计等。利用高斯模型和高斯混合模型可以在不同颜色空间中建立肤色模型来进行人脸检测。通过提取彩色图像中的面部区域以实现人脸检测的方法能够处理多种光照的情况，但该算法需要在固定摄像机参数的前提下才有效。Comaniciu 等学者利用非参数的核函数概率密度估计法来建立肤色模型，并使用mean-shift 方法进行局部搜索实现了人脸的检测和跟踪。这一方法提高了人脸的检测速度，对于遮挡和光照也有一定的鲁棒性。该方法的不足是和其他方法的可结合性不是很高，同时，用于人脸检测时，处理复杂背景和多个人脸时存在困难。

为了解决人脸检测中的光照问题，可以针对不同光照进行补偿，然后再检测图像中的肤色区域。这样可以解决彩色图像中偏光、背景复杂和多个人脸的检测问题，但对人脸色彩、位置、尺度、旋转、姿态和表情等具有不敏感性。

2.2.2.2 基于边缘特征的检测

利用图像的边缘特征检测人脸时，计算量相对较小，可以实现实时检测。大多数使用边缘特征的算法都是基于人脸的边缘轮廓特性，利用建立的模板（如椭圆模版）进行匹配。也有研究者采用椭圆环模型与边缘方向特征，实现简单背景的人脸检测。Fröba 等采用基于边缘方向匹配（Edge-Orientation Matching，EOM）的方法，在边缘方向图中进行人脸检测。该算法在复杂背景下误检率比较高，但是与其他的特征相融合后可以获得很好的效果。

2.2.2.3 基于统计理论方法

本文重点介绍基于统计理论方法中的Adaboost人脸检测算法。Adaboost算法是通过无数次循环迭代来寻求最优分类器的过程。用弱分类器Haar特征中任一特征放在人脸样本上，求出人脸特征值，通过更多分类器的级联便得到人脸的量化特征，以此来区分人脸和非人脸。Haar功能由一些简单黑色白色水平垂直或旋转45°的矩形组成。目前的Haar特征总的来说广义地分为三类：边缘特征、线特征以及中心特征2。

这一算法是由剑桥大学的Paul Viola 和Michael Jones 两位学者提出，该算法优点在于不仅计算速度快，还可以达到和其他算法相当的性能，所以在人脸检测中应用比较广泛，但也存在着较高的误检率。因为在采用Adaboost 算法学习的过程中，最后总有一些人脸和非人脸模式难以区分，而且其检测的结果中存在一些与人脸模式并不相像的窗口。

2.2.3 人脸特征提取

目前主流的人脸识别系统可支持使用的特征通常可分为人脸视觉特征、人脸图像像素统计特征等，而人脸图像的特征提取就是针对人脸上的一些具体特征来提取的。特征简单，匹配算法则简单，适用于大规模的建库；反之，则适用于小规模库。特征提取的方法一般包括基于知识的提取方法或者基于代数特征的提取方法。

以基于知识的人脸识别提取方法中的一种为例，因为人脸主要是由眼睛、额头、鼻子、耳朵、下巴、嘴巴等部位组成，对这些部位以及它们之间的结构关系都是可以用几何形状特征来进行描述的，也就是说每一个人的人脸图像都可以有一个对应的几何形状特征，它可以帮助我们作为识别人脸的重要差异特征，这也是基于知识的提取方法中的一种。

2.2.4 人脸识别

我们可以在人脸识别系统中设定一个人脸相似程度的数值，再将对应的人脸图像与系统数据库中的所有人脸图像进行比对，若超过了预设的相似数值，那么系统将会把超过的人脸图像逐个输出，此时我们就需要根据人脸图像的相似程度高低和人脸本身的身份信息来进行精确筛选，这一精确筛选的过程又可以分为两类：其一是一对一的筛选，即对人脸身份进行确认过程；其二是一对多的筛选，即根据人脸相似程度进行匹配比对的过程。

2.2.5 活体鉴别

生物特征识别的共同问题之一就是要区别该信号是否来自于真正的生物体，比如，指纹识别系统需要区别带识别的指纹是来自于人的手指还是指纹手套，人脸识别系统所采集到的人脸图像，是来自于真实的人脸还是含有人脸的照片。因此，实际的人脸识别系统一般需要增加活体鉴别环节，例如，要求人左右转头，眨眼睛，开开口说句话等。

2.3 人脸识别主要方法

人脸识别技术的研究是一个跨越多个学科领域知识的高端技术研究工作，其包括多个学科的专业知识，如图像处理、生理学、心理学、模式识别等知识。在人脸识别技术研究的领域中，目前主要有几种研究的方向，如：一种是根据人脸特征统计学的识别方法，其主要有特征脸的方法以及隐马尔科夫模型（HMM，Hidden Markov Model）方法等；另一种人脸识别方法是关于连接机制的，主要有人工神经网络（ANN，Artificial Neural Network）方法和支持向量机（SVM，Support Vector Machine）方法等；还有一个就是综合多种识别方式的方法。

2.3.1 基于特征脸的方法

特征脸的方法是一种比较经典而又应用比较广的人脸识别方法，其主要原理是把图像做降维算法，使得数据的处理更容易，同时，速度又比较快。特征脸的人脸识别方法，实际上是将图像做Karhunen-Loeve 变换，把一个高维的向量转化为低维的向量，从而消除每个分量存在的关联性，使得变换得到的图像与之对应特征值递减。在图像经过K-L 变换后，其具有很好的位移不变性和稳定性。所以，特征脸的人脸识别方法具有方便实现，并且可以做到速度更快，以及对正面人脸图像的识别率相当高等优点。但是，该方法也具有不足的地方，就是比较容易受人脸表情、姿态和光照改变等因素的影响，从而导致识别率低的情况。

2.3.2 基于几何特征的方法

基于几何特征的识别方法是根据人脸面部器官的特征及其几何形状进行的一种人脸识别方法，是人们最早研究及使用的识别方法，它主要是采用不同人脸的不同特征等信息进行匹配识别，这种算法具有较快的识别速度，同时，其占用的内存也比较小，但是，其识别率也并不算高。该方法主要做法是首先对人脸的嘴巴、鼻子、眼睛等人脸主要特征器官的位置和大小进行检测，然后利用这些器官的几何分布关系和比例来匹配，从而达到人脸识别。

基于几何特征识别的流程大体如下：首先对人脸面部的各个特征点及其位置进行检测，如鼻子、嘴巴和眼睛等位置，然后计算这些特征之间的距离，得到可以表达每个特征脸的矢量特征信息，例如眼睛的位置，眉毛的长度等，其次还计算每个特征与之相对应关系，与人脸数据库中已知人脸对应特征信息来做比较，最后得出最佳的匹配人脸。基于几何特征的方法符合人们对人脸特征的认识，另外，每幅人脸只存储一个特征，所以占用的空间比较小；同时，这种方法对光照引起的变化并不会降低其识别率，而且特征模板的匹配和识别率比较高。但是，基于几何特征的方法也存在着鲁棒性不好，一旦表情和姿态稍微变化，识别效果将大打折扣。

2.3.3 基于深度学习的方法

深度学习的出现使人脸识别技术取得了突破性进展。人脸识别的最新研究成果表明，深度学习得到的人脸特征表达具有手工特征表达所不具备的重要特性，例如它是中度稀疏的、对人脸身份和人脸属性有很强的选择性、对局部遮挡具有良好的鲁棒性。这些特性是通过大数据训练自然得到的，并未对模型加入显式约束或后期处理，这也是深度学习能成功应用在人脸识别中的主要原因。

深度学习在人脸识别上有7 个方面的典型应用：基于卷积神经网络(CNN)的人脸识别方法，深度非线性人脸形状提取方法，基于深度学习的人脸姿态鲁棒性建模，有约束环境中的全自动人脸识别，基于深度学习的视频监控下的人脸识别，基于深度学习的低分辨率人脸识别及其他基于深度学习的人脸相关信息的识别。

其中，卷积神经网络（Convolutional Neural Networks,CNN）是第一个真正成功训练多层网络结构的学习算法，基于卷积神经网络的人脸识别方法是一种深度的监督学习下的机器学习模型，能挖掘数据局部特征，提取全局训练特征和分类，其权值共享结构网络使之更类似于生物神经网络，在模式识别各个领域都得到成功应用。CNN 通过结合人脸图像空间的局部感知区域、共享权重、在空间或时间上的降采样来充分利用数据本身包含的局部性等特征，优化模型结构，保证一定的位移不变性。

利用CNN 模型，香港中文大学的Deep ID 项目以及Facebook 的Deep Face 项目在LFW数据库上的人脸识别正确率分别达97.45%和97.35%只比人类视觉识别97.5%的正确率略低。在取得突破性成果之后，香港中文大学的DeepID2 项目将识别率提高到了99.15%。Deep ID2通过学习非线性特征变换使类内变化达到最小，而同时使不同身份的人脸图像间的距离保持恒定，超过了目前所有领先的深度学习和非深度学习算法在LFW 数据库上的识别率以及人类在该数据库的识别率。深度学习已经成为计算机视觉中的研究热点，关于深度学习的新算法和新方向不断涌现，并且深度学习算法的性能逐渐在一些国际重大评测比赛中超过了浅层学习算法。

2.3.4 基于支持向量机的方法

将支持向量机（SVM）的方法应用到人脸识别中起源于统计学理论，它研究的方向是如何构造有效的学习机器，并用来解决模式的分类问题。其特点是将图像变换空间，在其他空间做分类。

支持向量机结构相对简单，而且可以达到全局最优等特点，所以，支持向量机在目前人脸识别领域取得了广泛的应用。但是，该方法也和神经网络的方法具有一样的不足，就是需要很大的存储空间，并且训练速度还比较慢。

2.3.5 其他综合方法

以上几种比较常用的人脸识别方法，我们不难看出，每一种识别方法都不能做到完美的识别率与更快的识别速度，都有着各自的优点和缺点，因此，现在许多研究人员则更喜欢使用多种识别方法综合起来应用，取各种识别方法的优势，综合运用，以达到更高的识别率和识别效果。

3 论文阅读：

ArcFace: Additive Angular Margin Loss for Deep Face Recognition

4 论文介绍

4.1 摘要

论文下载地址Paper

使用Deep Convolutional Neural Networks进行大规模人脸识别的特征学习的主要挑战之一是设计适当的损失函数来增强鉴别能力。Centre loss通过惩罚深层特征和它们对应类中心之间的欧氏距离，以实现类内紧凑性。SphereFace假设最后一个全连接层中的线性变换矩阵可以用作角空间中类中心的表示，并以乘法方式对深度特征及其相应权重之间的角度进行惩罚。最近，一个流行的研究方向是在已成熟的损失函数中加入margin，以最大化人脸类别的可分性。在本文中，我们提出了Additive Angular Margin Loss（ArcFace）来获得人脸识别的高分辨特征。由于与超球面上的测地距离精确对应，所提出的ArcFace具有清晰的几何解释。我们在10多个face recognition benchmarks上对所有SOTA人脸识别方法进行了最广泛的实验评估，包括一个新的具有万亿对级别的大规模图像数据库和一个大规模视频数据集。作者表明，ArcFace始终优于SOTA，并且在计算开销可以忽略不计的情况下轻松实现。

4.2 现有方法缺陷

使用Deep Convolutional Neural Network嵌入的人脸表示是人脸识别方案之一。典型地，在姿态标准化处理之后，DCNNs将人脸图像映射成具有小的类内距和大的类间距特征。训练用于人脸识别的DCNNs主要有两条研究路线。那些训练多分类的分类器可以分离训练集中的不同身份，例如通过使用softmax分类器，以及那些直接学习嵌入的分类器，如triplet loss。基于大规模训练数据和精心设计的DCNN结构，基于softmax loss和triplet loss的方法都可以在人脸识别上获得优异的性能。然而，softmax loss和triplet loss都有一些缺点。

对于softmax loss：

(1)线性变换矩阵的尺寸$W \in \mathbb{R}^{d \times n}$随$n$线性增加；

(2)对于闭集分类问题，学习的特征是可分离的，但对于开集人脸识别问题，学习的特征并没有足够的区分度。

对于triplet loss：

(1)face triplets的数量存在组合爆炸，特别是对于大规模数据集，这导致迭代步骤数量显著增加；

(2)semi-hard样本挖掘对于有效的模型训练是一个相当困难的问题。

4.3 提出的方法

最广泛使用的分类损失函数softmax损失如下：

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}$$

其中$x_i \in \mathbb{R}^d$表示属于第y_i个类别的第i个样本的深度特征，嵌入的特征维度d被设为512，$W_j \in \mathbb{R}^d$表示权重$W \in \mathbb{R}^{d\times n}$的第$j$列，$b_j \in \mathbb{R}^n$则是偏置项，$N$代表batchsize，$n$代表类别数。传统的softmax广泛应用于深度人脸识别。然而，softmax损失函数没有明确地优化特征嵌入，以加强类内样本的更高相似性和类间样本的多样性，这导致在大的类内外观变化(例如姿势变化和年龄差距)和大规模测试场景（例如百万对或万亿对）下深度人脸识别的性能差距。

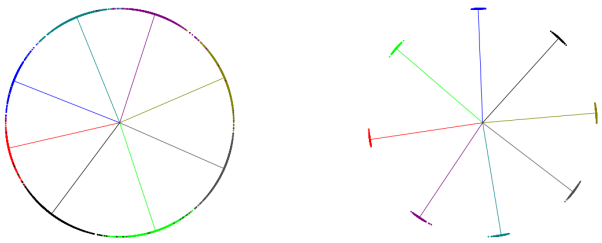
为简单起见，固定$b_j = 0$，使$W_{y_i}^T x_i = \|W_j\|\,\|x_i\|\cos\theta_j$，其中$\theta_j$是权重$W_j$与特征$x_i$。利用$l_2$正则化，固定$\|W_j\| = 1$，$\|x_i\| = s$。在特征以及权重上的正则化步骤使得预测仅依赖于特征和权重之间的角度。因此，所学习的嵌入特征分布在半径为s的超球面上。

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{yi}}}{e^{s\cos\theta_{yi}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_{yi}}}$$

由于嵌入特征分布在超球面上的每个特征中心周围，我们在W_{y_i}和x_i之间增加了一个additive angular margin惩罚m以同时增强类内紧密度和类间差异。由于提出的additive angular margin惩罚等于在标准化超球面中geodesic distance margin惩罚，因此将提出的方法命名为ArcFace。

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_{y_i}}}$$

我们从包含足够样本（约1500个图像/类）的8个不同身份中选择人脸图像，分别使用软softmax和ArcFace训练2D特征嵌入网络。如下午所示，softmax提供了粗糙的可分离的特征嵌入，在决策边界中则产生了明显的模糊性，而所提出的ArcFace显然可以在最相近的类之间形成更明显的差距。



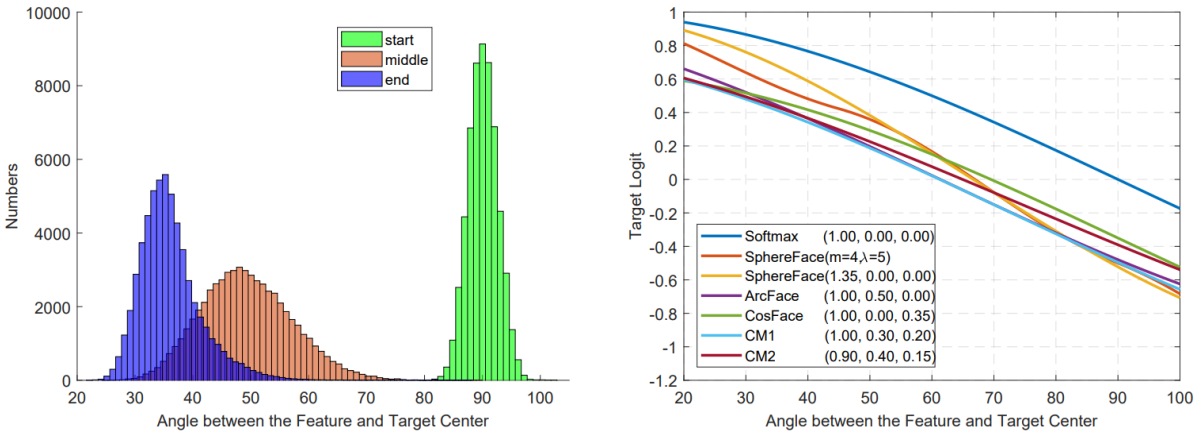
(a) Softmax (b) ArcFace

Toy examples under the softmax and ArcFace loss on 8 identities with 2D features. Dots indicate samples and lines refer to the centre direction of each identity. Based on the feature normalisation, all face features are pushed to the arc space with a fixed radius. The geodesic distance gap between closest classes becomes evident as the additive angular margin penalty is incorporated.

4.4 SphereFace与CosFace的比较

Numerical Similarity：SphereFace、ArcFace和CosFace中，提出了三种不同的裕度惩罚（margin penalty），例如multiplicative angular margin m_1、additive angular margin m_2和additive cosine margin m_3。从数值分析的角度来看，不同的裕度惩罚，无论是增加角度还是余弦空间，都通过惩罚目标logit来加强类内紧凑性和类间多样性。如下图所示，我们绘制了SphereFace、ArcFace和CosFace在最佳边距设置下的

目标逻辑曲线。我们只在[20°，100°]内显示这些目标逻辑曲线，因为在ArcFace训练期间，W_{y_i}与x_i之间的角度从大约90°（随机初始化）开始，并在大约30°结束。直觉上，目标logit曲线中有三个因素会影响性能，即起点、终点和斜率。



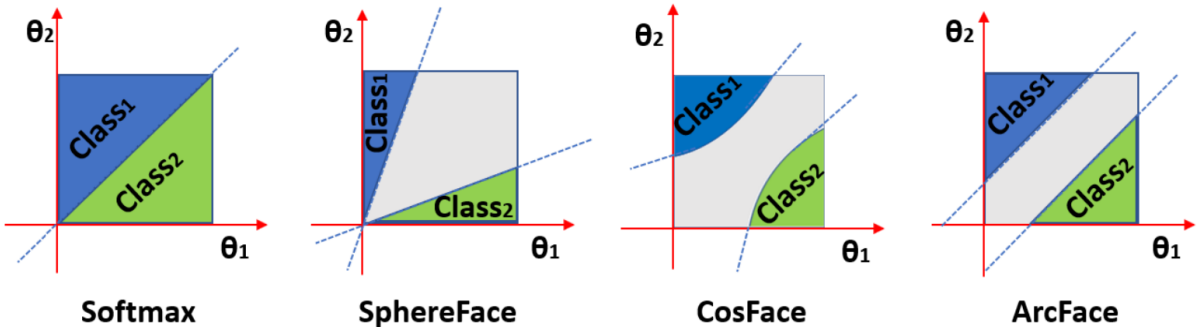
Target logit analysis. (a) distributions from start to end during ArcFace training. (2) Target logit curves for softmax, SphereFace, ArcFace, CosFace and combined margin penalty.

通过结合所有的margin penalties，我们在一个统一的框架中实现了SphereFace、ArcFace和CosFace，其中m_1、m_2和m_3是超参数。

$$L_4 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos\left(m_1\theta_{y_i}+m_2\right)-m_3)}}{e^{s(\cos\left(m_1\theta_{y_i}+m_2\right)-m_3)}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}}$$

如上图(b)所示，通过组合上述所有margins $(cos(m_1\theta + m_2) - m_3)$，我们可以很容易地得到一些其他j具有很高性能的目标logit曲线。

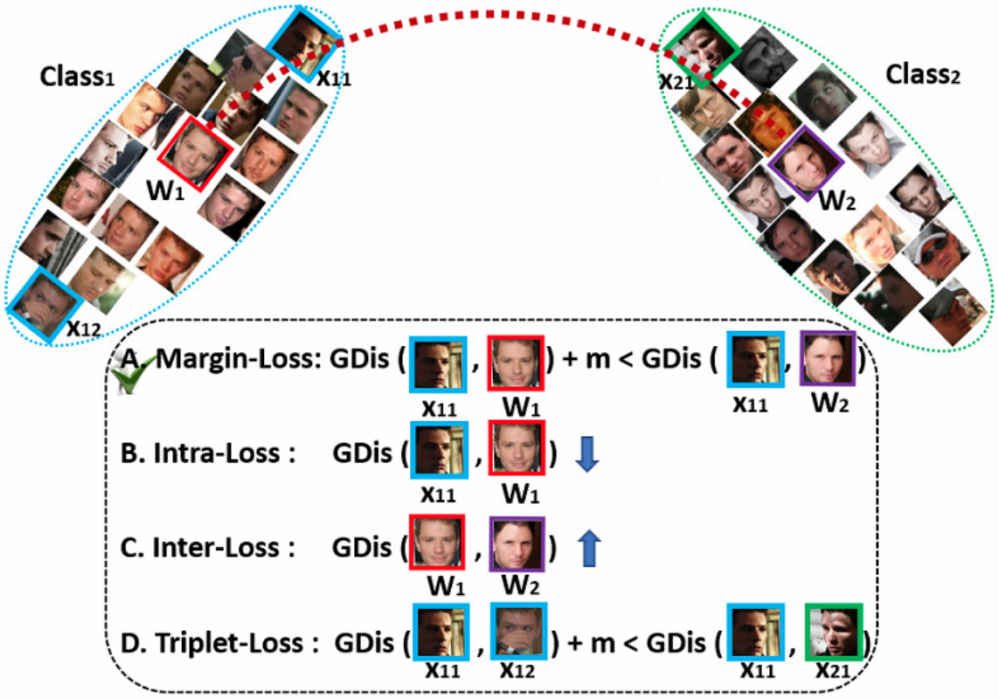
Geometric Difference： 尽管ArcFace和以前的工作在数值上有相似之处，但所提出的additive angular margin具有更好的几何属性，因为additive angular margin与测地距离有精确的对应关系。如下图所示，我们比较了二分类情况下的决策边界。所提出的ArcFace在整个区间内具有恒定的linear angular margin。相比之下，SphereFace和CosFace只有一个nonlinear angular margin。



Decision margins of different loss functions under bi- nary classification case. The dashed line represents the decision boundary, and the grey areas are the decision margins.

margin设计的微小差异会对模型训练产生“蝴蝶效应”。例如，最初的SphereFace采用了退火优化策略。为了避免训练开始时的发散，在SphereFace中使用softmax的联合监督来削弱multiplicative margin惩罚。通过应用反余弦函数，而不是使用复杂的双角度公式，我们实现了一个不需要在margin上使用整数的SphereFace。在我们的实现中，我们发现$m = 1.35$可以获得与原始SphereFace相似的性能且没有任何收敛困难。

其他损失函数可以基于特征和权重向量的角度表示来设计。例如，我们可以设计一个损失来加强超球面上的类内紧性和类间差异。如下图所示，我们比较了其它三种损失。



Based on the centre and feature normalisation, all identities are distributed on a hypersphere. To enhance intra-class compactness and inter-class discrepancy, we consider four kinds of Geodesic Distance (GDis) constraint. (A) Margin-Loss: insert a geodesic distance margin between the sample and centres. (B) Intra-Loss: decrease the geodesic distance between the sample and the corresponding centre. (C) Inter-Loss: increase the geodesic distance between different centres. (D) Triplet-Loss: insert a geodesic distance margin between triplet samples. In this paper, we propose an Additive Angular Margin Loss (ArcFace), which is exactly corresponded to the geodesic distance (Arc) margin penalty in (A), to enhance the discriminative power of face recognition model. Extensive experimental results show that the strategy of (A) is most effective.

Intra-Loss： 旨在通过减小样本和地面真实中心之间的角度/弧度来提高类内紧密度。

$L_5 = L_2 + \frac{1}{\pi N}\sum_{i=1}^{N}\theta_{y_i}$

Inter-Loss： 目标是通过增加不同中心之间的角度/弧度来增强类间差异。

$$L_6 = L_2 - \frac{1}{\pi N(n-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq y_i}^{n}\arccos\left(W_{y_i}^T W_j\right)$$

这里的Inter-Loss是Minimum Hyper-spherical Energy（MHE）方法的一个特例。在这个特例中，隐藏层和输出层都由MHE正则化。在MHE论文中，还提出了一个特殊的损失函数的例子，它在网络最后一层将SphereFace损失和的MHE损失结合起来。

Triplet-loss： 旨在扩大三个样本之间的角度/弧度余量。在FaceNet中，Euclidean margin被应用于归一化的特征。在这里，我们采用triplet-loss作为特征的角度表示$\arccos\left(x_i^{pos}x_i\right) + m \leqslant \arccos\left(x_i^{neg}x_i\right)$

4.5 部分数据集评估结果

Results on LFW, YTF, CALFW and CPLFW： LFW 和 YTF 数据集是在图像和视频上不受约束面部验证最广泛使用的基准。如下表所示，在 MS1MV2 上使用 Resnet100 培训的 ArcFace 在 LFW 和 YTF 以显著的 margin 击败了 baseline （ Sphereface 和 Cosface ），这表明 additive angular margin 惩罚可以显着提高深度学习特征的辨别力，这展示 ArcFace 的有效性。

Method	#Image	LFW	YTF
DeepID [32]	0.2M	99.47	93.20
Deep Face [33]	4.4M	97.35	91.4
VGG Face [24]	2.6M	98.95	97.30
FaceNet [29]	200M	99.63	95.10
Baidu [16]	1.3M	99.13	-
Center Loss [38]	0.7M	99.28	94.9
Range Loss [46]	5M	99.52	93.70
Marginal Loss [9]	3.8M	99.48	95.98
SphereFace [18]	0.5M	99.42	95.0
SphereFace+ [17]	0.5M	99.47	-
CosFace [37]	5M	99.73	97.6
MS1MV2, R100, ArcFace	5.8M	**99.83**	**98.02**

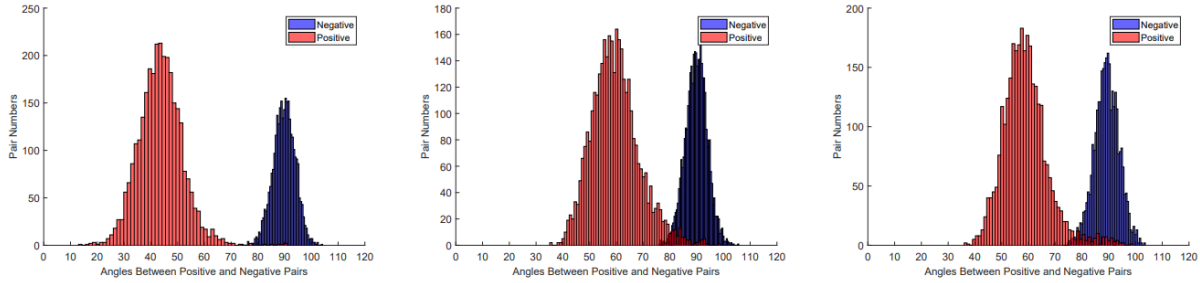
V erification performance (%) of different methods on LFW and YTF.

除了 LFW 和 YTF 数据集外，我们还报告了最近引入的数据集（例如 CPLFW 和 CALFW ）上 ArcFace 的性能，其显示了与 LFW 相同身份的更广泛的姿态和年龄变化。如下表所示，在所有开源面部识别模型中， ArcFace 模型以优于同行明显的·margin·被评估为顶级人脸识别模型。

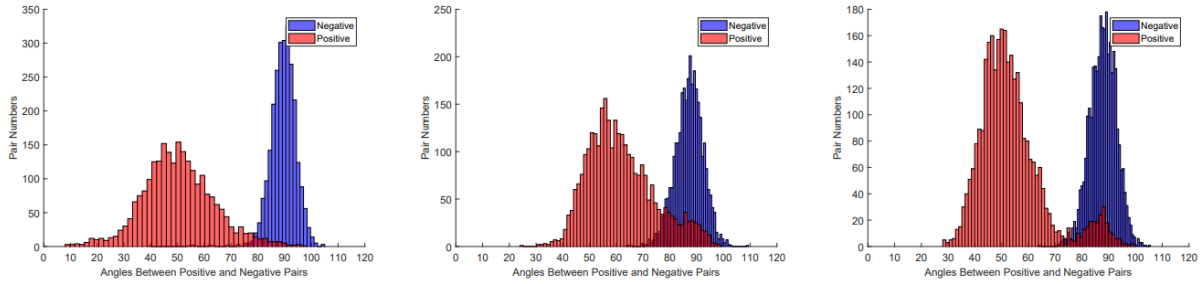
Method	LFW	CALFW	CPLFW
HUMAN-Individual	97.27	82.32	81.21
HUMAN-Fusion	99.85	86.50	85.24
Center Loss [38]	98.75	85.48	77.48
SphereFace [18]	99.27	90.30	81.40
VGGFace2 [6]	99.43	90.57	84.00
MS1MV2, R100, ArcFace	**99.82**	**95.45**	**92.08**

Verification performance (%) of open-sourced face recog- nition models on LFW, CALFW and CPLFW.

如下图所示，我们说明了在 LFW 、 CFP-F P、 AgeDB-30 、 YTF 、 CPLFW 和 CALFW 上正负对的角度分布（由在 MS1MV2 数据集上使用 ResNet100 训练的 ArcFace 模型进行预测）。我们可以清楚地发现，由于姿态和年龄间隔引起的帧内方差显着增加了正对之间的角度，从而使得面部验证的最佳阈值增加并且在直方图上产生更多的混乱区域。



(a) LFW (99.83%) (b) CFP-FP (98.37%) (c) AgeDB (98.15%)



(d) YTF (98.02%) (e) CPLFW (92.08%) (f) CALFW (95.45%)

Angle distributions of both positive and negative pairs on LFW, CFP-FP , AgeDB-30, YTF, CPLFW and CALFW. Red area indicates positive pairs while blue indicates negative pairs. All angles are represented in degree. ([MS1MV2, ResNet100, ArcFace])

Results on MegaFace.

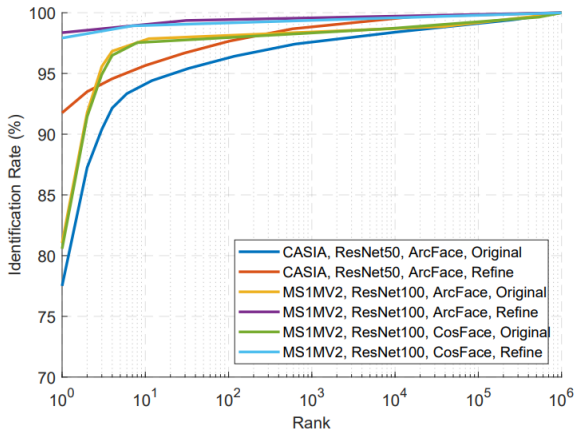
MegaFace DataSet包括1M张图像，其中包含690k独特个体作为gallery set，来自Facescrub的530独特个体的100k照片作为probe set。在Megaface上，在两个协议（大型或小型训练集）下有两个测试场景（识别和验证）。如果它包含超过0.5米的图像，则定义训练集。对于公平的比较，我们分别在小型协议和大协议下培训CAISA和MS1MV2的ArcFace。在表6中，Casia培训的ArcFace培训了最佳的单模识别和验证性能，不仅超越了强的基线（例如，Sphereface [18]和Cosface [37]），还优于其他公开的方法[38,17]。两个协议（大/小训练集）下有两个测试场景（识别 和 验证 ）。如果训练集包含超过0.5M图像则被定为 大数据集。为了公平的比较，我们分别在小协议和大协议下在CAISA和MS1MV2上训练ArcFace。在下表中，CAISA上训练的ArcFace实现了最佳的单模识别和验证性能，不仅超越了强大的基线（Sphereface和Cosface），还优于其他公开的方法。

Methods	Id (%)	Ver (%)
Softmax [18]	54.85	65.92
Contrastive Loss[18, 32]	65.21	78.86
Triplet [18, 29]	64.79	78.32
Center Loss[38]	65.49	80.14
SphereFace [18]	72.729	85.561
CosFace [37]	77.11	89.88
AM-Softmax [35]	72.47	84.44
SphereFace+ [17]	73.03	-
CASIA, R50, ArcFace	77.50	92.34
CASIA, R50, ArcFace, R	91.75	93.69
FaceNet [29]	70.49	86.47
CosFace [37]	82.72	96.65
MS1MV2, R100, ArcFace	81.03	96.98
MS1MV2, R100, CosFace	80.56	96.56
MS1MV2, R100, ArcFace, R	98.35	98.48
MS1MV2, R100, CosFace, R	97.91	97.91

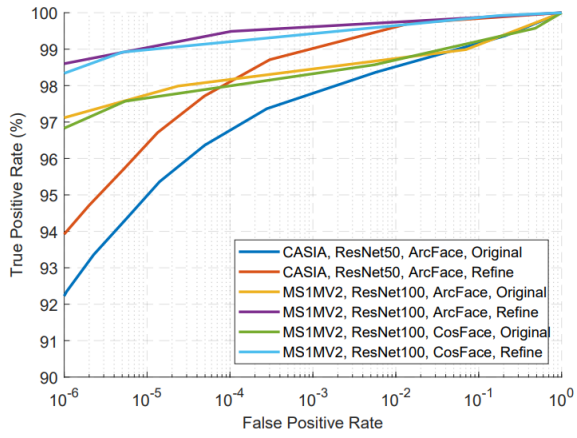
Face identification and verification evaluation of different methods on MegaFace Challenge1 using FaceScrub as the probe set. “Id” refers to the rank-1 face identification accuracy with 1M distractors, and “Ver” refers to the face verification TAR at 10−6 FAR. “R” refers to data refinement on both probe set and 1M distractors. ArcFace obtains state-of-the-art performance under both small and large protocols.

由于我们在识别和验证之间观察到明显的性能差距，我们在整个MegaFace数据集中进行了彻底的手动检查，并发现了许多具有错误标签的面部图像，这将明显影响测试性能。因此，我们手动改进了整个MegaFace数据集，并在MegaFace上报告了ArcFace的正确表现。在数据清洗后的MegaFace，ArcFace仍然显著优于CosFace并实现了验证和识别方面的最佳性能。

在大协议下，ArcFace通过明确的margin超越Faceget，与CosFace相比，在识别上获得了可比较的结果，在验证上获得了更好的结果。由于CosFace采用私人的训练数据，我们在MS1MV2数据集上将CosFace联合Resnet100重新训练。在公平比较下，ArcFace在CosFace上显示出优越性，并在识别和验证场景下形成压倒性优势，如下图所示。



(a) CMC

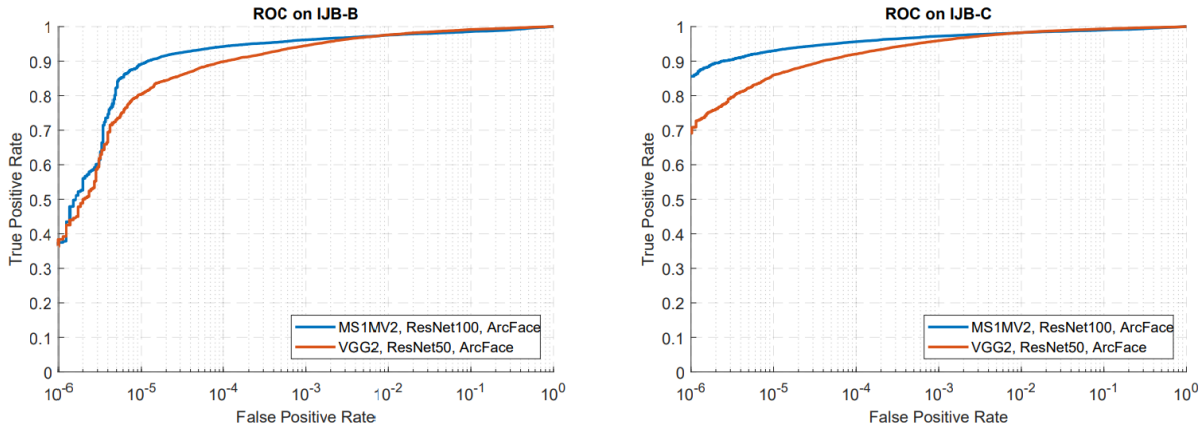


(b) ROC

CMC and ROC curves of different models on MegaFace. Results are evaluated on both original and refined MegaFace dataset.

Results on IJB-B and IJB-C: IJB-B数据集包含$1,845$个主题，共有$21.8K$静止图像和来自 $7,011$个视频的$55K$帧。总共有$12,115$个模板，具有$10,270$个真实匹配和$8M$的冒名匹配。IJB-C数据集是IJB-B的另一个延伸，具有$3,531$个受试者，具有$31.3k$静态图像和$117.5k$帧，来自$11,779$个视频。总共有$23,124$个模板，$19,557$个真实匹配和$15,639K$冒名匹配。

在IJB-B和IJB-C数据集上，我们使用VGG2数据集作为训练数据，Reset50作为嵌入式网络来训练ArcFace，以便与最近的方法进行公平比较。在下表中，我们将ArcFace的TAR（@$FAR = 1E-4$）与先前的最先进模型进行比较。ArcFace可以显然提高IJB-B和IJB-C（约3∽5% 的性能，这是错误的显着减少）。从更多训练数据（MS1MV2）和更深的神经网络（Resnet100）中，ArcFace可以在IJB-B和IJBC上进一步将TAR（@$FAR = 1E-4$）改善为94.2%和95.6%。在下图中，我们在IJB-B和IJB-C上显示了所提出的ArcFace完整ROC曲线，即使在FAR= 1E-6，ArcFace也可以实现令人印象深刻的性能并设置一个新的baseline。



(a) ROC for IJB-B (b) ROC for IJB-C

ROC curves of 1:1 verification protocol on the IJB-B and IJB-C dataset.

Results on Trillion-Pairs. Trillion-Pairs数据集提供了来自Flickr的$1.58M$图像作为gallery set以及来自于$5.7k$ LFW身份的$274K$图像作为probe set。gallery set和probe set之间的每一对都用于评估（总共0.4万亿对）。在下表中，我们比较了在不同数据集上训练的ArcFace性能。与CASIA相比，所提出的MS1MV2数据集明显提高了性能，甚至略优于具有双身份的DeepGlint-Face数据集。当结合MS1MV2和DeepGlint的亚洲名人的所有身份时，ArcFace实现了84.840（@$FPR = 1e-3$）的最佳识别性能，并且其验证性能能够与lead-board最新提交(CIGIT IRSEC)不相上下。

Method	Id (@FPR=1e-3)	Ver(@FPR=1e-9)
CASIA	26.643	21.452
MS1MV2	80.968	78.600
DeepGlint-Face	80.331	78.586
MS1MV2+Asian	**84.840** (1st)	80.540
CIGIT_IRSEC	84.234 (2nd)	**81.558** (1st)

Identification and verification results (%) on the Trillion-Pairs dataset. ([Dataset*, ResNet100, ArcFace])

Results on iQIYI-VID.:
iQIYI-VID挑战赛包含来自爱奇艺综艺节目、电影和电视剧的4934个身份的565,372个视频剪辑（训练集219,677、验证集172,860和测试集172,835）。每个视频的长度从1到30秒不等。该数据集提供了多模态线索，包括人脸、布料、声音、步态和字幕，用于字符识别。iQIYI-VID数据集采用MAP@100作为评价指标。MAP（Mean Average Precision）指的是总体平均准确率，是测试集中检索到的人物ID对应视频对训练集中每个人物ID（作为查询）的平均准确率的均值。

如下表所示，在MS1MV2和Asian数据集上使用ResNet100训练的ArcFace设置了一个高baseline（$MAP = (79.80)$）。基于每个训练视频的嵌入特征，我们训练了一个附加的三层全连通网络，该网络带有一个分类损失，以获得iQIYI-VID数据集上的自定义特征描述符。MLP在iQIYI-VID训练集上的学习显著提高了6.60的平均成绩。借助模型集成的支持和现成的对象和场景分类器的上下文特征，我们的最终结果明显优于亚军（0.99%））。

Method	MAP(%)
MS1MV2+Asian, R100, ArcFace	79.80
+ MLP	86.40
+ Ensemble	88.26
+ Context	**88.65 (1st)**
Other Participant	87.66 (2nd)

MAP of our method on the iQIYI-VID test set. “MLP” refers to a three-layer fully connected network trained on the iQIYI-VID training data.

4.6 Conclusions

在本文中，我们提出了一个Additive Angular Margin损失函数，对于人脸识别，可以有效增强通过DCNN学习的特征嵌入的判别能力， 在文献报道的最全面的实验中，证明了我们的方法始终优于最先进的方法。

5 回答题目中的问题

5.1 人脸识别模型：

ARCFACE

5.2 实现方式：

见以上分析，在经过将fc层权重和输出feature进行规范化处理后，两者的点积就可以看做是深度卷积网络输出的人脸feature。使用了arc-consine 函数来计算输出feature与目标权重的角度。然后作者在目标角度上增加了附加角边距(additive angular margin)。最后作者通过固定的特征规范化将所有的logits进行重新缩放，剩下的步骤就与基于标准softmax loss的步骤一致。由于其是当前最为先进的人脸识别算法，但依然存在训练时模型收敛速度不够快的问题。

5.3 问题二：二者的异同

相同：

5.3.1 人工神经网络结构是基于生物学观察

人工神经网络中最小也是最重要的单元叫神经元。与生物神经系统类似，这些神经元也互相连接并具有强大的处理能力。一般而言，ANNs试图复现真实大脑的行为和过程，这也是为什么他们的结构是基于生物学观察而建模的。

每个神经元都有输入连接和输出连接。这些连接模拟了大脑中突触的行为。与大脑中突触传递信号的方式相同——信号从一个神经元传递到另一个神经元，这些连接也在人造神经元之间传递信息。每一个连接都有权重，这意味着发送到每个连接的值要乘以这个因子。再次强调，这种模式是从大脑突触得到的启发，权重实际上模拟了生物神经元之间传递的神经递质的数量。所以，如果某个连接重要，那么它将具有比那些不重要的连接更大的权重值。

由于可能有许多值进入一个神经元，每个神经元便有一个所谓的输入函数。通常，连接的输入值都会被加权求和。然后该值被传递给激活函数，激活函数的作用是计算出是否将一些信号发送到该神经元的输出。

不同：

5.3.2 人的视觉系统与机器的差别

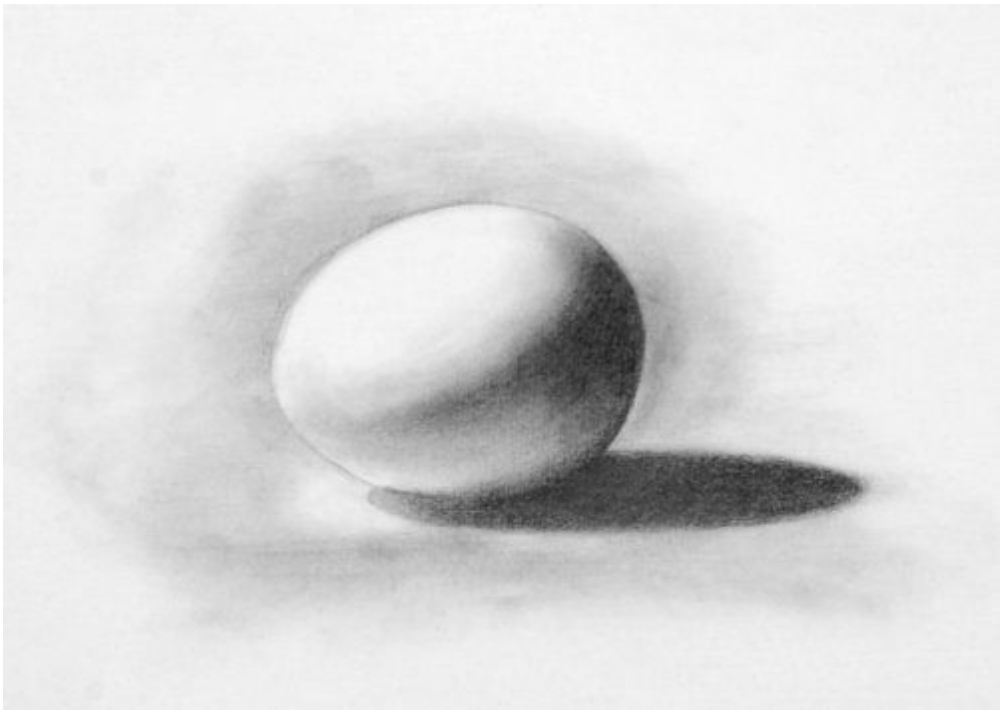
人的眼睛与摄像头有着本质的差异。眼睛的视网膜中央非常小的一块区域叫做“fovea”，里面有密度非常高的感光细胞，而其它部分感光细胞少很多，是模糊的。可是眼睛是会转动的，它被脑神经控制，敏捷地跟踪着感兴趣的部分：线条，平面，立体结构…… 人的视觉系统能够精确地理解物体的形状，理解拓扑，而且这些都是 3D 的。人脑看到的不是像素，而是一个 3D 拓扑模型。

眼睛观察的顺序，不是一行一行从上往下把每个“像素”都记下来，做成 6000x4000 像素的图片，而是聚焦在重点上。它可以沿着直线，也可以沿着弧线观察，可以转着圈，也可以跳来跳去的。人脑通过自己的理解能力，控制着眼睛的运动，让它去观察所需要的重点。由于视网膜中央分辨率极高，所以人脑可以得到精度非常高的信息。然而由于不是每个地方都看的那么仔细，所以眼睛采集的信息量可能不大，人脑需要处理的信息也不会很多。

人的视觉系统能理解点，线，面的概念，理解物体的表面是连续的还是有洞，是凹陷的还是凸起的，分得清里和外，远和近，上下左右…… 他能理解物体的表面是什么质地，如果用手去拿会有什么样的反应。他能想象出物体的背面大概是什么样子，他能在头脑中旋转或者扭曲物体的模型。如果物体中间有缺损，他甚至能猜出那位置之前什么样子。

人的视觉系统比摄像头有趣的多。很多人都看过“光学幻觉”（optical illusion）的图片，它们从一个角度揭示了人的视觉系统背后在做什么。比如下图本来是一个静态的图片，可是你会感觉有很多暗点在白线的交叉处，但如果你仔细看某一个交叉处，暗点却又不见了。这个幻觉很经典，被叫做 Herman grid，在神经科学界被广泛研究。

5.3.3 人脑会构造事物的 3D 模型



靠着光和影的组合，人和动物能得到很多信息。比如上图，我们不但看得出这是一个立体的鸡蛋，而且能推断出鸡蛋下面是一个平面，可能是一张桌子，因为有阴影投在了上面。

但神经网络根本不知道影子是什么。早就有人发现，Tesla 基于图像识别的 Autopilot 系统会被阴影所迷惑，以为路面上的树影是一个障碍物，试图避开它，却差点撞上迎面来的车。

现在很多自动驾驶车用激光雷达构造 3D 模型，可是相对于人类视觉形成的模型，真是太粗糙了。激光雷达靠主动发射激光，产生一个扫描后的“点云”，分辨率很低，只能形成一个粗糙的 3D 轮廓，无法识别物体，也无法理解它的结构。我们应该好好思考一下，为什么人仅靠被动接收光线就能构造出如此精密的 3D 模型，理解物体的结构，而且能精确地控制自己的动作来操作这些物体。

现在的深度学习模型都是基于像素的，没有抽象能力，不能构造 3D 拓扑模型，甚至连位置关系都分不清楚。缺乏人类视觉系统的这种“结构理解”能力，可能就是为什么深度学习模型需要那么多的数据，那么多的计算，才勉强能得出物体的名字。而小孩子识别物体根本不需要那么多数据和计算，看一两次就知道这东西是什么了。

人脑提取了物体的要素，所以很多信息都可以忽略了，所以人需要处理的数据量，可能比深度学习模型小很多。深度学习领域盲目地强调提高算力，制造出越来越大规模的计算芯片，GPU，TPU…… 可是大家想过人脑到底有多大计算能力吗？它可能并不需要很多计算。

5.3.4 大脑的神经元比计算机的集成电路慢得多。

大脑的力量来自于它是执行大规模并行处理的机器。大脑没有CPU。相反，它具有数百万个同时合并信号的神经元。在任何给定时间，大脑的许多大型专业区域并行运行以执行各种任务，例如处理视觉或听觉信息或计划动作。即使在这些区域中的每个区域中，信息也会通过没有重要序列结构的神经网络流动。

5.3.5 确定性与非确定性

从给定输入的意义上说，计算机是确定性机器，它们将始终产生相同的输出。

但这并不意味着该输出总是可预测的。例如，计算机可以通过引入伪随机变量来模拟非确定性系统。计算机还可以应用来自混沌物理学的方程，其中确定性过程的结果可能会受到初始条件中微小变化的极大影响。

整个大脑被认为是非确定性系统，原因很简单：一个时刻到下一个时刻永远不会完全相同。

它不断地形成新的突触，并根据其用法来增强或削弱现有的突触。因此，给定的输入将永远不会产生完全相同的输出两次。但是，脑活动的生理化学过程被认为是确定性的。