東南大學

毕业设计(论文)报告

题 目： 基于深度学习的遮挡人脸识别算法研究

学　　号：	04019509
姓　　名：	张宸婷
学　　院：	信息科学与工程学院
专　　业：	信息工程
指导教师：	罗琳
起止日期：	2023.01.01-2023.05.30

东南大学毕业（设计）论文独创性声明

本人声明所呈交的毕业（设计）论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者签名：张宸婷 日期：2023 年 5 月 28 日

东南大学毕业（设计）论文使用授权声明

东南大学有权保留本人所送交毕业（设计）论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权东南大学教务处办理。

论文作者签名：张宸婷 导师签名：[illegible]

日期：2023 年 5 月 28 日 日期：2023 年 5 月 28 日

摘　　要

随着科技的进步，人脸识别算法也不断进行着迭代更新，截止目前，一般情况下的人脸识别技术已经趋于成熟，遮挡人脸识别作为人脸识别的分支，旨在口罩、眼镜、光照等人脸关键特征信息被破坏的情境下仍保持较高的识别率。本文主要工作包括：

首先使用 MobileNet 网络架构实现人脸 128 维特征向量的提取，MobileNet 是卷积神经网络（CNN）架构的分支，MobileNet 是在传统卷积神经网络上的提升改进，其使用可分离卷积块，构建轻量型网络。除此之外，MobileNet 还跳过了某些层之间的连接以改善网络中的信息流，以及包含了网络末端的全局池化层，以产生最终输出特征。

在模型的训练阶段，本文使用了用于度量学习的三元组损失作为损失函数，其目标是最小化同类特征数据之间的距离，同时最大化异类特征数据之间的距离。其训练神经网络模型参数学习率的更新方法为余弦退火学习率调度，模型参数的更新算法为小批量随机梯度下降优化算法，其训练数据集来自 WebFace 经典人脸数据集。

在模型的验证阶段，使用在常规人脸识别数据集 LFW（Labelled Faces in the Wild）的基础上加入随机生成口罩遮挡生成 OCC-LFW（Occluded Labelled Faces in the Wild）数据集测试人脸识别算法的性能。该模型在遮挡验证数据集上的准确率为 97.867%，表明该模型在处理遮挡人脸的情况下表现良好。

最后，加入了人脸识别网页可视化模块，使得用户上传照片加入人脸数据库，再用摄像头实时捕捉人脸，并在画面中显示用户身份信息。并以多位同学的各种遮挡情况下的人脸照片进行了测试，测试结果表明，该遮挡人脸识别模型能用于现实生活人脸遮挡情景。

关键词：遮挡人脸识别，机器学习，卷积神经网络，三元组损失函数

ABSTRACT

With the advancement of technology, face recognition algorithms are constantly being iteratively updated. As of now, general face recognition technology has become mature. Occluded face recognition, as a branch of face recognition, aims to maintain a high recognition rate even when key facial features such as masks, glasses, and lighting are disrupted.

This paper presents the implementation of a MobileNet network architecture for the extraction of 128-dimensional facial feature vectors. MobileNet is a branch of the convolutional neural network (CNN) architecture. Compared to traditional CNN architectures, MobileNet aims to reduce model complexity by using separable convolution blocks, skips connections between certain layers to improve information flow in the network, and includes a global pooling layer at the end of the network to produce the final output features.

During the training phase, this study employed the triplet loss as the loss function for metric learning, aiming to minimize the distance between samples of the same class and maximize the distance between samples of different classes. The learning rate for training the neural network model was updated using the cosine annealing learning rate schedule, and the optimization algorithm used for updating model parameters was mini-batch stochastic gradient descent. The training dataset was sourced from the WebFace dataset, a well-known face dataset.

During the validation phase, the performance of a face recognition algorithm was tested using the OCC-LFW (Occluded Labelled Faces in the Wild) dataset, which was created by adding randomly generated mask occlusions to the conventional face recognition dataset, LFW (Labelled Faces in the Wild). The accuracy achieved by the model on the occlusion validation dataset was found to be 97.867%, indicating excellent performance in handling occluded faces.

Finally, a web-based visualization module for face recognition was developed, allowing users to upload their photos to the face database. The system can then capture real-time face images through a camera and display the corresponding user identity information on the screen. The model was tested using face images of multiple individuals under various occlusion scenarios, and the results showed that the proposed occluded face recognition model is applicable to real-life situations.

In summary, this study presents a scientific approach for occluded face recognition. It utilizes the triplet loss for training, cosine annealing learning rate schedule for learning rate updates, and

mini-batch stochastic gradient descent for optimization. The model demonstrates promising performance on benchmark datasets and includes a web-based visualization module for real-time face recognition.

KEY WORDS: Occluded Facial Recognition, Machine Learning, Convolutional Neural Network, Triplet Loss

目　　录

第一章 绪论

1.1 研究背景

遮挡人脸识别是人脸识别领域的一个重要研究方向，它的研究背景可以从以下几个方面进行探讨。首先，人脸识别技术在过去几十年取得了显著的进展，成为了计算机视觉和人工智能领域的热门研究方向之一。人脸识别技术广泛应用于安全、监控、人机交互、身份认证等领域，为社会生活带来了许多便利和安全性。然而，在实际应用中，人脸识别技术仍然面临许多挑战。其中之一就是遮挡问题。在日常生活中，人们经常面对光照条件的变化、佩戴口罩、戴眼镜、帽子、围巾等遮挡物的存在，这些遮挡因素会严重影响人脸识别的准确性和鲁棒性。特别是在当前全球范围内新冠疫情的影响下，人们在公共场所普遍要求佩戴口罩，进一步加剧了遮挡问题对人脸识别的挑战。

遮挡问题对人脸识别的影响主要体现在两个方面。首先，遮挡会导致人脸关键特征信息的丢失或不完整，使得传统的人脸识别算法无法准确提取和匹配特征，从而导致识别性能下降。其次，遮挡会引入干扰信息，使得人脸图像的质量下降，可能导致误识别或错误拒识。针对遮挡问题，传统的人脸识别方法通常采用人工设计的特征提取方法，如局部纹理特征、形状信息等。然而，这些方法在遮挡情况下往往表现不佳，难以捕捉到遮挡区域之外的有效信息。

随着深度学习的快速发展，特别是卷积神经网络（Convolutional Neural Network，CNN）的广泛应用，深度学习方法成为解决遮挡人脸识别问题的新方向。深度学习模型具有强大的学习能力和对深层特征的有效提取能力，可以自动学习具有鲁棒性的特征表示，从而提高遮挡人脸识别的准确性和鲁棒性。近年来，研究者们提出了许多基于深度学习的遮挡人脸识别算法。这些算法主要关注于如何在遮挡环境下提取鲁棒的人脸特征表示，包括利用多尺度特征融合、引入注意力机制、采用生成对抗网络等方法。这些方法在不同程度上缓解了遮挡对人脸识别的影响，提高了遮挡人脸识别的准确性和鲁棒性。尽管深度学习方法在遮挡人脸识别中取得了一定的进展，但仍然存在一些挑战。例如，遮挡的种类和程度多样，如何对各种类型的遮挡进行建模和处理仍然是一个难题。此外，深度学习模型在遭遇遮挡时仍然容易受到干扰，对于复杂遮挡情况下的人脸识别仍然存在困难。

因此，进一步的研究仍然需要关注遮挡人脸识别算法的鲁棒性、泛化能力和实时性等问题。通过研究遮挡人脸识别的研究背景和挑战，可以为后续的研究提供指导，并促进遮挡人脸识别技术的进一步发展和应用。

1.2 国内外研究现状

传统人脸识别算法发展到现在已较为成熟，其中最先进的为基于弹性边际损失的深度人脸识别模型（ElasticFace）[1]。在传统的人脸识别模型中，通常在常用的分类损失函数（Softmax Loss）上加入一个固定的惩罚幅度，通过最小化类内聚合和最大化类间差异来提高人脸识别模型的判别能力。引入边际惩罚的 Softmax 损失，比如 ArcFace 和 CosFace，假设不同身份之间和之内使用固定的惩罚幅度平等地学习[1]，ElasticFace 算法在前人的基础上，放宽了固定的惩罚幅度约束，提出弹性惩罚边际损失（ElasticFace）的概念，灵活地推动类的可分离性[1]。其在每次训练迭代中利用从正态分布中抽取的随机边际值，给决策边界变化的空间，以达到灵活分类学习的目的[1]。其人脸特征提取部分利用了以 ElasticArcFace 损失训练的 ResNet-100 模型，数据集选取了传统人脸数据集 MS1MV2，其模型相较于以 ArcFace 和 CosFace 损失训练，类内聚合度以及类间差异度性能更好[1]。该模型其在传统无遮挡人脸数据集中表现优异，但由于其未考虑遮挡条件，在遮挡人脸数据集中的性能损失较快。

针对遮挡场景，文献[2],[3]将现有的遮挡人脸识别算法进行了归类。文献[2]将各类算法以减少未被遮挡区域的特征所受影响、修复被遮挡区域的固有特征以及基于特征融合三种处理遮挡的不同思路比较了各类算法并分析了其性能；其中，文献[2]又根据遮挡方式的不同将修复类算法细分为随机遮挡修复和规则遮挡修复，又根据算法中预测生成网络的不同，将其进一步分为基于卷积神经网络（CNN）和基于生成对抗网络（GAN）。

具体而言，文献[4]提出了一种基于多模态的两阶段人脸修复算法，该算法旨在利用原始图片、人脸属性信息三方面模态进行融合，并将融合向量输入到两阶段人脸修复网络，分别生成粗略的修复结果和精细的最终修复结果。此外，文献[4],[5]均在其网络结构中嵌入注意力机制网络，使其算法能将真实的未遮挡区域特征赋予较高的权重，实现对面部的深度特征提取，以提高训练效果。文献[6]对现有的 Inception 网络结构进行改进，结合空间特征和通道特征提出一种融合特征的局部遮挡人脸识别算法。文献[7]提出了局部线性嵌入式卷积神经网络 LLE-CNN，该模型包含三个模块，第一个模块 Proposal Module 利用 CNN 网络提取局部面部特征块并用含噪声的描述符刻画；第二个提取特征模块通过由人脸字典和非人脸字典组成的特征子空间训练出的近邻向量细化描述符；最后验证模块进行人脸区域验证。在图像特征提取的子领域，矩阵因子分解作为数据降维的方法之一，在海量数据处理中具有重要意义。文献[8]比较了不同的矩阵因子分解方法，并得出岭正则化的复杂矩阵

分解（SCMF-L2 范数）在遮挡人脸识别上具有最快的收敛率和最高的准确度。

文献[9]从万物互联时代衍生的安全监控问题的角度，针对 ATM 智能监控系统的目标检测和识别问题，提出了一种新颖的遮挡人脸识别框架。该框架包含三个步骤：人脸区域检测，人脸追踪以及人脸验证。该论文首先提出了一种基于能量函数的算法用于人脸区域检测，再采用了一种利用 CNN 从几何对称的人脸图像中提取深度特征的有效字典学习算法，最后使用基于深度学习的稀疏分类模型用于检测人脸是否被遮挡，该框架在精度和速度上均取得了不错的成果。

在数据集创建方面，相较于已被成熟应用的传统人脸数据集，遮挡人脸数据集由于其稀缺性和情景的复杂性，也广泛引起了研究者的关注。文献[5],[10]均提出了利用卷积神经网络在人脸图像上叠加人工合成口罩的方法，文献[5]使用 Dlib 算法为人脸特征划定遮挡空间，辅以仿射变换方法，将遮挡掩模块映射至人脸预定位置，生成遮挡人脸数据集 Maskface；文献[10]利用深度卷积神经网络 FaceNet 来提取面部嵌入，并用支持向量机进行面部分类。其开发的遮挡人脸数据集被用于后续研究中；文献[11]提出的 Occ-LFW 数据集是一种合成遮挡人脸数据集的方法，这些图像是通过在随机的人脸位置上添加一个随机的遮挡物图像来合成的；文献[12]提出的 AR 数据集包含了来自 126 个个体的 4000 张人脸图像，其遮挡来源为口罩或围巾两种情况，其相较于人工合成的遮挡物更真实自然。

由于近年疫情的影响，口罩的普及使得遮挡人脸识别领域面临新的挑战，文献[13]将研究重点放在口罩人脸识别（MFR）上，该论文比较了最先进的口罩识别模型 FocusFace、ElasticFace-Arc-Aug、EUM，遮挡人脸识别模型 FROM 和传统通用人脸识别模型 ElasticFace 在经典的数据库上的性能表现，并论证了针对口罩人脸识别开发的解决方案对一般遮挡类型的数据集的泛用性。

其中，ElasticFace-Arc-Aug 在多个数据集上都具有优良的表现，其原理是在最先进的 ElasticFace 模型基础上加入模板级的知识提馏（KD）的方法[14]，旨在产生与无遮挡的人脸相似的遮挡人脸嵌入。该模型包含一个预先训练好的教师网络，学生网络用遮挡和无遮挡的人脸图像训练，在训练过程中，ElasticFace-Arc-Aug 的损失函数包含两部分：教师网络和学生网络产生的嵌入的均方差（L_{KD}），以及原始 ElasticArc 模型的损失函数（$L_{Elastic}$），其总损失函数又作为反馈影响学生网络的学习过程，最终达到口罩一致性（Mask-invariant）的效果。

总的来说，目前已有的遮挡人脸识别模型在数据库建立、预测生成网络模型的选择和修复算法的设计等方面取得了一定的成果。然而，在未来的研究中仍存在一些关键问题需

要解决。首先，由于遮挡类型的多样性，构建复杂的遮挡人脸检测数据集对于进一步研究至关重要。这样的数据集可以涵盖各种实际情境中可能遇到的遮挡情况，为算法的训练和评估提供更真实和全面的基础。此外，还需要进一步建立完善的中国人脸数据集，以更好地适应国内的人脸识别需求。其次，设计轻量级网络架构是未来发展高效训练算法的关键。随着遮挡人脸识别算法的复杂性增加，为了节省计算资源并提高实时性能，研究者需要对现有模型进行优化和重组，探索轻量级网络结构的设计，以提高算法的效率和性能。最后，除了业界广泛使用的算法评价指标，引入新的专门针对人脸修复算法的评价指标是未来研究的趋势之一。传统评价指标难以完全准确地衡量修复算法的效果，因此需要开发更具针对性和可靠性的评价指标，以更好地评估算法的修复能力和性能。

综上所述，未来遮挡人脸识别领域的研究方向包括构建复杂遮挡人脸检测数据集、设计轻量级网络架构以提高效率，以及引入新的专门评价指标。这些努力将推动遮挡人脸识别技术的发展，使其能够更好地适应复杂的现实应用场景，并为实际应用带来更高的准确性和可靠性。

1.3 研究内容概述

本文采取的 MobileNet 网络架构已成为近年来越来越流行的人脸识别方法。它是卷积神经网络的一个分支，相较于传统的卷积神经网络架构，MobileNet 是卷积神经网络（CNN）架构的分支，也是在传统卷积神经网络上的提升改进，其使用可分离卷积块，构建轻量型网络。其具有训练速度更快，更易于进行参数调控和实验的特点。训练数据集为用于人脸识别和验证的公共数据集——WebFaces 数据集。该数据集包含了超过 10000 个人的从互联网上收集而来的人脸图像，这些图像包含了多种种族和年龄段的人，且人脸姿势和光照条件各异。

本文采用了用于度量学习的三元组损失作为损失函数进行参数更新和模型训练，使同一身份的人脸图像在特征空间中的距离更近，而不同身份的人脸图像在该空间中的距离更远，通过优化该损失函数，我们可以学习到一个高效的特征提取器，用于将人脸图像映射到高维特征空间，并计算不同人之间的相似性。该模型用 MaskTheFace 方法生成的遮挡 LFW 验证数据集进行验证，其验证准确率为 97.867%，说明模型表现良好。

此外，本文还设计了一个网页化的人脸识别界面，通过摄像头实时捕捉人脸，在对人脸区域进行动态监测后，计算人脸与已录入数据库里的身份信息的相似度进行人脸验证，最后我们选取了 3 位同学在口罩、围巾、手部分遮挡等情况下的人脸进行了实验，以及区

域内多张人脸情况下也进行了实验，其结果表明模型性能表现好，在一定程度上具有稳定性。

1.4 本文组织结构

第一章为绪论。介绍了遮挡人脸识别的研究背景与发展前景。总结了近年来国内外研究成果和未来研究趋势，简要概括了论文的研究内容和论文结构。

第二章为模型建立。首先介绍了该实验使用的 MobileNet 的网络架构，详细介绍了 MobileNet 的核心层——深度可分离卷积块以及整个网络架构和参数训练。之后介绍了整个遮挡人脸检测模型的技术金字塔并对每个单元分别进行了介绍，着重强调了三元组损失函数。

第三章为模型验证及实验。用 LFW 数据集对实验使用的遮挡人脸识别模型进行了验证。首先介绍了 LFW 数据集以及 MaskTheFace 方法生成的口罩遮挡 LFW 数据集版本，进行验证后计算出了验证准确率和错误接受率，绘制出了 ROC 以及 AUC 曲线。并给定两张特定的人脸图片，通过上述深度学习网络分析出了两张人脸图片的特征以及计算出特征之间的欧式距离。之后介绍了可视化模块的设计，以及通过现实生活中采样的人脸进行了验证。最后分析了 MobileNet 模型的优缺点以及可以改进的方向。

第四章为总结与展望。对全文的研究内容进行总结并且对未来 MobileNet 轻量型模型在其他领域的发展和应用进行了讨论与展望。

第二章 模型建立

2.1 平台介绍

在这个实验中，选择了 PyCharm 作为集成开发环境（IDE）来进行代码编写和调试。PyCharm 提供了丰富的功能和工具，使开发者能够更高效地进行深度神经网络的开发和调试工作。对于深度神经网络框架，使用了 Tensorflow 1.12.2 版本作为主要的开发平台。Tensorflow[20]是一个广泛使用的深度学习框架，提供了丰富的 API 和工具来构建、训练和部署神经网络模型。此外，还使用了 Keras 库的 2.2.4 版本，Keras 是一个高级神经网络 API，可以与 Tensorflow 等后端框架结合使用[21]，Keras 简化了深度学习模型的构建过程，提供了简洁而灵活的接口。另外，为了进行图像处理和分析任务，使用了 OpenCV 库的 3.4.3.18 版本，OpenCV 是一个强大的计算机视觉库[22]，提供了各种图像处理和计算机视觉算法，方便进行图像预处理、特征提取和结果可视化等操作。

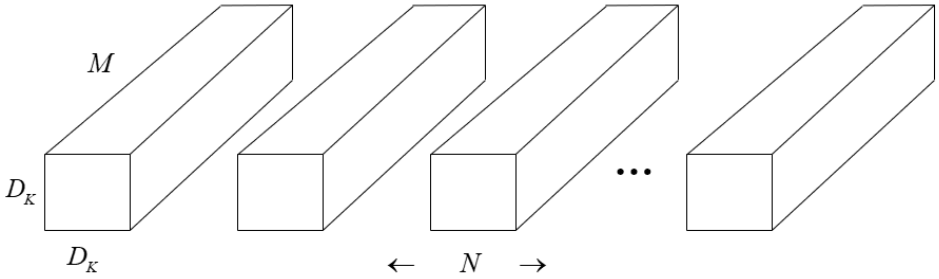
由于 MobileNet 是一种轻量级的网络模型，训练过程没有使用 GPU，而是利用默认的 CPU 进行训练。尽管在 CPU 上训练的速度相对较慢，一次训练大约需要 5 小时左右，但这种设置可以在资源受限的环境中进行训练，并在移动设备上实现高效的实时推理。通过以上的环境配置和选择，该实验能够在 PyCharm 中使用 Tensorflow 和 Keras 开发深度神经网络模型，并利用 OpenCV 进行图像处理和分析。同时，使用默认的 CPU 进行 MobileNet 的训练，确保在轻量化的网络设计下实现了高效的模型训练和推理。

2.2 MobileNet 网络架构

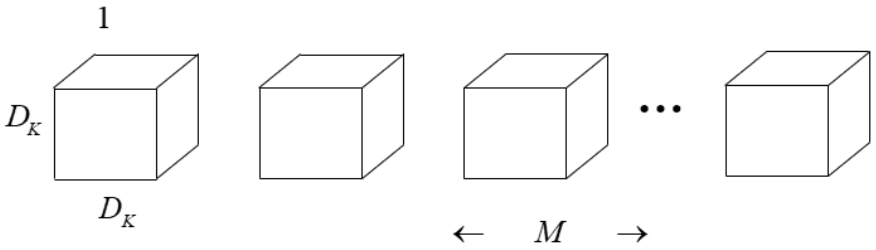
在深度学习网络的选取上，深度卷积神经网络（CNN）被广泛用于人脸识别等各领域，在此，提出一种轻量化模型 MobileNet，这个网络架构是基于卷积神经网络（CNN）的一种高效设计。它被称为“轻量”，意味着它在计算资源上的需求比较低，不会给设备带来太大的负担。同时，它还具有低延迟的特性，意味着它可以快速地进行图像处理并给出结果。这个网络的训练速度也很快，这意味着可以更快地训练出一个准确的模型。这个网络架构非常适合应用在移动设备和嵌入式系统中，因为它满足了这些应用的设计要求，既省资源又能高效地完成视觉任务。下面对 MobileNet 的架构进行介绍描述。

首先介绍 MobileNet 的核心层——深度可分离卷积块。MobileNet 模型使用了一种特殊的卷积技术，称为深度可分离卷积。这种卷积操作将大的过滤器分解成两个较小的步骤。首先，深度卷积会分别处理输入图像的每个通道，以提取不同通道的特征。接着，逐点卷积会将这些通道的特征进行组合，形成最终的输出。图 2-1 展示了一个标准卷积被分解为

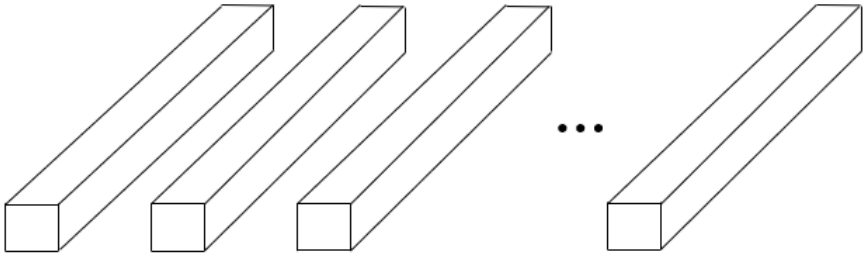
深度卷积和逐点卷积的过程。



（a）标准卷积滤波器



（b）深度卷积滤波器



（c）逐点卷积

图 2-1 标准卷积分解为深度卷积和逐点卷积的过程

一个标准的卷积层，假定其输入特征维度映射 F 为 $D_F \times D_F \times M$ ，其输出特征维度 $D_F \times D_F \times N$ ，其中 D_F 为输入特征的空间宽度和高度，M 是输入通道的深度，N 为输出通道的深度。一个标准卷积核 K 的大小为 $D_K \times D_K \times M \times N$ ，其中 D_K 是卷积核的空间维度。步长为 1 的标准卷积的输出特征计算公式为：

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \bullet F_{k+i-1,l+j-1,m} \tag{2.1}$$

其计算成本为 $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$ ，计算成本随着通道数量乘积式增加。标准卷积操作的作用是根据卷积核来提取图像中的特征，并将这些特征混合在一起以创建新的图像

表示。而分解卷积则将这个提取和混合的过程分成两个独立的步骤，以减少计算的复杂性。

深度可分离卷积由两层组成：深度卷积层和逐点卷积层。深度卷积层会对每个输入通道分别应用一个滤波器，就像是对每个通道单独进行筛选，提取出各自通道的重要特征。接下来，在逐点卷积层中，使用一个简单的卷积操作，将深度卷积层的输出进行逐点的线性组合，就像是把不同通道的特征信息合并在一起。每个输入通道单独滤波的深度卷积过程可以写为：

$$\hat{G}_{k,l,m}=\sum_{i,j}\hat{K}_{i,j,m}\bullet F_{k+i-1,l+j-1,m} \tag{2.2}$$

其中深度卷积核 $\hat{K}$ 的大小为 $D_K\times D_K\times M$，特征 F 的第 m 通道，通过滤波器 $\hat{K}$ 的第 m 分量，生成输出特征的第 m 通道。深度可分离卷积通过两个步骤完成特征提取和组合的过程。首先，深度卷积层对每个通道进行滤波，捕捉各个通道的重要特征。然后，逐点卷积层将这些特征进行线性组合，生成新的特征表示。其总的计算成本为 $D_K\cdot D_K\cdot M\cdot D_F\cdot D_F+M\cdot N\cdot D_F\cdot D_F$。

相较于标准卷积，深度可分离卷积大大降低了计算复杂度，MobileNet 网络使用了3×3的深度可分离卷积。深度可分离卷积可以减少计算量和模型大小，从而使 MobileNet 模型具有更高的效率和更少的存储需求。其在移动设备和嵌入式设备中具有很高的应用价值，因为这些设备通常具有较少的计算资源和存储容量。

MobileNet 在某种程度上对遮挡具有一定的鲁棒性。正是因为 MobileNet 采用了深度可分离卷积块的操作，这意味着它可以同时在不同的尺度上提取特征。这种多尺度感受野的设计使得 MobileNet 能够在一定程度上捕捉到物体的局部特征，即使部分区域被遮挡也能从其他可见的区域中获取有用的信息。

本文使用的深度神经网络的网络架构如表 2.1 所示。

表 2.1 网络结构图

层类型	输出维度	参数数量
输入层	160×160×3	0
普通卷积块 1	80×80×32	992
深度卷积块 1	80×80×32	416
逐点卷积块 1	80×80×64	2304
深度卷积块 2	40×40×64	832

逐点卷积块 2	40×40×128	8704
深度卷积块 3	40×40×128	1664
逐点卷积块 3	40×40×128	16896
深度卷积块 4	20×20×128	1664
逐点卷积块 4	20×20×256	33792
深度卷积块 5	20×20×256	3328
逐点卷积块 5	20×20×256	66560
深度卷积块 6	10×10×256	3328
逐点卷积块 6	10×10×512	133120
深度卷积块 7	10×10×512	6656
逐点卷积块 7	10×10×512	264192
深度卷积块 8	10×10×512	6656
逐点卷积块 8	10×10×512	264192
深度卷积块 9	10×10×512	6656
逐点卷积块 9	10×10×512	264192
深度卷积块 10	10×10×512	6656
逐点卷积块 10	10×10×512	264192
深度卷积块 11	10×10×512	6656
逐点卷积块 11	10×10×512	264192
深度卷积块 12	5×5×512	6656
逐点卷积块 12	5×5×1024	528384
深度卷积块 13	5×5×1024	13312
逐点卷积块 13	5×5×1024	1052672
全局池化层	1024	0
随机丢弃层	1024	0
瓶颈层	128	131072
批量归一化瓶颈层	128	384
全连接层	10575	1364175
Softmax 层	10575	0
嵌入层	128	0

批量归一化（Batchnorm）和 ReLU 非线性激活函数来处理每一层的输出。除了最后的全连接层，在每一层的卷积层之后，都连接着批量归一化层（Batchnorm）和 ReLU 激活函数层，其中，ReLU 激活函数是一种常用于神经网络的激活函数，全称为修正线性单元。它的作用是在神经网络中引入非线性变换，以增加网络的表达能力。它是一种简单而高效的非线性激活函数，有助于神经网络学习和建模输入与输出之间的复杂关系。ReLU 激活函数形式为：

$$f(x)=\max(0,x) \tag{2.3}$$

图 2-2 为其函数形状，其中 x 是激活函数的输入，$f(x)$ 是激活函数的输出。该函数返回输入的正值，如果输入为负，则返回零。

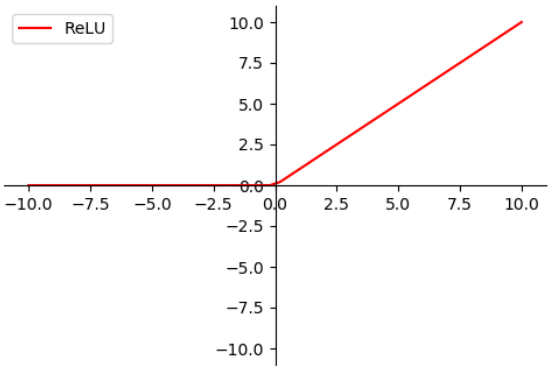


图 2-2 ReLU 函数图像

换句话说，它是一个“门槛”函数，只有当输入值大于零时才会激活神经元。ReLU 函数使用广泛，可以减轻梯度消失问题，有助于加速神经网络的训练；可以使得神经网络更加稀疏，从而减少参数的数量，防止过拟合。图 2-3 表明了标准卷积层以及深度可分离卷积层与批量归一化层和激活函数层的连接图。

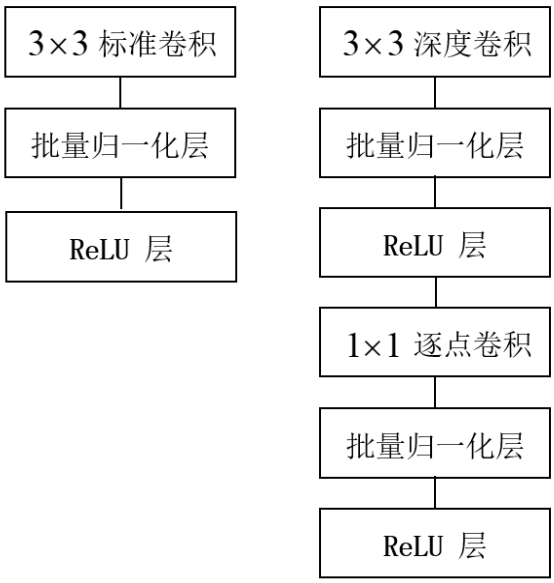


图 2-3 标准卷积块与深度可分离卷积块

在经过第一层标准卷积块和众多深度可分离卷积块堆叠后，特征通过了一个全局池化层（Global pooling）。在全连接层之前，最后的平均池化会将空间分辨率降至 1。特别的，MobileNet 模型使用了全局平均池化（GlobalAveragePooling2D）方法。这种池化层常用于卷积神经网络（CNN）中的图像分类任务。它的作用是对卷积层的输出特征图进行操作，并在每个特征图上应用平均池化操作。全局平均池化是一种降维操作，它将卷积层输出的特征图转换为一个固定长度的向量。具体地，对于每个特征图，它会计算该特征图上所有像素值的平均值，得到一个单一的数值。然后，对所有特征图进行类似的操作，得到一组池化后的数值。最终，这组数值被用作分类器的输入。全局平均池化的优点是它能够保留重要的特征信息，同时减少特征图的维度。这样做有助于减少模型的计算复杂度和参数数量，使得模型更加轻巧和高效。此外，全局平均池化还具有一定程度的位置不变性，对输入图像的平移和缩放具有鲁棒性。正是因为 MobileNet 在卷积层中使用了池化操作，这导致了特征在空间上的位置不变性，这意味着即使某个区域被遮挡，MobileNet 仍然能够在其他可见区域中检测到相同的特征。这使得 MobileNet 对于一定程度的遮挡具有一定的容忍性。

在通过全局池化层后，再连接了一个随机丢弃层（dropout）防止网络过拟合。过拟合会导致模型对于新的数据无法进行准确预测。造成过拟合的原因可能是模型过于复杂，拥有过多的参数和层数，使得模型可以对训练数据中的噪声和异常值进行拟合，但对于测试数据却无法推广。此外，如果训练数据太少，也容易导致过拟合现象，因为模型无法从有

限的数据中获取足够的信息来进行准确预测。

在训练 MobileNet 的时候，使用了防止网络参数过拟合的随机丢弃的正则化技术。在训练过程中，随机丢弃层（dropout 层）会以一定的概率随机丢弃一部分神经元。每次训练时，只有部分神经元参与计算，其他神经元则被“关闭”。在本次实验中，随机丢弃概率为 40%。这样，网络不再依赖于单个神经元的输出，而是需要多个神经元共同发挥作用来进行预测，从而降低了模型的复杂性。随机丢弃层可以有效地减少过拟合并提高模型的泛化能力，从而提高模型在测试数据上的性能。此外，随机丢弃层还可以帮助网络学习到更加鲁棒和通用的特征，从而提高模型的整体性能。

随机丢弃层之后连接着瓶颈层，它是指在神经网络模型中，将输入数据通过一系列较高维度的特征表示转换为较低维度的特征表示的层。瓶颈层常用于降低模型的复杂性和计算开销，同时提取输入数据的关键特征。在此实验中深度神经网络的瓶颈层将 1024 维的输入维度降低到了 128 维的目标输出维度。

之后紧接着连接的是批量归一化瓶颈层。批量归一化用于规范化每个特征通道的输入，以加速训练过程和提高模型的泛化能力。而瓶颈结构则是指通过使用较少的参数和计算量，从而在神经网络中形成一个瓶颈，使得网络可以更高效地学习和表示复杂的特征。

全连接层是神经网络中的一个重要组成部分，它通常出现在卷积层和输出层之间。在卷积神经网络中，全连接层的作用是将卷积层输出的高维特征图压缩成一个一维向量，并将其连接到输出层，以进行分类或回归任务。全连接层的结构很简单，每个神经元都与前一层的所有神经元相连接，形成一个完全连接的网络。这意味着全连接层中的每个神经元都会考虑前一层的所有输入信息。由于这种连接方式，全连接层的参数数量通常很大，因此容易导致过拟合的问题。在本次实验中，全连接层的网络参数数量为 1364175，占全部网络参数的 28.87%。全连接层输出 128 维特征向量。

全连接层在神经网络中没有非线性特征，因此它通常与 Softmax 层结合使用，以进行最终的分类任务。Softmax 层是一种常见的神经网络输出层，用于解决多分类问题。它能够将神经网络的输出转化为一个概率分布，其中每个输出单元表示样本属于某个类别的概率。人们希望具有更大特征维度的模型能够与具有较小特征维度的模型表现相当好。然而，为了达到相同的准确性，可能需要更多的训练数据。值得注意的是，在训练过程中，使用了 128 维的浮点向量表示特征，但实际上可以将其压缩为 128 字节而不会丢失准确性。因此，每张人脸可以用一个紧凑的 128 维字节向量表示。这对于大规模的聚类和识别任务非常有效，可以取得理想的效果。通过使用这样的表示方式，我们可以在计算和存储资源方面更

加高效地处理人脸数据较低的特征维度可能会略微降低准确性，但可以在移动设备上使用，具有实用性。在此实验构建的深度神经网络中，包含各种卷积块中的归一化层和 ReLU 层的话，总共包含 88 层，合计 4724495 个参数，其中可训练的参数数量为 4702351 个，不可训练的参数数量为 22144 个。

2.3 遮挡人脸识别模型架构

图 2-4 为本文设计的遮挡人脸识别模型架构的缩略图，表 2.2 为对应模型架构技术金字塔，列举了该模型里每一步采用的各类算法。下文对相应算法及模型结构进行进一步说明。

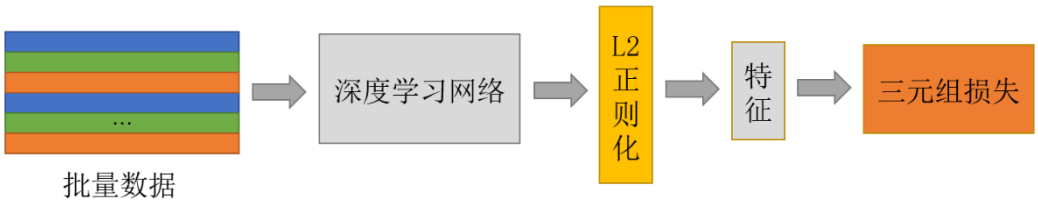


图 2-4 模型结构图

表 2.2 模型架构技术金字塔

网络模型	损失函数	学习率优化算法	学习率调整算法
MobileNet 网络	三元组损失函数	小批量随机梯度下降算法	Warm Cosine 学习率调度方法

网络由批量输入层和深度卷积神经网络组成，其中的深度神经网络架构详情如前文所示，在提取特征后，随后进行 L2 正则化，生成面部特征。

L2 正则化（L2 Regularization）是一种正则化方法，用于在机器学习中降低模型的复杂度并防止过拟合。其通过在损失函数中添加正则化项来实现。这个正则化项是模型权重的平方和与一个正则化参数的乘积。它的作用是惩罚权重较大的特征，鼓励模型使用较小的权重，从而避免过度依赖少数特征。在 L2 正则化中，正则化项的计算公式如下：

$$\text{正则化项} = \lambda \times \|w\|^2, \tag{2.4}$$

其中，λ 是正则化参数，控制正则化项的权重；$\|w\|^2$ 是模型的权重向量的 L2 范数的平方。L2 正则化是一种通过在模型的损失函数中添加一个额外的项来降低模型复杂度的技术。这个额外的项与模型的权重向量有关，它是权重向量的每个元素的平方和的乘以一个正则化参数。可以将 L2 正则化看作是一种“惩罚”机制，它鼓励模型选择较小的权重值。这样做的目的是防止模型过度依赖于训练数据中的噪声和细节，从而提高模型的泛化能力。通过

添加 L2 正则化项到损失函数中，训练过程会尽量减小权重向量的 L2 范数的平方，从而使模型的权重趋向于较小的值。

在训练模型参数时，本文选择使用三元组损失函数（Triplet loss）进行模型训练，并将结果直接反映在人脸验证识别以及聚类的过程中。图 2-5 是三元组损失函数的训练过程，下面进行介绍三元组损失函数的原理介绍。

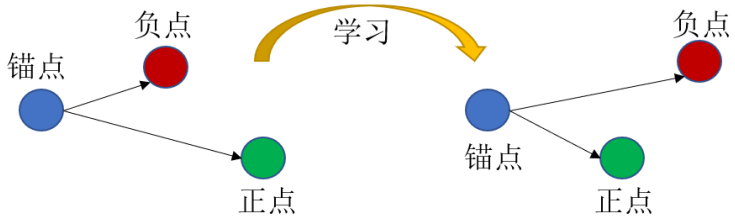


图 2-5 三元组损失函数训练过程

我们希望从一张图片中提取图像特征 $f(x)\in R^d$，使其从图片数据映射到 d 维的特征空间，并且我们希望特征向量存在于 d 维超球面上，满足 $\|f(x)\|_2=1$。该方法将一组训练样例分成三个数据点：锚点 x_i^a（anchor）、正点 x_i^p（positive）和负点 x_i^n（negative）。锚点和正点来自于同一个人，负点来自于不同人，我们希望其特征满足如下关系：

$$\begin{aligned}&\left\|f(x_i^a)-f(x_i^p)\right\|^2+\alpha<\left\|f(x_i^a)-f(x_i^n)\right\|^2,\\&\forall(f(x_i^a),f(x_i^p),f(x_i^n))\in\Omega,\end{aligned}\tag{2.5}$$

其中，α 是边际距离，也称为余量，是正样本和负样本之间的最小距离，其应该大于一个预设的阈值，这个值在该项(2.5)目中为人为指定的超参数，在模型训练前进行人为宏观调控。Ω 表示在训练集里面的所有可能的三元组组合，里面包含 N 个元素。则损失函数定义为：

$$L(x)=\sum_i^N\left\|f(x_i^a)-f(x_i^p)\right\|^2-\left\|f(x_i^a)-f(x_i^n)\right\|^2+\alpha\tag{2.6}$$

在实验中，超参数边际距离 α 取值为 0.2，在对三元组的选择上，为了保证较快的收敛，对于指定的锚点 x_i^a，我们想要选取相应的正点 x_i^p（hard positive）和负点 x_i^n（hard negative）满足：

$$\begin{aligned}x_{iopt}^{\ p}&=\arg\max_{x_i^p}\left\|f(x_i^a)-f(x_i^p)\right\|_2\\x_{iopt}^{\ n}&=\arg\min_{x_i^n}\left\|f(x_i^a)-f(x_i^n)\right\|_2\end{aligned}\tag{2.7}$$

但在实际中，在整个训练数据集上遍历计算 $\arg\max$ 和 $\arg\min$ 是不可取的，并且这样也可能导致训练效果不佳，可能导致错误标记和较差的样本主导正点和负点的决策。可以采取小批量（mini-batch）生成正点和负点的方法取代遍历所有样本点。从每个身份中选择了一个小批量，其中包含了 32 张脸。除了这些正样本之外，还随机选择了一些负面样本脸，并将它们添加到每个小批量中。正确的三元组选择可以使收敛速度更快。

本实验采取的学习率优化算法是经典的小批量随机梯度下降算法（Mini-batch SGD）。小批量随机梯度下降（Mini-batch SGD）是一种在机器学习中常用的优化算法，用于训练模型参数[15]。

随机梯度下降（Stochastic Gradient Descent）是一种常用的优化算法，用于训练机器学习模型中的参数。在传统的梯度下降算法中，每次更新模型参数时都需要计算整个训练数据集的梯度，这在大规模数据集上可能非常耗时。相比之下，SGD 每次更新只使用单个样本的梯度来估计整体梯度方向，从而大大减少了计算开销。具体而言，SGD 的工作方式是：对于每个训练样本，计算其损失函数关于模型参数的梯度，并使用该梯度来更新模型参数。这样，通过迭代处理所有训练样本，逐步调整模型参数以最小化损失函数[15]。

小批量随机梯度下降算法是在随机梯度下降算法上的进一步提升，小批量随机梯度下降算法每次更新模型参数时，不是使用整个训练数据集的梯度，而是仅使用小部分数据，称为小批量（Mini-batch）。具体而言，小批量随机梯度下降将训练数据集分成若干个大小相等的小批量，每次迭代时，从这些小批量中随机选择一个作为当前迭代的训练样本。然后，计算该小批量样本的损失函数梯度，并使用该梯度来更新模型参数[15]。这样可以降低计算梯度的开销，并且可以在一定程度上减少梯度的方差，有助于更稳定地优化模型。小批量采样的方法也用于在随机梯度下降（Stochastic Gradient Descent）的模型参数更新过程中提高收敛速度[15]。

本实验的使用的学习率调整算法是 Warm Cosine 学习率调度方法。Warm Cosine 学习率调度是一种结合了温启动和余弦退火策略的方法。在训练的早期阶段，通过温启动逐渐增加学习率，从一个极小值逐步提升至初始学习率。这样做的目的是避免网络在训练初期就陷入不稳定状态。温启动阶段完成后，学习率按照余弦退火策略进行调整，随着训练的进行，学习率逐渐减小。这个策略有助于网络在训练过程中逐渐趋于稳定，并最终收敛到更好的最优解。Warm Cosine 学习率调度可以用以下科学表达式描述，在温启动阶段，学习率（Learning Rate）逐渐增加：

$$LR = LR_0 \times (1 - \exp(-\alpha \times current\_step)), \tag{2.8}$$

在余弦退火阶段，学习率按照余弦函数进行调整：

$$LR = 0.5 \times LR_0 \times (1 + \cos(\pi \times current\_step / total\_step)), \tag{2.9}$$

其中学习率用 LR 表示，初始学习率用 LR_0 表示，当前步数，即训练进行的步数用 $current\_step$ 表示，总共的训练步数用 $total\_step$ 表示，α 是一个控制温启动速度的超参数。这两个阶段的学习率调整方式结合在一起，形成了 Yolox 的 Warm Cosine 学习率调度策略。

2.4 本章小结

本章节介绍了该实验采用的 MobileNet 深度神经网络的架构。着重介绍了其核心成分——深度可分离卷积块，以及对网络架构里的 ReLU 非线性激活函数、全局池化层、随机丢弃层、瓶颈层、批量归一化瓶颈层、全连接层、Softmax 层分别进行了介绍。在此实验中，输出的人脸特征维度为 128 维，其构建的深度神经网络在包含各种卷积块中的归一化层和 ReLU 层的情况下，总共有 88 层，合计 4724495 个参数。

之后介绍了整个遮挡人脸识别模型的架构。着重介绍了三元组损失函数的原理，绘制了技术金字塔并对其学习率更新算法小批量随机梯度下降算法以及 Yolox 的 Warm Cosine 学习率调度策略进行了介绍。并完善了整个遮挡人脸识别模型用于进行后续验证工作。

第三章 模型验证及实验

3.1 数据集选取

在训练完深度神经网络之后，首先在 LFW 数据集上对模型进行了验证，LFW（Labelled Faces in the Wild）是一个广泛用于人脸识别领域的公开数据集，如图 3-1 所示为数据集中随机挑选的样本。该数据集由来自互联网的人脸图像组成，包含超过 1 万个身份的 13000 多张人脸图像，其中一些身份有多张图像。这些图像具有不同的姿态、表情、光照条件和背景，并且来自于各种年龄、种族、性别和职业。该数据集旨在评估算法在实际场景中的性能表现。

图 3-1 LFW 数据集样本示例

LFW 数据集在人脸识别领域广泛使用，既可用于评估传统的机器学习方法，也可用于评估深度学习方法。由于该数据集具有挑战性和多样性，因此是评估算法在复杂场景下的鲁棒性和泛化能力的理想选择。

在使用 LFW 数据集验证后，本文使用 MaskTheFace 的方法人为将 LFW 的数据集样本加上了口罩遮挡，生成了 OCC-LFW 数据集，如图 3-2 所示。

图 3-2 OCC-LFW 数据集样本示例

下面简要介绍一下 MaskTheFace 人工口罩合成方法的原理。MaskTheFace 是一个基于计算机视觉的脚本，用于在图像中遮挡人脸。它使用基于 dlib 的人脸关键点检测器来确定人脸的倾斜角度和六个关键特征，以便应用遮挡。根据人脸的倾斜角度，从遮挡库中选择相应的模板遮挡物。然后，根据六个关键特征将模板遮挡物进行变换，使其完美地适应人脸。MaskTheFace 提供了多种可供选择的遮挡物。由于在各种条件下收集遮挡物数据集比较困难，因此 MaskTheFace 可以将任何现有的人脸数据集转换为遮挡人脸数据集。MaskTheFace 可以识别图像中的所有人脸，并根据用户选择的遮挡物对其进行处理，考虑到人脸角度、遮挡物适配度、光照条件等各种限制。其口罩类型包含手术口罩、N95 口罩、KN95 口罩、布口罩、防毒面具等口罩种类。在本文中，我们对每一个身份的每一张图像进行随机口罩遮挡类型的选择，并将重新生成的数据集命名为 OCC-LFW 数据集且用于后续的验证工作中。

3.2 模型验证

本文还在人脸验证任务上评估了此方法。即给定两张脸部图像，使用平方 L2 距离阈值 $D(x_i, x_j)$ 来确定同类或者异类的分类，所有同一身份的人脸对 (i, j) 用 P_{same} 标记，不同身份的人脸对用 P_{diff} 标记。我们定义真阳性和假阳性的概念，真阳性表示被正确分类为同类的人脸对样本 (i, j)，假阳性表示被错误分类为同类的人脸对样本 (i, j)，其阈值用 d 表示。定义如下：

$$\begin{aligned} TA(d) &= \left\{(i, j) \in \mathrm{P}_{same}, D(x_i, x_j) \le d\right\} \\ FA(d) &= \left\{(i, j) \in \mathrm{P}_{diff}, D(x_i, x_j) \le d\right\} \end{aligned} \tag{3.1}$$

在给定人脸距离阈值 d 下的验证准确率 $VAL(d)$ 和错误接受率 $FAR(d)$ 定义为：

$$\begin{aligned} VAL(d) &= \frac{|TA(d)|}{|\mathrm{P}_{same}|} \\ FAR(d) &= \frac{|FA(d)|}{|\mathrm{P}_{diff}|} \end{aligned} \tag{3.2}$$

在求出验证准确率和错误接受率之后，绘制出了 ROC（Receiver Operating Characteristic）曲线，如图 3-3 所示。ROC 是一种描述二元分类器性能的曲线，当分类器的预测结果不确定时，可以通过改变分类器的阈值来调整模型的验证准确率和错误接受率。ROC 曲线可以反映出在不同阈值下分类器的性能，同时通过计算 ROC 曲线下面积（AUC，Area Under

Curve）来综合评估分类器的性能，AUC 的取值范围在 0 到 1 之间，数值越接近 1，表示分类器的性能越好[16]。在本实验中，在对 OCC-LFW 数据集进行验证的过程中，分类器性能最好时对应的最佳的阈值选值为 1.1100。其分类准确率为97.867%±0.482%，验证率为82.833%±3.99%。其表明该模型既适用于正常无遮挡的数据集验证，也适用于随机生成口罩遮挡的数据集验证。

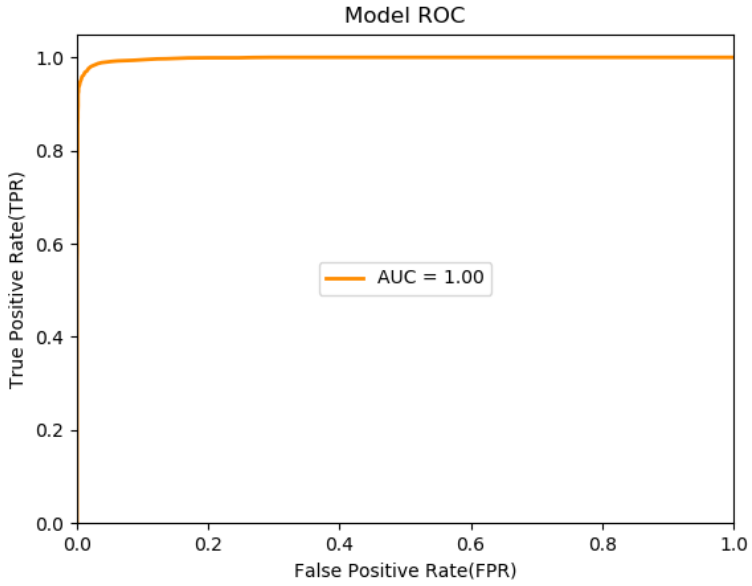
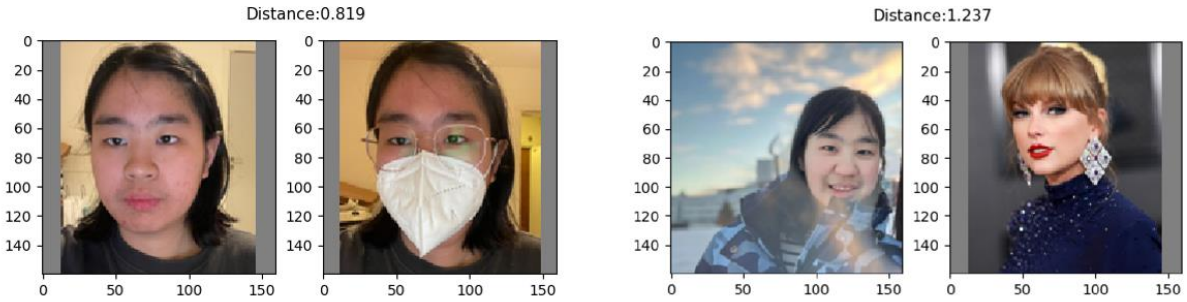


图 3-3 OCC-LFW 验证数据集下的 ROC 以及 AUC 曲线

3.3 身份判断

之后，给定两张特定的人脸图片，先通过上述深度学习网络分析出两张人脸图片的特征，之后计算特征之间的欧式距离，并给定判断阈值。如果两张人脸图片之间的差异小于某个预设的阈值，那么可以认为它们属于同一个身份；而如果两张人脸图片之间的差异大于该阈值，那么可以认为它们属于不同的身份。阈值取值越小表明标准越严苛。例如，在图 3-4 的情况中，“正常状态的 Chenting”和“戴口罩和眼镜的 Chenting”之间的特征距离为 0.819，“Chenting”和“Taylor Swift”之间的特征距离为 1.237。在给定阈值为 1.0 的情况下，第一种情况会被判断为是同一身份，第二种情况被判断为不同身份。



（a）遮挡状态和正常状态下的 Chenting　　（b）Chenting 和 Taylor Swift

图 3-4 计算两张人脸图像之间的距离并进行阈值比较示例

3.4 本章小结

本章介绍了不含遮挡的 LFW 数据集和使用 MaskTheFace 的方法加上口罩遮挡的 OCC-LFW 数据集，并绘制出了 OCC-LFW 验证数据集下的表示二分类器性能的模型 ROC 下的 AUC 曲线。其 AUC 值为 1 表明该模型适用于随机生成口罩遮挡的数据集验证。

之后通过分析两张人脸图片的特征向量以及特征向量之间的距离，与给定的阈值进行比较后进行身份判断。阈值取值越小表明标准越严苛。此判断方法也用于了后续的可视化模型的功能一里，是用于判断人脸图片身份的核心。

第四章 可视化模块

4.1 平台介绍

在这个实验中，选择了 PyCharm 作为集成开发环境（IDE）来进行代码编写和调试。PyCharm 提供了丰富的功能和工具，使开发者能够更高效地进行深度神经网络的开发和调试工作。对于深度神经网络框架，使用了 Tensorflow 1.12.2 版本作为主要的开发平台。Tensorflow[20]是一个广泛使用的深度学习框架，提供了丰富的 API 和工具来构建、训练和部署神经网络模型。此外，还使用了 Keras 库的 2.2.4 版本，Keras 是一个高级神经网络 API，可以与 Tensorflow 等后端框架结合使用[21]，Keras 简化了深度学习模型的构建过程，提供了简洁而灵活的接口。另外，为了进行图像处理和分析任务，使用了 OpenCV 库的 3.4.3.18 版本，OpenCV 是一个强大的计算机视觉库[22]，提供了各种图像处理和计算机视觉算法，方便进行图像预处理、特征提取和结果可视化等操作。

由于 MobileNet 是一种轻量级的网络模型，训练过程没有使用 GPU，而是利用默认的 CPU 进行训练。尽管在 CPU 上训练的速度相对较慢，一次训练大约需要 5 小时左右，但这种设置可以在资源受限的环境中进行训练，并在移动设备上实现高效的实时推理。通过以上的环境配置和选择，该实验能够在 PyCharm 中使用 Tensorflow 和 Keras 开发深度神经网络模型，并利用 OpenCV 进行图像处理和分析。同时，使用默认的 CPU 进行 MobileNet 的训练，确保在轻量化的网络设计下实现了高效的模型训练和推理。

在训练并使用标准验证数据集验证之后，本文还设计了一个可视化的遮挡人脸识别网页化界面。为了实现这一目标，使用了 Flask 集成包。Flask 是一个轻量级的 Python Web 应用框架，它的设计简单易用且具有良好的可扩展性，Flask 类的对象被视为 WSGI 应用程序。WSGI（Web 服务器网关接口）已成为 Python Web 应用程序开发的标准，它定义了 Web 服务器和 Web 应用程序之间的通用接口，并使用了 Werkzeug 作为 WSGI 工具包的一部分，它实现了请求和响应对象以及其他实用函数的功能[17]。这使得可以在 Werkzeug 的基础上构建出强大的 Web 框架。

通过将人脸识别模型与 Flask 框架结合，并结合使用 HTML 语言构建的简易模板，我们成功构建了一个可视化的人脸识别模块。该网页具备两个主要功能。首先，它可以判断两张人脸图片是否来自同一个人，从而实现人脸比对功能。其次，它能够实时调用摄像头进行人脸检测，并进行身份验证。这意味着用户可以通过该网页实时地使用摄像头进行人脸识别，验证其身份。利用此功能，我们能够将模型应用于实际的网络应用程序中，并通

过网页界面提供友好的用户体验。这种可视化的方式使得人脸识别模块更易于使用和操作，进一步推动了人脸识别技术在各个领域的应用。同时，这也为后续的功能扩展和改进提供了基础。通过不断改进和优化这个可视化的人脸识别模块，我们可以使其在实际应用中发挥更大的作用，并满足不同场景下的需求。

本可视化模块设计了两个功能。“判断图像身份”功能与“摄像头检测识别”功能，下面分别对两个功能进行介绍。

4.2 判断图像身份功能

第一个功能的原理与上文描述相同，通过比较输入人脸图像和数据库里所有注册身份的欧式距离，若数据库里所有身份与输入人脸特征之间的距离均超过阈值，则表明无对应身份；若数据库里有小于阈值的潜在身份，则选择距离最小的身份作为最终选择。若两张人脸图片的距离大于阈值，则判断为不同身份。阈值取值越小表明标准越严苛。

第二个功能为实时摄像头人像提取识别，其人脸检测模块调用的是 Retinaface[18]的人脸检测模块。RetinaFace 是一种人脸检测算法，旨在图像中检测多个人脸。它基于一种称为单阶段人脸检测器的深度神经网络架构。下面对 Retinaface 的多图人脸检测原理作简要分析。

RetinaFace 的原理是利用卷积神经网络（CNN）在多个尺度和位置上对输入图像进行分析。该网络在大型带有标注人脸图像的数据集上进行训练，以学习可以区分人脸和非人脸的判别特征。RetinaFace 的架构包括一个主干网络、一组中间特征图和一组任务特定的边界框回归分支。主干网络负责提取图像特征，中间特征图用于在不同尺度上检测人脸，边界框回归分支用于精确地定位人脸位置和姿态。此外，它还通过在特定尺度上进行密集的候选框生成和边界框回归来提高检测的准确性。

首先，经过多尺度金字塔特征提取，RetinaFace 首先使用一个主干网络（如 ResNet）来提取输入图像的特征。然后，通过多尺度特征金字塔[18]的方式，从主干网络的不同层级获得一系列特征图。这些特征图具有不同的尺度，用于检测不同大小的人脸。之后经过特征融合，将不同尺度的特征进行结合，以综合不同尺度上的信息。通常，这可以通过将低层级特征与高层级特征进行连接或级联的方式实现。特征融合有助于提高对不同尺度人脸的检测能力和定位精度。在检测头部阶段，每个特征金字塔层级上，RetinaFace 使用一组卷积和全连接层来进行人脸的边界框回归和分类。边界框回归任务用于预测人脸框的位置和姿态，分类任务则用于判断框内是否包含人脸。之后高效候选框生成阶段，为了提高效

率，RetinaFace 使用了一种称为“anchor”的策略来生成候选框。Anchor 是预定义的一组框，以不同尺度和长宽比分布在特征图上。对于每个 anchor，RetinaFace 通过回归网络对其位置和大小进行微调，以更好地匹配真实人脸。而为了消除重叠的框并提取最佳检测结果。为了解决在不同层级和尺度上进行了多次检测，同一人脸可能会被多个候选框检测到的问题，RetinaFace 使用非极大值抑制算法（NMS）[19]对候选框进行筛选和合并。最后，RetinaFace 根据经过 NMS 处理后的边界框和对应的置信度，输出检测到的人脸位置和相关信息，如人脸框的坐标和人脸的关键点位置。通过以上步骤，RetinaFace 能够在一张图像中同时检测和定位多个人脸，并给出相应的人脸检测结果。

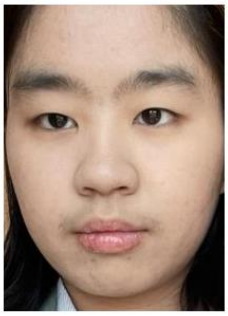
在检测到人脸后，用本文设计的模型进行身份验证识别，返回人脸图像特征间距离最小的身份名称，并实时显示在摄像头监控界面上，且在终端输出其对应的特征距离大小。

下面介绍网页化界面的实操演示，下面简要介绍流程。

首先，运行该可视化子函数后，在 Python 终端界面显示分配的 IP 地址。其网页在端口号 5000 处运行，点进去后进入图 4-1 所示的“人脸识别在线服务窗口”主界面。在“选择文件”命令下将人脸图片传入数据库。“点击这里立即开始人脸识别”按钮表示的是第一个功能——进行人脸图片身份验证；“点击这里进入实时人脸识别系统！”按钮表示的是第二个功能——进行实时摄像头检测。

图 4-1 可视化网页主界面

在实验中，我上传了“Chenting”，“Haihong”，“Taylor Swift”等人的人脸注册图像，并保存在文件夹里，分别命名为“chenting1”，“haihong”，“taylor1”。其上传的图片检测到人脸图像后进行了部分裁剪，其效果如图 4-2 所示。其文件夹里保存的同名注册身份图像如图 4-3 所示：

(a)上传的原图　　（b）裁剪后的注册身份图

图 4-2 图像裁剪示例



（a）Chenting　　（b）Haihong　　（c）Taylor Swift

图 4-3 注册身份图像

点击“点击这里立即开始人脸识别”按钮后，进入图 4-4（a）所示界面，作为示例，在我上传 Taylor Swift 的第二张人脸图片，如图 4-4（b）之后，网页搜寻数据库并显示验证结果。

（a）加载待验证身份的网页界面　　（b）上传的身份图像

图 4-4 人脸身份识别界面

若注册数据库里没有相应的人脸身份，则表示人脸身份匹配失败，显示界面如图 4-5 所

示。

（a）表明无匹配身份的网页界面　　（b）上传的身份图像

图 4-5 人脸身份匹配失败界面

4.3 实时摄像头检测功能

点击“点击这里进入实时人脸识别系统!”按钮后进入第二个功能——实时摄像头检测。图 4-6 为测试存在 N95 口罩、带上镜框眼镜、用手捂住整张脸、用手捂住嘴巴和半只眼睛等不同遮挡情况下的人脸检测以及身份识别情况。

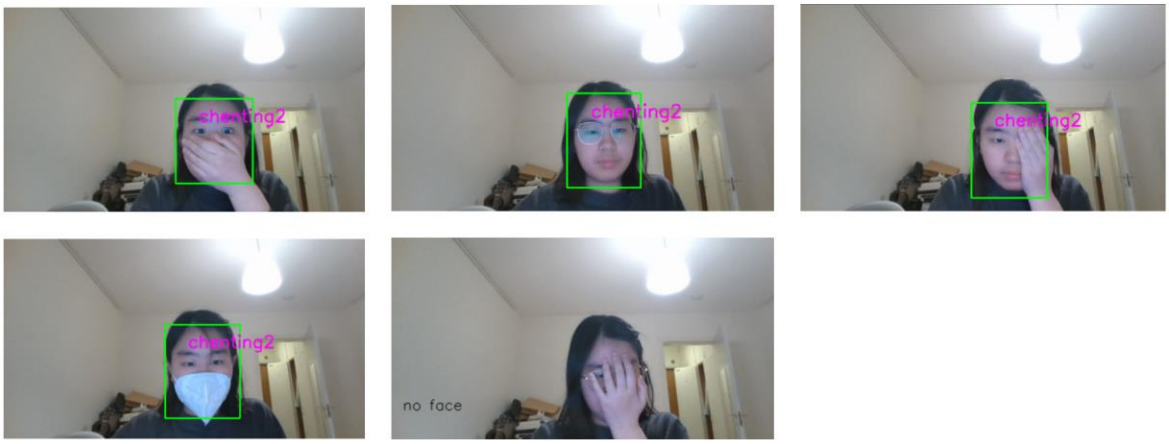


图 4-6 不同遮挡情况下的实时人脸身份识别

由于实时摄像头人像提取时的人脸检测模块调用的是 Retinaface 的人脸检测模块，其功能支持多图人脸识别。如图 4-7 为 Chenting 和 Haihong 同时在摄像头镜头捕捉范围内的系统人脸识别情况。

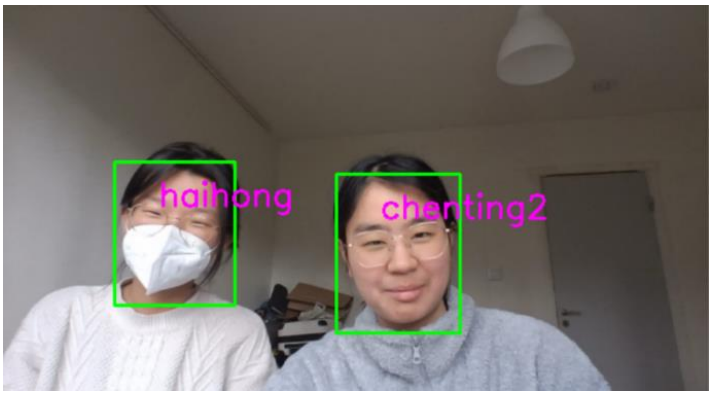


图 4-7 多身份同框的实时人脸身份识别

4.4 模型优缺点分析

MobileNet 是一种轻量级的深度神经网络架构，旨在于移动设备和嵌入式系统等资源受限的环境下实现高效的计算和推理。但 MobileNet 也具有其局限性。

相对于一些更大的模型，MobileNet 在模型大小和计算效率上表现出色，适用于资源受限的环境，如移动设备和嵌入式系统。且由于参数和计算量的减少，MobileNet 的推理速度非常快速。这使得 MobileNet 在实时应用和边缘设备上能够实现高效的实时图像处理和分析，如实时目标检测、人脸识别等。此外，MobileNet 的架构设计使得它可以根据需要进行扩展和调整。通过改变模型的宽度（width multiplier）参数，可以在模型大小、计算复杂度和准确性之间进行权衡，以满足不同应用场景的需求。且 MobileNet 专门设计用于移动设备和嵌入式系统上的应用。它的轻量化特性使得它能够在移动端设备上进行本地计算，避免了对云端的依赖，提高了隐私性和实时性。MobileNet 在模型设计中注重参数的压缩和精简，从而导致较小的模型体积。这对于模型的存储和部署非常有利，特别是在资源受限的设备上。尽管 MobileNet 在模型大小和计算效率方面取得了很大成功，但也存在一些局限性。

MobileNet 网络的特征表示能力相对较弱。相比于一些更大的模型，MobileNet 可能难以捕捉更复杂、细粒度的特征。且模型具有相对较高的误差率，由于模型的轻量化设计，MobileNet 在一些具有挑战性的任务上可能会表现出相对较高的误差率。尤其是在一些需要更高级别语义理解或更精细的视觉分析的任务中，可能需要更复杂的模型架构来达到更好的性能。前两个点在用现实生活中的人脸身份验证时也有所体现，例如在给定阈值过高的情况下无法捕捉双胞胎之间的差异，以及在数据集注册身份缺失的情况下，有一定的几率会误识别为长相相似的另一个人而不是显示“未存在注册身份”选项等等。且由于 MobileNet 的设计注重模型的轻量化，它对于小目标的检测和识别相对困难。这是因为较

小的目标通常具有更少的上下文信息和特征可用于区分，而 MobileNet 在减少参数和计算的同时也限制了其感受野的大小和特征表示的能力。MobileNet 的性能在很大程度上依赖于输入图像的分辨率。较低的分辨率可以减少计算和内存需求，但同时也可能导致细节丢失和性能下降。提高分辨率可以改善细节的保留和任务的准确性，但会增加计算和内存负担。

此外 MobileNet 并不具备专门处理遮挡的能力，而且它的遮挡处理能力有限。它包含的深度可分离卷积块以及全局池化的操作，使得该模型对一定程度的遮挡具有鲁棒性。然而对于严重的遮挡或遮挡导致关键信息完全丢失的情况，MobileNet 可能无法准确地进行分类或提取有用的特征。对于特定的遮挡场景和任务需求，可能需要使用更专门针对遮挡问题进行设计的算法或模型来获得更好的性能。

综上所述，MobileNet 在轻量化和计算效率方面具有优势，但在特征表示能力、任务复杂性和小目标检测等方面存在一定的局限性。根据具体的应用需求，开发者需要权衡这些因素并选择适合的模型架构。

4.5 本章小结

本章介绍了可视化模块，将遮挡人脸识别网络的功能网页化。

首先介绍了搭建网页的平台，将人脸识别模型与 Flask 框架结合，构建可视化模块。其模块包含“判断图像身份”功能与“摄像头检测识别功能”两个功能，在简要介绍了实现流程后使用现实生活中的各种遮挡情况下的人脸样本对其进行了验证和分析，其实验结果显示模型能识别遮挡情况下的人脸样本并进行正确的身份标识。

最后，该章讨论了以 MobileNet 网络架构为基础的遮挡人脸识别模型的优缺点以及适用的领域范围和局限性。

第五章 总结与展望

5.1 工作总结

本文设计了以 MobileNet 深度神经网络为架构的遮挡人脸识别模型，该网络架构可以从人脸照片中提取 128 维特征向量并进行后续的分类训练。MobileNet 建立在卷积神经网络的基础上，其深度可卷积块的设计减轻了模型的训练负担，通过卷积块的堆叠以及网络末端池化层的设计输出最终特征。

在模型的训练阶段，其核心为三元组损失函数的选取，该损失函数旨在拉近同类特征数据之间的距离而拉远异类特征数据之间的距离。配合小批量随机梯度下降优化算法和余弦退火学习率调度方法，实现遮挡人脸的识别和分类。

在模型的验证阶段，使用 LFW 数据集和通过 MaskTheFace 生成的 OCC-LFW 数据集对遮挡人脸识别模型进行验证，其实验数据表明该模型表现良好。之后举例说明了两张相同身份的人脸图片和两张不同身份的人脸图片其特征之间的距离，和在给定阈值情况下的分类情况。

最后设计了网页可视化模块，将人脸识别模块与 Flask 包结合，在注册人脸数据集样本集里加载身份图片，可以实现身份验证和实时摄像头人脸检测识别两个功能，该实验选取了不同身份个体在佩戴眼镜，手遮挡脸不同部位，佩戴 N95 口罩等各类遮挡情况下的实时身份识别情况，其实验结果表明，该模型性能良好，能应付一定情况下的遮挡人脸识别任务。

5.2 工作展望

未来展望方面，针对 MobileNet 或轻量化模型的发展，有以下几个方向值得关注。可以进行模型改进和优化，继续改进和优化轻量化模型的设计，提升其特征表示能力和性能。可以通过引入更复杂的模块或结构变体，加强模型的表达能力，并通过模型搜索和自动化方法来进一步改进模型架构。并且特定任务的优化：针对特定任务进行模型优化。不同的任务可能对模型的特征表示能力和计算效率有不同的要求。因此，可以通过模型剪枝、蒸馏和定制化等技术手段，针对特定任务进行模型的定制和优化，以提高性能和效率。利用跨模态和多模态融合的方法，将轻量化模型扩展到跨模态和多模态领域，以处理多源数据的融合和跨模态任务。例如，结合图像和语音数据进行人机交互或跨领域的智能应用，需要开发适应性强、计算效率高的轻量化模型。引入增强学习和自监督学习。利用增强学习和自监督学习的方法，进一步提高轻量化模型的性能。通过引入增强学习的技术，模型可

以通过与环境的交互来进行自我优化，适应不同的应用场景。自监督学习则可以利用无监督的数据来进行模型训练，减少对标注数据的依赖。也可以尝试从联邦学习和边缘计算的角度进行拓展，结合轻量化模型与联邦学习和边缘计算的技术，以实现在分布式环境中进行模型训练和推理的能力。通过在本地设备上进行模型计算和推理，可以保护用户隐私，并减少与云端的通信开销。

综上所述，未来的研究方向包括改进模型设计、特定任务优化、跨模态和多模态融合、增强学习和自监督学习以及联邦学习和边缘计算等方面。这些方向将进一步推动轻量化模型的发展和应用，并满足不同领域和场景的需求。

参考文献

[1] Boutros F, Damer N, Kirchbuchner F, Kuijper A. Elasticface: Elastic margin loss for deep face recognition [J]. CoRR, 2021, vol. abs/2109.09416.

[2] 徐遐龄, 刘涛, 田国辉, 于文娟, 肖大军, 梁陕鹏. 有遮挡环境下的人脸识别方法综述[J]. 计算机工程与应用, 2021, 57(17): 46-60.

[3] 刘颖, 张艺轩, 佘建初, 王富平, 林庆帆. 人脸去遮挡新技术研究综述[J]. 计算机科学与探索, 2021, 15(10): 1773-1794.

[4] 王晨博. 基于深度学习的口罩遮挡人脸识别算法研究与实现[D]. 北京: 北方交通大学, 2021.

[5] 袁德飞. 基于深度神经网络的遮挡人脸识别算法研究[D]. 内蒙古: 内蒙古大学, 2022.

[6] 张刚. 局部遮挡的人脸识别深度学习算法改进[D]. 宁夏: 宁夏大学, 2022.

[7] GE S, LI J, YE Q, et al. Detecting Masked Faces in the Wild with LLE-CNNs[C]. In Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2682-2690.

[8] D. U. K. Putri et al. Occluded Face Recognition Using Sparse Complex Matrix Factorization with Ridge Regularization[C]. 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien City, Taiwan, 2021, pp. 1-2, doi: 10.1109/ISPACS51563.2021.9651107.

[9] Li Mao, Fusheng Sheng, and Tao Zhang. Face Occlusion Recognition With Deep Learning in Security Framework for the IoT[J]. IEEE Access, vol. 7, pp. 174531-174540, Nov. 2019.

[10] Recto, I. J. H. & Devaraj, M. (2022). Synthetic Occluded Masked Face Recognition using Convolutional Neural Networks[C]. In 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT) (pp. 124-129). BALI, Indonesia: IEEE. doi: 10.1109/IAICT55358.2022.9887517.

[11] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu. Occlusion robust face recognition based on mask learning with pairwise differential Siamese network[C]. In Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 773–782.

[12] Phelpsmemo. AR face database (128x128) [DB/OL]. https://www.kaggle.com/datasets/phelpsmemo/ar-face-database-128x128.

[13] Neto, P. C, et al. Beyond Masks: On the Generalization of Masked Face Recognition Models to Occluded Face Recognition[J]. IEEE Access, vol. 10, pp. 86224-86233, Jul. 2022, doi: 10.1109/ACCESS.2022.3044337.

[14] Huber M, et al. Mask-invariant face recognition through template-level knowledge distillation[C]. In Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Dec. 2021, pp. 1-8.

[15] Raschka, S. Stochastic Gradient Descent - Methods and Variations. Sebastian Raschka - Personal website [EB/OL]. https://sebastianraschka.com/faq/docs/sgd-methods.html.

[16] Google Developers. An ROC Curve (Receiver Operating Characteristic Curve) Definition. Google Machine Learning Crash Course [EB/OL]. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

[17] Pallets. "Flask." GitHub [EB/OL]. https://github.com/pallets/flask.

[18] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, RetinaFace: Single-shot multi-level face localisation in the wild[C] In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020.

[19] S. Singh, Non-Maximum Suppression (NMS) Towards Data Science [EB/OL]. https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c.

[20] TensorFlow: An Open-Source Machine Learning Framework for Everyone [EB/OL]. https://www.tensorflow.org/overview.

[21] Keras: Deep Learning for Humans [EB/OL]. https://keras.io/.

[22] OpenCV: Open Source Computer Vision Library [EB/OL]. https://opencv.org/.

致　　谢

感谢罗琳老师以及沙路为学长在研究方向的选取以及技术指导方面给予我的支持。在开题的时候我本来选取的技术方向是在 ElasticFace 模型基础上加入模板级的知识提馏（KD）的方法，但后面在实践中发现模型的体量过大，而且模型结构也比较复杂，没法在有限的时间内以及在我目前的电脑配置上进行实验。感谢沙路为学长在我困顿之际为我提供一个新的思路，以及介绍一条新的轻量化模型的实验路线，使得实验可以继续进行。以及感谢沙路为学长在我刚入门这个领域的时候给我指引的深入学习方向以及要点，使我能快速上手并理解原理和继续进行实验以及代码调试。

感谢我的朋友于海泓接受我的请求并愿意提供的人脸样本进行后续测试，使得可视化模块注册人脸识别样本具有多样性，丰富了实验数据。

感谢在我四五月进行本科毕设的同时完成 KTH 研究生第一学年下半学期课程的期间给予我心理支持的各位亲人和朋友。