

# 深圳大学考试答题纸

(以论文、报告等形式考核专用)

二〇二四~二〇二五 学年度第 一 学期

课程编号	150329	课序号	01	课程名称	智能识别系统设计	主讲教师	沈琳琳	评分
	0001						文嘉俊	
	20221550							
学 号	33	姓名	洪子敬	专业年级	2022 级软件工程（腾班）			

教师评语：

题目： 基于多方位图像增强和特征提取方法的人脸识别系统

**摘要**--人脸识别技术是计算机视觉和模式识别领域的重要研究方向，广泛应用于安全监控、身份验证、社交网络等多个领域。而人脸特征提取是人脸识别系统中的关键步骤，它的目的是从原始人脸图像中提取出具有区分性的特征，以便进行后续的分类和识别。此外，由于采集数据时存在的各种环境因素，图像存在检测和识别的困难。针对这些问题，本文提出了多方位特征提取方法，包括不局限于机器学习、深度网络以及前沿方法 ArcFace，在图像进行多方位图像增强和人脸检测的基础上，用提取的特征训练不同的分类器以求达到最好的分类性能，从而形成一个较为完备、有较好的鲁棒性和准确性的人脸识别系统。

**关键词**—多方位、图像增强、特征提取、机器学习、深度网络、ArcFace

正文：

## I. 系统框架介绍

本系统整体可分为数据预处理、人脸检测、特征提取与分类器训练和分类四个部分。数据预处理方面主要是对图像可能存在的不同角度、光照变化、遮挡干扰和噪声干扰等问题进行处理；人脸检测方面主要是对处理过后图像使用预训练的 Haar 人脸检测器进行人脸检测，剪裁出检测到的人脸区域便于后续的特征提取；特征提取方面主要是采用多种特征提取方法提取人脸特征，其中特征提取方法包括机器学习、深度网络和前沿算法；分类器的训练和分类方面主要是使用训练集提取的特征进行多分类器训练，使用测试集来验证精度，选择最好的分类器。

## II. 数据预处理

现代图像获取设备在不同环境和条件下拍摄会得到不同质量的图像，图像中存在的问题包括不局限于角度变化、光照变化、遮挡干扰和噪声干扰等问题。本实验选用训练和测试的数据集是 10 个视频帧集，每个视频帧集是在不同环境下连续获取得到的，也存在前面所说的几个问题，本文也主要针对这一个问题进行探讨。

### 1. 高斯噪声的处理：

在图像中，高斯噪声会影响每个像素点的颜色或亮度值，图像中的每个像素都会加上一个从高

斯分布中采样的噪声值，从而导致图像出现颗粒感或模糊；对此的应对方法可以使用均值滤波或高斯滤波进行处理，然而传统的高斯滤波通常需要进行参数调整，因此本文采用了一种 **自适应高斯滤波** [1]的方法，具体来说：

基本思想是根据图像局部的标准差调整滤波窗口的大小和高斯核的标准差，以便在图像中不同区域采用不同强度的高斯滤波，从而更好地处理图像中的噪声，特别是应对不同噪声强度的区域，具体来说：

- a. 在噪声较强的区域，标准差会较大，滤波器的窗口会增大，进行更多的平滑。
- b. 在噪声较弱或者细节较清晰的区域，标准差较小，窗口大小也会相应较小，以保留更多的细节信息。

因此借助这种思想，我们先从 3x3 的窗口大小开始，逐步遍历，计算标准差（局部标准差 `local_std` 较大时，标准差 `sigma` 会增大，滤波器的窗口也会增大，更多的邻域信息将被引入以平滑噪声；标准差较小时，窗口和滤波器的标准差会较小，从而保留更多的细节信息。最后用 OpenCV 的 `filter2D` 实现平滑处理，具体用增率 `k`（设定为 2）来调节平滑程度，用公式  $\sigma = k * local\_std$  来设定平衡参数，将窗口从小遍历到大，尝试从 3,5,7 的窗口大小视角来平滑，从而实现对不同区域噪声的合理平滑，下面是个示例结果：



图 1：高斯噪声；（左：去噪前；右：去噪后）

然而，在实际应用情景中，该算法的时间复杂度较大，因此我们后续对平均标准差采样，即统计每个窗口内的标准差，取平均值进行滤波（因为在通常任务和数据集中，噪声的分布很少具有局部性，这是在缺乏算力硬件下的折中方案考虑）；经过多次实验验证，表现最好的是 3\*3 的高斯核

## 2. 椒盐噪声的处理

中值滤波是一种常用的去噪方法，特别适合去除椒盐噪声，与传统的中值滤波不同，自适应中值滤波会根据图像局部区域的特征动态调整滤波器的窗口大小，从而更好地保留图像细节，其具体实现是以 3\*3 的窗口开始，逐步增大窗口，直到达到最大窗口大小 `max_window_size`（设定为 7），这样可以保证滤波器根据图像的局部情况动态调整；

具体先用 `np.median(window)` 计算窗口内的中值 `median_val`，然后与中心像素值 `pixel_val` 比较以及最大、最小值，如果  $min\_val < pixel\_val < max\_val$ ，说明当前像素值是正常的，不需要更改，否则，认为当前像素值是噪声，使用窗口的中值 `median_val` 替换当前像素值，遍历 3,5,7 的窗口依次处理。示例效果如下：



图 2：椒盐噪声；（左：去噪前；右：去噪后）

然而，在实际应用情景中，该算法的时间复杂度同样较大，但是经过测试发现，中值滤波对于窗口大小并不敏感，用固定窗口进行滤波同样具有相同好的效果，因此后续实验为节约时间，采用窗口大小  $\text{win\_size} = 5$  进行中值滤波。

### 3. 亮暗区的处理

图片在自然情景下，有可能出现过度曝光（过亮）以及过暗的情景，而  $\gamma$  变换可以很好地处理这方面识别困难的问题，同样采用自适应策略，手动统计图像中的亮区，暗区的比例（设 0-85 为亮区，170-255 为暗区），用较多的那方执行对应的指令执行反向（暗区增强、亮区增强）即可。下面同样是个示例展示：



图 3：过亮 / 过暗；（左：增强前；右：增强后）

#### 4. 对比度的处理

直方图均衡化 (Histogram Equalization) 是一种常见的图像对比度增强方法, 尤其适用于那些对比度较低、亮度分布不均的图像。其主要思想是通过重新分配图像的像素值, 使得图像的灰度级分布尽可能均匀, 从而提高图像的对比度, 使得细节更为清晰, 其基本步骤为:

- ① 计算图像的灰度直方图;
- ② 计算累积直方图 (累积直方图 (Cumulative Distribution Function, CDF) 是直方图的累积和, 反映了灰度值的分布情况);
- ③ 重分配像素值: 将图像中的每个像素值映射到新的灰度级, 使得图像的灰度值分布尽可能均匀, 从而提高图像的对比度。

同样地其示例效果如下:



图 4: 低对比度; (左: 增强前; 右: 增强后)

然而, 在多重处理中需要慎用这种方法, 因为有时候直方图均衡化会使图像的灰度值分布过度均匀, 对于一些已经有良好对比度的图像, 直方图均衡化可能会导致过度增强, 出现图像细节丢失或噪声放大的情况, 特别是在图像本身的对比度较高时, 过度均衡可能会使亮区和暗区过度饱和, 造成图像的视觉失真, 这点会在后续的讨论中详细解释。

#### 5. 多角度旋转处理

由于本项目选用的是 Haar 人脸检测器, 该方法对旋转的容忍度低; 为了增强对旋转图像的鲁棒性, 可以采用多选择角度的策略, 其具体实现为先实现一个旋转函数, 实现对图像进行指定角度 `angle` 的旋转, 然后在主函数中, 可以选取不同的旋转刻度 `angle_scale` (例如  $5^\circ$ ,  $10^\circ$ ,  $15^\circ$  等), 测试能否在旋转后的图像中检测到人脸, 从而实现对图片法向的人脸进行检测。

然而, 经过测试, Haar 对旋转的鲁棒性较差, 需要较小的刻度 `angle_scale` 才能有较好的识别结果, 这意味着它将至少检测  $\text{int}(360^\circ / \text{angle\_scale}) - 1$  次, 若 `angle_scale` 较小, 可能会严重增加检测时间, 因此我们用对旋转鲁棒性较好的 HOG 的方法, 在检测的基础上进行对齐, 经测试, `angle_scale = 45^\circ` 时, HOG 方法就能很好的人脸检测效果, 在此基础上进行水平对齐处理, 可以提高直接用小 `angle_scale` 用 Haar 检测的效率。示例效果如下:



图 5：旋转图像；（左：水平对齐前；右：水平对齐后）

## 6. 遮挡处理

对于图像中存在的遮挡区域，我们的处理是首先是分块投票，这种方法是将图像分成若干个小块，并对每个小块进行预测，然后将这些预测结果进行投票合并，但是只适用于一般的方法（比如结合 PCA 使用，以及非深度的机器学习预测方法使用），因为大部分预测方法（例如网络）是结合整体的图片特征进行预测的，对小块特征并不具有较好的预测性。

然后是如何检测遮挡，这方面可以考虑用 PCA 的线性残差，通过建立一个线性模型来拟合图像的非遮挡部分，并根据残差来判断图像的遮挡区域：

此方法提供了检测、以及分块处理，二者结合可以实现通过残差的检验是否有遮挡然后在有遮挡的情况下，采取分块预测，投票综合的方式整合信息，可以在后续提取 PCA 特征的时候使用，但是需要在（haar）提前使用，来进行人脸轮廓边缘的提取。

## 7. 处理顺序讨论

接下来我们综合上述各种提及的难度图像进行综合起来，有一个**关键问题**是无法提前得知图片事先存在什么问题，而应对的处理方法**使用顺序不同时**，有可能有不同的修复效果，且有可能导致**前后影响**，因此需要对各种处理方法的顺序进行讨论，同时我们将以图 6 的两个例子为例，讨论各方法处理顺序的逻辑合理性。

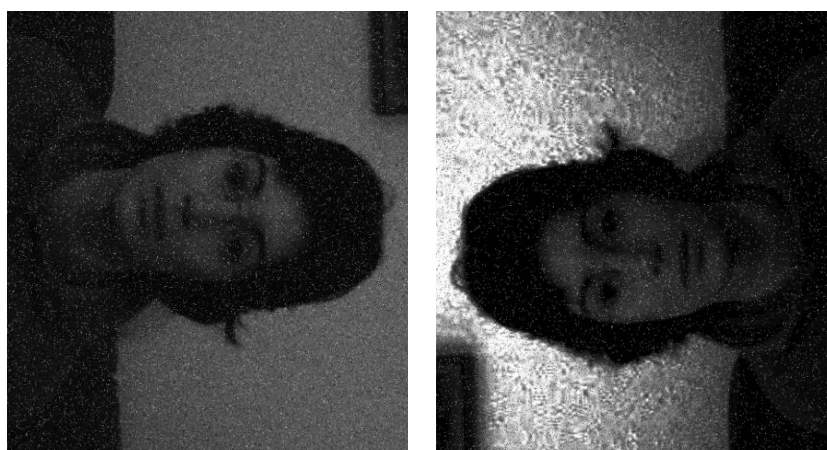


图 6：图像增强的例子；（左：Exp1；右：Exp2）

图 6 展示了两个例子，通过实验发现，进行不同的图像处理顺序，得到的图像是不一样的，而最佳处理顺序和具体要点是：直方图均衡化放在最后，将旋转（复杂度较大的排在后面，这样可以达到节省时间效果）放在次最后，而椒盐噪声优先处理，其他在中间过程的处理顺序调整对整体的影响不大，对此，进行处理的顺序伪代码如下：

对 Img 正常检测

If 未检测到人脸：# 一次探测

    Img ← 椒盐噪声去噪 # 首要

    If 未检测到人脸：# 二次探测

        Img ← 高斯噪声去噪 # 可与 gamma 顺序调换

        If 未检测到人脸：# 三次探测

            Img ← 自适应 gama 变换 # 可与高斯去噪顺序调换

            If 未检测到人脸：# 四次探测

                Img ← 旋转角度 # 次后

                If 未检测到人脸：# 五次探测

                    Img ← 直方图均衡化 + 旋转角度 # 放最后

下面以图 6 的 Exp1 为例来对进行解释，下图展示的是四次探测和五次探测的图像修复区别：

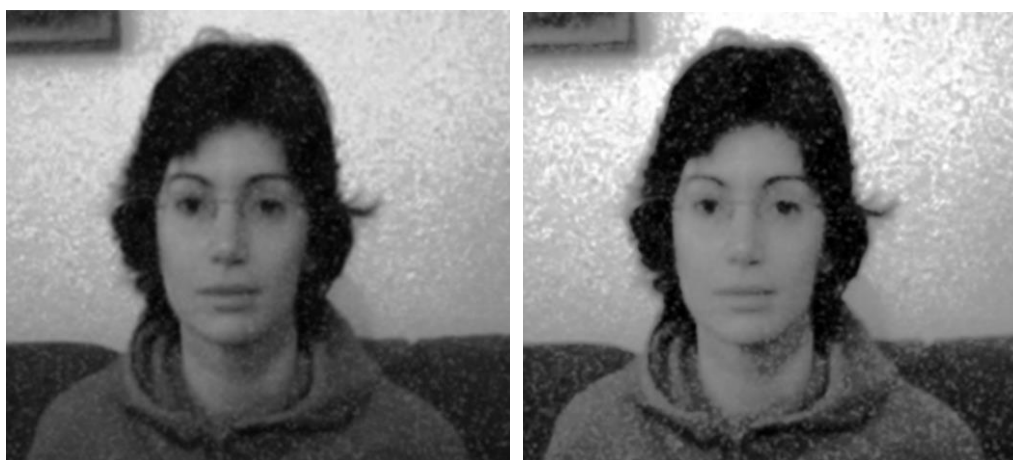


图 7：Exp1 修复处理；（左：直方图均衡化前；右：直方图均衡化后）

如图 7 所示，由于直方图均衡化适用于在极弱对比度下的修复，使用后有可能丢失局部细节（具有较明显的局部相似像素区块），若提前使用可能导致后面的关键解决方法失效，因此需要在最后使用，而如图 7 左可以看出，修复算法能有较好的修复性能，可以在特征预处理的时候使用，增强数据的质量，有助于后续提高算法性能。

如图 8 所示，图 6 的 Exp2 是局部对比度较弱的例子（脸部），此时最后才使用直方图均衡化有更好的效果：





## IV. 特征提取和分类

在对预处理后的数据检测出人脸后，接下来就是进行特征提取操作了，下面介绍多方位的特征提取方法，均是采用给定的默认划分训练集和测试集的视频帧数据集进行训练和测试，将采样结果均划分为  $150 \times 150$  的样本数据。

### 1. PCA 和分类器的组合

首先是测试传统机器学习方法，我们  $150 \times 150$  维的人脸采样为例，先将其展平，用 PCA 进行降维，选取前  $k$  个主成分的特征，经过多方面测试，选用 28 维特征可以取得较好的实验效果；提取特征后，可以配合传统的分类方法进行分类，其中我们选用传统的和先进的非深度分类器，包括：K-邻近分类（KNN），支持向量机分类（SVM），高斯朴素贝叶斯分类器（GNB），以及当前较为先进的基于梯度提升集成的方法 XGBoost 和基于树集成的随机森林（RF）方法进行精准度测试。使用各分类器模块的默认参数，特征降到 28 维，得到的实验结果如下（以精准度 ACC 为指标）：

表 1: Haar + PCA 的各类分类方法在视频帧数据集的精准率

	SVM	KNN	GNB	XGBoost	RF
ACC	<b>0.8881</b>	0.8461	0.5616	0.7748	0.8143

### 2. 现代卷积深度网络

现代在处理图像数据，基于深度卷积网络的方法较为流行，在和传统方法对比的基础上，我们用现代较为传统的卷积架构进行实验对比（均采用 10 轮训练，用 Adam 优化器，步幅为 0.001），实验结果如下：

表 2: Haar + 常用深度卷积网络方法在视频帧数据集的精准率

	VGG19_bn	ResNet101	DenseNet121	GoogleNet
ACC	<b>0.923</b>	0.8761	0.894	0.9146

在现代深度卷积网络的方法方面，有几个值得注意的现象：

- ① 大多数网络都在 4~5 轮就已经有 100% 的验证集精度了，有可能导致过拟合。
- ② 每次运行得到的精度度（ACC）标准差较大，可能是因为和网络随机化有关。

对比传统和深度网络，我们发现两者性能还是比较接近的，但深度网络更胜一筹，这与其网络分类器性能更好有关，下面我们将增加实验对比并探讨测试精准度未达成的原因分析。

### 3. 初步结果讨论

接下来我们讨论网络在 VidData 的性能较差的问题，我们以 GoogleNet 为例，输出具体每个分类指标的结果，如图 11 所示。

可见在类别 6，8 表现不好，结合先前的图 10（图 10 中的图片均采样自类别 6 和类别 8，含遮挡测试）得知表现不好的原因极有可能是遮挡区域太大的问题，当遮挡或未露面区域较大时，该方法即使是用前述的 PCA+分区域投票策略也无法很好地解决，因此该测试是合理的，接下来我们对两种方法（基于 PCA+非深度分类器和深度卷积网络）进行更多的性能对比，以及能提高人脸类别分辨率策略的讨论。



Test Accuracy: 91.46%				
	precision	recall	f1-score	support
0	0.91	1.00	0.95	284
1	0.94	1.00	0.97	230
2	0.99	0.92	0.95	263
3	1.00	0.71	0.83	808
4	0.96	1.00	0.98	360
5	0.99	1.00	1.00	140
6	0.50	0.91	0.64	236
7	0.91	0.89	0.90	140
8	0.78	1.00	0.88	147
9	1.00	0.97	0.98	1290
accuracy			0.91	3898
macro avg	0.90	0.94	0.91	3898
weighted avg	0.94	0.91	0.92	3898

图 11：不同类别的分类指标结果（以 GoogleNet 为例）

#### 4. 采样策略调整

实际上一种最合理的猜想是采样的策略（划分训练集、测试集的策略）导致了如上的情况，例如以类别 6 的数据集为例，该视频中前期（训练集）图片中的任务并未进行过多的遮挡和地法方向的旋转，然而在后期（测试集）却增加了遮挡、地法方向旋转的行为，因此在分布上并不能同比抽样，为了证明这点，我们改变了训练集、测试集的采样策略，具体来讲，我们使用了新的数据集 sample\_new，不用与原有的视频采样（直接将视频前半段作为训练集，后半段作为测试集）而是对视频的帧数对半分，分别将奇、偶分别到训练集、测试集进行采样，然后以效果较好的深度卷积神经网络为例，得到的效果如下（受网络随机影响）：

表 3：Haar + 常用深度卷积网络方法在调整采样下视频帧数据集的精准率

	VGG19_bn	ResNet101	DenseNet121	GoogleNet
ACC	<b>1.0</b>	<b>0.9782</b>	<b>1.0</b>	<b>0.9949</b>

可见调整策略后，分类性能均能接近于 100% 的效果，可见在人脸识别系统中，训练过程中对人脸的采样也特别重要。

#### 5. 更多类别的测试（以 AR 数据集为例）

由于视频帧数据集仅有 10 个数量的类别，为了测试在更多类别和涵盖光照情景下的不同算法性能，我们用 AR 数据集（一个涵盖光照、遮挡的人脸数据集），进行更加好地验证在更多情景、更多类别的算法性能，由于该数据集均是黑白灰度图像，包括 120 个人的类别（每个类别有 26 张图像，有可能含遮挡、不同光照下的），以卷积深度网络的为例，采用三折交叉验证的方法，固定划分顺序的随机种子（固定为 42），轮数固定为 10 轮，均采用 Adam 优化器，步幅为 0.001，统计精准率，其实验结果如下：

表 3：深度卷积网络在 ARFace 的精准率 (3Fold-CV)

	VGG19_bn	ResNet101	DenseNet121	GoogleNet
Fold 1	0.7962	0.4317	1.0	1.0
Fold 2	0.9365	0.9952	1.0	0.9808
Fold 3	0.9298	0.8337	1.0	1.0
avg ACC	0.8875	0.7535	<b>1.0</b>	0.9936

可见深度卷积算法通常具有较好精准率，通常是人脸识别系统的首选（具体应用场景需要根据硬件设备、算力要求等考虑）。

## 6. 前沿复现: ArcFace 进行人脸特征提取

在现代人脸识别领域，准确性和鲁棒性是衡量算法性能的关键指标。近年来，深度学习技术的迅猛发展推动了人脸识别技术的显著进步，其中，基于深度卷积神经网络的特征学习方法尤为突出。其中 **ArcFace** 是一种基于角度距离的损失函数，在人脸识别任务中表现出色，成为了学术界和工业界广泛关注的焦点。

在人脸识别领域，随着深度学习技术的迅速发展，基于深度卷积神经网络（CNN）的人脸识别方法逐渐成为主流。**ArcFace** 是一种高效且精度较高的深度人脸识别模型，它采用了基于角度度量的损失函数，从而能够实现更为精确的特征嵌入与人脸匹配。**ArcFace** 的核心思想是通过将人脸图像映射到一个高维的特征空间，并通过角度距离来度量不同人脸之间的相似性，实验表明该方法能大幅提升了在大规模人脸数据集上的识别精度。

因此本节旨在复现 ArcFace 的核心思想，并通过实验对比分析其与其他流行方法的性能差异。通过对 ArcFace 的详细复现和方法对比，首先该方法可以在 Python 的 insightface. app.FaceAnalysis 模块中下载相应权重进行调用，本文也是调用该库进行实现。

该方法最终将图片提取为一个 **512** 大小的向量，我们将其与 **Haar** 进行横向对比，为了体现其鲁棒性（其自带抗噪、光照对比度的处理），我们不进行类似于 **Haar** 的预处理，而是进行人脸检测，以视频帧数据集为例，检测结果下：



图 12: train samples, test samples 文件夹的人脸检测数目结果 (ArcFace)

对于两种检测的方式，汇总图 9、图 12 的检测结果，收集信息列表如下：

表 4: 人脸检测率对比 (Haar &amp; ArcFace)

	train_samples & Ratio(%)		test_samples & Ratio(%)	
预处理 + Haar	2182 / 2400	90.92	3898 / 5194	75.05
ArcFace	2332 / 2400	<b>97.17</b>	5004 / 5194	<b>96.34</b>

可见 ArcFace 具有显著的检测优势，接下来用 ArcFace 进行分类测试，用先前的五种非深度方法进行检验，效果对比如下：

表 5: 不同特征的 PCA 分类方法在 VidData 的精准率对比 (ACC)

	PCA+SVM	PCA+KNN	PCA+GNB	PCA+XGBoost	PCA+RF
预处理 + Haar	<b>0.8881</b>	0.8461	0.5616	0.7748	0.8143
ArcFace	0.9554	<b>0.9586 (best)</b>	0.8891	0.9133	0.9490

由此可见，最终用 ArcFace + PCA + KNN 的方法，在这个数据集上获得了最好的精准率，为了猜测其效果和高维空间的标准有关，我们用 t-SNE 降维的方法，对视频帧数据集进行二维可视化表示，如图 13 所示：

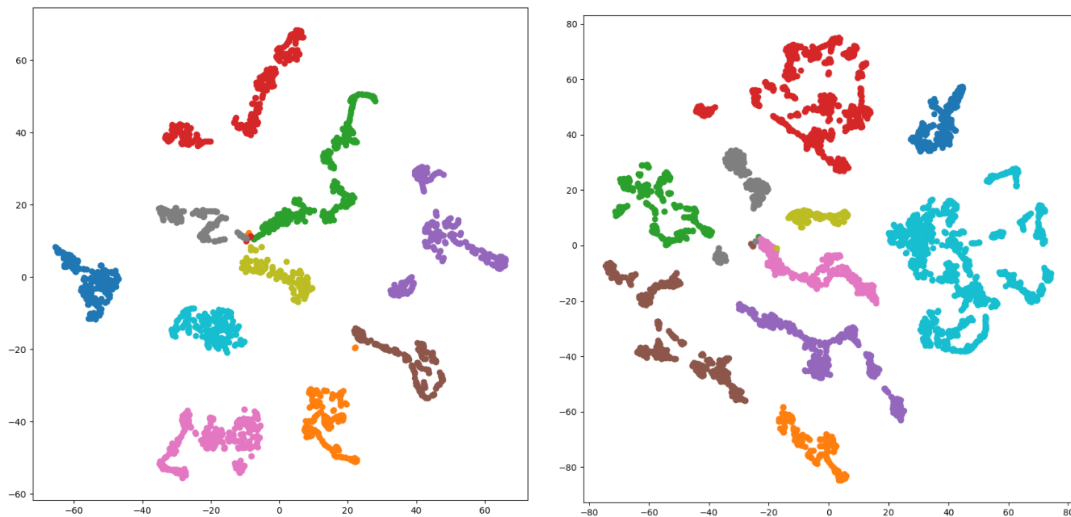


图 13: ArcFace 处理的 VidData 在 t-SNE 的降维可视化结果 (左: 训练集; 右: 测试集)

其中相同的颜色的点代表同一种类别, 由此可见, 该无监督方法提取的特征在表征上能相对较好地体现相同样本之间的相似性以及不同样本的差异性, 因此用 KNN、SVM 在这种同类相似性、异类差异性较好的数据集的预测表现较好。

## V. 总结与应用场景讨论

我们在本实验分别对实际场景的视频帧数据集 VidData(10 类别)和实验室场景的数据集 ARFace(120 类别)这两个数据集上进行了大量的方法实验, 包括深度和非深度方法, 我们发现在实验室环境与自然环境下, 识别算法的设计存在显著差异。这些差异主要源于数据采集条件、环境复杂度以及目标应用需求的不同。在我们结合 VidData 和 ARFace 这两个数据集的实验, 以及查阅了相关资料, 统计了一些关键的设计考量与创新方向。

- 1. 实验室环境下的识别算法设计:** 实验室环境通常具有受控的光照、背景和姿态, 这使得数据集集中的图像具有较高的一致性和较低的噪声水平。例如, ARFace 数据集在受控条件下采集, 因此算法在处理这些数据时可以更专注于特征的精确建模和提取。传统的特征提取方法 (如旋转不变性的 LBP、HOG) 在此类环境中仍能提供稳健的性能。此外, 深度学习方法在充分利用卷积神经网络 (CNN) 等模型的强大特征学习能力时, 也能在实验室环境中实现高精度识别。
- 2. 自然环境下的识别算法设计:** 自然环境下的数据集, 如 VidData, 往往包含多变的光照、复杂的背景、不同的拍摄角度以及可能的遮挡。这些不确定因素增加了识别任务的难度。因此, 算法设计需要具备更强的鲁棒性和泛化能力。深度学习方法在此环境中显示出显著优势, 尤其是通过大规模数据训练的神经网络能够自动学习到鲁棒的特征表示。数据增强技术和迁移学习策略也常被用于提高模型在自然环境中的表现。

而对于提高算法创新的策略, 通过文献综述[2][3], 目前该领域大概有如下的研究方向:

- 1. 多模态融合:** 结合多种传感器数据 (如 RGB 图像、深度信息、红外数据) 可以增强算法对环境变化的适应能力, 特别是在自然环境中。
- 2. 自监督学习:** 利用无标签数据进行预训练, 使模型能够自动学习数据的潜在结构, 从而提

升在不同环境中的泛化能力。

3. **领域自适应**：通过领域自适应技术，模型可以在训练阶段学习如何在不同数据分布之间迁移，增强其在实验室与自然环境之间的适应性。
4. **实时处理与优化**：在自然环境中，算法需要具备实时处理能力。优化模型结构以减少计算复杂度和提高处理速度是一个重要的创新方向。
5. **数据合成与增强**：例如可以用生成对抗网络（GAN）等技术生成多样化的训练样本，增加数据集的多样性和丰富性，有助于提升模型在自然环境中的表现。

综上所述，识别算法在不同环境下的设计需要综合考虑环境特征和应用需求，通过引入多模态数据、优化模型结构和利用先进的学习技术，可以有效提高算法的创新性和适用性。

#### 附加说明：

详细实验代码可见 github 链接：<https://github.com/breakthrougher/FaceRecognition/>

实验过程可见提交的“FaceRecognition.ipynb”文件

参数测试过程可见附件“附加实验结果说明”文件

#### 参考资料：

- [1] Deng G, Cahill L W. An adaptive Gaussian filter for noise reduction and edge detection[C] 1993 IEEE conference record nuclear science symposium and medical imaging conference. IEEE, 1993: 1615-1619.
- [2] Taskiran M, Kahraman N, Erdem C E. Face recognition: Past, present and future (a review)[J]. Digital Signal Processing, 2020, 106: 102809.
- [3] Boutros F, Struc V, Fierrez J, et al. Synthetic data for face recognition: Current state and future prospects[J]. Image and Vision Computing, 2023, 135: 104688.、