

德州学院计算机与信息学院

课程设计报告

实习名称 人工智能应用开发课程设计

设计题目 基于改进 CBAM 注意力机制风扇异常状况的识别

实习时间 2024.12.23--2024.12.29

专业班级 22 软件（智能开发）本 2

指导老师 赵一

教学单位 计算机与信息学院 （盖章）

小组分工情况：

学号	姓名	分工
202201011070	刘明	代码实现
202201011083	聂睿	文档撰写
202201011052	李璐宇	代码纠错
202201011065	刘高政	数据集处理
202201011068	刘凯	实验结果测试

二 0 二 四 年 六 月 二 十 三 日

目录

一、设计目标..... 1

二、设计内容..... 1

 2.1 模型搭建..... 1

 2.2 Conv_Block 2

 2.3 改进 CBAM 注意力模块..... 2

 2.4 反向残差模块..... 4

 2.5 主网络架构..... 4

 2.6 数据处理与增强..... 5

 2.7 迁移学习应用..... 6

 2.8 实验数据..... 7

 2.9 实验结果..... 8

三、技术要求..... 10

四、功能要求..... 11

五、评估标准..... 11

参考文献： 12

基于改进 CBAM 注意力机制风扇异常状况的识别

一、设计目标

本研究旨在通过改进的 CBAM（Convolutional Block Attention Module）注意力机制，结合 MobileNetV3 模型，提高工业风扇异常状况识别的准确性和效率。通过分析在-6dB 噪声环境中故障风扇发出的声音，提取声音样本的语谱图，并利用深度学习技术，实现对风扇异常状况的快速、准确识别，以保障工业生产的安全和效率。

Cong Li(2016)等人利用风扇速度测量值高度相关的经验结果设计了风扇的异常检测方法，在不同的环境下均取得了较高的准确率 Error! Reference source not found.。但是仅通过风扇速度无法全面了解风扇的运行状态，例如，风扇可能会因为轴承磨损、电机故障等问题而产生异常声音，这些问题可能不会影响风扇的速度。利用现有工厂传感器作为数据来源，使用 RELAP5-3D/LSTM 模型对数据进行异常检测，取得了较高的准确率，但是在风扇上直接安装传感器可能会对机器造成一些影响，导致产生异常数据。Qian Yang 等人(2015)用自相关函数指标（ACFI）来确定连续小波变换（CWT）的系数 Error! Reference source not found.。通过该方法得到的频谱可以直观地描述风扇故障的过程和范围。由于该方法需要选择合适的小波函数和尺度，否则可能会导致故障特征频率的误判，因此不能应对多种风扇故障情况。

针对上述问题，本研究使用 MIMII 数据集（Purohit 等人，2019）作为数据来源，对声音进行特征提取得到语谱图作为输入数据对风扇进行故障检测。只需要一个声音采集设备，就可以进行风扇故障的诊断，操作简单，方便快捷，无需打开或接触风扇本身，不会对设备造成任何损伤。X. Chu 等人(2020)采用 MoblieNetV3 作为基本模型，该模型通过引入倒残差(Bottleneck)和 Hard Swish 激活函数等，有效提高了模型在资源受限环境下的性能和效率，采用了深度可分离卷积和特征压缩等优化策略，使得模型在推理阶段拥有更快的速度和更低的延迟 Error! Reference source not found.。本研究使用迁移学习，利用预训练模型的知识 and 泛化能力，提高模型的性能并加速训练过程，有效应对数据量不足的情况。另外，MobileNetV3 可以通过改变模型的分辨率、宽度和数量等参数来适应不同的硬件和应用需求，可对应多种风扇故障情况。

二、设计内容

2.1 模型搭建

本研究对 MobileNetV3 模型中的 SE（Squeeze-and-Excitation）注意力模块进行了改进，采用空洞卷积优化的 CBAM 模块替代，以增强模型对语谱图特征的识别能力。

MobileNetV3 结合了挤压-激励块和反转残差等技术，结合改进的 CBAM 以提高模型分类能力，如图 1。

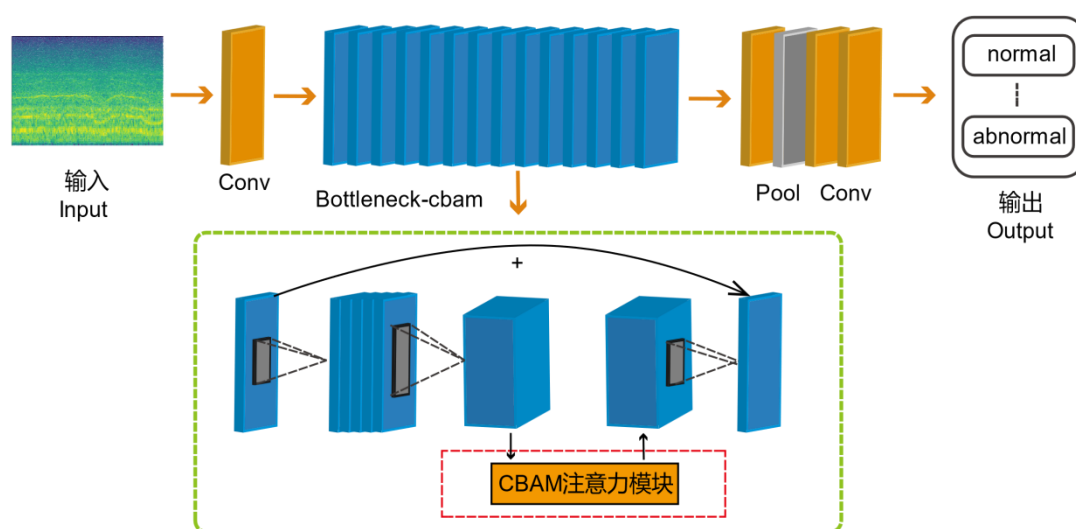


图 1 模型整体网络结构^[4]

2.2 Conv_Block

卷积块（Conv_Block）是神经网络中的基本组成部分，由卷积、批量归一化（Batch Normalization）和激活函数 ReLU/Hardswish 组成，如图 2。卷积是对图像进行特征提取的主要步骤，批量归一化能使网络训练更快更稳定，激活函数能引入非线性特性使神经网络能拟合复杂问题。

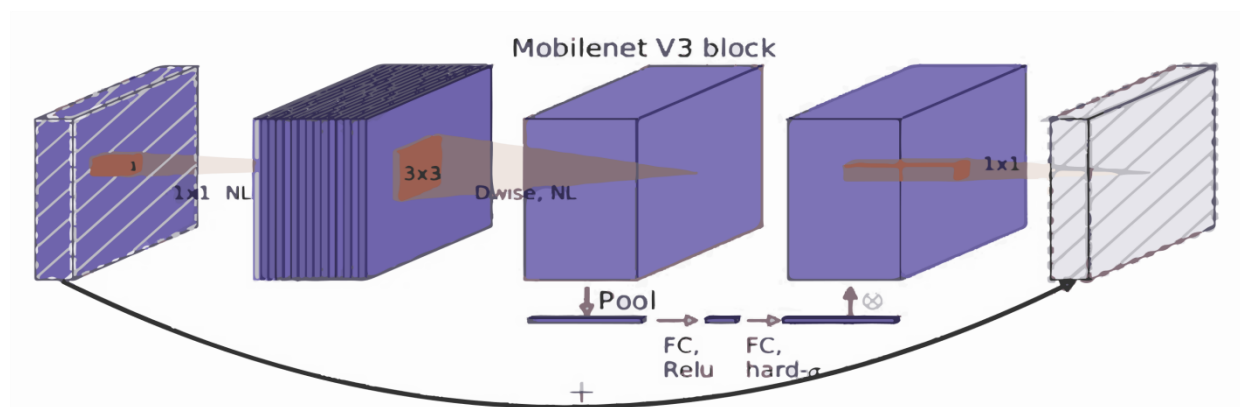


图 2 MobileNetV3 block 结构

2.3 改进 CBAM 注意力模块

在本研究中，基于 CBAM 注意力机制的改进是其核心创新点之一。然而，为了增强

研究的系统性和多样性，本研究参考了以下几项使用 MobileNetV3 的最新研究成果，并将其与本研究的方法进行了对比。

Jia, L.等人（2021）结合了 MobileNetV3 和 CBAM 模块，用于解决竹条计数问题^[5]。他们的方法重点优化了 CBAM 的通道注意力部分，通过减少冗余特征，提高了模型的精度和速度。在对比中，本研究发现他们的优化策略更适用于视觉特征中存在重复性较高的任务。

Zhang 等人（2023）针对车道和车辆的同时检测，提出了基于 MobileNetV3 的改进模型，并引入动态通道调整机制^[6]。通过结合 CBAM 模块，提升了特征提取在复杂场景下的鲁棒性。相比之下，本研究的模型重点在于声音信号的特征提取，而非视觉信号，这使得本研究采用空洞卷积改进的 CBAM 模块具有更广泛的上下文捕获能力。

Park 等人（2023）提出了一种改进的 MobileNetV3 模型，结合了轻量化的注意力机制以降低计算复杂度，同时保留了模型的高精度^[7]。在他们的模型中，CBAM 模块进行了简化，减少了通道注意力和空间注意力的计算量。本研究的研究在背景噪声较大的环境下仍保持高精度，而他们的模型更适用于计算资源有限的场景。

基于上述研究，本研究的改进 CBAM 方法具有以下对比优势：

采用空洞卷积替代传统 CBAM 的空间注意力模块，使得模型对声音特征的提取更加细致，适用于语谱图中稀疏信息的识别。本研究重点在工业噪声环境中的风扇异常检测，而上述方法更多地专注于视觉或计算资源优化任务。本研究在迁移学习和数据增强的基础上，利用 MobileNetV3 的硬件适应性，针对风扇故障声音数据进行了特定优化。

"Channel Attention"和"Spatial Attention"是 CBAM(Convolutional Block Attention Module)的两个核心部分。

Channel Attention: 通道注意力模块用于捕获特征通道之间的依赖关系。它的操作方法是先通过全局平均池化和全局最大池化来获得每个通道的描述，然后通过 MLP（Multi-Layer Perceptron，多层感知器）操作，得到每个通道的权重。最后，通过这些权重来重新调整每个通道的特征。

Spatial Attention: 空间注意力模块是用来捕获特性在空间上的依赖性。该模块首先对输入特性的每个通道进行平均池化和最大池化以获得两个二维的特性图，表示每个位置在所有通道上的平均响应和最大响应。然后把这两个图在通道维度进行连接并通过一个卷积操作，获得最终的空间注意力图。这个注意力图会被用来调整输入特性的每个位置

的响应。

本研究提出了将空间注意力模块（Spatial Attention）替换为空洞卷积（Dilated Convolution）的新方法。Dilated Convolution 提供了更广泛的视野，使得模型可以捕获更大范围的上下文信息。改进后的空间注意力模块包括了三个卷积层，两个 1×1 的卷积层用于降维和升维，中间的一个卷积层用于进行空间的特征提取，如图 3。

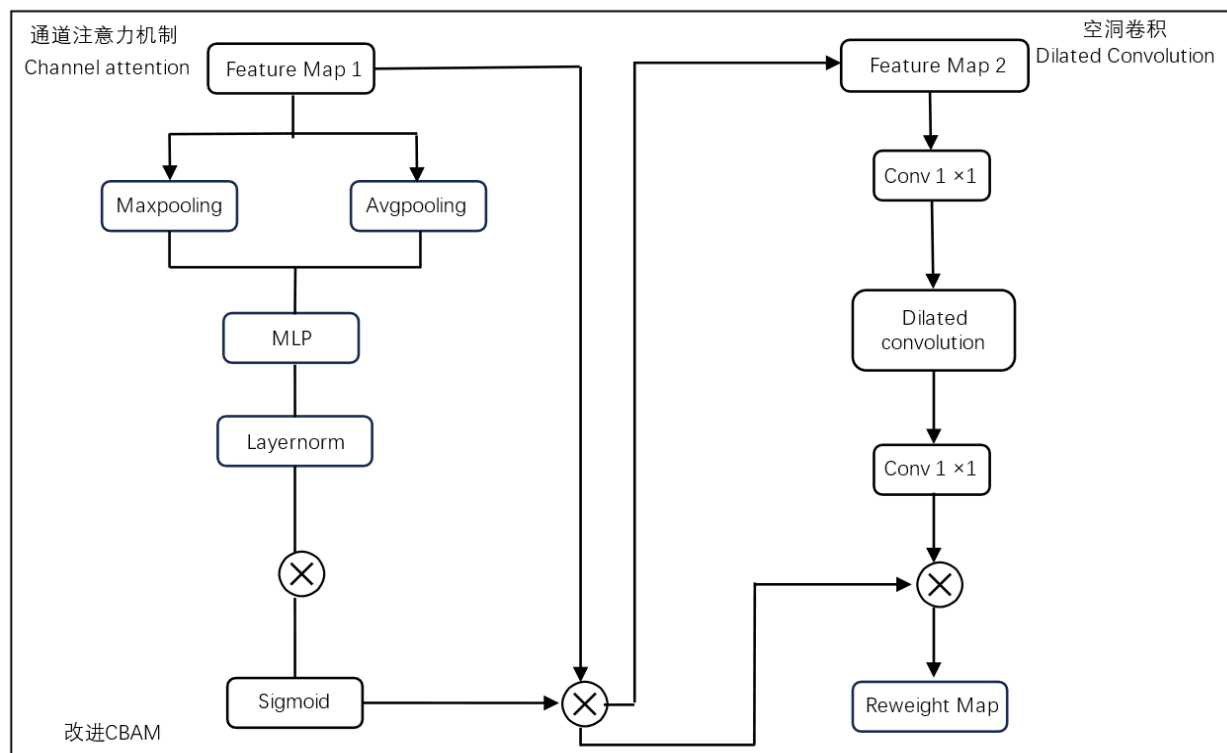


图 3 改进后的 CBAM 结构图

2.4 反向残差模块

反向残差模块是 MobileNetV3 的核心部分，它使用 1×1 卷积进行通道升维，然后使用 3×3 深度可分离卷积进行特征提取，最后再使用 1×1 卷积进行通道降维，间接使用批量归一化和激活函数进行非线性变换和特征归一化。

每个反向残差模块包含一个可选的 CBAM_Block，使得模型能够根据每个通道和空间位置的特征重要性进行自我调整，这种自适应性的特性增强了网络的预测性能。

2.5 主网络架构

MobileNetV3 由一个卷积块，14 个反向残差模块，一个卷积块，一个全局平均池化层和分类器组成，如图 4。全局平均池化层用于减少参数数量和防止过拟合，分类器用于输出对应的类标签。

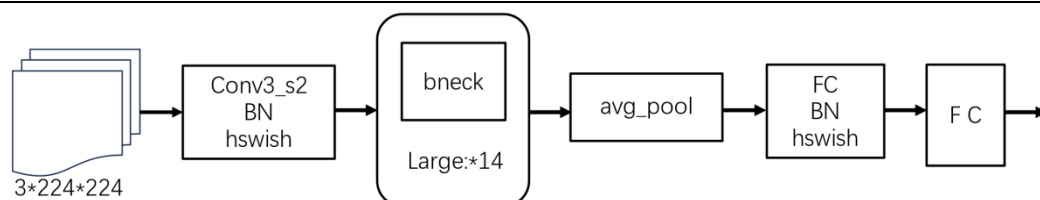


图 4 MobileNetV3 的网络结构图

2.6 数据处理与增强

使用 MIMII 数据集进行预处理和数据增强，包括调整语谱图的亮度、进行波纹位移、添加噪声和高音修正等，以提高模型的泛化能力和鲁棒性。

2.6.1 语谱图 (Spectrogram)

在机器学习中，首先需要提取出适合模型学习的特征。对于图像数据，例如 28x28 的图片，可以将每个像素视为一个特征，其对应的像素值大小表示颜色强度，作为特征值。这样的特征提取方式将图像转换为特征向量，使得模型可以从中学习到图像的特征信息。

Greenberg 等人通过语谱图 (Spectrogram) 将音频数据转换为二维图像，如图 5，能够更直观地表现音频信号的时频特征 **Error! Reference source not found.**。这样的数据表示形式有助于模型更好地理解 and 捕捉音频信号的特征，提高了模型对音频内容的识别和理解能力。其次，语谱图能够减少数据的维度，从而降低了模型的复杂度和训练的计算成本，使得模型更容易进行训练和部署。故本研究采用语谱图作为数据集进行模型训练。

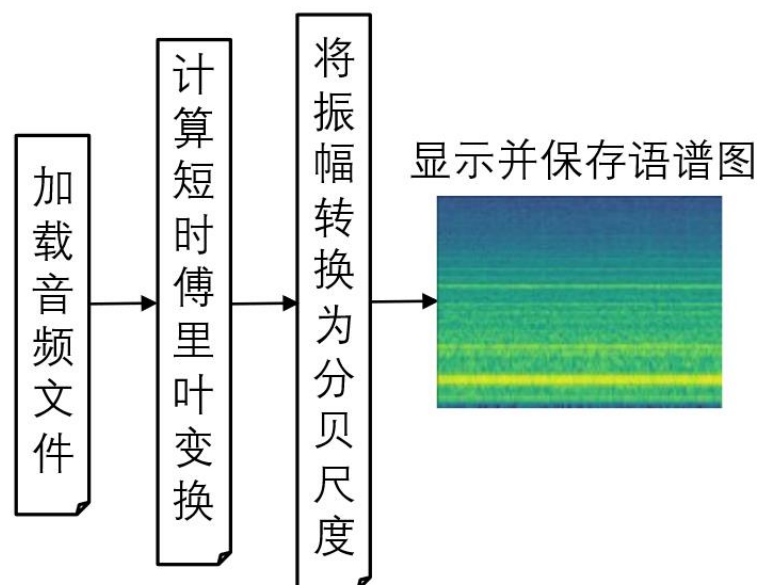


图 5 示意图

2.6.2 数据增强

数据增强是通过对原始数据集进行一系列改变和变换，生成新的数据集来增加训练数据的数量和多样性。数据增强旨在减少过拟合，提高模型的泛化能力，同时可以帮助模型更好地学习数据的特征和模式。

本研究对部分语谱图整体明亮度调低，对部分音频文件进行波纹位移，添加噪声，高音修正，效果如图 6。

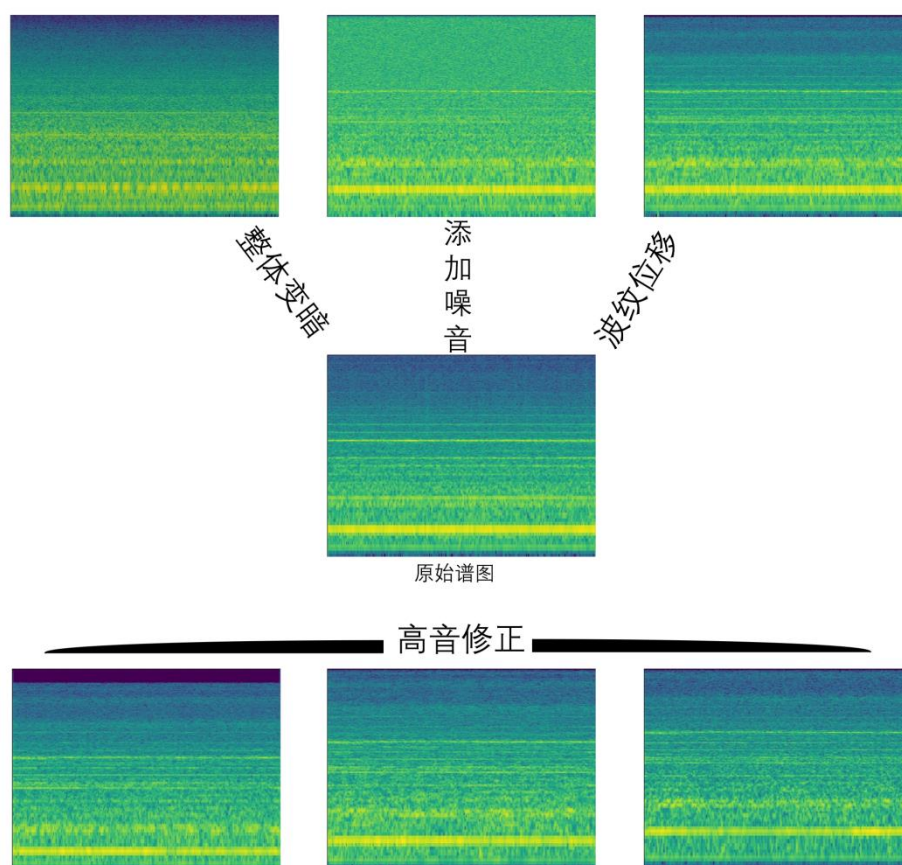


图 6 数据增强示意图

2.7 迁移学习应用

利用预训练的 MobileNetV3 模型，通过迁移学习技术，对模型的全连接层进行重新构建，并使用目标风扇的音频数据对模型进行微调，以适应新的分类任务。

传统深度学习侧重在优化原有结构，专注于神经网络结构上的突破。迁移学习更趋向于在已有网络结构的基础上，研究领域适应性问题。通常将源域（Source Domain）与目标域（Target Domain）分别表示为 D 和 T ； χ 与 y 分别表示特征空间与标签空间； $P(X)$ 表示边际概率分布，其中 $X=\{x_1, k, x_n\} \in \chi$ ； $f(\cdot)$ 表示，目标预测函数，迁移学习即借助有标注的 D 来学习 T 的知识，公式如下：

$$D = \{\chi, P(X)\} \quad (1)$$

$$T = \{y, f(\cdot)\} \quad (2)$$

深度迁移学习的基本思想将一个已经在大规模数据集上训练过的深度神经网络模型（源模型），应用于另一个相关任务的学习过程。这个过程中，源模型的一些或全部层被保留下来，并被用于初始化目标任务的新模型。这种方法可以帮助解决许多实际问题，尤其是当新任务的数据集非常小的时候。

本研究的实验模型在 MIMII 数据集（Purohit 等人，2019）上进行预训练，该数据集包涵各种异常情况下，如旋转不平衡、泄漏、污染等，记录泵、风扇、阀门和滑轨的健康状态音频及其故障状态 **Error! Reference source not found.**。他们被分为健康和 unhealthy 两类。将预训练得到的模型参数作为网络的初始化参数，之后依据经验定制全连接层，以此使感受野得到增大，进而学习到有风扇故障声音语谱图像更高细粒度特征模板。

通过设置不同的实验环境和参数，对改进后的 MobileNetV3 模型进行训练和测试，评估模型在风扇异常识别任务上的性能。

本实验采用 Python 编程语言，GPU 处理器为 NVIDIA GeForce GTX 3080Ti，内存为 16GB，操作系统为 ubuntu20.04，深度学习框架为 Pytorch。

2.8 实验数据

本研究使用 MIMII 数据集（Purohit 等人，2019）作为预训练模型的数据集。在各种异常情况下，例如旋转不平衡、泄漏和污染，记录了泵、风扇、阀门和滑轨的健康状态音频以及其故障状态，将它们分为健康和 unhealthy 两类。在距离风扇 500 毫米处放置了一个圆形麦克风阵列，其配备了八个独立麦克风。数据以 10 秒的文件形式记录，采样频率为 16kHz，使用 16 位音频信号。除了目标机器的声音外，还记录了几个真实工厂中的背景噪声，并将其与目标机器的音频相结合，以模拟自然环境。为了提高模型在各种环境下的准确率，本研究选择最具挑战性 -6db 的背景噪声。

表 1 数据集来源

Name	Audio duration(s)	Class	Number
MIMII Dataset	<=10	0_dB_fan	5550
MIMII Dataset	<=10	-6_dB_fan	5550
MIMII Dataset	<=10	6_dB_fan	4713

2.8.1 参数设置

选用改进 CBAM 注意力机制的 MobileNetV3 网络结构可以使得模型的精度和效率达到最优效果，同时也不会过度消耗浪费计算机资源。经多次实验后，得到了本模型的最优参数组合，设置图像标准化 Norm_size 为 224，批量大小 Batch_size 为 32，初始学习率为 0.01，共迭代了 Epoch 201 次，同时为进一步降低资源消耗，本研究在验证集损失 Best_loss 小于 2 时保存再保存权重。

2.9 实验结果

根据常用性能评价指标，准确率越高，模型识别越准确，越能有效避免风扇故障未被识别出的情况；损失越低，可说明模型鲁棒性的强大。实验结果表明：本实验最终验证集准确率、和损失率分别为 0.9879 和 0.0604，网络收敛性良好，如图 7。

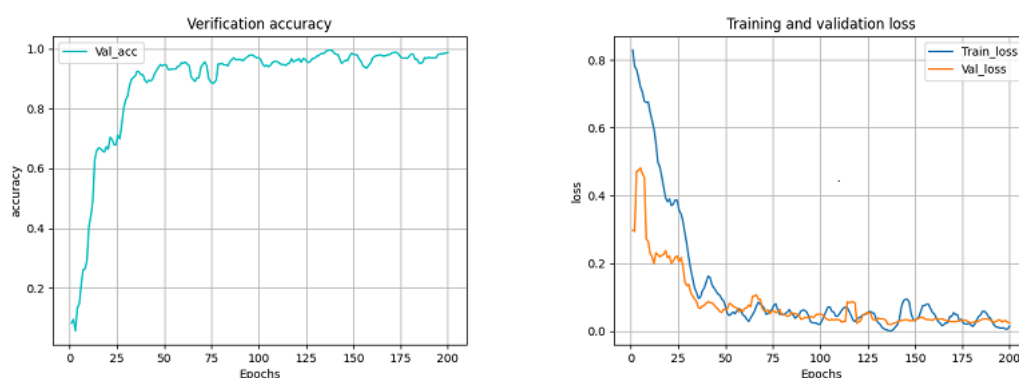


图 7 模型准确率和损失率

2.9.1 结果分析

本实验为了保证数据的独立性，避免某张图片扩充之后同时在训练集和测试集出现，从而生成虚假的准确度，将每类语谱图按照 7: 2: 1 随机划分为训练集、验证集、测试集，即训练集图像共 2020 张，验证集图像共 580 张，测试集图像共 348 张。为了更准确地看出识别效果，本研究对测试集上的图像进行预测，并用混淆矩阵查看预测结果，如图 8。

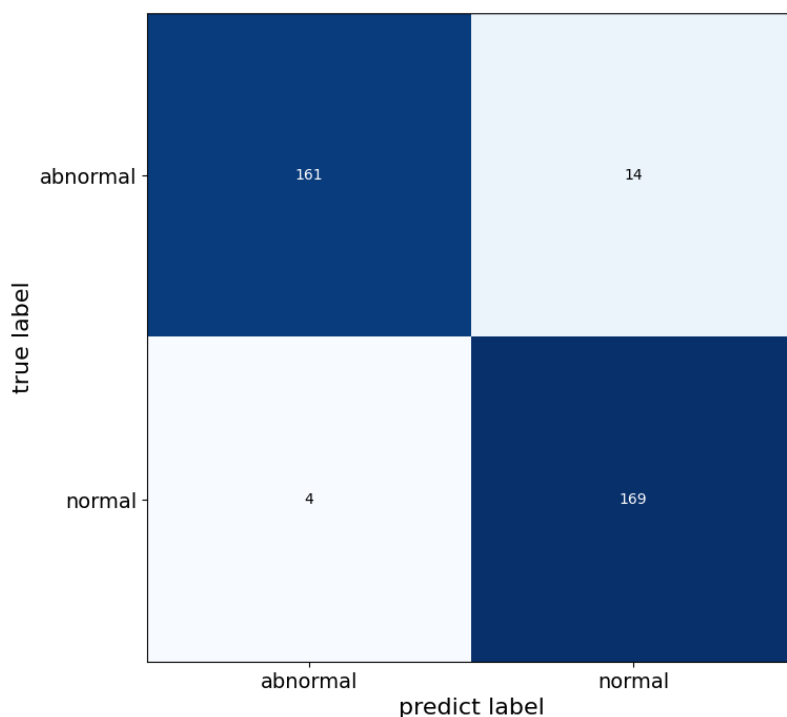


图 8 混淆矩阵

从图中可以看出，在测试集上，对于正常和不正常的风扇声音语谱图图像，准确率均达到 98%左右，可见本模型对于风扇声音语谱图图像分类效果理想。

相较于未改进的 MobileNetV3 模型和其他深度学习网络模型如 MobileVit，CNN，ResNet50，Conformer 对比数据见表 2^{Error! Reference source not found.}。

表 2 MIMII 数据集上不同模型的性能指标

Model	输入特征	Accuracy
改进 MobileNetV3	Spectrogram	0.9706
MobileVit	Spectrogram	0.9436
ResNet50	Log-mel energies	0.8787
Conformer	Log-mel energies	0.9029
CNN	Log-mel energies	0.9228
Transformer	Spectrogram	0.9217

在模型训练过程中，可以看出本模型改进的 CBAM 注意力相较于其他注意力机制准确率更高，例如 SOCA，SimAm，ECA，NAM。对比数据见表 3。

表 3 数据集上不同注意力的性能指标

注意力	输入特征	Accuracy	Loss
改进 CBAM	Spectrogram	0.9879	0.0604
ECA	Spectrogram	0.9758	0.3328
SCOA	Spectrogram	0.9692	0.0637
SimAm	Spectrogram	0.9689	0.0423
NAM	Spectrogram	0.9672	0.2736

本研究对比了四种音频特征提取方法 Spectrogram, Melspectrogram, Log-Frequency-Power-spectrogram, Liner-Frequency-Power-spectrogram。如图 9, 经过实验后发现 Spectrogram 谱图对模型有较高的准确性, 如表 4。

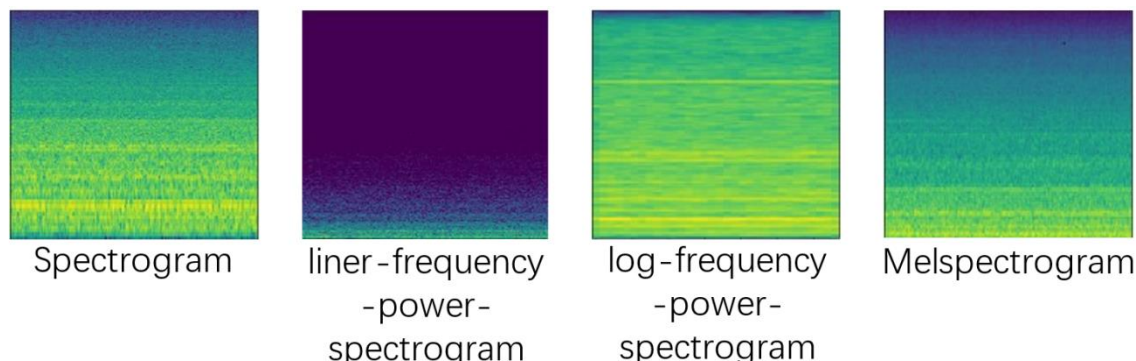


图 9 语谱图

表 4 特征对比

Model	输入特征	Accuracy	Loss
改进 MobileNetV3	Spectrogram	0.9879	0.0604
改进 MobileNetV3	Melspectrogram	0.9620	0.1152
改进 MobileNetV3	Log-Frequency-Power-spectrogram	0.9787	0.0595
改进 MobileNetV3	Liner-Frequency-Power-spectrogram	0.9448	0.1631

三、技术要求

1. **编程语言:** 项目采用 Python 作为编程语言, Python 因其丰富的生态系统和简单易用的语法而成为深度学习任务的主流选择。深度学习框架选用 Pytorch, Pytorch 以其动态计算图的特性、直观的调试能力和丰富的第三方工具库支持, 使其适合进行快速模型开发和优化实验。

2. **数据增强:** 为了提高模型的泛化能力和鲁棒性, 需对数据集进行多种增强操作: 波纹位移: 对音频数据进行微小的波形偏移, 模拟风扇音频信号中的相位偏移现象。添加噪声: 加入环境噪声 (如白噪声或工业现场背景噪声), 使模型能够在实际工业环境中表现出色。高音修正: 通过调整音频信号的高频部分, 增强音频特征对异常信号的敏感性。

3. **特征提取:** 音频数据的特征提取采用 Spectrogram 方法, 通过将时域音频信号转换为频域特征, 使模型能够更高效地捕获音频信号中的关键信息。使用短时傅里叶变换 (STFT) 生成谱图, 确保在时间和频率分辨率之间达到平衡。

4. **音频分类:** 分类模型基于 MobileNetV3 进行改进, 引入 CBAM 注意力机制以增强

模型对关键特征的提取能力。结合空洞卷积优化 CBAM 模块，扩大感受野，提高模型对异常信号的检测精度和效率。

四、功能要求

1. **噪声环境处理能力：**模型应能够识别并处理在-6dB 噪声环境下的风扇异常声音，体现其在复杂工业环境中的适应性。针对工业环境中常见的低信噪比情况，通过引入多样化的噪声增强数据和设计鲁棒的特征提取模块来提升模型的性能。

2. **高准确率分类能力：**模型应具备良好的分类能力，能够在正常和异常风扇声音之间实现高效区分。模型训练和验证阶段需严格监控准确率指标，确保模型的可靠性。

3. **灵活性和可扩展性：**模型设计需兼顾轻量化和高性能，能够部署于不同硬件环境（如嵌入式设备、服务器等）。使用模块化编程结构，方便后续功能扩展或迁移至其他音频分类任务。

五、评估标准

1. **功能完成度：**模型在实际工业环境中能够准确识别异常风扇声音，且在-6dB 噪声条件下仍具有较高的识别准确率，满足应用需求。

2. **准确率：**在实验验证集上的准确率应达到或超过 0.9879，损失率应低于 0.0604；以此为衡量标准，确保模型的性能和稳定性。

3. **代码规范：**项目代码需遵循清晰的结构设计，确保模块分工明确，逻辑清晰。提供详细的注释和文档说明，便于团队协作和后续维护。

4. **创新性：**将 CBAM 注意力机制与 MobileNetV3 结合，提高了模型对关键特征的提取能力。通过引入空洞卷积优化 CBAM 模块，增强了模型的感受野，有效提升了对风扇异常声的分类能力。利用迁移学习技术，充分利用预训练模型的知识，加速收敛并提高小样本情况下的模型性能。数据增强策略的创新应用（如高音修正）进一步增强了模型的泛化能力，使其能够适应不同工业环境中的变化和挑战。

参考文献：

- [1] Cong Li, A. Daniel and N. Ahuja, "A new approach to predict fan failures with fan speed correlation," 2016 32nd Thermal Measurement, Modeling & Management Symposium (SEMI-THERM), San Jose, CA, 2016, pp. 90-94, doi: 10.1109/SEMI-THERM.2016.7458451.
- [2] Qian Yang, Qiang Yang, Wenjun Yan and Tiankun Wang, "PSO based LS-SVM approach for fault prediction of primary air fan," 2015 Chinese Automation Congress (CAC), Wuhan, 2015, pp. 75-80, doi: 10.1109/CAC.2015.73824722.
- [3] X. Chu, B. Zhang and R. Xu, "MoGA: Searching Beyond Mobilenetv3," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 4042-4046, doi: 10.1109/ICASSP40776.2020.9054428.
- [4] 刘明, 王荣燕, 王汝旭等人《基于改进 CBAM 注意力机制的 MobileNetV3 风扇异常状况识别研究》编号:32106, 《工业控制计算机》2025（待发表）
- [5] Jia, L., et al. MobileNetV3 with CBAM for Bamboo Stick Counting [J]. IEEE Journals & Magazine, 2021.
- [6] Zhang, F., et al. Simultaneous Vehicle and Lane Detection via MobileNetV3 in Car Following Scene [J]. PLOS ONE, 2023.
- [7] Park, J., et al. (2023). Parking Lot Occupancy Detection with Improved MobileNetV3. *Sensors*, 23(17), 7642. DOI: 10.3390/s23177642.
- [8] Greenberg, S. , and B. E. D. Kingsbury . "The modulation spectrogram: in pursuit of an invariant representation of speech." IEEE International Conference on Acoustics IEEE, 1997.
- [9] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi, "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection," arXiv preprint arXiv:1909.09347, 2019.
- [10] Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., ... & Le, Q. V. (2020). Conformer: Convolution-augmented Transformer for Speech Recognition. In Proceedings of Interspeech 2020 (pp. 5036-5040).

指导教师评语	<p>该小组在人工智能应用开发课程设计中选择了基于改进 CBAM 注意力机制风扇异常状况的识别作为选题，整体表现优秀。该组同学在本课程设计和实施中展现了较好的实践能力和团队协作精神。课程设计文档规范，内容详实，图标格式正确，体现了较高的文档整理水平。综合而言，小组成员在技术和团队协作方面表现较好。</p> <p style="text-align: right;">指导教师（签名） <u>赵一</u></p>																				
成绩	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">学号</th> <th style="width: 33%;">姓名</th> <th style="width: 33%;">成绩</th> </tr> </thead> <tbody> <tr> <td>202201011070</td> <td>刘明</td> <td>优秀</td> </tr> <tr> <td>202201011083</td> <td>聂睿</td> <td>良好</td> </tr> <tr> <td>202201011052</td> <td>李璐宇</td> <td>优秀</td> </tr> <tr> <td>202201011065</td> <td>刘高政</td> <td>良好</td> </tr> <tr> <td>202201011068</td> <td>刘凯</td> <td>良好</td> </tr> </tbody> </table>			学号	姓名	成绩	202201011070	刘明	优秀	202201011083	聂睿	良好	202201011052	李璐宇	优秀	202201011065	刘高政	良好	202201011068	刘凯	良好
学号	姓名	成绩																			
202201011070	刘明	优秀																			
202201011083	聂睿	良好																			
202201011052	李璐宇	优秀																			
202201011065	刘高政	良好																			
202201011068	刘凯	良好																			
备注	无																				

注：实习成绩由指导教师评定出成绩，分优秀、良好、中等、及格、不及格五级，分别给小组的每个成员打分。