



本科毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

题 目	基于深度典型相关分析的多 视角学习研究
学 院	机电工程与自动化
专 业	自动化
学 号	19121458
学生姓名	温春晖
指导教师	修贤超
起讫日期	2023.2.15 ~ 2023.6.1



姓 名: 温春晖

学号: 19121458

论文题目: 基于深度典型相关分析的多视角学习研究

原创性声明

本人声明: 所呈交的论文是本人在指导教师指导下进行的研究工作。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发表或撰写过的研究成果。参与同一工作的其他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。

签名: 温春晖 日期: 2023.5.26

本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定。即: 学校有权保留论文及送交论文复印件, 允许论文被查阅和借阅; 学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

签名: 温春晖 指导教师签名: 倪良超 日期: 2023.5.26



上海大学

SHANGHAI UNIVERSITY

本科毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

基于深度典型相关分析的多
视角学习研究

学 院	机电工程与自动化
专 业	自动化
学 号	19121458
学生姓名	温春晖
指导教师	修贤超
起讫日期	2023.2.15 ~ 2023.6.1

摘 要

随着各类数据采集与处理设备的先进化与迅速迭代,模式识别所依赖的信息数量经历了指数级增长,同时信息的种类也在不断地丰富和拓宽。伴随着用户所持设备的计算性能的逐步提升,运用多视角数据进行机器学习的方法逐渐显现出其实质性的优势。对比于仅使用单一视角数据进行学习,多视角学习能有效发掘各个视角数据中的潜在关联性和互补性信息,整合特征,填补单视角数据中的信息缺失,从而有可能更精准地完成识别与分类任务,同时显著提升模型的稳定性。本文基于深度典型相关对多视角学习进行研究,具体如下:

- (1) 针对深度典型相关分析(DCCA)的原始结构较为朴素,难以更有效地应对分类任务的问题,提出了基于卷积与残差块的深度典型相关分析模型 ConvCCA 与 ResCCA。实验结果揭示,这些改进后的模型在多种数据集中的分类任务准确率上具有显著的提高。与此同时,当视角的目标表示更多元化时,神经网络能够获取更丰富的 CCA 指导信息,进而更有效地寻找优化的映射参数。进一步地,本文通过超参数实验,表明了模型性能的提升在很大程度上取决于网络架构、数据样本特性、数量及质量。
- (2) 针对 DCCA 与深度典型相关自编码器(DCCAE)中深度核的局限性以及无监督学习的问题,提出了半监督深度典型相关分析的结构以及 SConvCCAE 与 SResCCAE 模型。实验结果揭示,这些新方法以多集典型相关分析(MCCA)作为目标函数的情况下,在三视角数据集中的性能显著提升。此外,进一步的实验探索揭示,不同标签比例对模型性能的影响显著,改进后的模型能在最低的标签比例下达到较高的识别准确度。实验结果显示,改进后的模型能在需要最低的标签比例下达到较高的识别准确度。最后,本研究在两种数据集下探讨了不同的相关性算法对各模型性能的影响程度,结果显示基于广义典型相关分析(GCCA)的深度模型在无监督学习的表现优异,而基于张量典型相关分析(TCCA)的深度模型在完全监督的条件下性能表现显著提升。

关键词: 深度学习; 多视角学习; 典型相关分析; 深度典型相关自编码器

ABSTRACT

With the advanced and rapid iterations of various data collection and processing devices, the amount of information relied on for pattern recognition has experienced exponential growth, and the variety of information has also been continuously enriched and broadened. Along with the gradual improvement of the computing performance of the user's device, the method of machine learning using multi-view data is gradually showing its substantial advantages. Compared with learning using only single-view data, multi-view learning can effectively explore potential correlations and complementary information in each view data, integrate features, and fill in the missing information in single-view data, thus making it possible to complete the recognition and classification tasks more accurately and significantly improve the stability of the model. In this paper, we investigate multi-viewpoint learning based on deep canonical correlation as follows:

- (1) To overcome the problems raised by the somewhat simplistic original structure of Deep Canonical Correlation Analysis (DCCA), which results in challenges when dealing with classification tasks, we introduce ConvCCA and ResCCA - novel models of deep canonical correlation analysis based on convolution and residual blocks. Experimental results reveal a significant improvement in classification accuracy across multiple datasets following these modifications. Moreover, when the target representations from different perspectives are more diversified, the neural network can acquire richer guidance information from CCA, enabling more effective optimization of mapping parameters. Further, through hyperparameter experimentation, we demonstrate that the enhancement in model performance is largely contingent upon network architecture, the characteristics, quantity, and quality of data samples.
- (2) To overcome the limitations of the depth kernel in DCCA and Deep Canonical Correlation Autoencoder (DCCAE), as well as the problems in unsupervised learning, we propose a semi-supervised deep canonical correlation analysis structure and SConvCCA and SResCCA models. Experimental results show a significant improvement in performance across tri-view datasets when these new methods use Multiset Canonical Correlation Analysis (MCCA) as the objective function. Furthermore, experimental exploration demonstrates that the label ratio significantly influences model performance, with the improved models achieving high recognition accuracy even at the lowest label ratios. Lastly, this research examines the impact of different corre-

lation algorithms on model performance across two datasets, revealing that the deep models based on Generalized Canonical Correlation Analysis (GCCA) perform excellently in unsupervised learning, while those based on Tensor Canonical Correlation Analysis (TCCA) exhibit significantly enhanced performance under fully supervised conditions.

Keywords: Deep Learning; Multi-view Learning; Canonical Correlation Analysis; Deep Canonically Correlated Auto-Encoders

目 录

摘 要	I
ABSTRACT	II
第一章 绪论	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.2.1 非线性典型相关分析	3
1.2.2 深度典型相关分析	3
1.2.3 监督型深度典型相关分析.....	4
1.2.4 应用场景示例	5
1.3 主要内容安排	5
第二章 相关概念与理论基础	7
2.1 典型相关分析相关概念	7
2.1.1 典型相关分析	7
2.1.2 多集典型相关分析	10
2.1.3 广义典型相关分析	12
2.1.4 张量典型相关分析	12
2.2 深度神经网络	13
2.2.1 人工神经网络	13
2.2.2 卷积神经网络	15
2.3 分类算法	15
2.3.1 支持向量机	15
2.3.2 K-近邻	16
2.4 数据降维可视化.....	17
2.5 本章小结	17
第三章 基于深度典型相关分析的算法研究	18
3.1 深度典型相关分析	18
3.2 学习过程	19
3.3 模型介绍	20
3.4 数值实验	21
3.4.1 实验环境配置	21

3.4.2	分类准确度分析	21
3.4.3	网络复杂度与输出特征数分析	26
3.4.4	样本批大小分析	28
3.5	本章小结	29
第四章	基于深度典型相关自编码器的算法研究	30
4.1	判别式深度典型相关自编码器	30
4.2	学习过程	31
4.3	模型介绍	32
4.4	数值实验	32
4.4.1	分类准确度分析	32
4.4.2	判别标签比例分析	36
4.4.3	相关性算法分析	38
4.5	本章小结	39
第五章	总结与展望	41
5.1	总结	41
5.2	展望	42
插图索引		44
表格索引		46
参考文献		47
致 谢		51

第一章 绪论

1.1 研究背景与意义

图像识别作为人类观察与辨识事物的一种重要能力，在日常生活和工作具有重要意义。例如，可以通过观察电脑相册的照片，迅速联想到对应的人或事物；还可以通过其他感官如气味和声音，来更加精确地感知和了解物体的属性。与人类观察事物的方式相似，计算机进行图像识别的步骤主要是对图像中的特征进行提取，并利用这些特征分辨所期望得到的对象与其他对象，预测图像中所包含的对象类别，从而实现对所期望目标的最大可能性估计，如图 1.1 所示。图像识别技术通过模拟人类观察事物的过程，实现了对图像中特征的提取和对象的识别。这种技术在计算机视觉领域已经取得了重大进展，广泛应用于人脸识别、物体检测与自动驾驶等领域。

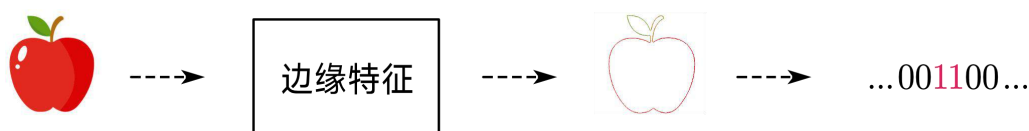


图 1.1 图像识别中的特征提取（边缘特征）

若仅使用一种类型的数据源来表示物体，则称为单视角数据。在过去采集与计算设备匮乏的情况下，许多纷繁枯燥的识别与分类任务需要大量的人力。随着设备逐步发展，获得单视角数据所需要的成本在逐渐下降，一些场景的识别与分类任务可以尝试交由机器完成。在单视角数据中的特征进行提取的众多算法中，较为主流的是主成分分析（Principal Component Analysis, PCA）^[1-4]，其在人脸识别^[5]、遗传学^[6]、生物医药^[7]、心电分析^[8]与碰撞检测^[9]等场景广泛使用。除此之外，线性判别分析（Linear Discriminant Analysis, LDA）^[10,11]与局部保持投影（Locality Preserving Projection, LPP）^[12,13]也都非常具有代表性的单视角数据处理方法。然而，单视角数据在描述目标时存在局限性，即只能捕捉到目标的部分信息，难以对目标进行全面地刻画。因此，在识别和分类任务中单视角数据无法显著提高准确度。

随着现代硬件设施的飞速发展，人们能够以较低的成本获取目标丰富的多视角数据，以便更为全面地理解目标。如图 1.2 所示，多视角数据指的是从不同角度、不同模态或不同传感器获取的关于同一目标的数据。这类数据可以为研究和开发人员提供目标的更丰富信息，有助于提高计算机模型的识别准确性。例如，对于同一物体，可以拍摄不同角度的照片，以获得更全面的三维信息；在实验室环境中，可以利

用各种传感器收集多种生物特征,包括人脸特征、声音、指纹和虹膜等,从而实现更为准确和安全的身份验证任务。

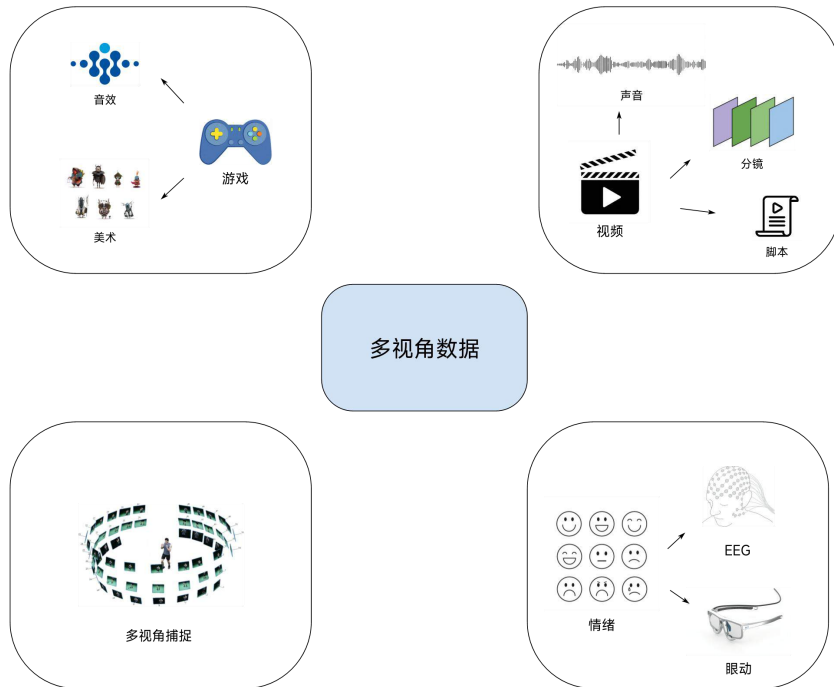


图 1.2 多视角数据示例

作为机器学习领域的一个重要分支,多视角学习 (Multi-view Learning) 致力于最大程度地挖掘多视角数据中的特征,它充分利用了不同视角数据的描述信息以实现信息互补,并有效减少对冗余数据的计算处理,从而促进模型性能的提升^[14,15]。典型相关分析 (Canonical Correlation Analysis, CCA)^[16] 是一种经典的多视角学习方法,其主要目标在于找到两个视角数据中共享的一个子空间,通过这种方式实现高维数据到低维数据的映射。CCA 的核心思想是对两个多元随机变量进行线性转换,从而最大化变量间相关性,揭示两个视角数据之间的内在联系,有助于理解和解释数据的内在结构和规律。

然而,CCA 只能处理两个视角的数据,而且仅限于线性关系。对于现实世界中的复杂问题,多视角的数据和非线性关系常常是无法避免的。因此,如何扩展 CCA 以处理多视角数据和非线性关系,以及如何在保证计算效率的前提下实现这些扩展,是当前研究的重要方向。

1.2 国内外研究现状

1.2.1 非线性典型相关分析

为了解决 CCA 无法处理多个视角数据的问题, Kettenring 等人^[17] 提出了多集典型相关分析 (Multiset Canonical Correlation Analysis, MCCA)。MCCA 目的是求得一组权重向量, 使得所有视角在经过权重向量投影后具有最大相关性。当视角数为 2 时, Vía 等人^[18] 证明了此时 MCCA 变为 CCA。Carroll 等人^[19] 提出了广义典型相关分析 (Generalized Canonical Correlation Analysis, GCCA)。相比于 MCCA, GCCA 的目标是为所有的视角寻找一个共享的线性子空间。此外, Luo 等人^[20] 提出了张量典型相关分析 (Tensor Canonical Correlation Analysis, TCCA), 通过计算多个视角的协方差张量来直接最大化多个视角的相关性。值得注意的是, 这三种对 CCA 扩展到多个视角的方法都是线性的。

为了解决 CCA 无法处理非线性数据的问题, Melzer 等人^[21] 提出了基于核函数技巧 (Kernel Trick) 的核典型相关分析 (Kernel Canonical Correlation Analysis, KCCA)。KCCA 通过选取不同的核函数, 将原始数据非线性映射到高维特征空间, 因此 KCCA 的目标是求两个视角的数据在核空间中的最大相关性。KCCA 已经在人脸表情识别^[22]、基因检测^[23]、功能性磁共振造影^[24] 等方向大量应用。

1.2.2 深度典型相关分析

随着深度学习的发展, Andrew 等人^[25] 提出了深度典型相关分析 (Deep Canonical Correlation Analysis, DCCA), 将深度神经网络 (Deep Neural Network, DNN) 视为一种“深度核”^[26], 利用深度核对两个视角的数据进行非线性映射, 完成特征提取与数据降维后再利用 CCA 计算输出层的特征从而得到最大相关性, 进而指导深度核的优化。相比于 KCCA, DCCA 能够尽可能地降低核函数的选择成本, 提高核函数在不同的场景的通用性。

将深度学习与 CCA 结合的思路启发了众多的研究者, 许多的工作喷涌而出。Wang 等人^[27] 基于 DCCA 与自编码器 (Auto-Encoders) 提出了深度典型相关自编码 (Deep Canonically Correlated Auto-Encoders, DCCAE), 通过最大化两个视角编码后特征的相关性的同时, 最小化特征经解码后与原视角数据相比的重构损失, 使视角之间能相互学习到有用的信息。Yang 等人提出了 CCANet^[28,29] 与 TCCANet^[28]。CCANet 利用多层 CCA 来指导多个视角中滤波器的学习, 尽可能地使滤波器提取图像更全面的特征。TCCANet 则将 CCANet 推广到多视角, 利用了所有视角中的高阶相关性, 相比于 MCCA^[30] 在多视角遥感领域具有显著提升。除此之外, Benton 等人^[31] 将深度学习与 GCCA 结合, 提出的深度广义典型相关分析 (Deep Generalized

Canonical Correlation Analysis, DGCCA), 相比于 DCCA 在社交媒体用户推荐主题标签的任务中较为显著的改进。Wong 等人^[32] 提出深度张量典型相关分析 (Deep Tensor Canonical Correlation Analysis, DTCCA), 通过实验表明了分类性能相比于 TCCA 与 DGCCA 在百科分类与场景识别等任务中有明显提升。

1.2.3 监督型深度典型相关分析

然而, 上述 DCCA 并未利用标签信息, 这激发了研究者们探索融入判别性信息以实现其性能的优化。Liu 等人^[33] 提出了监督型深度典型相关分析 (Supervised Deep Canonical Correlation Analysis, SDCCA), 该方法旨在确保非线性映射后特征的最大相关性, 同时最小化两个视角中训练样本的类内散度。Elmadany 等人^[34] 引入了深度判别式典型相关分析 (Deep Discriminative Canonical Correlation Analysis, Deep DisCCA), 此方法能在最大化类内相关性的同时, 最小化类间的相关性。基于前述研究, Zhang 等人^[35] 进一步提出了深度学习监督型典型相关分析 (Deep Learning Supervised Canonical Correlation Analysis, DL-SCCA)。在此方法中, 两个视角的数据经过深度核特征抽取后, 最大化特征相关性同时并入公共全连接层。此外, 该方法引入标签信息以计算交叉熵, 以此在优化深度核过程中引导其共同作用。

在图 1.3 中给出了基于 DCCA 的分类任务的处理流程, 其步骤可以概述如下: 首先, 利用深度核对输入的各个视角的数据进行映射, 以获取源视角的高级抽象表示。其次, 应用无监督或监督的相关性分析算法对这些表征特征进行处理。在反向传播的过程中, 深度核的参数得到优化, 使得相关性分析的输出更接近 CCA 方法的优化目标。经过多轮迭代学习后, 深度核将原始视角的数据映射到子空间。最后, 采用各种分类算法以完成识别任务。

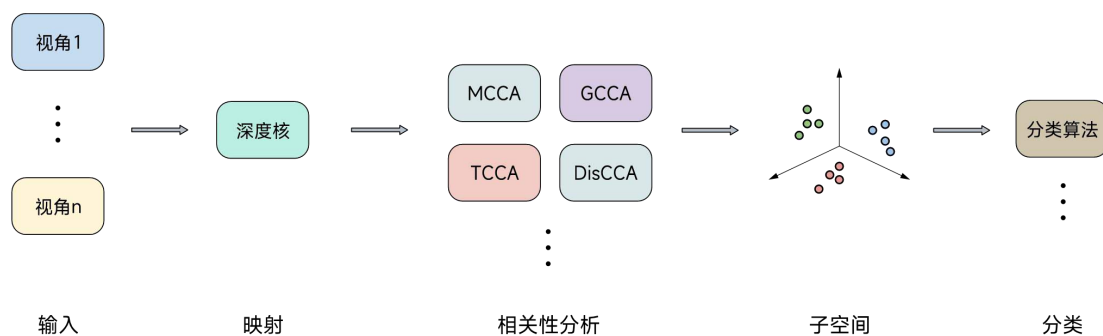


图 1.3 基于 DCCA 的分类任务工作流

1.2.4 应用场景示例

深度典型相关分析已被广泛应用于多种领域。例如, Yuan 等人^[36]成功地将卷积神经网络与 CCA 特征融合,提出了一种创新的渔业预测模型。他们的模型有效地剔除了冗余信息,从而在相较于传统深度学习算法中显著提高了预测精度。Isobe 等人^[37]也利用了 DCCA,在噪声环境下研发了自动语音识别(Automatic Speech Recognition, ASR)系统。相较于传统的纯音频 ASR 和视听融合的 ASR,基于 DCCA 的 ASR 具有更优的性能表现,尤其是在训练数据较少的情况下,采用音频 DCCA 与视觉 DCCA 的特征增强,可以进一步提升系统的表现。此外, Sun 等人^[38]将 DCCA 引入多模态情感识别领域,通过融合不同模态的信息,成功提高了情感识别的准确性。类似地, Qiu 等人^[39]采用了脑电图(Electroencephalogram, EEG)和眼动数据作为 DCCA 的双视角输入,相比于传统的 CCA 方法,他们的情绪分类准确度获得了显著的提升。在另一个实例中, Yang 等人^[40]提出了一种新颖的基于 DCCA 的跨传感器变化检测方法。他们在交叉传感器图像数据集上进行实验,结果表明,该方法相比传统的 KCCA 和 CCA 方法具有显著的优势。以上实例都表明了 DCCA 在各领域的广泛应用及其优越性。

1.3 主要内容安排

本篇论文主要基于现有的深度典型相关分析与深度典型相关自编码算法,对引入判别式信息、不同的网络结构与优化目标函数等改进策略进行探讨,并在不同的数据集上进行分析。本文的结构安排如下:

第一章为绪论。首先简要阐述了多视角学习的研究背景与意义,其次概括了典型相关分析与深度典型相关分析的研究现状,最后说明了本文的内容安排。

第二章介绍与典型相关分析有关的理论基础。首先对 CCA、MCCA、GCCA 与 TCCA 的原理进行讨论,其次介绍深度神经网络的相关理论基础,最后介绍本文实验部分使用的两种分类算法的基本原理与高维数据降维可视化算法。

第三章对深度典型相关分析算法进行研究。首先对 DCCA 算法的相关知识进行说明,其次引入不同的神经网络模型对 DCCA 进行改进,提出卷积典型相关分析(Convolutional Canonical Correlation Analysis, ConvCCA)与残差典型相关分析(Residual Canonical Correlation Analysis, ResCCA)两种模型。通过一系列数值实验来验证改进后模型的有效性。最后探究了模型中不同的超参数对分类结果的影响。

第四章对深度典型相关自编码器算法进行研究。首先介绍 DCCAE 的基本结构和原理,并在此基础上提出监督深度典型相关自编码器(Supervised Deep Canonically Correlated Auto-Encoders, SDCCAE)、监督型卷积典型相关自编码器(Supervised

Convolutional Canonically Correlated Auto-Encoders, SConvCCAE) 与监督型残差典型相关自编码器 (Supervised Residual Canonically Correlated Auto-Encoders, SResCCAE) 三种模型。其次对 DCCA 与 DCCAE 的扩展模型进行分类任务数值实验与半监督超参数实验, 分析各个模型的结果。最后使用 GCCA 与 TCCA 来替代 MCCA 算法, 探究不同相关性损失函数对分类表现的影响程度。

第五章为整篇论文的内容总结与对未来研究的展望。总结了本文的主要贡献, 并针对当前研究存在的局限和未来可能的研究方向进行了展望, 旨在为基于深度学习的多视角典型相关分析算法的改进和优化提供有益的启示和参考。

第二章 相关概念与理论基础

2.1 典型相关分析相关概念

2.1.1 典型相关分析

典型相关分析 (CCA) 通过寻找一组投影方向, 将两个视角的数据投影到共同空间, 使得投影后两个视角的数据相关性最大。具体而言, 设有两个视角的数据, 分别表示为 V_1 和 V_2 , 两个视角的均值分别为 \bar{V}_1 与 \bar{V}_2 , 样本数量与维度都相同。对 V_1 与 V_2 去均值后得到 \hat{V}_1 与 \hat{V}_2 :

$$\begin{aligned}\hat{V}_1 &= V_1 - \bar{V}_1 \\ \hat{V}_2 &= V_2 - \bar{V}_2\end{aligned}\tag{2.1}$$

设有一组向量为 p_1 和 p_2 , 并对 \hat{V}_1 与 \hat{V}_2 进行投影:

$$\begin{aligned}\hat{V}'_1 &= p_1^\top \hat{V}_1 \\ \hat{V}'_2 &= p_2^\top \hat{V}_2\end{aligned}\tag{2.2}$$

计算投影后 \hat{V}'_1 与 \hat{V}'_2 的相关系数:

$$\rho = \frac{Cov(\hat{V}'_1, \hat{V}'_2)}{\sqrt{D(\hat{V}'_1)}\sqrt{D(\hat{V}'_2)}}\tag{2.3}$$

CCA 的目标是求 ρ 为最大时的投影向量 p_1 与 p_2 , 也就是求解以下优化问题:

$$\max_{p_1, p_2} \frac{Cov(\hat{V}'_1, \hat{V}'_2)}{\sqrt{D(\hat{V}'_1)}\sqrt{D(\hat{V}'_2)}}\tag{2.4}$$

根据协方差的计算得到:

$$\begin{aligned}Cov(\hat{V}_1, \hat{V}_2) &= E(\hat{V}_1 \hat{V}_2^\top) \\ D(\hat{V}_1) &= E(\hat{V}_1 \hat{V}_1^\top) \\ D(\hat{V}_2) &= E(\hat{V}_2 \hat{V}_2^\top)\end{aligned}\tag{2.5}$$

为了进一步化简式子, 将协方差 $E(\hat{V}_1 \hat{V}_2^\top)$ 、 $E(\hat{V}_1 \hat{V}_1^\top)$ 与 $E(\hat{V}_2 \hat{V}_2^\top)$ 分别记为 C_{12} 、 C_{11} 与 C_{22} 。变量替换后的优化目标表示为:

$$\max_{p_1, p_2} \frac{p_1^\top C_{12} p_2}{\sqrt{p_1^\top C_{11} p_1} \sqrt{p_2^\top C_{22} p_2}}\tag{2.6}$$

引入约束条件 $\sqrt{p_1^\top C_{11} p_1} \sqrt{p_2^\top C_{22} p_2} = 1$ ，使得上述式子分母为 1，最终的优化问题表示为：

$$\begin{aligned} \max_{p_1, p_2} \quad & p_1^\top C_{12} p_2 \\ \text{s.t.} \quad & p_1^\top C_{11} p_1 = 1, \quad p_2^\top C_{22} p_2 = 1 \end{aligned} \quad (2.7)$$

求解上述问题可以使用拉格朗日函数，将目标改写为：

$$L(p_1, p_2, \lambda, \mu) = p_1^\top C_{12} p_2 + \lambda (p_1^\top C_{11} p_1 - 1) + \mu (p_2^\top C_{22} p_2 - 1) \quad (2.8)$$

对函数内的各个变量求偏导后令偏导数为 0：

$$\begin{cases} \frac{\partial L(p_1, p_2, \lambda, \mu)}{\partial p_1} = C_{12} p_2 + \lambda C_{11} p_1 + \lambda C_{11}^\top p_1 = 0 \\ \frac{\partial L(p_1, p_2, \lambda, \mu)}{\partial p_2} = C_{12}^\top p_1 + \mu C_{22} p_2 + \mu C_{22}^\top p_2 = 0 \\ \frac{\partial L(p_1, p_2, \lambda, \mu)}{\partial \lambda} = p_1^\top C_{11} p_1 - 1 = 0 \\ \frac{\partial L(p_1, p_2, \lambda, \mu)}{\partial \mu} = p_2^\top C_{22} p_2 - 1 = 0 \end{cases} \quad (2.9)$$

协方差矩阵 C_{11} 、 C_{12} 与 C_{22} 是对称矩阵，故整理上述方程组：

$$\begin{cases} C_{12} p_2 + 2\lambda C_{11} p_1 = 0 \\ C_{12} p_1 + 2\mu C_{22} p_2 = 0 \\ p_1^\top C_{11} p_1 = 1 \\ p_2^\top C_{22} p_2 = 1 \end{cases} \quad (2.10)$$

利用约束条件提取出 λ 与 μ ：

$$\begin{cases} p_1^\top C_{12} p_2 + 2\lambda = 0 \\ p_2^\top C_{12} p_1 + 2\mu = 0 \end{cases} \quad (2.11)$$

对上式移项后可知 -2λ 为优化目标，即相关系数。与此同时，观察第二个式子可推出 $(p_1^\top C_{12} p_2)^\top = p_2^\top C_{12} p_1$ ，故得到 $\lambda = \mu$ 。对公式 (2.10) 前两个式子分别乘以 C_{11}^{-1} 与 C_{22}^{-1} 得到：

$$\begin{cases} C_{11}^{-1} C_{12} p_2 = -2\lambda p_1 \\ C_{22}^{-1} C_{12} p_1 = -2\mu p_2 \end{cases} \quad (2.12)$$

对第一个式子乘以 -2μ 并代入 $\lambda = \mu$ 后消元得到:

$$C_{11}^{-1}C_{12}C_{22}^{-1}C_{12}p_1 = 4\lambda^2 p_1 \quad (2.13)$$

令 $M = C_{11}^{-1}C_{12}C_{22}^{-1}C_{12}$, 得到 $Mp_1 = 4\lambda^2 p_1$, 故优化问题转化为求特征分解, 矩阵 M 最大特征值的平方根就是问题所期望的最大相关系数, 其对应的特征向量就是主子空间的投影方向。

相应地, 使用辅助变量凑出 SVD 分解^[41,42] 的格式同样能够解决问题。设投影向量 q_1, q_2 为:

$$\begin{cases} q_1 = C_{11}^{1/2} p_1 \\ q_2 = C_{22}^{1/2} p_2 \end{cases} \quad (2.14)$$

代入约束条件后得到:

$$\begin{aligned} q_1^\top C_{11}^{-1/2} C_{11} C_{11}^{-1/2} q_1 &= 1 \Rightarrow q_1^\top q_1 = 1 \\ q_2^\top C_{22}^{-1/2} C_{22} C_{22}^{-1/2} q_2 &= 1 \Rightarrow q_2^\top q_2 = 1 \end{aligned} \quad (2.15)$$

因此公式 (2.7) 等价于:

$$\begin{aligned} \max_{q_1, q_2} \quad & q_1^\top C_{11}^{-1/2} C_{12} C_{22}^{-1/2} q_2 \\ \text{s.t.} \quad & q_1^\top q_1 = 1, q_2^\top q_2 = 1 \end{aligned} \quad (2.16)$$

令 $M = q_1^\top C_{11}^{-1/2} C_{12} C_{22}^{-1/2} q_2$, 对矩阵 M 进行奇异值分解, 得到 $M = U\Sigma V^\top$, 故优化目标变为:

$$\begin{aligned} \max_{q_1, q_2} \quad & q_1^\top U\Sigma V^\top q_2 \\ \text{s.t.} \quad & q_1^\top q_1 = 1, q_2^\top q_2 = 1 \end{aligned} \quad (2.17)$$

设 $y_1 = U^\top q_1, y_2 = V^\top q_2$, 因 U, V 为正交矩阵, 约束条件变为 $y_1^\top y_1 = q_1^\top U U^\top q_1 = 1, y_2^\top y_2 = q_2^\top V V^\top q_2 = 1$, 最终优化问题转变为:

$$\begin{aligned} \max_{y_1, y_2} \quad & y_1^\top \Sigma y_2 \\ \text{s.t.} \quad & y_1^\top y_1 = 1, y_2^\top y_2 = 1 \end{aligned} \quad (2.18)$$

其中 y_1, y_2 为单位向量, Σ 为对角矩阵, 对角线上为降序排列的奇异值。所求的最大相关系数就对应矩阵 Σ 对角线上最大奇异值, 将奇异值对应的 y_1, y_2 经反向变

换后的 p_1, p_2 构成目标投影方向:

$$\begin{cases} p_1 = C_{11}^{-1/2} U y_1 \\ p_2 = C_{22}^{-1/2} V y_2 \end{cases} \quad (2.19)$$

总之, 两种方法都可以求得最终的结果, 相较而言后者的计算会更简洁, 因此多数 CCA 的算法实现均采用 SVD 进行求解。值得注意的是, 两种方法都涉及到计算协方差矩阵 C 的 k 次方的操作。实现矩阵的乘方一般是对矩阵进行特征值分解后, 取对角矩阵的 k 次方后乘以特征向量矩阵。当 $k < 0$ 时, 需要注意对角矩阵是否存在小于或等于 0 的元素。在 CCA 的计算中, 协方差矩阵 C 是半正定的, 也就意味着它的对角矩阵会出现值接近或者为 0 的元素。因此, 在实际的算法实现中, 通常会引入一个非常小的正则化参数 γ , 将它的标量矩阵 γE 加到对角矩阵中。这种做法确保不会出现除以 0 引起的非数 (Not a Number, NaN) 传播, 提高数值稳定性。

2.1.2 多集典型相关分析

多集典型相关分析 (MCCA) 将 CCA 的思想扩展到多个视角, 也就意味着寻找一组向量, 将投影后的各个视角对的最大相关系数之和最小。设有三个视角的数据, 分别表示为 V_1, V_2, V_3 , 它们的均值分别为 $\bar{V}_1, \bar{V}_2, \bar{V}_3$, 对各个视角的数据进行去均值后得到 $\hat{V}_1 = V_1 - \bar{V}_1, \hat{V}_2 = V_2 - \bar{V}_2, \hat{V}_3 = V_3 - \bar{V}_3$ 。设有一组投影向量为 p_1, p_2, p_3 , 对 $\hat{V}_1, \hat{V}_2, \hat{V}_3$ 进行投影:

$$\begin{aligned} \hat{V}'_1 &= p_1^\top \hat{V}_1 \\ \hat{V}'_2 &= p_2^\top \hat{V}_2 \\ \hat{V}'_3 &= p_3^\top \hat{V}_3 \end{aligned} \quad (2.20)$$

类似地, 计算投影后 \hat{V}'_1, \hat{V}'_2 与 \hat{V}'_3 每两对视角的相关系数 ρ_{12}, ρ_{13} 与 ρ_{23} 。根据公式 (2.5) 得到不同视角之间的协方差矩阵 C_{12}, C_{13} 与 C_{23} 。与 CCA 相似, 可以将公式 (2.7) 扩展为:

$$\begin{aligned} \max_{p_1, p_2, p_3} \quad & p_1^\top C_{12} p_2 + p_1^\top C_{13} p_3 + p_2^\top C_{23} p_3 \\ \text{s.t.} \quad & p_1^\top C_{11} p_1 = 1, p_2^\top C_{22} p_2 = 1, p_3^\top C_{33} p_3 = 1 \end{aligned} \quad (2.21)$$

这是一个多元特征值问题, 求解这个问题依旧可以使用拉格朗日函数。然而, Chu 等人^[43] 证明了 MCCA 在多于三个视角的情况下没有解析解。该问题的数值解往往受限于初始向量的选择^[44], 故多次迭代重新选择初始向量^[45,46] 成为逼近该问题局部最优解的一种选择。

将 MCCA 中的多个约束改写成一个约束, 变为:

$$\begin{aligned} \max_{p_1, p_2, p_3} \quad & p_1^\top C_{12} p_2 + p_1^\top C_{13} p_3 + p_2^\top C_{23} p_3 \\ \text{s.t.} \quad & p_1^\top C_{11} p_1 + p_2^\top C_{22} p_2 + p_3^\top C_{33} p_3 = 1 \end{aligned} \quad (2.22)$$

仿照公式 (2.8) 写出拉格朗日函数:

$$L(p_1, p_2, p_3, \lambda) = p_1^\top C_{12} p_2 + p_1^\top C_{13} p_3 + p_2^\top C_{23} p_3 + \lambda (p_1^\top C_{11} p_1 + p_2^\top C_{22} p_2 + p_3^\top C_{33} p_3 - 1) \quad (2.23)$$

对函数的各个自变量求偏导后令其为 0:

$$\begin{cases} \frac{\partial L(p_1, p_2, p_3, \lambda)}{\partial p_1} = C_{12} p_2 + C_{13} p_3 + \lambda (C_{11} p_1 + C_{11}^\top p_1) = 0 \\ \frac{\partial L(p_1, p_2, p_3, \lambda)}{\partial p_2} = C_{12}^\top p_1 + C_{23} p_3 + \lambda (C_{22} p_2 + C_{22}^\top p_2) = 0 \\ \frac{\partial L(p_1, p_2, p_3, \lambda)}{\partial p_3} = C_{13}^\top p_1 + C_{23}^\top p_2 + \lambda (C_{33} p_3 + C_{33}^\top p_3) = 0 \\ \frac{\partial L(p_1, p_2, p_3, \lambda)}{\partial \lambda} = p_1^\top C_{11} p_1 + p_2^\top C_{22} p_2 + p_3^\top C_{33} p_3 - 1 = 0 \end{cases} \quad (2.24)$$

整理方程组为矩阵的形式得到:

$$\begin{bmatrix} 0 & C_{12} & C_{13} \\ C_{12}^\top & 0 & C_{23} \\ C_{13}^\top & C_{23}^\top & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = -2\lambda \begin{bmatrix} C_{11} & 0 & 0 \\ 0 & C_{22} & 0 \\ 0 & 0 & C_{33} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \quad (2.25)$$

用 A, B 代替等式两边的矩阵, 用 x 代替目标投影向量, 矩阵方程可以简写为:

$$Ax = -2\lambda Bx \quad (2.26)$$

公式 (2.26) 是一个广义特征值问题, 现有的数值代数计算库已经能够非常快速且准确的求解此类问题。通常会对矩阵 B 进行正则化, 做法与上一小节的 CCA 相同。此外, 当矩阵规模过大时, 直接使用特征值分解会导致计算过慢, 可以使用数值迭代法求解, 如偏最小二乘法^[18]。

MCCA 优化多对视角的相关系数之和, 因此也被称为 SUMCOR-CCA^[47-49]。将

视角数推广到 n ，即可得到单约束 MCCA 优化目标的一般形式：

$$\begin{aligned} \max_{p_1, p_2, \dots, p_n} \quad & \sum_{i=1}^n \sum_{j=i+1}^n p_i^\top C_{ij} p_j \\ \text{s.t.} \quad & \sum_i p_i^\top C_{ii} p_i = 1 \end{aligned} \quad (2.27)$$

2.1.3 广义典型相关分析

广义典型相关分析 (GCCA) 优化所有视角的投影向量，使得投影后尽可能地接近共同的子空间。具体而言，假设有两个已经去均值的视角数据为 \hat{V}_1 与 \hat{V}_2 ，投影方向为 p_1 与 p_2 ，子空间为 G 。投影后的视角与子空间 G 的距离可以使用 Frobenius 范数的平方来度量，因此 GCCA 的目标函数可以表示为：

$$\begin{aligned} \min_{p_1, p_2, G} \quad & \|G - p_1^\top \hat{V}_1\|_F^2 + \|G - p_2^\top \hat{V}_2\|_F^2 \\ \text{s.t.} \quad & G^\top G = I \end{aligned} \quad (2.28)$$

要解决这个优化问题，可以求解下面矩阵的特征分解^[50]得到最终的结果：

$$\left[\left(\hat{V}_1^\top (\hat{V}_1 \hat{V}_1^\top)^{-1} \hat{V}_1 \right) + \left(\hat{V}_2^\top (\hat{V}_2 \hat{V}_2^\top)^{-1} \hat{V}_2 \right) \right] G^\top = G^\top \Lambda \quad (2.29)$$

将双视角扩展到三个及以上的视角，GCCA 的目标函数可以表示为：

$$\begin{aligned} \min_{p_1, p_2, \dots, p_n, G} \quad & \sum_{i=1}^n \|G - p_i^\top \hat{V}_i\|_F^2 \\ \text{s.t.} \quad & G^\top G = I \end{aligned} \quad (2.30)$$

特征分解问题则推广为：

$$\left[\sum_{i=1}^n \hat{V}_i^\top (\hat{V}_i \hat{V}_i^\top)^{-1} \hat{V}_i \right]^{-1} G^\top = G^\top \Lambda \quad (2.31)$$

在公式 (2.31) 中， Λ 是特征值对角矩阵，投影向量 p_i 可以通过计算 $(\hat{V}_i \hat{V}_i^\top)^{-1} \hat{V}_i G^\top$ 来求得。值得注意的是，在计算中涉及到矩阵求逆。若矩阵的规模很大，特征分解的运行速度将会被严重拖慢。因此，Fu 等人^[51]提出了一种通过结构正则矩阵与交替优化的方法来帮助 GCCA 扩展规模。

2.1.4 张量典型相关分析

与 MCCA 估计成对视角的相关性不同，张量典型相关分析 (TCCA) 通过构建协方差张量直接最大化所有视角的高阶相关性。具体而言，设有 n 个视角分

别为 V_1, V_2, \dots, V_n ，去均值后为 $\hat{V}_1, \hat{V}_2, \dots, \hat{V}_n$ 。因此优化目标变为找到投影方向 p_1, p_2, \dots, p_n 使得投影后的视角 $\hat{V}'_1 = p_1^\top \hat{V}_1, \hat{V}'_2 = p_2^\top \hat{V}_2, \dots, \hat{V}'_n = p_n^\top \hat{V}_n$ 的相关性最大。因此，TCCA 的优化问题表示为：

$$\begin{aligned} \max_{p_1, p_2, \dots, p_n} \quad & \text{corr}(\hat{V}'_1, \hat{V}'_2, \dots, \hat{V}'_n) \\ \text{s.t.} \quad & (\hat{V}'_1)^\top \hat{V}'_1 = 1, (\hat{V}'_2)^\top \hat{V}'_2 = 1, \dots, (\hat{V}'_n)^\top \hat{V}'_n = 1 \end{aligned} \quad (2.32)$$

n 个视角的相关程度 $\text{corr}(\hat{V}'_1, \hat{V}'_2, \dots, \hat{V}'_n)$ 可以表示为：

$$\text{corr}(\hat{V}'_1, \hat{V}'_2, \dots, \hat{V}'_n) = C_{12\dots n} \bar{\times}_1 p_1^\top \bar{\times}_2 p_2^\top \dots \bar{\times}_n p_n^\top \quad (2.33)$$

在公式 (2.33) 中， $\bar{\times}_i$ 代表 i 模态积。 $C_{12\dots n}$ 表示所有视角构成的张量协方差，通过对所有视角各自样本的张量积求和得到：

$$C_{12\dots n} = \frac{1}{N} \sum_{j=1}^N \hat{V}'_{1,j} \circ \hat{V}'_{2,j} \circ \dots \circ \hat{V}'_{n,j} \quad (2.34)$$

其中， $\hat{V}'_{i,j}$ 表示视图 i 的第 j 个样本。运算符 \circ 为张量外积。

TCCA 的目标函数可以通过交替最小二乘法 (Alternating Least Squares, ALS) 来计算。首先随机初始化所有投影方向 p_1, p_2, \dots, p_n ，固定除了 p_i 以外的所有投影方向，其次使用梯度下降等优化算法来优化 p_i 。对所有投影方向完成一轮优化后，可以使用一些收敛条件来判断是否满足要求，例如达到最大迭代次数或投影方向变化过小等。

2.2 深度神经网络

2.2.1 人工神经网络

人工神经网络 (Artificial Neural Networks, ANN) 以自然界中神经生物学为启示，构建了一种仿生的计算模型。如图 2.1 所示，类似于生物神经细胞通过突触进行信息传递，ANN 里的神经元同样会接收前层神经元的输出，对其进行加权累积并添加偏差，然后通过激活函数来确定对下一层神经元的激活程度。随着科技的进步，ANN 在其早期的单层感知机形态基础上逐渐发展为多层感知机 (Multilayer Perceptron, MLP)，并引入了用于调整神经元间权重的反向传播算法，实现网络的学习。

神经网络通过多次迭代前向与反向传播来对输入的数据进行建模。由图 2.2 可知，在前向传播中，输入数据经神经网络中不同层的加权求和与激活后抵达输出层。

神经元中的激活函数是神经网络非线性的一种体现，图 2.3 分别展示了 \tanh 、 sigmoid 、 ReLU 和 leakyReLU 。在早期的研究中， \tanh 与 sigmoid 被广泛应用。然而，

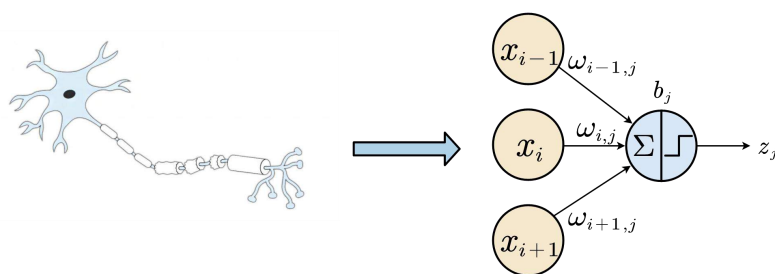


图 2.1 人工神经网络

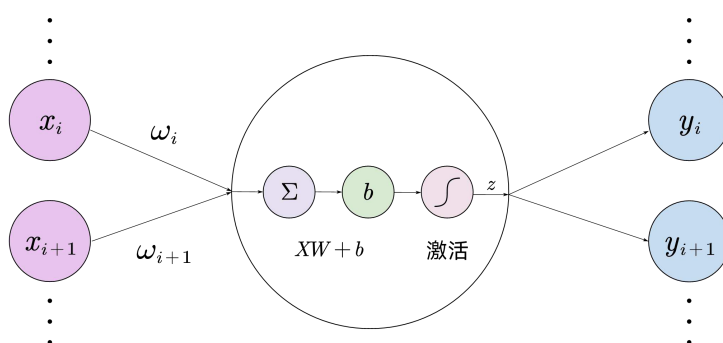


图 2.2 前向传播

研究者们发现它们的导数计算复杂度较高，并且在数据过大或过小时会趋于 0。经过发展，研究者们引入了 ReLU 来缓解计算复杂度于梯度消失的问题。如今，较为常用的是 leakyReLU。它保留了 ReLU 的优点，通过在负值区域引入较小的非零梯度来缓解神经元死亡的问题。

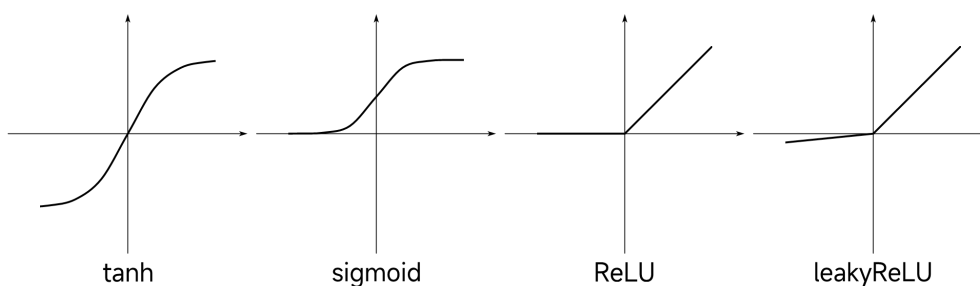


图 2.3 常见的四种激活函数

根据图 2.4 所描述的内容，反向传播首先需要度量输出层与目标的距离，即通过损失函数来计算输出层的结果相对于目标的偏差，其次利用链式法则计算损失函数相对于各个权重参数的梯度。具体而言，首先计算损失函数 L （如均方误差，交叉熵等）与 L 对于输出层的梯度。然后从输出层出发，利用链式法则逐层向前计算梯度，直到输入层。得到所有的梯度后，使用优化算法（如梯度下降，Adam 等）来更新权重，更新幅度可以通过学习率来控制。

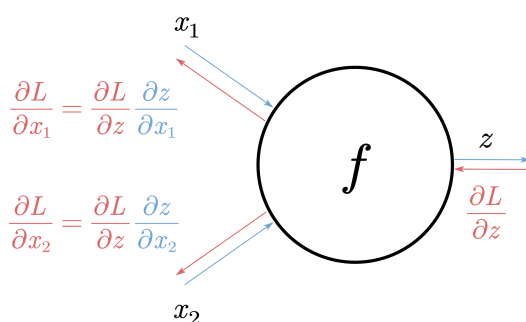


图 2.4 反向传播

2.2.2 卷积神经网络

卷积神经网络 (Convolutional Neural Network, CNN) 是一种专门针对二维数据, 如图像, 设计的深度神经网络。其设计思路类似于人类对物体的观察机制, CNN 通过利用卷积核 (也称为滤波器) 在输入图像上滑动, 提取出各种特征图; 随后, 通过汇聚层 (或称子采样层) 保留最显著的特征; 最终, 通过全连接层进行分类。CNN 的出现源于 LeCun 等人提出的 LeNet-5^[52], 其具体结构在图 2.5 中进行了展示。后续的 AlexNet^[53] 和 ResNet^[54] 进一步推动了 CNN 的发展和应用。

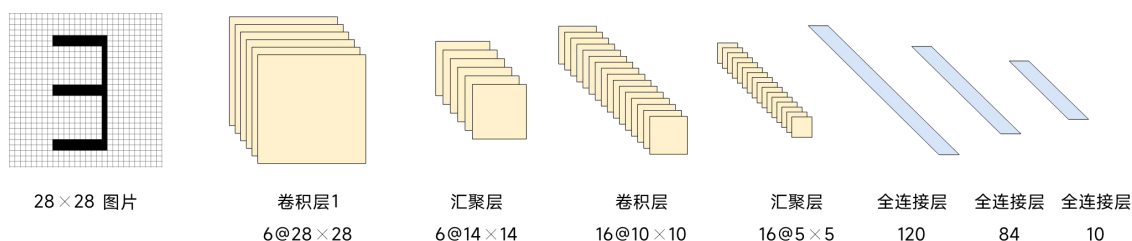


图 2.5 LeNet-5 结构

2.3 分类算法

为了更全面地评估模型的性能表现, 本文在所有数值实验中, 将同时采用支持向量机与 K-近邻作为最终的分类算法。在后续部分, 将对这两种算法的基本原理进行简要介绍。

2.3.1 支持向量机

支持向量机 (Support Vector Machine, SVM) 是一类广泛应用的监督学习算法。正如图 2.6 所展示的内容, SVM 目标是找到一个最优的超平面 (在二维数据中变为直线) 以最大程度区分不同类别的数据。在数据严格线性可分的情况下, SVM 通过最

大化数据点到超平面的距离来完成分类任务。而在数据线性不可分的情况下，SVM 引入了松弛变量和惩罚系数 c 。设置小的 c 值可以放宽分类间隔的限制，从而允许一定的分类误差；为了追求更高的分类准确度，可以适当提高 c 值，从而缩小分类间隔。对于非线性问题，可以使用核函数将数据映射到新的空间，在新的空间中寻找最优分类超平面。本文中所使用的 SVM 为线性 SVM，惩罚系数 c 设为 0.01。

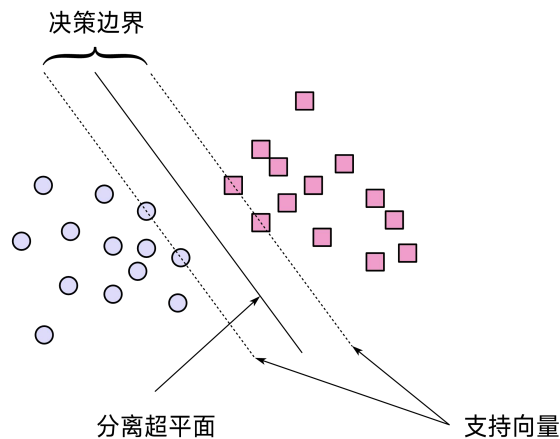


图 2.6 SVM

2.3.2 K-近邻

K-近邻 (K-Nearest Neighbors, KNN) 是一种在分类任务中广泛应用的算法。KNN 的基本理念是根据待分类样本距离最近的 K 个已知样本的类别，来判断待分类样本的类别。如图 2.7 所示，假设待分类样本为三角形，当 $K = 3$ 或 $K = 5$ 时，待分类的三角形将会被划归为最近的 3 或 5 个样本中最多的粉色圆形类别。本文所使用的 KNN 算法中，参数 K 设为 1，样本间距离的度量采用欧氏距离。

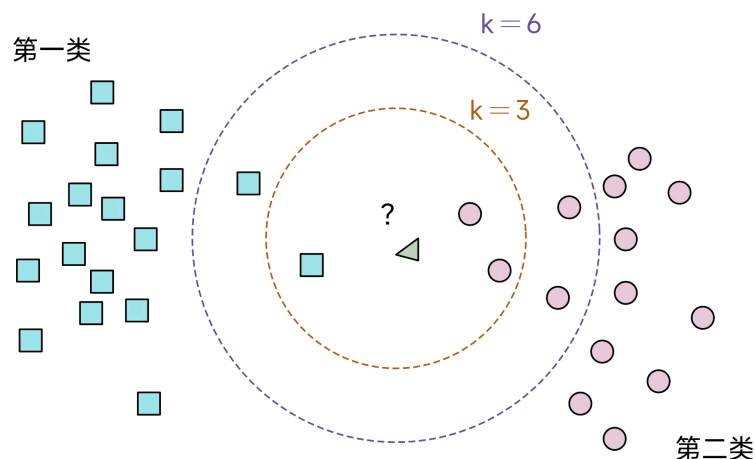


图 2.7 KNN

2.4 数据降维可视化

通常情况下，仅凭眼睛无法直接体会到 3 维以上的数据分布情况，并且高维数据在可视化时通常无法直接反映其内在结构。如图 2.8 所示，为了更好观察高纬度样本的具体情况，t 分布-随机邻近嵌入 (t-distributed Stochastic Neighbor Embedding, t-SNE)^[55] 作为一种广泛认可的高维数据降维并可视化的方法，有效地解决了观察高维特征的问题。

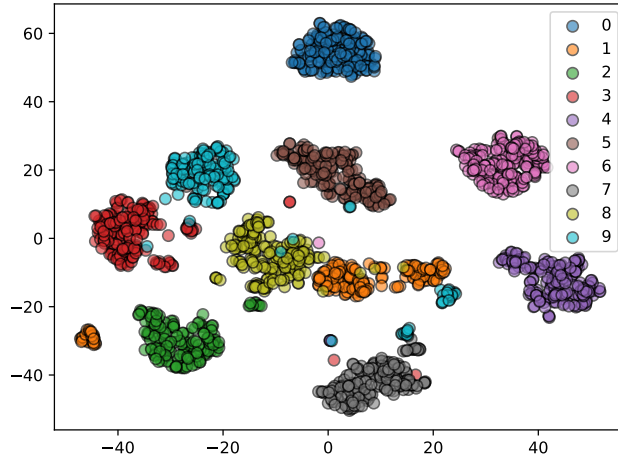


图 2.8 2D t-SNE

t-SNE 具有较强的局部结构保留能力，这使得它能够以视觉形式更真实地反映数据的分布，从而帮助研究人员直观地理解数据类别间的聚类程度。然而，t-SNE 同时面临一些挑战。首先，t-SNE 的计算复杂度较高，其计算速度受数据原始维度和计算迭代次数的影响。其次，学习率的设定对于最终结果是否收敛起着决定性的作用。最后，降维效果在很大程度上受随机种子以及“perplexity”参数的影响，需要通过多次尝试和调整才能达到较好的可视化效果。本文选择将“perplexity”参数设定为 30，并进行了 1000 次迭代计算，同时将收敛阈值设定为 10^{-7} 。

2.5 本章小结

本章首先详细说明了关于典型相关分析的数学推导，阐述算法实现的注意事项，并说明基于 CCA 的三个多视角扩展方法 MCCA、GCCA 与 TCCA 的各个特点。其次简要介绍了深度学习的基本内容和卷积神经网络的结构。最后，介绍了常见的分类算法与高维数据降维可视化算法。

第三章 基于深度典型相关分析的算法研究

3.1 深度典型相关分析

深度典型相关分析 (DCCA) 是一种将深度学习中的神经网络引入到核典型相关分析 (KCCA) 的方法, 该方法通过使用 MLP 来替代核函数对两个视角的数据进行非线性映射。如图 3.1 所示, 在进行降维映射并输出潜在的特征维度 (Latent Dimensions) 之后, DCCA 利用 CCA 最大化两个潜在特征的相关性。

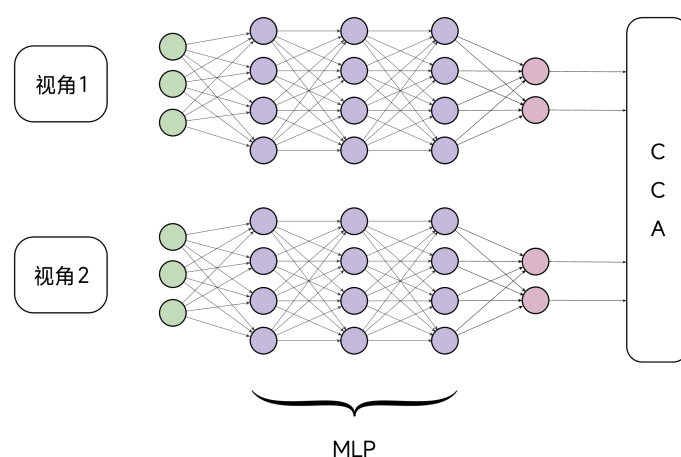


图 3.1 DCCA 结构

在深度学习中, 神经网络可以是任意的线性层或卷积层的排列组合。因此, 将 DCCA 中的 MLP 与 CCA 分别视作为深度核与相关性算法后, 可以得到图 3.2 中 DCCA 的抽象结构。

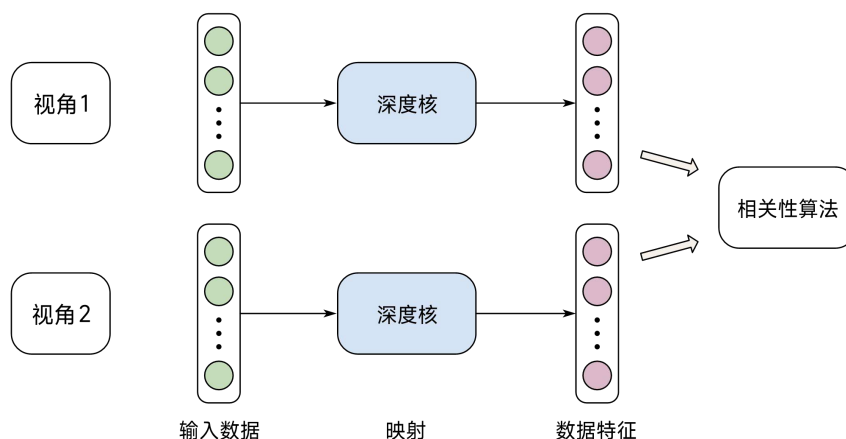


图 3.2 DCCA 抽象结构

以双视角模型为例, 设去均值后的视角 1 与视角 2 分别为 \hat{V}_1, \hat{V}_2 , 深度核映射函数为 $f_1(X), f_2(X)$, DCCA 的优化目标可以表示为:

$$\begin{aligned} \max_w \quad & f_1(\hat{V}_1)^\top f_2(\hat{V}_2) \\ \text{s.t.} \quad & f_1(\hat{V}_1)^\top f_1(\hat{V}_1) = 1, f_2(\hat{V}_2)^\top f_2(\hat{V}_2) = 1 \end{aligned} \quad (3.1)$$

上述的优化计算将通过机器学习完成, 优化的结果是神经网络各层之间的权重参数 w 。在优化完成之后, 两个视角的潜在特征 $f_1(\hat{V}_1)$ 与 $f_2(\hat{V}_2)$ 将通过 SVM 与 KNN 算法进行分类预测结果的相互校验。

3.2 学习过程

在通常的 CCA 计算过程中, 可以通过直接计算矩阵从而得到映射向量。当映射向量转化为所需要获得的对象时, 需要通过相关程度来反馈到神经网络进行学习。因此, 获得潜在特征相关程度构成了 DCCA 学习过程中的损失函数, 它的实现是通过计算在相关性最大化后的前 n 个潜在维度的特征值之和而得到。

算法 - 1: DCCA 中的相关性计算

输入: 视角 1 与视角 2 的数据 V_1, V_2 , 潜在特征维数 m

输出: 相关系数 ρ

- 1 计算样本数 n ;
 - 2 对 V_1, V_2 去均值, 得到 \hat{V}_1, \hat{V}_2 ;
 - 3 通过 $C_{ij} \leftarrow \frac{1}{n-1} \hat{V}_i^\top \hat{V}_j$ 计算协方差矩阵 C_{11}, C_{12}, C_{22} ;
 - 4 添加标量矩阵 γE 以正则化 C_{11} 与 C_{22} ;
 - 5 利用特征分解计算 $C_{11}^{-1/2}$ 与 $C_{22}^{-1/2}$, 添加误差系数 $\epsilon = 10^{-3}$ 增强数值稳定性;
 - 6 计算目标矩阵 $M \leftarrow C_{11}^{-1/2} C_{12} C_{22}^{-1/2}$;
 - 7 为增强数值稳定性, 对 $M^\top M + \gamma E$ 进行 SVD 分解, 得到奇异值矩阵 Σ ;
 - 8 对 Σ 进行数值增强, 小于 ϵ 的奇异值限制到 ϵ ;
 - 9 计算相关性 ρ 。对 Σ 对角线元素降序排列, 取前 m 个奇异值之和;
-

在机器学习中, 损失函数的值通常是非负的, 用于度量与目标的距离。为了匹配 DCCA 的需求, 在得到相关系数 ρ 后对其取反, 期望通过损失函数的最小化得到最大的 ρ 。将上述计算包装为函数后, 使用 PyTorch 内 functional 软件包中的 partial 函数来对输出特征维度求偏导。该偏导数是神经网络的目标函数, 其结果作为反向传播的起点, 在多轮迭代学习中不断使得神经网络各个参数能够最大程度地提取各

个视角的特征。

3.3 模型介绍

原有的 DCCA 仅使用 MLP 作为深度核对原视角进行映射，因此本文对 DCCA 进行了一定程度的改进，包括将深度核扩展为更深层 MLP 的 DCCA，添加卷积核或残差块（Residual Block）增强特征提取的 ConvCCA 与 ResCCA。本节实验中，为便于将 CCA 作为基准算法进行对比，所有模型均设为双视角模型。

ConvCCA 的深度核包含若干个卷积序列层。卷积序列层基于 VGGNet 进行修改，每个卷积序列层均由一个卷积核尺寸为 3×3 ，步长（Stride）为 1，填充（Padding）为 1 的卷积层与二维批标准化（2D Batch Normalization）层构成，激活函数选择斜率为 0.01 的 leakyReLU。ConvCCA 在最后一个卷积序列层之后使用二维自适应均值汇聚层（2D Adaptive Average Pooling）来结合最终的全连接层，最后输出具有期望维度数的潜在特征。

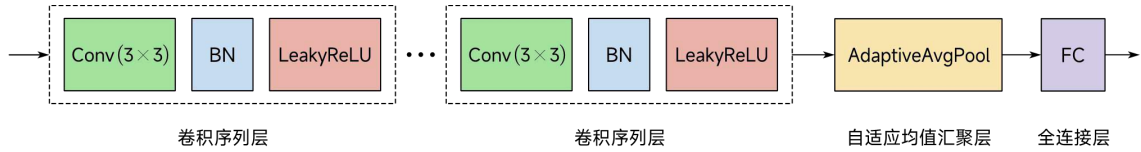


图 3.3 ConvCCA 深度核结构

ResCCA 的残差块基于卷积序列层进行拓展，在激活函数之后继续沿用了同样规模的卷积层与批标准化层，并额外将输入数据经过卷积核为 1 的卷积层，最后将两者数据求和后再次接入激活函数。ResCCA 中深度核的起始层是核尺寸为 7×7 的卷积层序列层，经过最大汇聚层之后接入两个相同规模的残差块，后续的网络结构与 ConvCCA 相同。

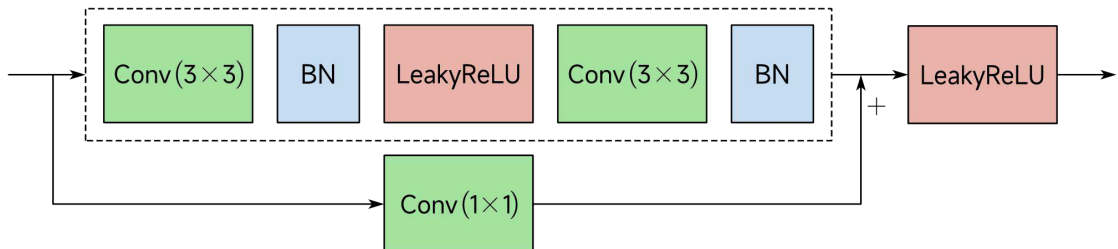


图 3.4 残差块结构

本文默认使用的 ConvCCA 为 4 层卷积序列层，卷积层输入通道数分别为 1、32、64 与 128，输出通道数为 32、64、128 与 128。ResCCA 起始卷积层的输出通道数为

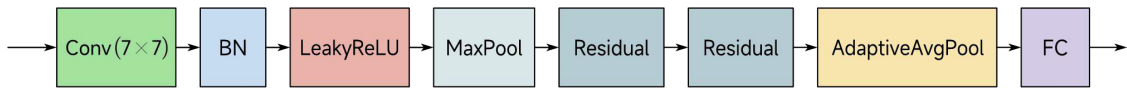


图 3.5 ResCCA 深度核结构

32，两个残差块的输入通道数分别为 32 与 64，输出通道数分别为 64 与 128。

3.4 数值实验

为了验证各个模型的性能，探究模型参数对最终结果的影响，本文将在不同的数据集来进行数值实验。如无特殊说明，所有的数据集均按照比例为 6:2:2 划分训练、验证与测试集，所有的图像缩放策略均采用双线性插值，优化器（Optimizer）采用 Adam 算法，学习率（Learning Rate）为 10^{-3} 。数据载入器在样本数未达到批大小时选择舍弃末尾（Drop Last）策略。此外，在表格中，使用红色表示最佳结果，其次结果使用蓝色表示。

3.4.1 实验环境配置

在实验过程中的硬件与软件环境配置如下：

表 3.1 实验环境

实验环境	描述
操作系统	Windows 11 22H2
CPU	Intel(R) Core(TM) i5-12600KF
显卡	NVIDIA RTX 3080
编程语言	Python 3.8
软件包	PyTorch,torchvision,numpy,mvlearn,opencv-python

3.4.2 分类准确度分析

3.4.2.1 Noisy MNIST/FashionMNIST/EMNIST 数据集

MNIST 是手写数字图片集合，有 60000 张训练与 10000 张测试样本，图片尺寸为单通道 28×28 。Fashion MNIST 是服装品类数据集，样本数与图片规模与 MNIST 相同。EMNIST 对 MNIST 进行扩展，引入了不同风格的手写字母与数字，样本数相较 MNIST 有巨大提升。其中 EMNIST 的 Letters 子类拥有 124800 个训练集样本与 20800 个测试集样本，总计 26 种分类。

本文对 MNIST、Fashion MNIST 与 EMNIST（仅 Letters 子类）进行噪声化，生成具有原始与噪声两个视角的 Noisy MNIST、Noisy Fashion MNIST 与 Noisy EMNIST

数据集。如图3.6、3.7与3.8所示，噪声样本是抽取同类别的另一个样本并加入均匀分布噪声。此外，原始与噪声样本都进行 -45° 到 45° 的随机旋转。在本文中，三个数据集的测试集沿用原始的数据，训练与验证集统一按照原始训练集的 8:2 进行划分。



图 3.6 Noisy MNIST 的部分样本

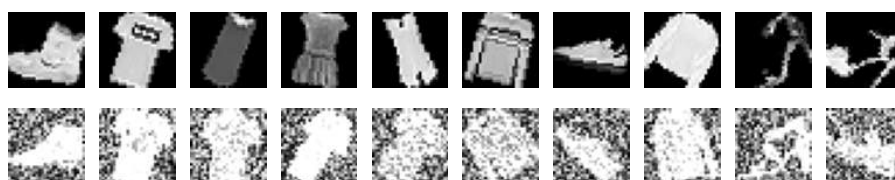


图 3.7 Noisy FashionMNIST 的部分样本



图 3.8 Noisy EMNIST 的部分样本

对于 Noisy MNIST 和 Noisy Fashion 数据集，设定所有模型输出潜在特征的维数为 10，并且深度模型 DCCA、ConvCCA 与 ResCCA 统一设置数据加载器的批尺寸 (Batch Size) 为 400。对于 Noisy EMNIST 数据集，考虑到其类别和样本量翻倍，令潜在特征维度为 26，批尺寸为 800。对于 DCCA，设其 MLP 的每层神经元数量为 1024。ConvCCA 和 ResCCA 采用本章介绍的默认结构。在学习完成后，将两个视角的潜在特征进行平均融合后视为最终的提取特征。所有深度模型的迭代学习次数 (Epoch) 统一设为 20 次。对于 CCA 和 DCCA，将图像展平为 1 维数据后输入模型。对于卷积模型，图像不进行额外处理。

根据表 3.2 的数据，可以观察到相比于 CCA 方法，DCCA 在三个数据集上均展现出一定程度的性能提升。其中在 Noisy Fashion MNIST 数据集上的提升幅度最小，

表 3.2 不同模型在 Noisy 数据集的分类结果

模型	Noisy MNIST		Noisy Fashion MNIST		Noisy EMNIST	
	KNN	SVM	KNN	SVM	KNN	SVM
CCA	80.15%	84.67%	72.48%	77.94%	36.33%	53.04%
DCCA	89.51%	87.89%	76.95%	78.02%	75.77%	73.44%
ConvCCA	93.89%	92.37%	80.33%	80.51%	93.12%	93.29%
ResCCA	95.93%	90.69%	81.42%	81.95%	93.30%	90.85%

仅 0.08 %，而在 Noisy EMNIST 数据集上的性能则显著优于 CCA 方法，提升幅度高达 39.44%。所有数据集的最优结果均由使用卷积层的 ConvCCA 与残差块的 ResCCA 模型所取得，这一事实有效地证明了引入更为先进的深度核能显著提升提取的潜在特征。在这些模型中，ResCCA 的表现最为优异，其中在 6 次分解结果中取得 4 次最优，其次为 ConvCCA 模型。不同的分类算法具有显著的差异，CCA 方法在 Noisy EMNIST 上的两种算法结果存在明显的差别，最大差距达到 16.17 %，最优的 ResCCA 模型在两种分类结果中也存在最高 5 % 的差异。

为了进一步探究各个模型提取的结果在高维空间的局部结构，画出这四个模型在输出潜在特征后的 2D t-SNE 图来进行分析。

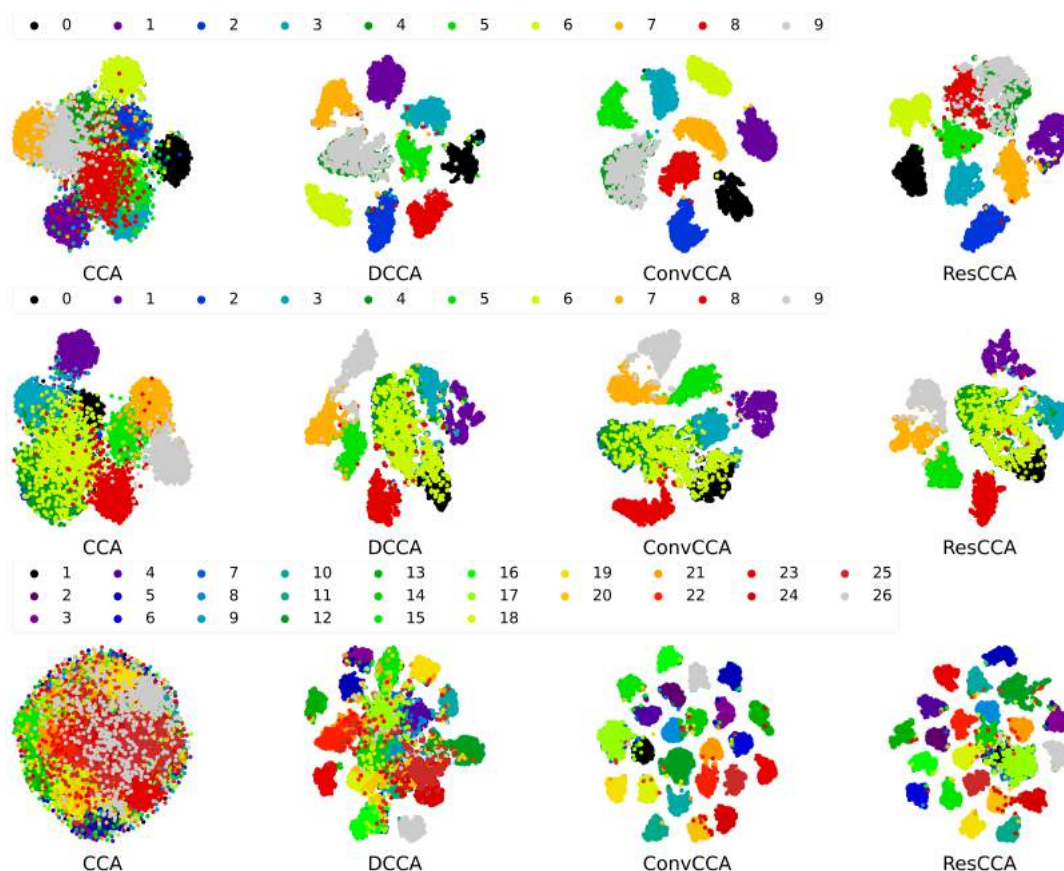


图 3.9 Noisy 测试集在各模型中的 2D t-SNE 图

从图 3.9 可以观察到, CCA 方法在三个数据集的特征降维结果分布相对较为紧凑。在输出潜在维度较低的分类任务中, CCA 方法能够对不同类别的样本进行初步区分。然而, 在输出维度较高的分类任务中, CCA 方法在将各个样本进行聚集与放大差别的能力上相对较弱, 这导致 KNN 算法在紧凑的分布中表现出较低的拟合结果。与此同时, 三种基于深度学习的 CCA 方法都较好地将不同类别的样本进行聚集, 并显著放大它们的差异。特别地, ConvCCA 在 Noisy EMNIST 数据集集中的表现最优, ResCCA 的结果与 ConvCCA 相似, 两者都优于仅使用 MLP 的 DCCA, 这一点也在分类算法的结果中得到了体现。此外, 分类结果也受到数据集本身特征的影响。在 Noisy Fashion MNIST 数据集中, 三个深度模型都无法较好地区分标签 0 与标签 6 的差异。在 Noisy MNIST 数据集中, 无论是基于 MLP、卷积或残差的深度核都在区分标签 4 与标签 9 时遇到困难, 两者的分布高度重叠, 这影响了最终的准确度。

进一步地, 本文还探索了将单视角的图片经过噪声加入和随机旋转后作为第二视角的 Noisy 数据集。在此数据集下, 所有的深度模型的表现都未能达到之前的结果。这可能是因为多视角学习的上下文中, 第二视角作为同类标签的其他样本更能满足“一个目标多个角度描述”的需求, 使得相关性算法能够在此基础上挖掘更多的关联信息。

3.4.2.2 COIL-20 数据集

COIL-20 数据集是一种物体图像数据集, 其中包含了 20 种不同的物体, 每种物体都有 72 张从不同角度拍摄的图片。该数据集总共包含 1440 个样本, 每个样本的尺寸为 128×128 , 如图 3.10 所示。在处理这个数据集的过程中, 所有的图像都被转换为灰阶, 并计算了图像的二次 Symlets、Haar、Coiflets 和 Daubechies 小波变换结果。在实验中, 灰阶图像与各种变换结果被联合作为两个视角的输入。

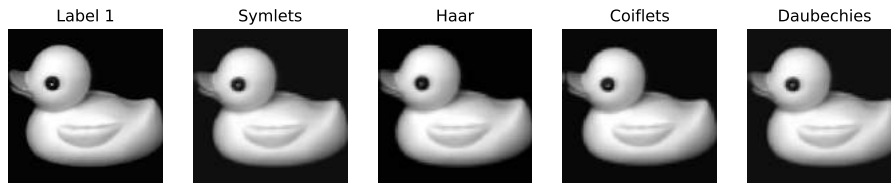


图 3.10 COIL-20 数据集

在处理 COIL-20 数据集时, 潜在特征输出维度被设定为 80。鉴于输入图像的尺寸最小为 64×64 , 如果将每个像素点视作一个维度, 总维数将超过样本数量, 进而引发“高维度小样本问题”。为了缓解这一问题, 每个输入图像都经过了 PCA 降维处理, 将其维度降低至 150。对于 ConvCCA 和 ResCCA, 输入图像的尺寸保持不变。此外, 训练的批大小被设定为 100, 迭代学习次数为 20。考虑到测试集中样本量相对

较少，为减轻过拟合的风险，训练、验证和测试集的比例被调整为 5:1:4。同时，在 DCCA 中，MLP 的神经元节点数量被调整至 250，层数调整为 2，以进一步减轻过拟合。

表 3.3 不同模型在 COIL-20 数据集的分类结果

视角 1	视角 2	CCA		DCCA		ConvCCA		ResCCA	
		KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
gray	coiflets	96.53%	94.97%	97.74%	93.75%	98.26%	94.10%	100.00%	97.05%
gray	daubechies	96.01%	94.27%	97.40%	94.79%	98.61%	93.75%	100.00%	96.53%
gray	haar	96.88%	95.31%	97.92%	94.27%	98.44%	95.83%	100.00%	97.22%
gray	symlets	96.70%	94.97%	98.09%	94.44%	98.44%	95.49%	100.00%	96.35%

表 3.3 揭示，在 COIL-20 测试集样本数量较少的条件下，利用 KNN 算法更有可能实现较高的准确性。随着模型复杂度的增加，采用 KNN 进行分类的结果也相应上升，并在 ResCCA 中出现了 100% 的准确率。在对比 CCA 与 DCCA 在 SVM 分类下的结果时，观察到基于 MLP 的深度核由于训练样本不足，所学习的参数并未能有效地挖掘视角间的关联特征，这间接揭示了深度学习模型对高样本量的依赖性。然而，基于 CNN 的 ConvCCA 与 ResCCA 在大部分情况下能够实现高于 DCCA 的 SVM 准确性。

同时，本文生成了各模型的 2D t-SNE 表示，以便对潜在特征的结构进行更深入地观察。

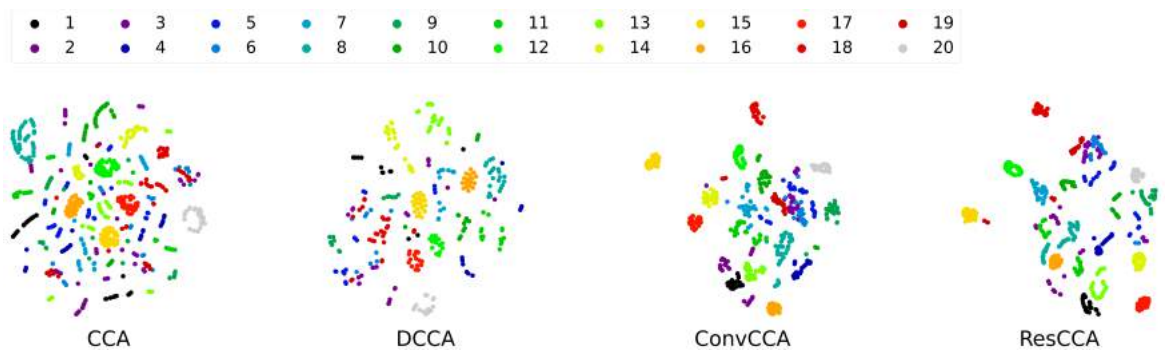


图 3.11 COIL-20 测试集在各模型中的 2D t-SNE 图

如图 3.11 所揭示的测试数据分布情况，可以观察到 CCA 使得各类别的分布相对疏散，而 DCCA 能够对相同标签的样本实现初步聚集，进一步地，ConvCCA 与 ResCCA 进行了更精细的类别分隔。然而，在低样本量及多标签的双重影响下，各类别的分布仍然相对混乱，这与 Noisy 数据集中各模型能显著呈现出不同标签数据的区域边界形成鲜明对比。此外，深度模型仍然受到学习次数等因素的限制。例如，如图 3.12 所示，在预处理过程中为 ResCCA 模型输入灰度与 Symlets 变换的两个视角，

设置学习次数为 800 次，最终得到的 KNN 与 SVM 的分类准确度分别为 99.83% 与 99.48%，这在一定程度上解释了在本研究所采用的 KNN 参数条件下，分类结果出现“完全无误”的偶然情况。

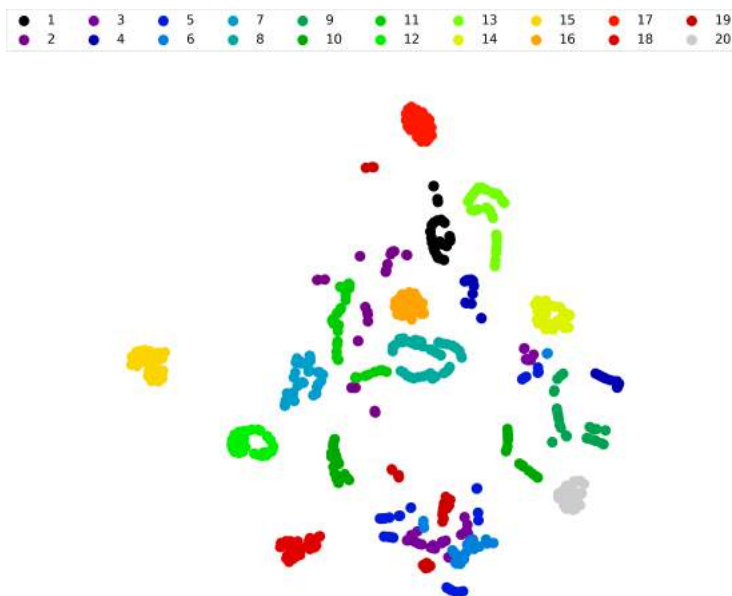


图 3.12 迭代 800 次的 ResCCA 模型

3.4.3 网络复杂度与输出特征数分析

为了深入探究深度模型在不同规模下对潜在特征挖掘的能力，本节在 Noisy MNIST 数据集上进行了数值实验。实验通过调整输出特征数和网络复杂度，研究这些变量对最终分类精度的影响。在输出特征数的实验中，设潜在特征维数从 10 到 100，步长为 10，同时保持 CCA、DCCA、ConvCCA 和 ResCCA 的规模不变。在网络复杂度实验中，对于 DCCA 和 ConvCCA 模型，将 MLP 和卷积序列层的层数逐步从 2 增加至 10，其中，ConvCCA 卷积层输出通道数在第 5 层及以后保持为 256 层。

图3.13 展示了输出特征数对分类精度的影响。除了使用 SVM 的 CCA 方法在大多数情况下保持稳定以外，其他模型组合的精度随着特征维度的增加而降低。特别地，使用 SVM 的 ConvCCA 在特征维度增至 60 时，分类精度出现了显著下滑。作为对比，ResCCA 在维度小于 70 时仍能保持 80% 的正确率。以神经元为基础的 DCCA 在两种分类算法组合下，其精度均随着潜在维度的增加而缓慢下降。此外，在特征维度低于 50 时，ConvCCA 和 ResCCA 的分类精度显著超过了 CCA 和 DCCA。在三个深度模型中，KNN 的表现在大部分情况下比 SVM 的结果更优，在 DCCA 的变化趋势中可以很明显的观察到两者之间的差别。

图 3.14 揭示了网络层数对分类精度的影响。在网络层数为 2 时，DCCA 在两种算

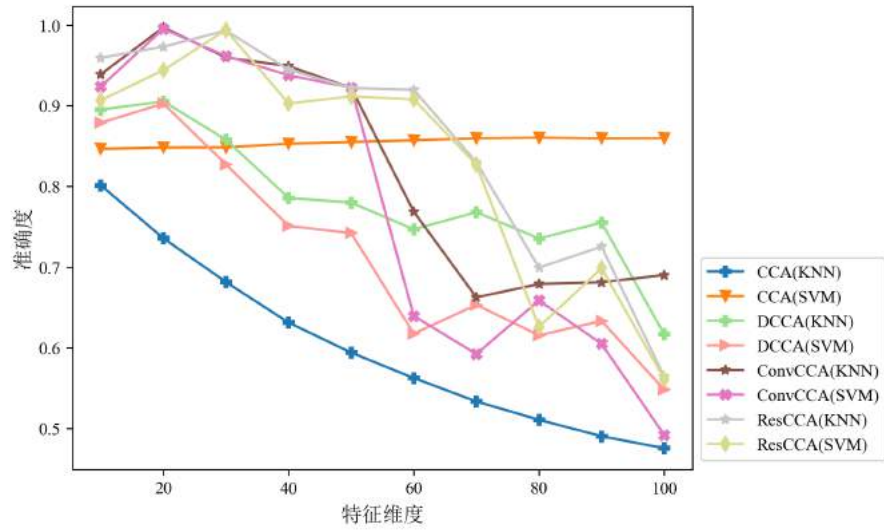


图 3.13 输出特征数对准确度的影响

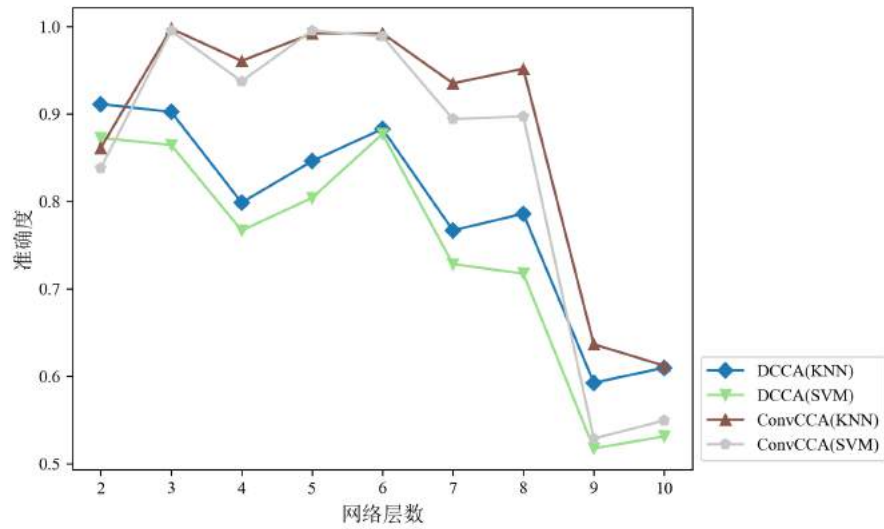


图 3.14 网络层数对准确度的影响

法下均优于 ConvCCA。然而，随着网络深度的增加，所有模型组合的精度都呈现下降趋势。在网络层数达到 9 时，DCCA 和 ConvCCA 的精度出现显著下滑。ConvCCA 在 3 到 6 层表现最佳，而 DCCA 则在 2 到 3 层表现最好。这些结果暗示，在大部分情况下，提高一定程度的网络层数，可以使基于卷积或残差的模型比感知机模型更好地挖掘两个视角的潜在关联。然而，复杂度的提升会带来训练参数量的翻倍，极大延长训练学习的时间，同时可能导致模型中存在更多的欠学习参数，对最终的表现提升相对有限，可能变为冗余参数。

3.4.4 样本批大小分析

批大小作为数据加载器的关键参数，决定了进行相关性计算的样本数量。为了深入研究其对最终识别结果的影响，本节以 Noisy MNIST 数据集为基础进行了一系列试验。在保持输出特征维度、训练次数和学习率不变的情况下，批大小从 100 调整至 2000，步长为 100。

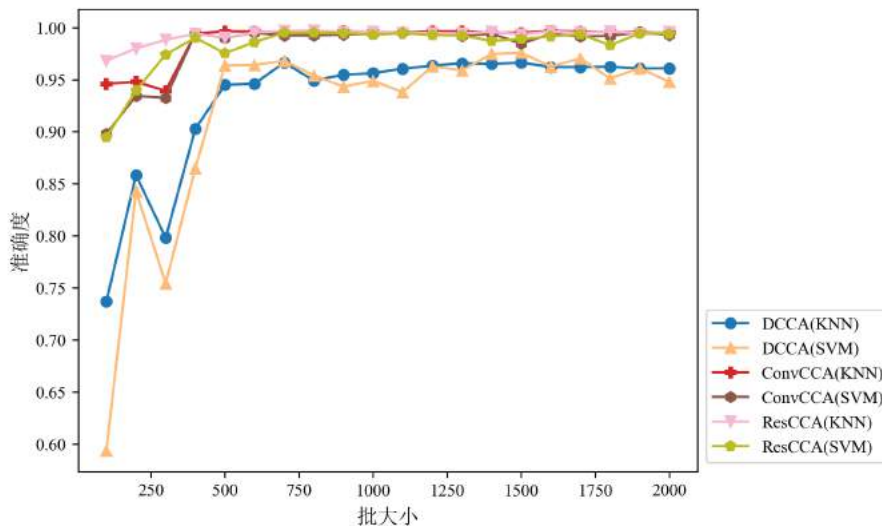


图 3.15 批大小对准确度的影响

图 3.15 显示了批大小对分类精度的影响。可以明显看出，批大小的增加在很大程度上提高了各个模型的识别精度。在 DCCA 中，批大小从 100 变更至 500 时，分类精度的提升最为显著，最高提升率达到 20.8%。相比之下，基于 CNN 的 ConvCCA 与 ResCCA 由于在较低批大小时就已经达到较高的准确率，所以随着批大小的增加，准确度的提升相对有限。然而，ConvCCA 与 ResCCA 在所有的批大小中都优于 DCCA，这一定程度上验证了这些改进方法的有效性。

3.5 本章小结

本章首先引入了深度多视角学习算法 DCCA 的基本结构和原理, 阐明了 DCCA 的目标函数在机器学习训练过程中的应用, 深入探讨了 DCCA 损失函数的具体实现。同时, 也讨论了利用 CNN 和残差块构建 DCCA 模型的可能性, 并在四个不同的数据集上进行了实验验证。实验结果表明, 经过 CNN 扩展后的 DCCA 能够超越原始的 CCA 和 DCCA 的表现。此外, 通过对数据进行 2D t-SNE 降维, 可观察到改进后的模型能使得相同标签的测试样本在局部空间中更紧密地聚集。最后, 本章还探讨了输出特征维度、深度模型的复杂度以及批大小选择对最终结果的影响。

第四章 基于深度典型相关自编码器的算法研究

4.1 判别式深度典型相关自编码器

深度典型相关自编码器 (DCCAE) 是对 DCCA 过程的一种扩展。DCCAE 不仅通过 MLP 对各个视角的数据进行编码, 得到各自的潜在特征后利用 CCA 最大化特征之间的相关性, 还额外添加若干 MLP 来对特征进行重建。DCCAE 在 DCCA 相关性约束的基础上, 使重建后的视角与原视角的偏差尽可能小, 从而最大保留原视角的固有特征。

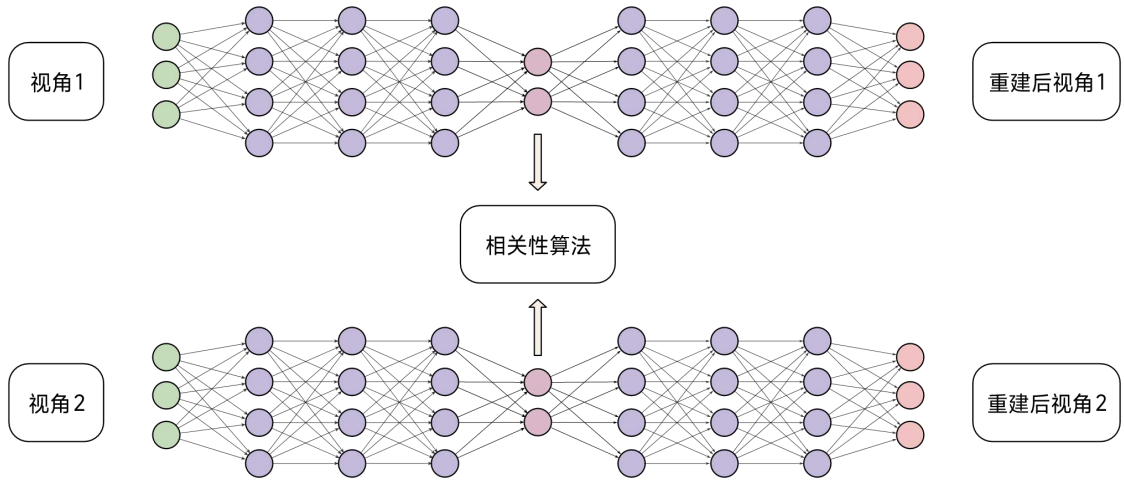


图 4.1 DCCAE 结构

设两个去均值的视角分别为 \hat{V}_1, \hat{V}_2 , 编码后的潜在特征为 $f_1(\hat{V}_1), f_2(\hat{V}_2)$, 解码后的视角重建为 $g_1(\hat{V}_1), g_2(\hat{V}_2)$ 。本文以均方误差 (Mean Squared Error, MSE) 来评估重建前后各个视角的差异。DCCAE 在 DCCA 的目标函数基础上引入了重建项与权重参数 λ , 最终形式如下:

$$\begin{aligned} \max_w \quad & -\frac{\lambda}{N} \left(\|\hat{V}_1 - g_1(\hat{V}_1)\|^2 + \|\hat{V}_2 - g_2(\hat{V}_2)\|^2 \right) + f_1(\hat{V}_1)^\top f_2(\hat{V}_2) \\ \text{s.t.} \quad & f_1(\hat{V}_1)^\top f_1(\hat{V}_1) = 1, f_2(\hat{V}_2)^\top f_2(\hat{V}_2) = 1 \end{aligned} \quad (4.1)$$

上述公式中, 前置于 MSE 的 λ 是一个需调整的超参数, 代表视角重建对整个学习过程的影响程度。经过实验调整后, 本文默认 $\lambda = 0.001$ 。在原优化目标的基础上增加带权重的 MSE 的相反数, 使得模型在学习过程中不断缩小重建距离。

鉴于 DCCA 与 DCCAE 本质上都是无监督的, 并未利用标签信息进行训练。为了探索引入标签信息对各个模型的提升, 同时尽可能保留无监督给予 DCCA 与 DCCAE

的优点，本文引入了弱监督学习（Weakly Supervised Learning）中的不完全监督学习（Incomplete Supervised Learning）。随机抽取训练集中一定比例的标签并设为“-1”，表示该样本标签无效。与其他有监督引入方式相同，将两个映射后的视角特征进行拼接，计算拼接视角与标签信息的交叉熵损失（Cross Entropy Loss），并与原损失函数共同引导每层网络参数的学习，忽略所有值为“-1”的标签。

在 DCCAE 中，深度核被抽象为一种编解码器，负责对目标数据进行特征提取与恢复。在原始结构的基础上对 DCCAE 结构引入标签，得到了监督深度典型相关编码器（Supervised DCCAE, SDCCAE）的抽象结构。

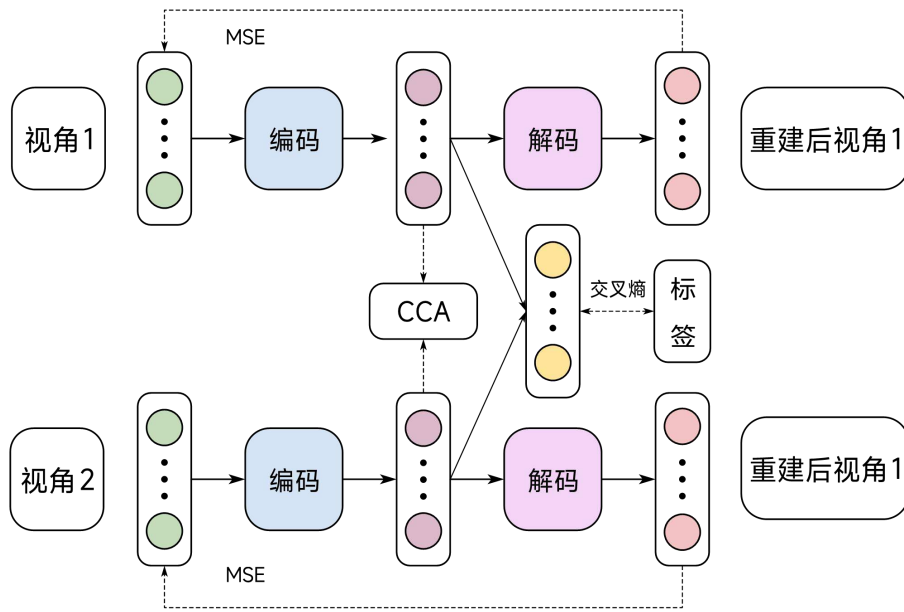


图 4.2 SDCCAE 结构

4.2 学习过程

DCCA 与 DCCAE 均使用双视角数据进行学习，然而，在视角数大于 2 时，基于 CCA 的损失函数将无法计算。为应对这种情况，本章默认采用 MCCA 作为所有模型损失函数中的相关性算法。此外，依旧使用相关系数取反的方法，以期望最大化编码后特征的相关性。同时，对 DCCAE 及其扩展模型均生成了一次学习次数为 20 次的预训练模型。在预训练的损失函数中，只计算了各个视角映射与重建前后的 MSE 之和，期望使编解码网络能够初步学习到视角变化的特征。

算法 - 2: 三视角的 MCCA 损失函数计算**输入:** 视角 1、视角 2 与视角 3 的数据 V_1, V_2, V_3 , 潜在特征维数 m **输出:** 相关系数 ρ

- 1 计算样本数 n ;
- 2 对 V_1, V_2, V_3 去均值, 得到 $\hat{V}_1, \hat{V}_2, \hat{V}_3$;
- 3 拼接矩阵 $\hat{V}_1, \hat{V}_2, \hat{V}_3$ 得到总视角矩阵 \hat{V} ;
- 4 计算所有视角的协方差矩阵 $C \leftarrow \frac{1}{n-1} \hat{V}^T \hat{V}$;
- 5 计算并正则化视角内的协方差矩阵 D , D 为矩阵 C 的对角块矩阵;
- 6 将 C 对角块的自协方差矩阵用 D 替代, 以增强数值稳定性;
- 7 计算 $R \leftarrow D^{-1/2}$, 误差系数 $\epsilon = 10^{-3}$;
- 8 计算 $\hat{C} \leftarrow RCR^T$ 并对 \hat{C} 特征分解, 得到特征值矩阵 Λ ;
- 9 对 Λ 对角元素降序排列并令其大于 ϵ ;
- 10 计算相关系数 ρ , ρ 为 Λ 中前 m 个元素的和;

4.3 模型介绍

在本研究的实验中, 继续采用了 CNN 的网络结构, 并将卷积核与残差块编码器引入到 DCCAE 中, 由此得到了 ConvCCAE 与 ResCCAE 两个模型。ConvCCAE 与 ResCCAE 的编码部分与 ConvCCA 和 ResCCA 的网络结构相同, 而解码部分与 DCCAE 一致, 皆为多层的 MLP。针对这些模型, 构建了相应的监督模型 SDCCA、SDCCAE、SConvCCAE 与 SResCCAE。

4.4 数值实验

为了对各模型的性能进行初步探索, 选择了 MNIST-USPS-SVHN 数据集和具有一定挑战性的 ETH-80 数据集进行实验。

4.4.1 分类准确度分析

4.4.1.1 MNIST-USPS-SVHN 三视角数字数据集

MNIST-USPS-SVHN 是一个三视角数字图像数据集, 由 MNIST、USPS 和街道数字数据集 SVHN 的组合构成。为了消除各数据集之间的差异, 对 SVHN 的三通道进行了压缩, 得到单通道灰度图, 并将所有样本的尺寸调整为 28×28 。鉴于 USPS 数据集的样本总数最少, 为了最大化数据的利用, 调整了其他两个数据集的样本数, 使之与 USPS 数据集相匹配, 并按照 6:4 的比例划分了训练集和验证集。在卷积模型

中将数据集中的图像恢复为原始尺寸。



图 4.3 MNIST-USPS-SVHN 数据集的部分样本

对于深度模型, 由于在本文的设置中, DCCAE 及其扩展模型的训练总次数设定为 40。为了尽可能保持不同模型之间的训练环境参数一致, 本章将其他基于 DCCA 的扩展模型的训练次数也设定为 40。同时, 设定批尺寸为 1024, 输出特征数为 10。DCCA 以及扩展模型的神经元数量为 1024, 层数为 3。DCCAE 的编码和解码部分与 DCCA 的网络结构相同。ConvCCAE 和 ResCCAE 的编码部分采用默认的 ConvCCA 和 ResCCA, 解码部分与 DCCAE 使用相同的结构。学习过程中的实验软件、硬件环境以及优化器参数等条件均保持一致。

表 4.1 不同模型在 MNIST-USPS-SVHN 数据集的分类结果

模型	无监督		完全监督	
	KNN	SVM	KNN	SVM
DCCA	75.34%	79.52%	84.85%	85.00%
DCCAE	79.12%	81.46%	86.70%	88.19%
ConvCCA	90.73%	82.61%	95.47%	96.16%
ResCCA	96.96%	91.98%	97.91%	98.06%
ConvCCAE	98.31%	96.31%	97.71%	94.87%
ResCCAE	97.36%	92.63%	97.96%	91.08%

从表格数据可以看出, ConvCCAE 模型在无监督模型中取得最高的识别率, ResCCA 在完全监督模型中取得了显著的提升, 分类准确率位列第一。与 DCCA 对比, 可以发现, DCCAE 在加入解码部分和重建约束后取得了一定的性能提升, 提升幅度在 1.94 % 到 3.19% 之间。在将 DCCAE 的编码部分使用基于卷积的模型进行改进后, 在两种算法的结果下都有明显的性能提升, 在无监督条件下提升幅度最高达到 19.19 %。相比于无监督情况, DCCA 与 DCCAE 在有标签的指导下, 两种分类算法的识别准确度都有显著的改善。对于卷积模型, 除了 ConvCCAE 以外, 其他模型都在一定程度上受益于标签约束的引入, 从而提高了识别的准确性。其中提升最显著的是使用 SVM 的 ConvCCA 模型, 提升幅度高达 13.55 %。

为了更深入地比较深度模型的结果, 画出各方法得到的特征在降维后的分布情

况。作为参考，额外采用 MCCA 对数据集进行拟合，将原始视角降到 10 维后使用平均融合。得到的结果是 MCCA 在 KNN 与 SVM 的准确度分别为 43.80% 与 54.70%。这能够辅助其他深度模型进一步解释各分类算法的结果。

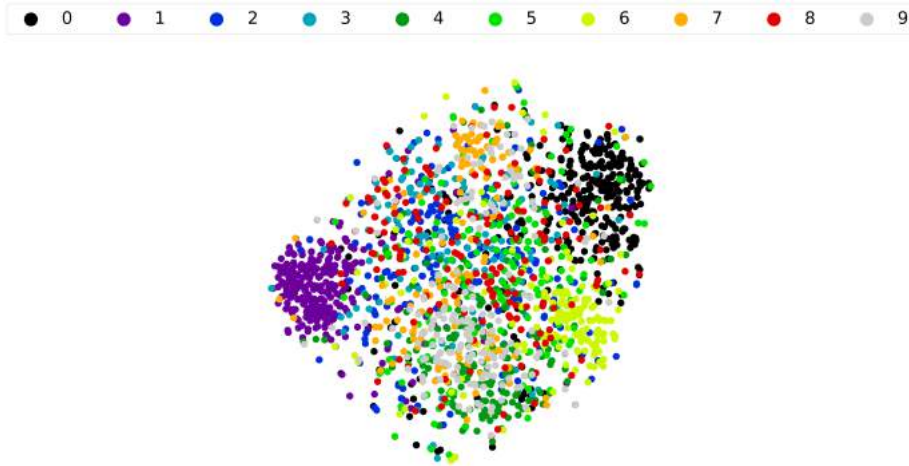


图 4.4 MNIST-USPS-SVHN 测试集在 MCCA 中的 2D t-SNE 图

由图 4.4 可知，MCCA 延续了 CCA 的特征提取分布规律，各类别数据分散，而且两种分类算法都取得了最低的识别准确率。根据图 4.5 可以观察到，DCCA 和 DCCAE 能在 MCCA 的基础上将标签 0 和 1 的样本尽可能地聚集在一起，DCCAE 相比 DCCA 更多地捕捉到了标签 2 的特征，能较好地进行分类。这两种模型在完全监督条件下，都能进一步学习到区分其他标签的映射方法，各类别的边界区域逐渐显现。而 ConvCCA 和 ResCCA 在无监督条件下已经能较好地地区分大部分测试样本。在加入监督信息后，ConvCCA 和 ResCCA 能在原本的基础上缓解标签 3、5、8 和 9 特征提取不足的问题，区分能力明显提升。表现优异的是 SResCCA 和 ConvCCAE，能明显观察到各类别的具体位置，这与它们最高识别率的实验结果一致。

4.4.1.2 ETH-80 物体数据集

ETH-80 数据集是一种常见物体的彩色图片数据集。该数据集包含 8 个大类别的物体，每个大类别下有 10 个小类别，每个小类别包含了从不同视角拍摄的 41 张图片，总共包含 3280 张图片。如图 4.6 所示，所有图片均为三通道的 128×128 图像。在本节实验中，选取图片的 RGB 通道作为三个输入视角。

本节实验选择对 8 大类物体进行识别。对于每个大类中的小类样本，在预处理中选择将其合并。为了尽可能地降低高维小样本的影响，将所有图片的尺寸调整为 32×32 。对于 MCCA、DCCA 和 DCCAE 方法，将图片的每个像素展开为一维向量后输入到网络中。而对于 ConvCCA、ResCCA、ConvCCAE 和 ResCCAE 方法，直接

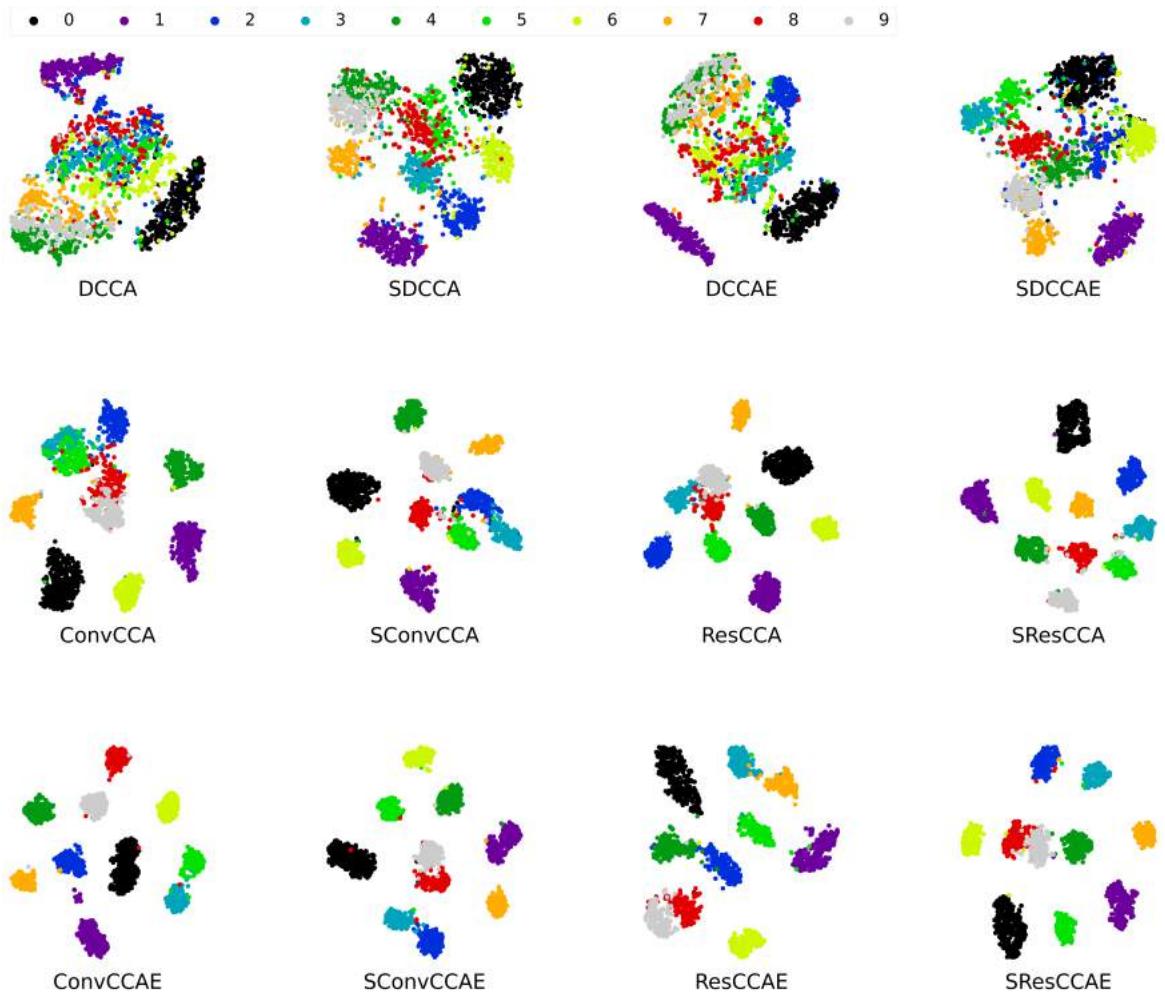


图 4.5 MNIST-USPS-SVHN 测试集在各模型中的 2D t-SNE 图

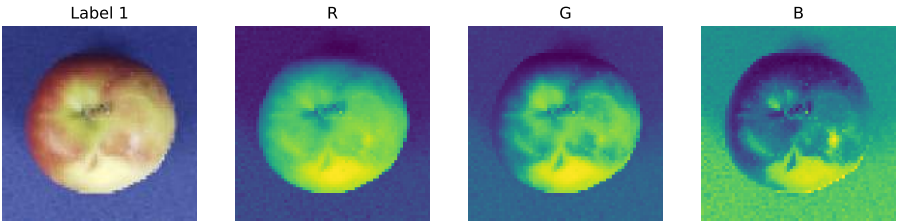


图 4.6 ETH-80 数据集

将原始图像作为输入。设定每个批次的大小为 300，网络的输出特征维数为 40。对于 DCCAE 的编解码部分与 ConvCCAE 和 ResCCAE 的解码部分，均使用了具有 3 层、每层 1024 个神经元的 MLP。

表 4.2 不同模型在 ETH-80 数据集的分类结果

模型	无监督		完全监督	
	KNN	SVM	KNN	SVM
DCCA	77.13%	80.34%	86.74%	84.45%
DCCAE	81.10%	78.81%	89.79%	87.65%
ConvCCA	88.72%	79.73%	91.16%	89.02%
ResCCA	91.62%	84.30%	94.66%	89.79%
ConvCCAE	90.09%	77.74%	90.55%	79.57%
ResCCAE	89.18%	79.73%	90.24%	82.93%

根据表 4.2 的结果可以观察到，所有模型在加入监督信息后的识别性能都有所提升，但提升幅度存在差异。在无监督学习条件下，DCCAE 和 DCCA 两种模型的表现各有千秋，然而在完全监督条件下，DCCAE 的提升幅度最为显著，最高达到 8.84%，显著优于 DCCA 的结果。ConvCCA 和 ResCCA 的提升也十分显著，尤其是在使用 SVM 时，ConvCCA 方法实现了所有模型中的最高提升幅度，达到了 9.29%。相对而言，ConvCCAE 和 ResCCAE 两种方法从标签中学习到的特征对最终结果的影响相对较小，最大的提升幅度仅为 3.2 %。从图 4.7 可以看出，加入监督信息后，类别 1 和类别 8 在 SDCCAE 和 SResCCA 方法下有明显的分界线，并且所有的模型在加入标签约束后，对类别 2 的区分都有一定程度的提升。

4.4.2 判别标签比例分析

为了具体的探究不完全监督对所有模型的影响程度，并同时提高训练的效率，本节将在 MNIST-USPS-SVHN 数据集上进行数值试验，设置有效标签的比例由 0% 提升到 25%，使用 TCCA 作为所有模型的目标函数，其余部分保持不变。

根据图 4.8 所揭示的各模型动态，可以明显观察到一个现象：随着有效标签比例的增加，所有模型在两种分类算法下的表现都体现出了上升的趋势。当有效标签比例超过 20% 时，所有模型的准确率均超过了 85%。其中，ResCCAE 表现显著，其在两种算法下，仅引入 1% 的标签信息就能达到大约 90% 的准确度。相比之下，传统的 DCCAE 模型需要引入超过 20% 的标签信息才能在两种分类准确度中达到 90%。在引入标签比例低于 10% 的情况下，通过比较 ResCCA 与 ResCCAE 可以发现在同等标签比例的前提下，后者能获得更高的准确度提升。在引入 20% 的标签时，ConvCCAE 在两种算法的表现均低于 ConvCCA。另外，无论在引入标签的情况下，ConvCCAE 与 ResCCAE 在多数情况下都表现优于 DCCA 与 DCCAE。

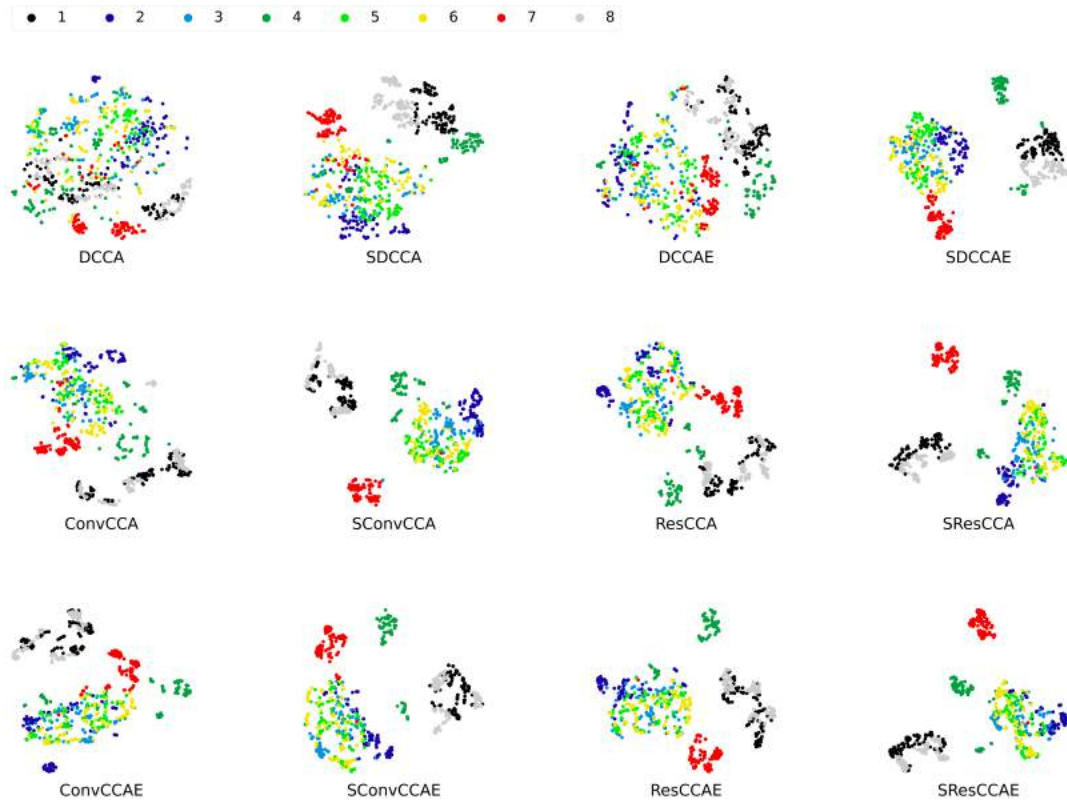


图 4.7 ETH-80 测试集在各模型中的 2D t-SNE 图

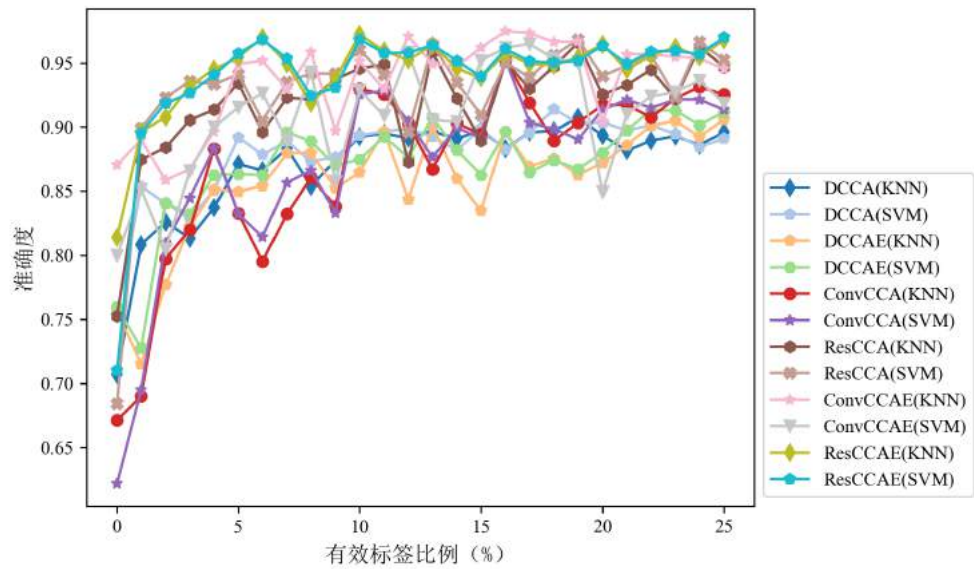


图 4.8 各个模型在不同有效标签比例的识别准确度

4.4.3 相关性算法分析

为了探究 GCCA 与 TCCA 引入深度模型后的性能表现与以及基于 MCCA 的深度模型的区别, 本节分别使用 MNIST-USPS-SVHN 与 MFD 两种数据集进行数值试验。本文对 GCCA 与 TCCA 的实现将基于文献 [56] 的研究进行修改。

对于 MNIST-USPS-SVHN 数据集, 表 4.1 提供了 MCCA 的结果。在本节实验中将进一步比较 GCCA 和 TCCA 算法在该数据集下的性能, 所有的模型均与 MCCA 的参数和批量大小保持一致。

表 4.3 GCCA 与 TCCA 算法在 MNIST-USPS-SVHN 数据集的表现

模型	无监督				完全监督			
	GCCA		TCCA		GCCA		TCCA	
	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
DCCA	80.02%	81.22%	70.70%	75.29%	78.72%	80.92%	84.35%	84.40%
DCCAE	80.37%	85.25%	75.69%	75.98%	80.57%	85.80%	90.38%	90.03%
ConvCCA	94.47%	83.31%	67.12%	62.18%	93.72%	84.50%	93.17%	92.53%
ResCCA	98.06%	92.63%	75.24%	68.41%	97.76%	92.92%	96.51%	96.26%
ConvCCAE	97.96%	94.12%	87.05%	79.97%	97.96%	88.69%	95.67%	93.97%
ResCCAE	98.51%	94.57%	81.37%	71.00%	98.11%	89.39%	96.81%	96.16%

如表 4.3 所示, 在无监督条件下, ResCCAE 在 KNN 和 SVM 下都获得了最高的准确率, 分别为 98.51% 和 94.57%。相较之下, ResCCA 和 ConvCCAE 分别在 KNN 和 SVM 下取得了次高的结果。值得注意的是, TCCA 的表现在 KNN 和 SVM 下的准确率分别达到了 87.05% 和 79.97%, 这是该算法在所有模型中的最高值。相比于 GCCA 在无监督的表现, TCCA 有着 4.68% 到 27.35 % 的差距。

在完全监督的情况下, ResCCAE 继续表现出色, 其在 KNN 下的准确率为 98.11%。然而, 尽管在无监督的条件下 TCCA 的表现不如 GCCA, 但在完全监督的情况下, TCCA 的提升幅度巨大, 在 KNN 和 SVM 下的准确率最高分别达到了 96.81% 和 96.16%。而 ResCCA 在 SVM 下达到了最高的准确率, 为 92.92%。

MFD (Multiple Feature Dataset) 是 UCI 机器学习库中的一个手写数字多视角数据集。它包含了 2000 个样本, 每个数字标签从 0 到 9, 每个标签有 200 个样本。这个数据集一共提供了 6 个视角, 每个视角各自提取了形态学、轮廓相关性和傅里叶系数等特征。在本次实验中, 将所有视角一并作为网络的输入, 并同时对比了 GCCA、TCCA 与 MCCA 分别作为目标函数的结果。鉴于所有视角的数据已经降至一维, 实验选择了 DCCA 和 DCCAE 及其完全监督模型的结果作为比较基准。

在无监督学习场景下, GCCA 在 DCCA 和 DCCAE 模型上均实现了较高的准确率, 分别达到 89.00 % 和 88.75 %, 与 MCCA 和 TCCA 方法间形成了显著的准确率差异。然而, 在完全监督学习环境中, TCCA 在 DCCA 与 DCCAE 两种模型下都获

表 4.4 无监督的 DCCA 与 DCCAE 模型在 MFD 数据集的分类结果

	MCCA		GCCA		TCCA	
模型	KNN	SVM	KNN	SVM	KNN	SVM
DCCA	81.25%	81.25%	89.00%	86.25%	81.00%	73.25%
DCCAE	85.00%	78.25%	88.75%	86.75%	86.50%	79.75%

表 4.5 完全监督的 DCCA 与 DCCAE 模型在 MFD 数据集的分类结果

	MCCA		GCCA		TCCA	
模型	KNN	SVM	KNN	SVM	KNN	SVM
DCCA	93.00%	91.75%	87.50%	83.50%	87.00%	81.75%
DCCAE	95.00%	94.25%	87.25%	84.00%	91.00%	88.25%

得了明显的性能提升，而 MCCA 则展现出了卓越的优势，其准确率分别达到 93.00% 和 95.00%。相较之下，GCCA 方法的表现相对有所下滑。

因此，通过整合表 4.3、4.4 与 4.5 中的数据，可以得出：GCCA 在无监督学习环境下的表现相对优秀，而在完全监督学习环境下，MCCA 的性能表现更加出色。另外，TCCA 在引入监督信息后能实现显著的性能提升。此外，无论在何种监督场景下，DCCAE 模型在大部分指标上的表现均略优于 DCCA，这证明了其融入的编码器结构使其能更为有效地完成多视角特征抽取任务。

此外，在所有分析过的算法中，MCCA 和 GCCA 展现出了最高的计算效率，而 TCCA 在计算优势上并未体现出明显优势。在无监督学习的条件下，TCCA 的表现相较于 MCCA 和 GCCA 显得相对较弱。然而，MCCA 和 GCCA 的性能并非始终优越，其有效性往往受数据集样本情况的影响。在某些情况下，MCCA 和 GCCA 在计算矩阵逆时可能会出现 NaN 传播，通过深入分析发现在某次训练过程中直接输出的维度即为 NaN，这暗示了这两种目标函数可能在某些情况下对神经网络的指导存在问题。相比之下，TCCA 尽管因其算法复杂度较高，计算速度相对较慢，但其稳定性卓越，并且在监督学习环境下能有效地学习并提取更多的数据集特征，这一点从前述表格中已得以证实。

4.5 本章小结

本章首先介绍了基于 DCCA 改进后的 DCCAE 的基本结构和原理，简要说明了 DCCAE 的目标函数与 DCCA 相比所作的变化。随后，讨论了在弱监督学习环境下的不完全监督学习，并将其引入到 DCCAE 的网络中，同时介绍了相关的网络结构。为了深入理解模型学习的过程，以三视角为例，阐述了多集典型相关分析（MCCA）算法的实施步骤。在此基础上，介绍了本章所提出的模型，该模型对 DCCAE 进行了一系列优化，主要包括利用卷积神经网络（CNN）和残差块进行改进。然后详细

介绍了本章所提出的模型，并针对 DCCAE 网络使用 CNN 与残差块进行了优化的 ConvCCAE 与 ResCCAE 以及对应的监督模型。为验证这些模型的性能，本章在多个数据集上进行了分类准确度实验。此外，本章进一步研究了不同标签比例对各算法模型性能的影响，并在两种数据集下探讨了不同的相关性算法对各模型性能的影响程度。最后，通过两种数据集考察了不同的相关性算法在两种数据集下对各种模型的影响程度。

第五章 总结与展望

5.1 总结

在当前的数据驱动时代，各类型数据以指数级别的速度增长，这使得多视角学习在数据处理和模型构建中的重要性日益凸显。以环境和可持续发展研究为例，大量的数据可能来自于不同的视角，如气候数据、经济数据、人口数据等。多视角学习，作为一种独特的机器学习方法，能够有效地利用这些多源数据，对目标进行精细的建模和预测，集成这些数据通常可以得到比单一视角更具鲁棒性的模型。其中，典型相关分析（CCA）作为一种统计方法，它通过计算线性投影向量以最大化两组数据在投影后的相关性，从而揭示数据的线性关系和潜在结构。

本文针对深度典型相关分析进行探究与分析，主要工作可以概括为两部分：

- (1) 深入探讨了深度典型相关分析（DCCA）算法的基本原理，并详细描述了 DCCA 的整体结构、优化目标以及损失函数的计算步骤。针对 DCCA 中深度核的局限性，本文借鉴深度学习的最新研究成果，提出了基于 VGGNet 与 ResNet 修改的卷积典型相关分析（ConvCCA）和残差典型相关分析（ResCCA）模型。在不同样本状况和样本数量的数据集上对 DCCA、ConvCCA 以及 ResCCA 进行了大量的比较实验，结果显示，利用更先进的网络作为深度核在对原始数据进行映射时，更能找到有利于 CCA 最大化相关性的特征。在此基础上，通过超参数实验进一步探究了网络复杂度和特征维度对最终目标识别的影响程度。
- (2) 详细讨论了深度典型相关分析自编码器（DCCAE）算法的基本原理，并讨论了 DCCAE 基于 DCCA 所做的扩展，详述了使用编解码器结构的 DCCAE 在优化目标上的差异。此外，本文在引入基于卷积的模型的同时，介绍了不完全监督的 DCCAE 模型及其抽象结构。接着，以多集典型相关分析（MCCA）为例，阐述了三视角数据映射后计算相关性的步骤。通过引入了一些数据集来验证各个模型在多视角条件下的性能表现，并创新性地探究了在不同不完全监督的条件下，各个模型的识别准确度。最后，为了验证不同相关性算法对网络指导能力的影响，在三视角与六视角的数据集上进行了比较实验。

将上述工作内容进行整理与汇总，对实验结果进行分析与讨论后，本文得出以下几点结论：

- (1) 相对于直接运用 CCA 与 MCCA 进行分类任务的传统方法，运用基于深度神经网络的多视角典型相关性分析模型往往能够实现更高的精确度。然而，这种性能提升在很大程度上受到神经网络的架构，以及数据集的样本特征、规模和质

量的影响。因此,寻找与数据集高度匹配的网络架构无疑构成了一个复杂而充满挑战的问题。

- (2) 当各个视角包含的目标表征更为多样化时,神经网络能够从 CCA 中获取更丰富的指导信息,以帮助寻找优化的映射参数。这一点在处理噪声手写数字、服装和手写字母数据集时得到了明显体现。当利用同类别的不同样本作为两个视角时,神经网络能更有效地学习到合适的映射策略。
- (3) 一般来说,批处理的规模越大,性能表现越好,而特征维度数量在大多数情况下,较小的规模表现更佳。批处理的大小决定了潜在特征的数量,这进一步决定了 CCA 能最大化利用的样本信息数量。样本数量越多,能够发掘的相关性就越丰富。然而,批处理的规模受到数据集和计算机硬件的限制。
- (4) 将基于深度学习的 CCA 与交叉熵形式的标签信息相结合,在大多数情况下能提升模型的性能表现。递增有效样本数量对不同模型进行训练,能够提升最终的分类结果。然而,这种改进程度依赖于数据集的选择以及目标函数的设定。在完全监督的条件下,使用 GCCA 作为目标函数的 DCCA 与 DCCAE 模型,其准确度表现几乎无明显改变,甚至有所下降。
- (5) 在对各相关性目标函数进行评估时,与 TCCA 相比,MCCA 和 GCCA 由于其实现的复杂度较低,因此更易于被计算机进行加速运算。在实际应用中,基于 MCCA 和 GCCA 的模型展示出了相对短的训练时间,但这可能会伴随着网络训练过程中的不稳定性,如非数字(NaN)的传播现象。相对而言,虽然 TCCA 的算法复杂度较高,这在一定程度上阻碍了计算机的加速运算,但它展示出了显著的稳定性,并且在引入监督信息后,其性能表现得到了明显提升。

5.2 展望

本文主要集中在 DCCA 和 DCCAE 的拓展与探索,仍存在一些局限性。对此,未来的研究可能需要考虑以下几个方向:

- (1) 针对 GCCA 模型中引入标签信息后对模型性能提升有限的问题,需要进一步地研究。考虑到 GCCA 本身的计算特性,损失函数的结果普遍较大,变化幅度剧烈。因此,如何更有效地引入标签信息到 GCCA 模型中是一个待解决的问题。
- (2) 在处理高维小样本数据时使用 CCA 的有效方法也需要进一步探讨。当数据规模不足时,可能需要通过数据增强或其他方法提高最终训练时能进入 CCA 的潜在特征数。
- (3) 对于在使用 MCCA 和 GCCA 对神经网络进行引导时出现的数值错误问题,亦需进一步研究。在实现相关性算法时,往往离不开矩阵的分解和幂运算。当出

现训练错误时，如何探讨网络训练的变化路径，并对网络结构优化与训练策略进行改进将是一个具有挑战性的问题。

此外，在环境科学和可持续发展研究中，可以采用基于先进深度网络的多视角典型相关分析模型以揭示更深层次的、非线性的关联性，从而对环境变化和可持续发展趋势进行更准确的理解与预测。具体而言，有以下几个方向值得进一步探讨：

- (1) 针对海温和大气温度的并行分析，可以利用基于深度的多视角 CCA 模型来揭示这两者之间的深层、非线性关系，为全球气候变暖研究提供独特的视角。
- (2) 可持续发展涉及社会、经济、环境等多个领域，基于深度的多视角典型相关分析模型有助于深入理解这些领域间的交互关系，为可持续发展策略提供强大的数据分析支持。

插图索引

图 1.1	图像识别中的特征提取（边缘特征）	1
图 1.2	多视角数据示例	2
图 1.3	基于 DCCA 的分类任务工作流	4
图 2.1	人工神经网络	14
图 2.2	前向传播	14
图 2.3	常见的四种激活函数	14
图 2.4	反向传播	15
图 2.5	LeNet-5 结构	15
图 2.6	SVM	16
图 2.7	KNN	16
图 2.8	2D t-SNE	17
图 3.1	DCCA 结构	18
图 3.2	DCCA 抽象结构	18
图 3.3	ConvCCA 深度核结构	20
图 3.4	残差块结构	20
图 3.5	ResCCA 深度核结构	21
图 3.6	Noisy MNIST 的部分样本	22
图 3.7	Noisy FashionMNIST 的部分样本	22
图 3.8	Noisy EMNIST 的部分样本	22
图 3.9	Noisy 测试集在各模型中的 2D t-SNE 图	23
图 3.10	COIL-20 数据集	24
图 3.11	COIL-20 测试集在各模型中的 2D t-SNE 图	25
图 3.12	迭代 800 次的 ResCCA 模型	26
图 3.13	输出特征数对准确度的影响	27
图 3.14	网络层数对准确度的影响	27
图 3.15	批大小对准确度的影响	28
图 4.1	DCCAE 结构	30
图 4.2	SDCCAE 结构	31
图 4.3	MNIST-USPS-SVHN 数据集的部分样本	33
图 4.4	MNIST-USPS-SVHN 测试集在 MCCA 中的 2D t-SNE 图	34
图 4.5	MNIST-USPS-SVHN 测试集在各模型中的 2D t-SNE 图	35
图 4.6	ETH-80 数据集	35
图 4.7	ETH-80 测试集在各模型中的 2D t-SNE 图	37

图 4.8	各个模型在不同有效标签比例的识别准确度	37
-------	---------------------------	----

表格索引

表 3.1	实验环境	21
表 3.2	不同模型在 Noisy 数据集的分类结果	23
表 3.3	不同模型在 COIL-20 数据集的分类结果	25
表 4.1	不同模型在 MNIST-USPS-SVHN 数据集的分类结果	33
表 4.2	不同模型在 ETH-80 数据集的分类结果	36
表 4.3	GCCA 与 TCCA 算法在 MNIST-USPS-SVHN 数据集的表现	38
表 4.4	无监督的 DCCA 与 DCCAE 模型在 MFD 数据集的分类结果	39
表 4.5	完全监督的 DCCA 与 DCCAE 模型在 MFD 数据集的分类结果	39

参考文献

- [1] Abdi H, Williams L J. Principal component analysis [J]. WIREs Computational Statistics, 2010, 2 (4): 433–459.
- [2] Wold S, Esbensen K, Geladi P. Principal component analysis [J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2 (1): 37–52. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.
- [3] Pearson K. On lines and planes of closest fit to systems of point in space [J]. Philosophical Magazine, 1901, 2 (11): 559–572.
- [4] Hotelling H. Analysis of a complex of statistical variables into principal components. [J]. Journal of educational psychology, 1933, 24 (6): 417.
- [5] Turk M A, Pentland A P. Face recognition using eigenfaces [C]. In Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition, 1991: 586–587.
- [6] Abeywardena V. An application of principal component analysis in genetics [J]. Journal of Genetics, 1972, 61 (1): 27–51.
- [7] Giuliani A. The application of principal component analysis to drug discovery and biomedical data [J]. Drug Discovery Today, 2017, 22 (7): 1069–1076.
- [8] Martis R J, Acharya U R, Mandana K, et al. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health [J]. Expert Systems with Applications, 2012, 39 (14): 11792–11800.
- [9] Erleben K, Sporring J, Henriksen K, et al. Physics-based animation [M]. Charles River Media Hingham, 2005.
- [10] Stolrasky M S, Jakov N B. Recognition Using Class Specific Linear Projection [J], 2015.
- [11] Fisher R A. The use of multiple measurements in taxonomic problems [J]. Annals of eugenics, 1936, 7 (2): 179–188.
- [12] He X, Yan S, Hu Y, et al. Face recognition using laplacianfaces [J]. IEEE transactions on pattern analysis and machine intelligence, 2005, 27 (3): 328–340.
- [13] Shikkenawis G, Mitra S K. 2D orthogonal locality preserving projection for image denoising [J]. IEEE transactions on Image Processing, 2015, 25 (1): 262–273.
- [14] Zhao J, Xie X, Xu X, et al. Multi-view learning overview: Recent progress and new challenges [J]. Information Fusion, 2017, 38: 43–54.
- [15] 唐静静, 田英杰. 多视角学习综述 [J]. 数学建模及其应用, 2017, 6 (3): 1–15.
- [16] Hotelling H. Relations between two sets of variates [J]. Breakthroughs in statistics: methodology and distribution, 1992: 162–190.
- [17] Kettenring J R. Canonical analysis of several sets of variables [J]. Biometrika, 1971, 58 (3): 433–451.
- [18] Vía J, Santamaría I, Pérez J. A learning algorithm for adaptive canonical correlation analysis of several data sets [J]. Neural Networks, 2007, 20 (1): 139–152.

- [19] Carroll J D. Generalization of canonical correlation analysis to three or more sets of variables [J], 1968.
- [20] Luo Y, Tao D, Ramamohanarao K, et al. Tensor canonical correlation analysis for multi-view dimension reduction [J]. IEEE transactions on Knowledge and Data Engineering, 2015, 27 (11): 3111–3124.
- [21] Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis [J]. Pattern recognition, 2003, 36 (9): 1961–1971.
- [22] Zheng W, Zhou X, Zou C, et al. Facial expression recognition using kernel canonical correlation analysis (KCCA) [J]. IEEE Transactions on Neural Networks, 2006, 17 (1): 233–238.
- [23] Larson N B, Jenkins G D, Larson M C, et al. Kernel canonical correlation analysis for assessing gene–gene interactions and application to ovarian cancer [J]. European Journal of Human Genetics, 2013, 22 (1): 126–131.
- [24] Haroon D R, Mourão-Miranda J, Brammer M, et al. Unsupervised analysis of fMRI data using kernel canonical correlation [J]. NeuroImage, 2007, 37 (4): 1250–1259.
- [25] Andrew G, Arora R, Bilmes J, et al. Deep Canonical Correlation Analysis [C] // Dasgupta S, McAllester D. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 17–19 Jun 2013, 28: 1247–1255.
- [26] 刘岩. 基于深度典型相关分析的特征提取算法研究 [D]. 扬州: 扬州大学, 2019.
- [27] Wang W, Arora R, Livescu K, et al. On deep multi-view representation learning [C]. In International conference on machine learning, 2015: 1083–1092.
- [28] Yang X, Liu W, Liu W. Tensor Canonical Correlation Analysis Networks for Multi-View Remote Sensing Scene Recognition [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34 (6): 2948–2961.
- [29] 杨兴浩. 基于典型相关分析的多视角图像识别算法研究 [D]. 青岛: 中国石油大学(华东), 2018.
- [30] Yang X, Liu W, Tao D, et al. Multiview Canonical Correlation Analysis Networks for Remote Sensing Image Recognition [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14 (10): 1855–1859.
- [31] Benton A, Khayrallah H, Gujral B, et al. Deep generalized canonical correlation analysis [J]. arXiv preprint arXiv:1702.02519, 2017.
- [32] Wong H S, Wang L, Chan R, et al. Deep tensor CCA for multi-view learning [J]. IEEE Transactions on Big Data, 2021, 8 (6): 1664–1677.
- [33] Liu Y, Li Y, Yuan Y-H, et al. Supervised deep canonical correlation analysis for multiview feature learning [C]. In Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part VI 24, 2017: 575–582.
- [34] Elmadany N E D, He Y, Guan L. Multiview learning via deep discriminative canonical correlation analysis [C]. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016: 2409–2413.
- [35] 张恒, 陈晓红, 蓝宇翔, 等. 基于深度学习的监督型典型相关分析 [J]. 计算机工程, 2022, 48 (222-228).

- [36] 袁红春, 刘慧, 张硕, 等. 基于深度学习和典型相关分析的渔情预测方法 [J]. 大连海洋大学学报, 2021, 36 (4): 670–678.
- [37] Isobe S, Tamura S, Hayamizu S. Speech Recognition using Deep Canonical Correlation Analysis in Noisy Environments. [C]. In ICPRAM, 2021: 63–70.
- [38] Sun Z, Sarma P K, Sethares W, et al. Multi-modal sentiment analysis using deep canonical correlation analysis [J]. arXiv preprint arXiv:1907.08696, 2019.
- [39] Qiu J-L, Liu W, Lu B-L. Multi-view Emotion Recognition Using Deep Canonical Correlation Analysis [C] // Cheng L, Leung A C S, Ozawa S. In Neural Information Processing, Cham, 2018: 221–231.
- [40] Yang J, Zhou Y, Cao Y, et al. Heterogeneous image change detection using Deep Canonical Correlation Analysis [C]. In 2018 24th International Conference on Pattern Recognition (ICPR), 2018: 2917–2922.
- [41] Golub G H, Reinsch C. Singular value decomposition and least squares solutions [J]. Linear algebra, 1971, 2: 134–151.
- [42] Mardia K V, Kent J T, Bibby J M. Multivariate analysis [J]. Probability and mathematical statistics, 1979.
- [43] Chu M T, Watterson J L. On a multivariate eigenvalue problem, Part I: Algebraic theory and a power method [J]. SIAM Journal on scientific computing, 1993, 14 (5): 1089–1106.
- [44] 徐莲花. 关于多元特征值问题的数值方法 [D]. 青岛: 中国海洋大学, 2008.
- [45] Horst P. Relations among m sets of measures [J]. Psychometrika, 1961, 26 (2): 129–149.
- [46] Rupnik J, Shawe-Taylor J. Multi-view canonical correlation analysis [C]. In Conference on data mining and data warehouses (SiKDD 2010), 2010: 1–4.
- [47] Horst P. Generalized canonical correlations and their application to experimental data [M]. Journal of clinical psychology, 1961.
- [48] Yang X, Liu W, Liu W, et al. A survey on canonical correlation analysis [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33 (6): 2349–2368.
- [49] 郭陈凤, 伍冬睿. 基于典型相关分析的多视图学习方法综述 [J]. 智能科学与技术学报, 2022, 4 (1): 14–26.
- [50] Velden M v d, Bijmolt T H. Generalized canonical correlation analysis of matrices with missing rows: a simulation study [J]. Psychometrika, 2006, 71 (2): 323–331.
- [51] Fu X, Huang K, Hong M, et al. Scalable and flexible multiview MAX-VAR canonical correlation analysis [J]. IEEE Transactions on Signal Processing, 2017, 65 (16): 4150–4165.
- [52] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278–2324.
- [53] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84–90.
- [54] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770–778.
- [55] Van der Maaten L, Hinton G. Visualizing data using t-SNE. [J]. Journal of machine learning research, 2008, 9 (11).

- [56] Chapman J, Wang H-T. CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework [J]. Journal of Open Source Software, 2021, 6 (68): 3823.

致 谢

感谢父母生养之恩，感谢一路陪伴过我的老师与同学，感谢腾讯 IEG 魔方工作室的实习机会。

在疫情占据一半的大学生活，虽有许多遗憾，但还是要到了毕业的时候。希望自己能学习更多想学的知识。

温春晖

上海大学

2023 年 5 月 24 日

