

分类号_____

编 号_____

U D C_____

密 级_____



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题 目：_____人脸表情检测 Web 应用开发_____

姓 名：_____张然_____

学 号：_____12011511_____

院 系：_____计算机科学与工程系_____

专 业：_____计算机科学与技术_____

指导教师：_____史玉回_____

2024 年 06 月 07 日

诚信承诺书

1. 笔者郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。
2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。
3. 笔者承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。
4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由笔者承担相应的法律责任。

作者签名：张然

2024 年 06 月 07 日

人脸表情检测 Web 应用开发

张然

(计算机科学与工程系 指导教师：史玉回)

[摘要]：随着深度学习技术的迅速发展，人脸表情检测成为了计算机视觉领域的热点问题，具有广泛的应用前景。本文旨在设计并实现一个基于多个神经网络的人脸表情检测系统，通过结合多个先进的神经网络模型，如 ResNet18、VGG 等，实现对人脸表情的准确识别。这些模型在 FER-2013 等公开数据集上进行了充分训练与验证，准确率均达到了 70% 多，有效地提高了表情识别的准确性和泛化能力。系统的前端界面采用了 Vue3 框架，不仅提供了良好的用户交互体验，而且借助 Vue3 的高效性能，实现了快速的渲染速度和轻量级的页面加载。后端采用 Flask 框架，提供了高效的模型预测和数据交互接口。此外，本文引入了 MTCNN 库进行人脸检测，精确识别图像中的人脸区域，进一步提高了识别准确性。面对模型部署和远程访问的挑战，本文采用了 Cpanel 内网穿透云服务，实现了从本地到公网的无缝连接，允许远程用户通过公网 IP 访问系统进行实时人脸表情识别。在实现过程中，本文面临了模型准确率、图片等挑战，通过数据增强、模型优化等方法进行了有效的解决。

[关键词]：人脸表情检测；Vue3；神经网络

[ABSTRACT]: With the rapid development of deep learning technology, facial emotion recognition has become a hot spot in the field of computer vision and has a wide application prospect. The aim of this thesis is to design and implement a facial expression detection system based on multiple neural networks by combining multiple advanced neural network models, such as ResNet18, VGG, etc., to achieve accurate recognition of facial emotion. These models have been fully trained and verified on public data sets such as FER-2013, and the accuracy rate has reached more than 70%, effectively improving the accuracy and generalization ability of expression recognition. The front-end interface of the system adopts the Vue3 framework, which not only provides a good user interaction experience but also realizes fast rendering speed and lightweight page loading with the efficient performance of Vue3. The Flask framework is adopted at the back end, which provides efficient model prediction and a data interaction interface. In addition, this thesis introduced the MTCNN library for face detection to accurately identify the face region in the image, further improving the recognition accuracy. Faced with the challenges of model deployment and remote access, this thesis adopted the Cpolars, an intranet penetrating cloud service to achieve seamless connection from local to the public network, allowing remote users to access the system through public IP addresses for real-time facial expression recognition. In the process of implementation, this thesis is

faced with challenges such as model accuracy and picture, which are effectively solved by data enhancement and model optimization.

[Keywords]: Facial Emotion Recognition; Vue3; Neural Network

目录

1. 研究背景	01
1.1 背景介绍	01
1.2 本文创新点及优势	02
1.2.1 准确性提升	02
1.2.2 用户体验的创新设计	02
1.2.3 技术实现的优化	02
1.2.4 创新点及优势总结	03
2. 数据集介绍	04
3. 技术栈介绍	05
3.1 人脸表情识别模型	05
3.2 人脸检测模型	07
3.3 Cppolar 内网穿透	08
4. 实验研究	10
4.1 总体结构展示	10
4.2 人脸表情识别模型性能展示	11
4.3 前端展示	15
4.3.1 前端性能分析	15
4.3.2 图片测试展示	17
4.3.3 实时视频分析展示	20
5. 总结	22

参考文献.....	23
致谢.....	25

1. 研究背景

1.1 背景介绍

人脸表情是人类社会中一种非语言交际方式。它可以分为许多类型，如厌恶、恐惧、愤怒、快乐、悲伤、惊讶和中立。Mehrabian 指出，7%的信息是通过口语传递的，38%的信息是通过语调传递的，55%的信息是通过人脸表情传递的[1]。由于人脸表情在交流中所占的信息比例很大，在日常生活中起着重要的作用。随着神经网络技术的快速发展，人脸表情识别技术已经成为计算机视觉领域内的重要分支，有相当一部分使用神经网络来发展人脸表情识别技术的应用[2]。具体来说，其在情感识别的领域发挥着很大的作用，如心理学、社交机器人、患者疼痛监测[3]和课堂状态检测[4]等。人脸表情识别大致包括四个步骤：首先是人脸检测，程序从静止图像或视频中检测人脸。第二个阶段是规范化阶段。程序将去除噪声并根据亮度对人脸进行归一化。在第三阶段，程序会识别相关特征，同时删除不必要的数据。然后，在最后一步，将相关特征归类到对应的表情下[5]。然而先前的研究大多都是基于单个方法的代码开源，缺少一个集成的、带前端界面的系统。基于这个背景，本文提出了一种基于多个神经网络模型的，容易使用的人脸表情识别系统，旨在通过技术手段来更准确且高效地理解人类的情感状态。

本文后端采用了多个领先的深度学习模型，包括但不限于 ResNet18、VGG 等，这些模型在人脸表情识别的准确性和效率上都有着显著的表现。深度学习的思想来源于神经网络，更具体地说来源于生物神经系统[6]。在 FER-2013 等知名公开数据集上的表现尤为突出，通过大量的实验训练与验证，这些模型的准确率均能达到 70%以上，这一成绩在当前的科研背景下具有较高的竞争力。此外，为了提高系统的泛化能力和适应不同环境下的表情识别需求，本文在数据预处理、模型训练过程中采用了多种策略和技术，如数据增强、模型微调等，确保了模型在不同的测试场景下都能保持较高的识别准确率。

本文的前端开发采用了当前流行的 Vue3 框架，该框架以其独特的响应式编程和组件化开发理念，为用户提供了流畅且直观的交互体验。在后端开发方面，系统选用了轻量级的 Flask 框架，它的简洁和灵活性使得后端能够快速搭建稳定的数据交互接口，同时兼顾了模型预测过程的高效率 and 低延迟。

为了进一步提升系统的表情识别准确性，本文引入了 MTCNN（Multi-task Cascaded Convolutional Networks）库进行高精度的人脸检测。MTCNN 库通过一系列卷积网络，能够在图片中准确地检测并定位到人脸区域，为后续的表情识别提供了精确的输入。这一步骤不仅显著提高了识别的准确率，也增强了系统对不同环境和姿态下人脸的适应性。

1.2 本文创新点及优势

尽管现有的许多项目已经实现了基本的人脸表情识别功能，但本文在准确性、用户易用性和技术实现方面都有显著的创新和改进。

1.2.1 准确性提升

在人脸表情识别领域，虽然现有的许多开源项目（如参考文献[7]和[8]所示）仍采用几年前的状态最优模型（State of the Art, SOTA），这些模型的识别准确率大致在 63%到 66%之间。本文通过参考 benchmark 网站 PapersWithCode[9]，选用了 4 个当前准确率超过 70%的先进模型，使得本文在识别性能上显著优于现有解决方案。此外，本文对模型进行了以下关键优化以适应项目的具体需求：

1. 模型选择：本文挑选了 4 种在 FER-2013 数据集上表现优异的模型，而不是单一依赖某一模型的架构。这种选择策略使得项目能够从各个模型的强项中受益，从而提高整体的表情识别准确率。

2. 集成学习：结合了多个高性能模型的预测结果，通过集成学习方法如投票和堆叠等，进一步提高了识别的准确性和鲁棒性。这种方法有效地缓解了单个模型可能的偏差和过拟合问题。

通过上述措施，本文不仅提高了表情识别的准确率，也确保了系统能在实际应用中达到快速且可靠的性能表现，满足了用户对高效交互和精确反馈的要求。

1.2.2 用户体验的创新设计

与传统的人脸表情识别项目相比，本文在用户体验方面进行了显著的创新。传统项目通常需要用户在本机或在线服务器上配置复杂的 Python 环境，如安装 Pytorch 或 TensorFlow 等深度学习框架。这一过程不仅耗时而且对于非技术用户来说门槛较高。

本文通过集成化的前后端设计克服了这一问题。具体来说，本文在启动之后，用户无需进行复杂配置，通过 IP 即可访问前端界面。用户可以通过简洁直观的

界面上传图片或进行实时视频分析，这极大地简化了操作流程并提升了用户的交互体验。此外，本文还基本上实现了实时反馈：即用户上传图片后，系统能够快速处理并显示表情识别结果，提供即时的反馈。

1.2.3 技术实现的优化

1. 后端优化：利用轻量级框架 **Flask** 配合高效的数据处理流程，确保了高速的数据传输和处理能力。

2. 负载均衡：通过服务器的负载均衡技术，有效管理用户访问压力，保障服务的稳定和可靠。

3. 并发处理：系统经过优化以支持并发请求处理，使用异步处理技术和多线程来提高应对高流量情况的能力。这允许多个用户同时上传图像并接收表情识别结果，降低了服务的响应时间，提升了用户舒适度。

1.2.4 创新点及优势总结

基于以上的技术创新和用户体验优化，本文在人机交互和情感分析等领域展示了广泛的应用潜力。这些创新和改进不仅提升了人脸表情识别系统的整体性能，也为未来的发展提供了坚实的基础。

首先，在准确性提升方面，本文通过选用 **PapersWithCode** 网站上 4 个当前准确率超过 70% 的先进模型，并结合多种优化策略如模型选择和集成学习，有效地提高了表情识别的准确率和系统的鲁棒性。相比现有的开源项目，本文的解决方案在识别性能上显著优于传统模型，能够更准确地处理复杂的表情数据。

其次，在用户体验的创新设计上，本文采用了集成化的前后端设计，极大地简化了用户操作流程，降低了技术门槛。用户无需复杂配置即可通过 **IP** 访问系统，并通过直观的界面进行操作，实现了实时的表情识别反馈。这种设计不仅提升了用户的交互体验，也增加了系统的易用性和吸引力。

在技术实现的优化方面，本文通过后端的高效数据处理流程、负载均衡技术和并发处理能力，确保了系统在高流量情况下的稳定性和响应速度。轻量级框架 **Flask** 的使用进一步提升了系统的灵活性和性能，使其能够快速处理用户请求并提供准确的识别结果。

展望未来，本文计划进一步探索跨文化表情识别的功能。这一研究方向旨在提高系统在不同文化背景下的适应性和识别准确性。通过收集和分析来自不同文

化背景的表情数据，本文期望开发出具有更高泛化能力的识别模型，推动人脸表情识别技术在更大范围内的应用。

2. 数据集介绍

FER-2013 数据集由 Kaggle 收集，该数据集由 Aaron Courville 和 Pierre-Luc Carrier 于 2013 年在国际机器学习会议 (International Conference on Machine Learning, ICML) 上展示。每张人脸基于不同的情感类别进行分类，数据集中的图像均为灰度。它有 35,887 张人脸情绪图像，包括七种不同的情绪：厌恶、恐惧、愤怒、快乐、悲伤、惊讶和不带感情色彩[10]。这些图片是从互联网上不同的地方收集的，每张图片都被贴上了一种情绪的标签。如图 1 所示。



图 1 FER-2013 数据集图片展示[10]

在人类表现方面，FER-2013 数据集上的平均准确率约为 65.5%[10]。由于其丰富的情感标签和广泛的图像来源，FER-2013 数据集已成为人脸情绪检测领域的重要基准数据集，广泛应用于人脸情绪检测、情绪分析和人机交互等多种研究与应用中。然而，FER-2013 数据集也存在图像分辨率和质量差异较大的问题，这增加了从图像中准确提取地标和其他特征的难度。尽管存在这些挑战，FER-2013 数据集仍然是评估和比较人脸表情检测算法性能的主要基准之一。它的多样性和复杂性为研究人员提供了一个严苛的测试环境，推动了该领域的技术进步和创新应用的发展。

3. 技术栈介绍

3.1 人脸表情识别模型

在开发一个高效的 Web 应用程序以实现人脸表情识别功能的过程中，选择适当的模型至关重要，尤其是在面对像 FER-2013 这样的挑战性数据集时。FER-2013 数据集因其包含多种复杂的表情类别和多样化的图像质量，已成为评价人脸表情识别算法性能的重要基准。因此，准确地挑选和应用合适的模型，直接决定了应用程序的最终性能和用户体验。曾经人们使用了各种各样的方法，包括隐马尔可夫模型[11]和朴素贝叶斯分类器[12]，并且已经成功地应用。同时在过去十年中，许多研究者发现多层感知器在 FER-2013 数据集中取得了很好的效果[13]。

为了确保所选模型在 FER-2013 数据集上的表现优异，本文参考了 PapersWithCode 网站[9]。该平台是一个集成了人工智能领域最新研究成果的综合性资源库，提供了详尽的基准测试系统，使研究人员能够对不同模型的性能进行比较，并紧跟最新的研究进展。PapersWithCode 网站不仅列出了当前最先进的模型及其在各种数据集上的表现，还详细描述了每个模型的实现细节和性能指标。这些信息为研究人员选择最适合自己的模型提供了宝贵的参考。

在本文中，笔者特别关注了 PapersWithCode 平台上关于 FER-2013 数据集的性能排行榜。该排行榜主要展示了不同模型在该数据集上的准确率性能指标，通过这些数据，可以清晰地看到各模型的优劣势，从而做出最优选择。图 2 展示了这些模型在 FER-2013 数据集上的具体表现，为本文后续的模型选择和优化提供了重要依据。

Rank	Model	Accuracy	Extra Training Data	Paper	Code	Result	Year	Tags
1	Ensemble ResMaskingNet with 6 other CNNs	76.82	✓	Facial Expression Recognition using Residual Masking Network	Code	Result	2021	
2	EmoNeXt	76.12	×	EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition	Code	Result	2023	
3	Segmentation VGG-19	75.97	✓	A novel facial emotion recognition model using segmentation VGG-19 architecture	Code	Result	2023	
4	Local Learning Deep+BOW	75.42	✓	Local Learning with Deep and Handcrafted Features for Facial Expression Recognition		Result	2018	
5	LHC-Net	74.42	✓	Local Multi-Head Channel Self-Attention for Facial Expression Recognition	Code	Result	2021	Self Attention ResNet
6	Residual Masking Network	74.14	✓	Facial Expression Recognition using Residual Masking Network	Code	Result	2021	
7	ResNet18 With Tricks	73.70	×	Fer2013 Recognition - ResNet18 With Tricks	Code	Result	2021	ResNet
8	VGGNet	73.28	×	Facial Emotion Recognition: State of the Art Performance on FER2013	Code	Result	2021	VGG
9	CNN Hyperparameter Optimisation	72.16	×	Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition	Code	Result	2021	
10	Ad-Corre	72.03	×	Ad-Corre: Adaptive Correlation-Based Loss for Facial Expression Recognition in the Wild	Code	Result	2022	

图 2 FER-2013 模型排行榜[9]

笔者在选择模型的过程中遇到了多项挑战。初始尝试复现排名较高的几个模型时，第二名和第三名模型因环境配置问题未能成功实现，第四名模型没有开放源代码，这导致本文无法直接采用。进一步的调查发现，第六名的模型代码与排名第一的模型属于同一篇研究论文，而第一名模型实际上是第六名模型的改进版本。因此，决定采用排名第一、第五、第七和第八的模型作为后端进行后续开发。它们分别为 Ensemble ResMaskingNet with 6 other CNNs, LHC-Net, ResNet18 With Tricks 和 VGGNet。以下是对这四个模型的简要介绍。

Ensemble ResMaskingNet with 6 other CNNs 模型采用了集成学习策略，将 ResMaskingNet 与其他六种卷积神经网络（CNN）相结合，实现了高效的人脸表情识别。通过整合多个模型的优势，该集成模型充分利用了每个独立模型在特定情境下的优点，通过策略融合显著提高了整体的识别准确率和系统的稳定性。同时其他六种 CNN 的计算效率和特征提取能力在图像分类中展现出了巨大的潜力[14]。集成学习方法通过结合不同模型的预测结果，有效减小了单一模型可能出现的偏差和过拟合问题，从而增强了模型的泛化能力。这使得 Ensemble ResMaskingNet 在学术研究和实际应用中备受关注和推崇[15]。

LHC-Net 模型是一种为轻量级操作和高效率设计的卷积神经网络,特别适用于移动设备和实时应用。该模型通过简化网络结构,显著减少了计算负担,使其能够在资源受限的环境中高效运行。同时,LHC-Net 通过采用创新的参数优化策略,在保证高准确性的同时,降低了计算复杂度。例如,利用高级优化算法和正则化技术,LHC-Net 在训练过程中有效避免了过拟合问题,并在测试阶段表现出良好的鲁棒性。该模型的设计特别考虑了移动设备的计算资源限制,使其在实时情感分析、移动健康监测和智能家居设备等领域具有广泛的应用前景[16]。

ResNet18 With Tricks 模型在标准的 ResNet18 基础上进行了多方面的优化,采用了诸如数据增强、正则化和早停法等先进的训练技巧,从而显著提升了模型的性能。数据增强策略通过增加训练数据的多样性,提高了模型的泛化能力;正则化方法有助于防止过拟合,使模型在不同数据集上均表现稳定;早停法则通过在训练过程中动态调整,防止模型过度训练。这些优化措施使得 ResNet18 With Tricks 在处理表情多样性的 FER-2013 数据集时,能够保持优异的表现和高效的识别能力[17]。

VGGNet 模型作为深度学习领域的经典模型之一,以其深层次的网络结构和优越的性能在各种计算机视觉任务中得到了广泛应用。在人脸表情识别任务中,VGGNet 通过其深层的卷积网络,能够捕捉到人脸表情的复杂特征,从而实现准确的表情识别。VGGNet 的多层次特征提取能力,使其能够有效处理不同情境下的表情变化,包括细微的面部动作和复杂的表情组合。这使得 VGGNet 在学术研究和实际应用中,特别是在情感计算和人机交互领域,展现出了强大的应用潜力[18]。

3.2 人脸检测模型

在人脸表情识别领域,实际应用中常常面临的一大挑战是上传图片与训练数据集(如 FER-2013)的不一致性。FER-2013 数据集只包含单一人脸的黑白图像,而实际情况远更复杂,例如,用户上传的图片可能为彩色,只有人脸的一部分,亦或是包含多张人脸,如图 3 所示。这些差异会显著影响人脸表情检测模型的判断,直接应用这些图片往往导致准确度大幅下降。

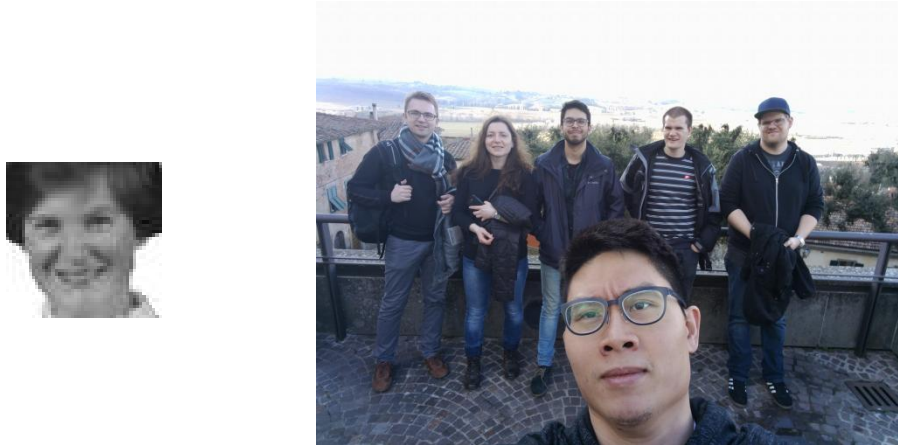


图3 数据集图片（左）和实际情况图片（右）

为解决这一问题，本文引入了多任务级联卷积网络（MTCNN），这是一种高效的人脸检测技术，特别适用于处理复杂的图像场景。MTCNN 通过其独特的三阶段检测流程（即 P-Net、R-Net 和 O-Net）逐步细化人脸检测的准确性。在第一阶段，P-Net 进行初步的人脸候选区域提取；随后，R-Net 评估这些候选区域的质量，筛选出潜在的真实人脸区域；最后，O-Net 对这些区域进行进一步的精细调整和关键点检测，如眼睛、鼻子和嘴巴的位置。MTCNN 的这一多阶段方法不仅提升了人脸检测的速度，还显著增强了模型在多尺度和部分遮挡条件下的表现。这种技术的引入，确保了即便在非理想环境中也能高效、准确地识别出人脸及其表情。特别是在人脸识别领域，MTCNN 因其出色的精确度和灵活性，已被广泛应用[19]。

在本文中，通过首先使用 MTCNN 进行人脸检测，本文能够精确地定位图像中的每个人脸区域，随后将这些区域作为输入数据提供给人脸表情检测模型。这样的处理流程不仅优化了人脸表情的识别准确率，也极大地提高了系统的实用性和可靠性。通过这种方法，本文能够有效应对来自不同来源和复杂背景下的图像，展现了在实际操作中的强大适应性和应用潜力。

3.3 Cpolar 内网穿透

在本文的开发过程中，本文已经基本完成了前后端的集成和框架搭建。然而，

本文面临一个关键问题：所有服务目前只能在本机上运行，无法被外部网络访问。为了克服这一限制，本文引入了 **Cpolar**，一个安全的内网穿透云服务，以实现服务的公网访问能力[20]。

Cpolar 提供了一种将内网下的本地服务器通过安全隧道暴露至公网的机制。这一技术特别适合于需要远程演示和测试的开发环境，因为它允许从任何地点通过生成的公共 URL 访问本地运行的应用。**Cpolar** 的核心功能是其能够创建安全的通道，这些通道使用了最新的加密标准，保障数据传输过程中的安全性。

进一步地，**Cpolar** 支持 HTTP/2 的优化技术版本，使得通过其隧道的连接可以实现更快的数据传输速度。这一特性尤其对于需要频繁更新数据的实时应用来说，提高了用户体验和响应速度。**Cpolar** 还避免了传统内网穿透方法中常见的问题，如需在路由器上进行繁琐的端口转发设置或依赖动态 DNS 解决方案。此外，即使设备网络环境发生变化，**Cpolar** 的连接仍然能够稳定工作，这一“零配置”功能显著降低了技术支持的复杂度和维护成本。

具体来说，笔者将本机的几个端口映射到公网 IP 上，使得外部设备可以通过公网 IP 来访问本机项目从而进行人脸表情识别。这种设置允许项目的利益相关者在没有直接访问内部网络的情况下测试和评估面部表情识别技术的有效性和准确性。通过使用 **Cpolar**，本文可以将面部表情识别应用的测试界面暴露给远程协作者。协作者可以实时上传他们的面部图像并接收系统分析后的表情识别结果。这提供了一个多样化的测试环境，来自不同地理位置的用户都可以参与到测试中。

4. 实验研究

4.1 总体结构展示

本文所开发的人脸表情识别系统的总体结构如图 4 所示。系统架构设计旨在实现高效、准确的实时人脸表情识别，通过前端和后端的紧密协作，实现数据的快速传递和处理。以下是对系统各组成部分及其工作流程的详细描述：

首先，系统的前端负责接收用户上传的图像。这些图像可以通过各种设备（如电脑、手机等）上传至前端界面。前端界面基于 Vue3 框架开发，提供了简洁直观的用户交互体验，并通过优化的界面设计确保图像上传过程的流畅性和便捷性。

在图像上传至前端后，图像数据会立即被发送至后端进行处理。后端采用 Flask 框架，确保了数据传输的高效性和系统的稳定性。后端处理的第一步是进行人脸检测，这一步骤由 MTCNN 模型完成。一旦 MTCNN 模型成功检测出人脸区域，系统将这些识别出的人脸图像传递给四个不同的表情识别模型进行进一步的表情判断。这四个模型分别是 Ensemble ResMaskingNet with 6 other CNNs、LHC-Net、ResNet18 With Tricks 和 VGGNet。每个模型通过独立的表情识别算法，对人脸图像进行情感分类，并返回其识别结果。各模型的设计和优化确保了在不同环境和表情复杂度下的高识别准确率和鲁棒性。

在完成表情识别后，系统将所有模型的识别结果汇总，并通过后端传回前端界面。前端界面实时显示这些识别结果，提供用户即时的反馈。

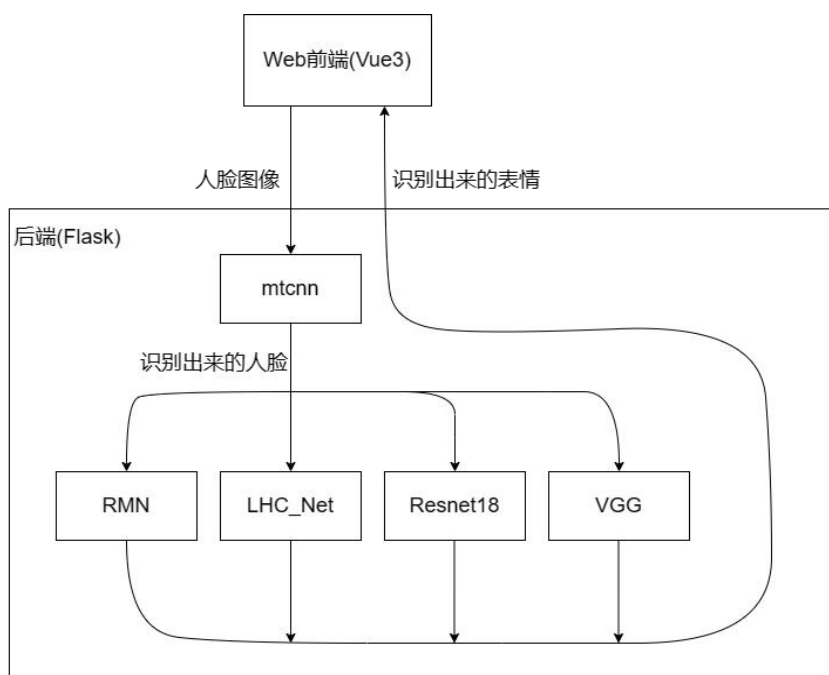


图 4 总体结构

4.2 人脸表情识别模型性能展示

本文所采用的四种人脸表情识别模型的性能评估结果，包括准确率，模型结构和混淆矩阵，分别如表 1、图 5、图 6、图 7、图 8、图 9、图 10、图 11 和图 12 所示。为了全面评估各模型的识别效果和性能，本文对每个模型在 FER-2013 数据集上的表现进行了详细的实验和分析。

表 1 不同模型的准确率

模型	准确率
Ensemble ResMaskingNet with 6 other CNNs	76.82%
LHC-Net	74.42%
ResNet18 With Tricks	73.70%
VGGNet	73.28%

数据来源：[9]

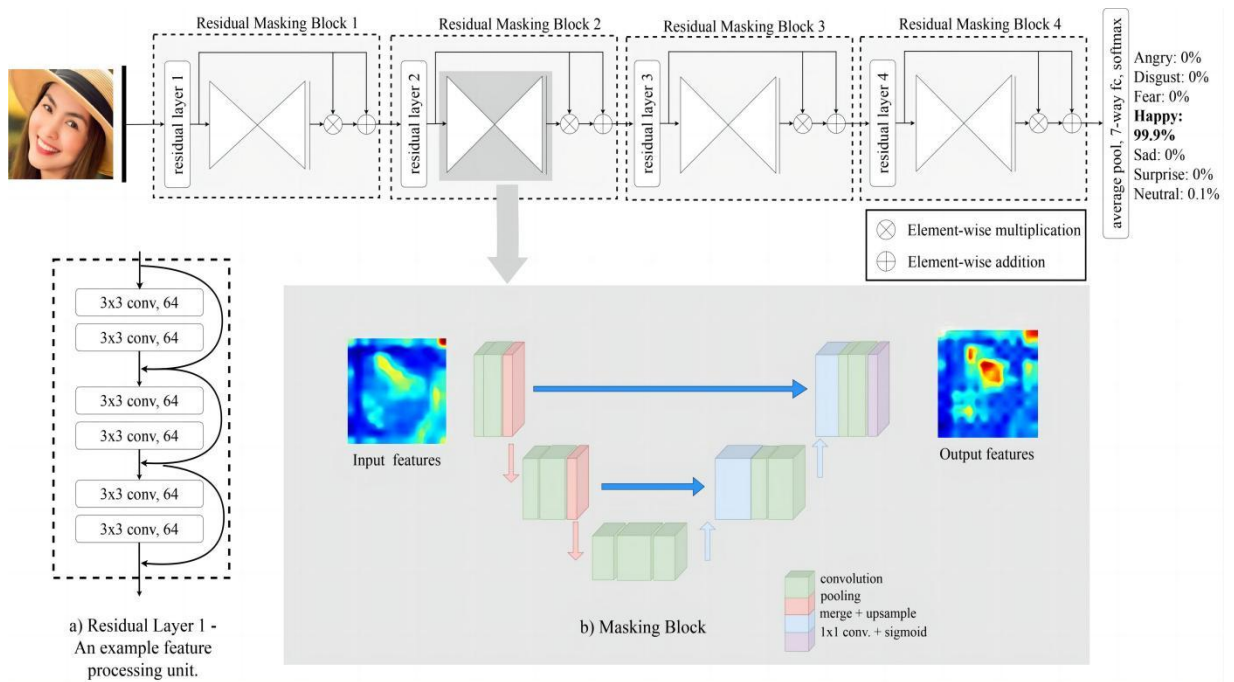


图5 Ensemble ResMaskingNet with 6 other CNNs 模型结构[15]

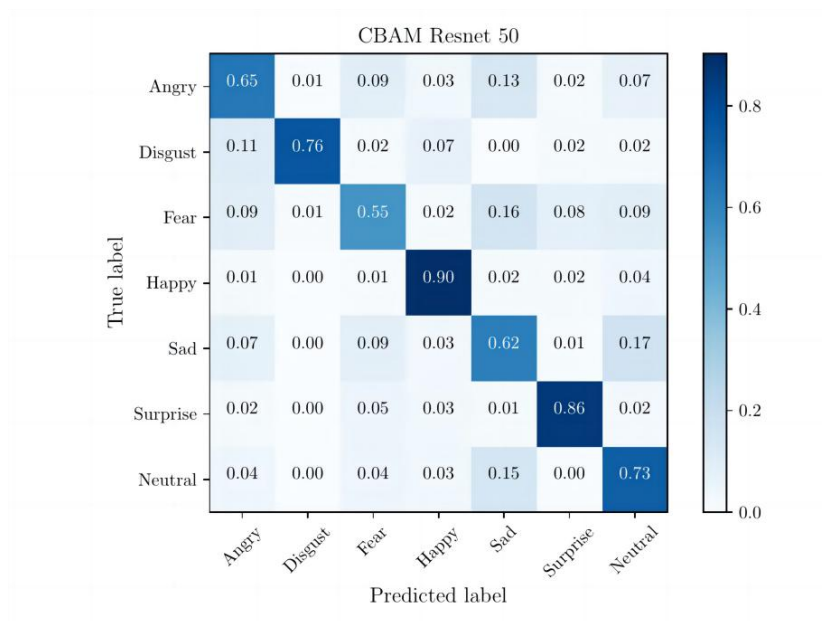


图6 Ensemble ResMaskingNet with 6 other CNNs 混淆矩阵[15]

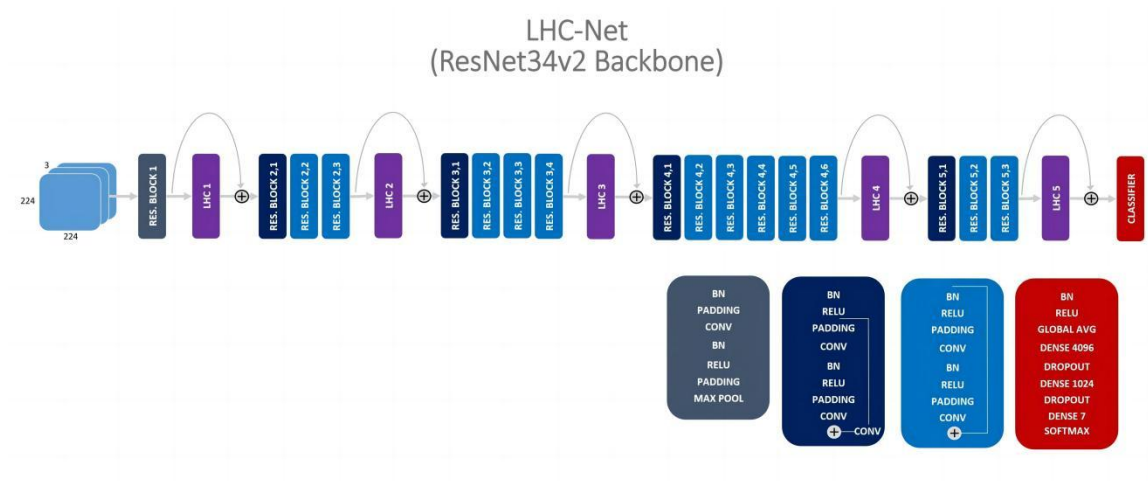


图 7 LHC-Net 模型结构[16]

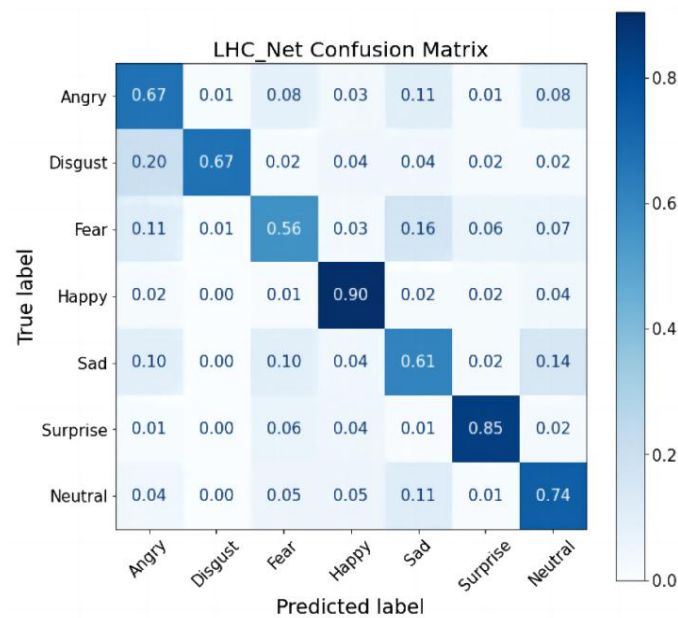


图 8 LHC-Net 混淆矩阵[16]

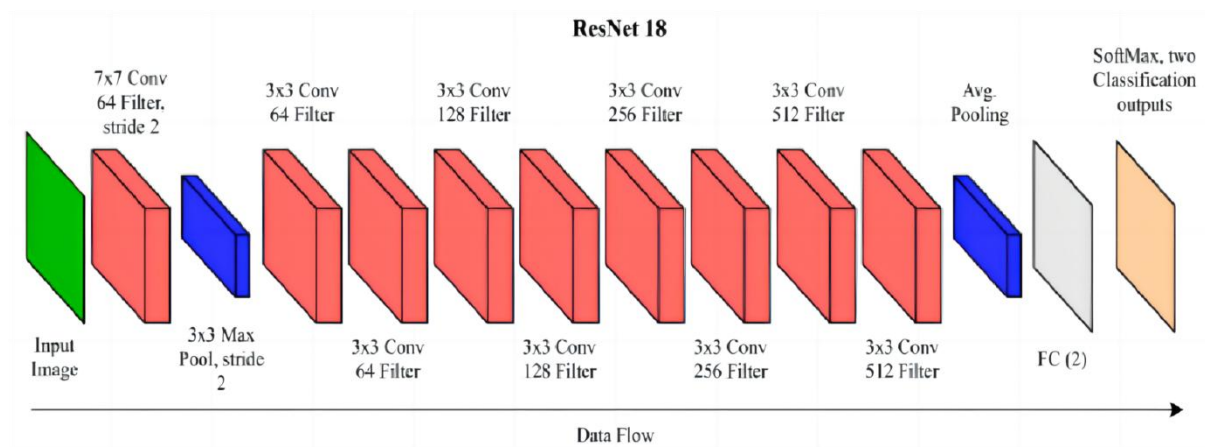


图 9 ResNet18 With Tricks 模型结构[17]

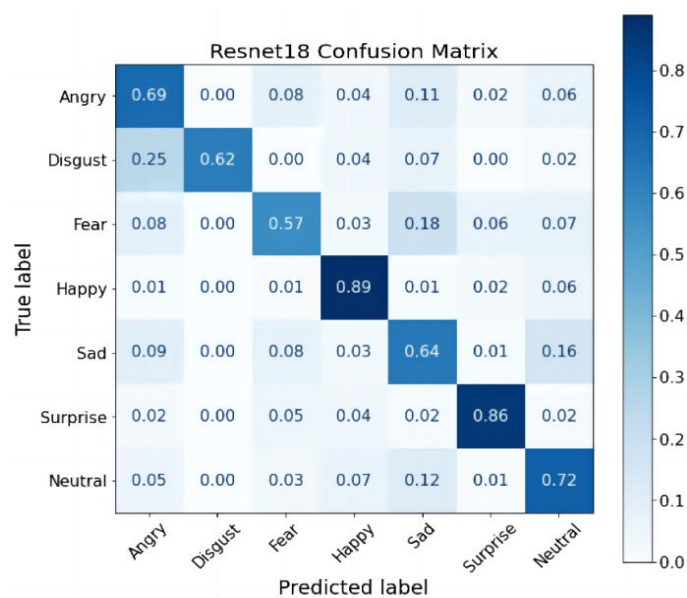


图 10 ResNet18 With Tricks 混淆矩阵[17]

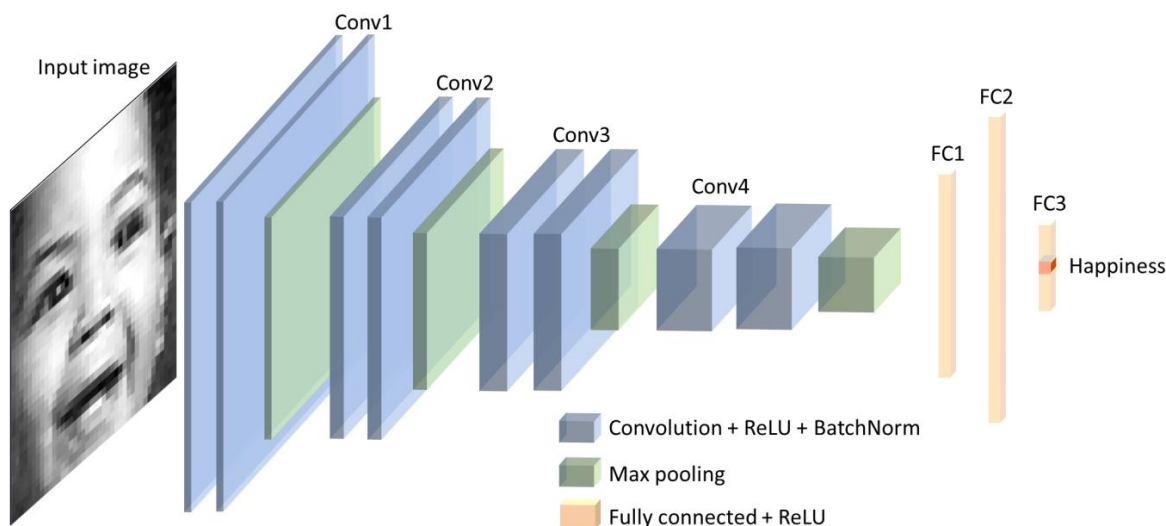


图 11 VGGNet 模型结构[18]

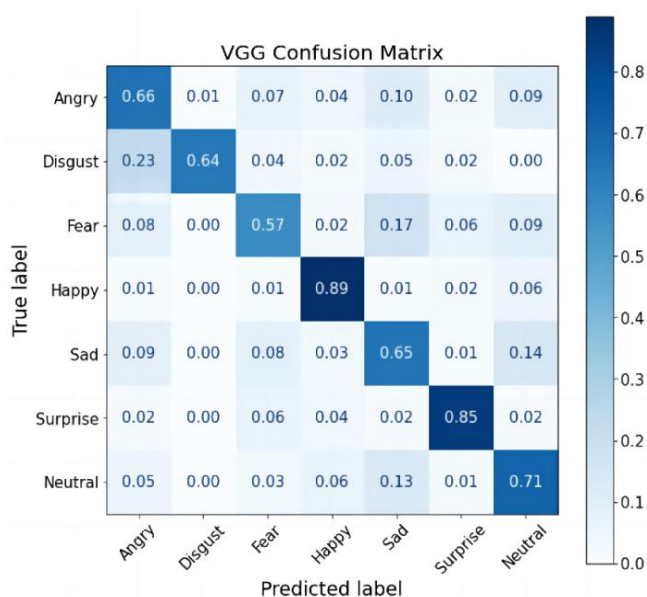


图 12 VGGNet 混淆矩阵[18]

4.3 前端展示

4.3.1 前端性能分析

在本文中，我们采用了几项关键的前端性能指标来评估网页加载性能和用户体验。这些指标包括首屏绘制 (First Paint, FP)、首屏内容绘制 (First Contentful Paint, FCP)、最大内容绘制 (Largest Contentful Paint, LCP) 以及页面加载完成时间 (DOMContentLoaded, DCL)。下面将详细介绍这些指标及其在本项目中的应用和分析。

- 1、首屏绘制（First Paint, FP）：页面开始绘制的时间点。
- 2、首屏内容绘制（First Contentful Paint, FCP）：页面上第一个内容元素绘制完成的时间点。
- 3、最大内容绘制（Largest Contentful Paint, LCP）：页面上最大内容元素绘制完成的时间点。
- 4、页面加载完成时间（DOMContentLoaded, DCL）：初始 HTML 文档完全加载和解析完成的时间点。

我们使用了 Chrome 开发者工具中的性能（Performance）选项卡来测试和分析页面的加载性能。具体操作步骤如下：首先打开 Chrome 开发者工具，选择“性能”选项卡。然后点击“开始录制”，然后刷新页面。停止录制后，查看生成的时间轴图表，获取各项性能指标的具体时间点，如图 13 所示。

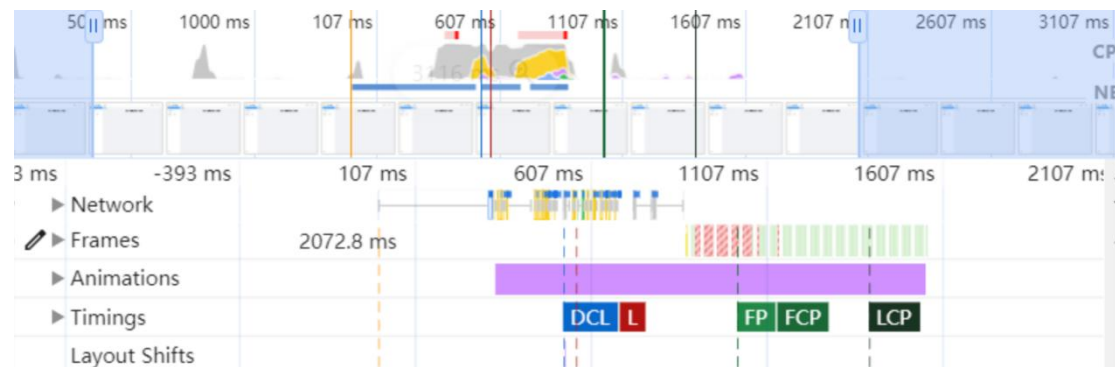


图 13 前端性能指标

根据测试结果，我们得到了以下性能数据：

- 1、首屏绘制：页面初始 HTML 文档加载和解析完成时间为 607 毫秒，显示了页面初始加载速度较快。
- 2、首屏内容绘制：首屏内容绘制时间为 1107 毫秒
- 3、最大内容绘制：时间为 1607 毫秒，表明页面的主要内容在 1.6 秒内加载完成。
- 4、页面加载完成时间：页面加载完成时间为 607 毫秒，显示出页面的初始 HTML 结构加载较快。

本次前端性能测试结果表明，页面加载的初始性能较好，大部分内容在短时间内加载完成。

4.3.2 图片测试展示

本文进行了详尽的图片测试，以展示系统在处理复杂情况方面的能力，特别是针对含有多个人脸且表情各异的图像时的表现。图 14 展示了系统的基本界面及用于测试的原始图片，这些图像被用作评估系统在不同模型下的表情识别性能的基准。

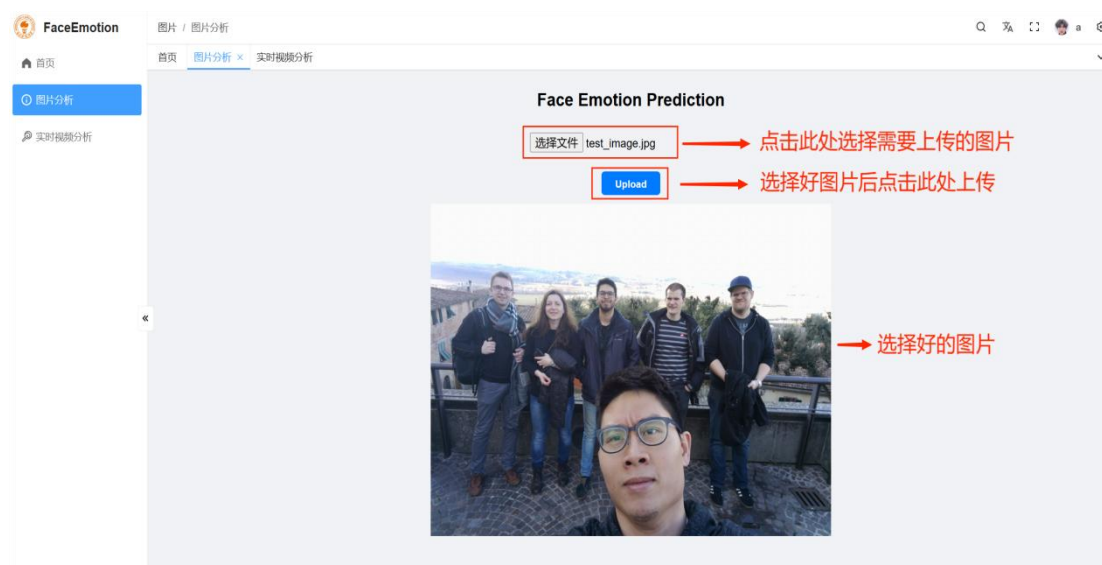


图 14 基本界面和测试原始图片

在测试过程中，本文使用了四种不同的模型对同一张图片进行了测试，以全面评估各模型在面部表情识别上的效能和准确性。以下为各模型的具体测试结果：

图 15 展示了 Ensemble ResMaskingNet 结合其他六种 CNN 模型的测试结果。通过集成学习方法，该模型结合了 ResMaskingNet 与其他六种卷积神经网络，显著提高了表情识别的准确率和鲁棒性。集成学习策略使得模型能够综合利用多个网络的优势，从而在复杂环境中表现出更高的稳定性和精确度。

图 16 展示了 LHC-Net 的测试结果。LHC-Net 特别设计用于处理人脸的局部特征，优化了对复杂表情的解析能力。通过简化网络结构和创新的参数优化策略，LHC-Net 在保持高识别准确率的同时，显著减少了计算资源的消耗，使其非常适用于移动设备和实时应用。

图 17 展示了 ResNet18 With Tricks 的测试结果。该模型在传统的 ResNet18 架构上进行了多项优化，引入了如数据增强和模型微调等技巧，从而在特定表情识别任务上表现出色。这些优化策略提高了模型的泛化能力，使其在处理多样化表

情数据时表现稳定。

图 18 为 VGGNet 的测试结果。作为深度学习领域的经典网络结构，VGGNet 在表情识别任务中同样表现出了强大的性能。通过其深层次的网络结构，VGGNet 能够准确捕捉和识别出人脸表情的复杂特征，从而在不同表情状态下均表现出良好的识别能力。

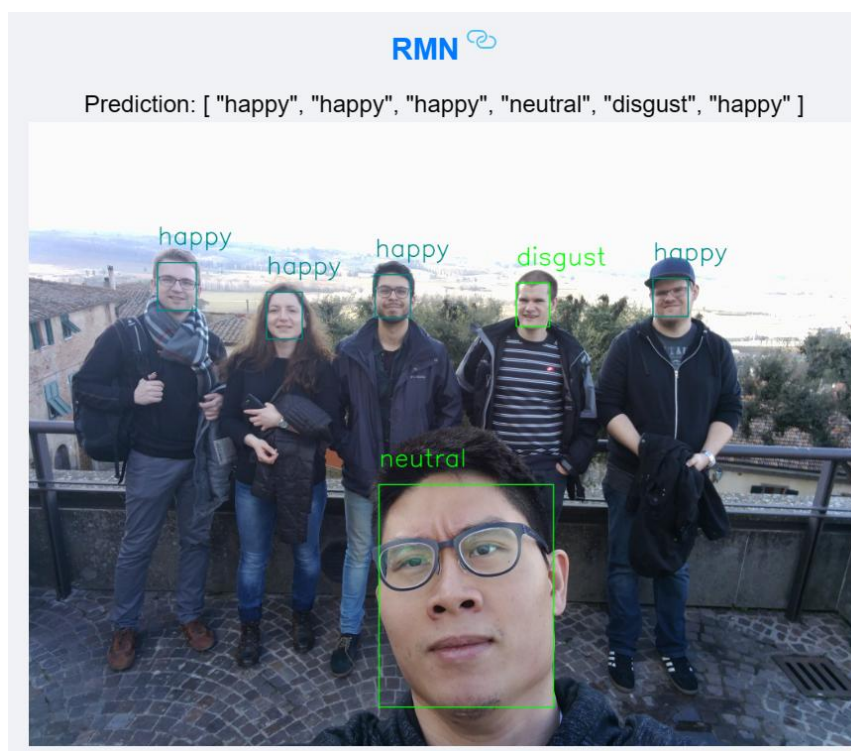


图 15 Ensemble ResMaskingNet with 6 other CNNs 测试结果



图 16 LHC-Net 测试结果

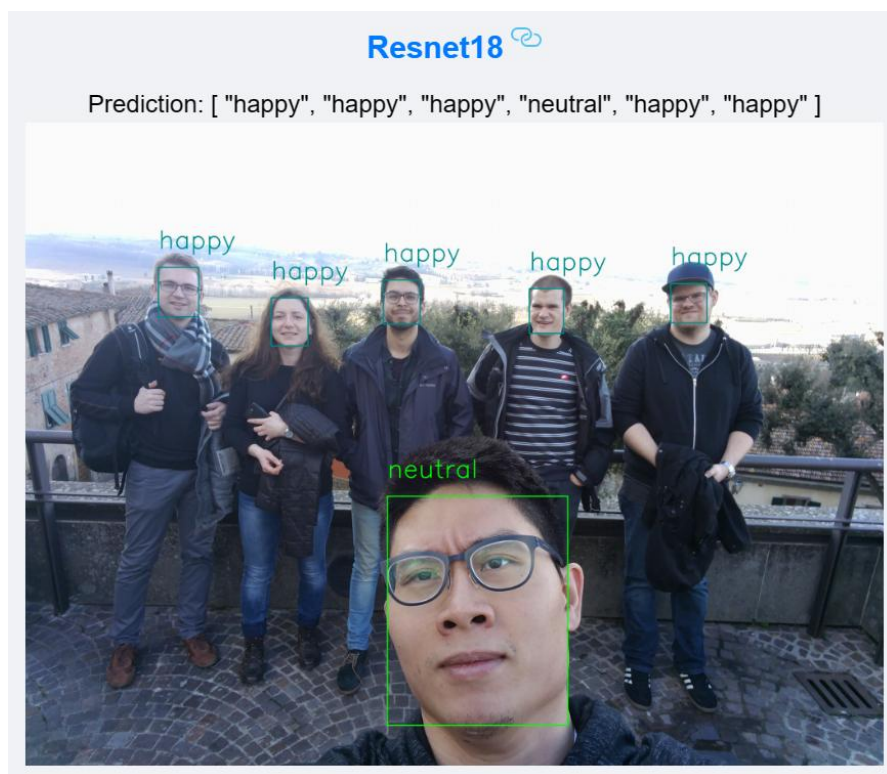


图 17 ResNet18 With Tricks 测试结果

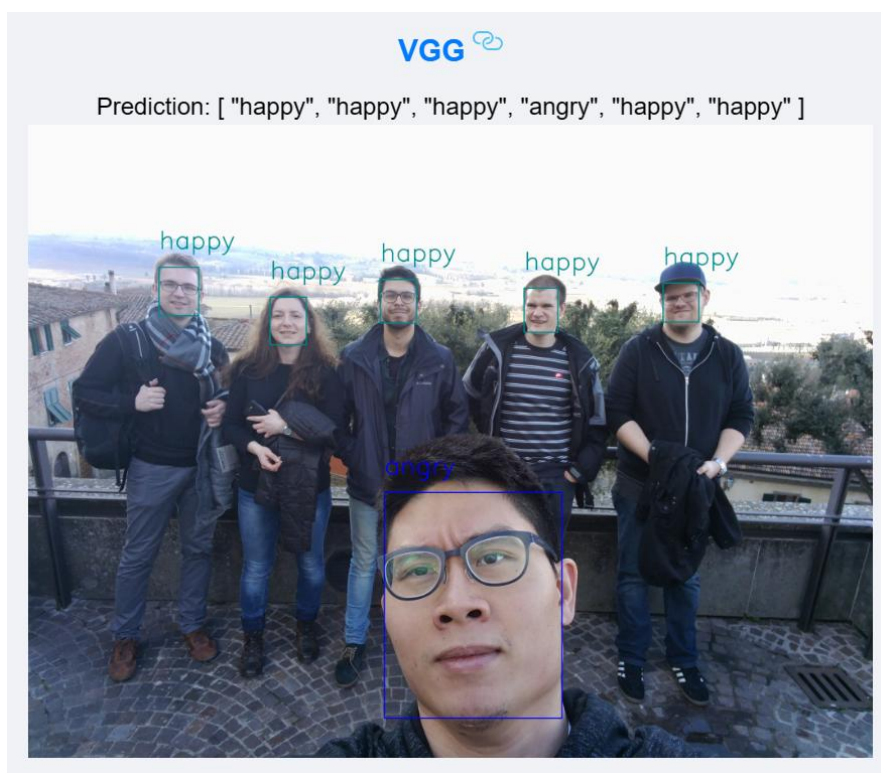


图 18 VGGNet 测试结果

4. 3. 3 实时视频分析展示

鉴于实时视频的动态特性，在静态的论文中无法完全展示视频内容的变化与效果，因此本文选取了笔者使用笔记本电脑摄像头进行实时视频分析测试过程中捕捉的两张截图，以直观体现系统在实际应用中的表现。以下是对这些截图的详细分析和解释：

图 19 为实时视频分析的界面的展示。在该界面中，用户可以实时查看摄像头的图像，并观察到系统对人脸的检测和表情识别后的反馈视频。用户界面设计直观，操作简便，方便用户的使用。

图 20 展示了实时视频分析的初步结果。在这张截图中，可以清晰地看到系统成功检测到了图像中的人脸，并对面部表情进行了初步识别。该结果展示了系统在实时视频流中对人脸进行准确检测和识别的能力，证明了其在动态环境中的实用性。

图 21 展示了实时视频分析的更进一步的结果，验证了系统在连续视频流中的表现稳定性和准确性。通过对不同视频图像的处理，系统能够保持一致的识别性能，准确检测和识别出视频中的人脸表情。



图 19 实时视频分析界面展示

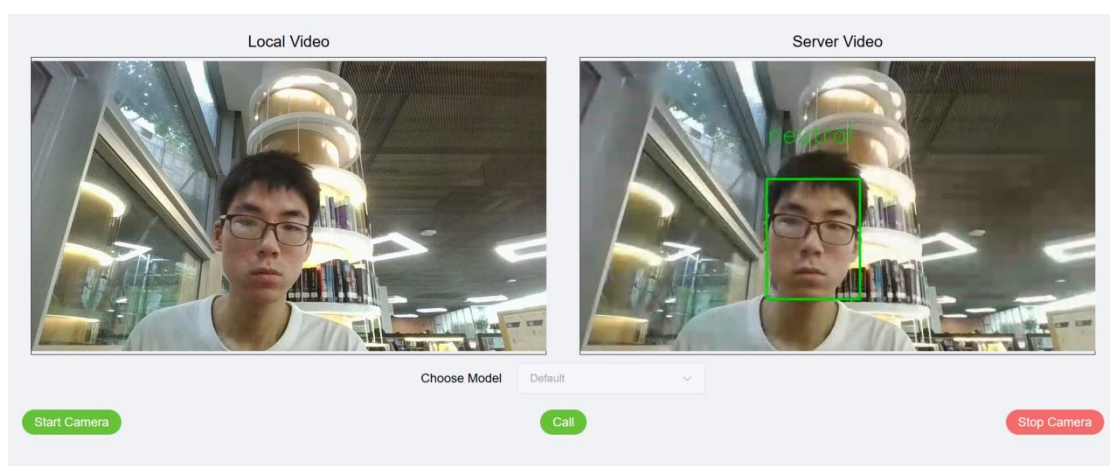


图 20 实时视频分析结果一

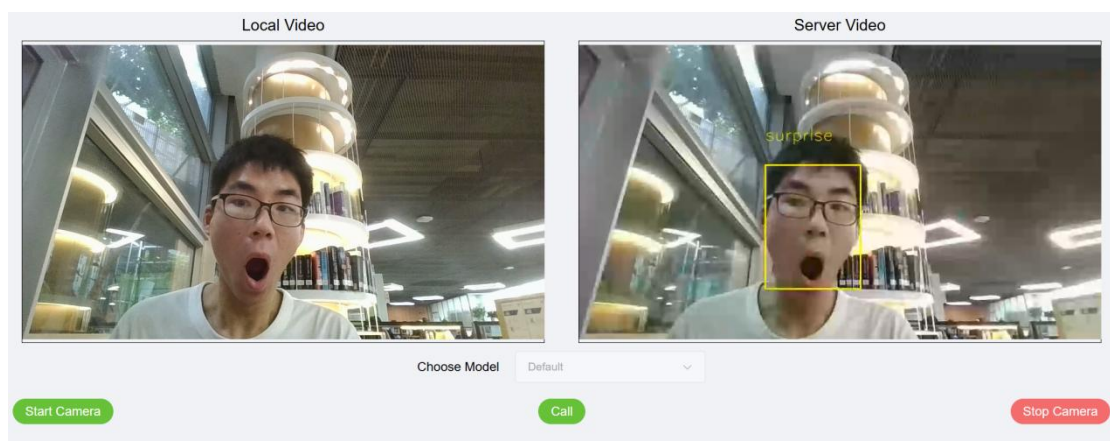


图 21 实时视频分析结果二

通过这些截图及其分析，本文展示了所开发的实时视频分析系统在实际应用中的良好性能。系统能够在动态视频环境中准确地进行人脸检测和表情识别，确保用户获得即时的情感反馈。这不仅证明了系统的技术先进性和实用性，也为其在更多实际应用场景中的推广提供了有力的支持。

5. 总结

通过实验和测试的结果，笔者成功开发了一种基于多种神经网络的人脸表情检测应用。该应用在模型准确率和数据处理效率方面表现出色，充分利用了深度学习技术的优势。同时，通过创新的用户交互设计，该应用不仅实现了高效的表情识别功能，还提供了优化的用户体验，显著提升了系统的易用性和用户满意度。

在测试过程中，应用展示了其在处理复杂表情识别任务中的卓越性能，尤其是在 FER-2013 数据集上的表现尤为突出。然而，本文在研究过程中也发现了一些不足之处。由于笔者在网络通信方面的知识储备不足，导致当前的系统无法在本机以外的设备上实现实时视频分析。这一限制在一定程度上影响了应用的广泛性和实用性。未来研究中，将针对这一问题进行优化，力求实现跨设备的实时视频分析功能，以满足更多用户的需求。

此外，随着技术的不断发展和应用场景的逐步拓展，人脸表情检测应用在未来预计将在人机交互和情感分析等领域发挥更为重要的作用。特别是在社交机器人、患者情感监测、教育等领域，该技术具有广阔的应用前景。因此，未来的研究将进一步探索跨文化表情识别等更多实用功能，旨在提升应用的泛化能力和实用价值。通过对跨文化表情识别的研究，笔者也希望能够开发出适用于不同文化背景下的表情识别模型，以提高系统在全球范围内的适应性和准确性。这不仅能够扩展应用的使用范围，也将为相关领域的研究提供新的数据和技术手段支持。

综上所述，本文开发的人脸表情检测应用在当前的技术背景下展示了强大的性能和广阔的应用前景。尽管存在一些不足，但通过持续的优化和改进，特别是对网络通信和跨文化识别功能的深入研究，本文的成果有望在未来的实际应用中发挥更大的作用，为人机交互和情感分析领域提供更加有效和可靠的解决方案。

参考文献

- [1] Mehrabian, Albert. "Communication without words." [M] Communication theory. Routledge, 2017. 193-200.
- [2] Kahou, Samira Ebrahimi, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video[J]. Journal on Multimodal User Interfaces, 2016, 10: 99-111.
- [3] Dubey, Monika, and Lokesh Singh. "Automatic emotion recognition using facial expression: a review." [J] International Research Journal of Engineering and Technology (IRJET) 3.2 (2016): 488-492.
- [4] 李燕苹, 李一凡. 基于表情识别的课堂状态检测分析 [J]. 工业控制计算机, 2023, 36 (01) : 116-118.
- [5] Rani, Jyoti, and Kanwal Garg. Emotion detection using facial expressions-A review[J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2014, 4(4).
- [6] Bishop, Chris M. Neural networks and their applications[J]. Review of Scientific Instruments, 1994, 65(6): 1803-1832.
- [7] Oarriaga. face_classification[OL]. [2024-05-01].
https://github.com/oarriaga/face_classification.
- [8] atulapra. Emotion-detection[OL]. [2024-05-01].
<https://github.com/atulapra/Emotion-detection>
- [9] PapersWithCode. Facial Expression Recognition on FER-2013[OL]. [2024-04-07].
<https://paperswithcode.com/sota/facial-expression-recognition-on-fer2013>.
- [10] Goodfellow, Ian J., et al. "Challenges in representation learning: A report on three machine learning contests." [C] Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20. Springer berlin heidelberg, 2013.
- [11] Lien, Jenn-Jier James. Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity[D]. University of Pittsburgh, 1998.

- [12] Azcarate, Aitor, et al. Automatic facial emotion recognition[D]. Universiteit van Amsterdam, 2005: 1-6.
- [13] Boughrara, Hayet, et al. Facial expression recognition based on a MLP neural network using constructive training algorithm[J]. Multimedia Tools and Applications, 2016, 75: 709-731.
- [14] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [15] Pham, Luan, The Huynh Vu, and Tuan Anh Tran. "Facial expression recognition using residual masking network." [C] 2020 25Th international conference on pattern recognition (ICPR). IEEE, 2021.
- [16] Pecoraro, Roberto, Valerio Basile, and Viviana Bono. "Local multi-head channel self-attention for facial expression recognition." [J] Information 13.9 (2022): 419.
- [17] Yuan, Xiaojian. Fer2013-Facial-Emotion-Recognition-Pytorch[OL]. (2021). [2024-04-07]. <https://github.com/LetheSec/Fer2013-Facial-Emotion-Recognition-Pytorch>.
- [18] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." [OL] arXiv preprint arXiv:2105.03588 (2021).
- [19] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." [J] IEEE signal processing letters 23.10 (2016): 1499-1503.
- [20] Cpolar. Cpolar: Secure tunnels to localhost[OL]. [2024-05-01]. <https://www.cpolars.com/>.

致谢

在毕业设计的工作中，我要感谢周围的同学、导师与学校。我很幸运能够拥有优秀的导师和在良好的学术环境下进行我的毕业设计。感谢史玉回教授、赵琪博士对本文的指导和建议，他们的指导对我的研究工作起到了巨大的帮助。我会更加努力工作、学习和研究，为更大的目标努力奋斗，希望能够在未来的研究中，尽可能地回馈社会。