

基于广义 Grassmann 流形的字典学习算法研究

马鑫¹ 李慧勇¹ 顾强¹ 邱枫¹ 廖亮^{2*} 赵杰² 陈花竹² 于洪年³ 谭恒⁴

MA Xin LI Huiyong GU Qiang QIU Feng LIAO Liang ZHAO Jie CHEN Huazhu YU Hongnian TAN Heng

摘要

为提升图像分类任务的准确性与鲁棒性,文章提出一种基于广义 Grassmann 流形的稀疏表示分类算法(H-gDL)。针对传统方法处理图像数据时忽略局部邻域信息的缺陷,通过引入广义矩阵模型与循环卷积乘法,将图像数据建模为广义 Grassmann 流形上的点,实现对图像结构特征的高效捕捉。在此基础上,结合字典学习与投影嵌入技术,将稀疏编码问题转化为广义流形上的优化问题,并通过离散余弦变换(DCT)进行频域优化,有效提升算法计算效率。在 ETH-80、RGB-D 及 YOLOv5 手势数据集上的实验结果表明,H-gDL 算法的分类准确率显著优于多种传统 Grassmann 流形基对比方法,且在复杂、非规则数据集上展现出更强的鲁棒性。

关键词

广义 Grassmann 流形;稀疏表示;图像分类;彩色视频;奇异值分解

doi: 10.3969/j.issn.1672-9528.2026.06.006

0 引言

分类任务作为机器学习和数据分析领域中最基础最常用的任务,核心在于根据不同数据的独有特征,通过预先设定好的分类方式自动划分到预先定义好的各个类别中。伴随着层出不穷的新技术,不同格式的图片、视频数据正在改变人们的生活。如果不能对它们进行有效的分类,大家在寻找各自感兴趣的内容时无异于大海捞针。因此,图像分类成为一项重要的任务。

在最新的研究中,神经网络和深度学习方法在图像分类任务中得到了广泛的应用。Kim 等^[1]提出了一种基于边缘标记图的神经网络(EGNN)用于小样本学习。通过在边缘标记图上适配深度神经网络,突破了以往图神经网络(GNN)在小样本学习中依赖节点标记框架的传统。该机制通过迭代更新边标签,直接利用类内相似性和类间差异性,从而提升了聚类质量。Li 等^[2]提出了一种新的深度度量学习方法来解决多标签图像分类问题。为了更好地学习图像之间的特征以及图像特征与标签之间的关联,尝试设计一个潜在空间,在

该空间中,图像和标签分别通过两个独立的神经网络进行嵌入。此外,再引入一个重构模块,用于恢复正确的标签,并将其作为正则化项融入整个框架,从而使标签嵌入空间更具代表性。

不同于神经网络类方法,线性子空间可以良好地表征复杂变化,因此使用流形来进行图像分类工作也获得了巨大的成功。Zhang 等^[3]提出一种格拉斯曼流形上的概率协同表示的新方法,从概率的视角阐释了图像集分类机制,通过实验证明了其在人脸识别与物体识别方面所具有的优异鲁棒性和有效性。Wang 等^[4]提出了一种多几何视角的图像集建模新算法,通过设计多核度量学习框架将混合核嵌入公共低维子空间以提升分类效能。Wei 等^[5]提出了一种面向图像集分类的 Grassmann 重建度量学习(GRML)算法,并以稀疏表示分类器(SRC)作为指导。并且进一步给出了 Grassmann 重建多核度量学习(GRMKML)算法。在其所选择的八个基准数据集上进行了大量实验,验证了所提算法的性能显著优于现有的最佳方法。

然而,上述方法在实际应用中,往往将图像中的每一个数据作为一个独立的元素进行运算,忽视了其内在包含的邻域约束信息,具有一定局限性。针对这个问题,本文提出了一种基于广义 Grassmann 流形的稀疏表示分类算法。利用广义 Grassmann 流形的数学模型来重建原始图像信息,并将基于 Grassmann 流形的 gDL 算法升阶为 H-gDL 算法。相对于传统算法,H-gDL 算法可以有效地利用邻域信息进行图像分类工作。

本研究通过实验验证了 H-gDL 算法相较于传统算法拥有

1. 中原工学院信息与通信工程学院 河南郑州 450007

2. 中原工学院数学与信息科学学院 河南郑州 450007

3. 爱丁堡龙比亚大学计算工程与建筑环境学院

苏格兰爱丁堡 EH105DT

4. 中原工学院经济管理学院 河南郑州 450007

[基金项目] 国家自然科学基金(62202513);中原工学院“机器智能与高维数据分析”科研团队发展计划(K2022TD001);河南省科学技术厅的河南省杰出外籍科学家工作室项目(GZS2022012)

更高的准确性，且在多个数据集上表现出了较好的稳定性和鲁棒性。

1 广义 Grassmann 流形

假设矩阵 $A \in \mathbb{C}^{D \times p}$ ($0 < p < D$) 各列相互正交，即满足 $A^H A = I_p$ ，则称 A 是 Stiefel 上的点。全部满足上述条件的 $\mathbb{C}^{D \times p}$ 矩阵的集合称为 Stiefel 流形，用符号 $\text{St}(D, p)$ 表示，即：

$$\text{St}(D, p) = \{A \in \mathbb{C}^{D \times p} : A^H A = I_p\} \quad (1)$$

假定有任意正交矩阵 $W \in \mathbb{C}^{p \times p}$ 对矩阵 A 做投影，则可以得到一个矩阵集合：

$$J(AW) = \{AW \in \mathbb{C}^{D \times p} : A^H A = I_p, WW^H = I_p\} \quad (2)$$

对于集合中的任意一个矩阵都可以张成同一个子空间，即集合中的所有矩阵都是 A 的等价类。

由此，可以得到 Grassmann 流形的定义，Grassmann 流形中的点由 Stiefel 流形中可以张成相同列空间的矩阵组成，即 Grassmann 流形是 Stiefel 流形的一种商流形。Grassmann 流形用符号 $\text{Gr}(D, p)$ 表示为：

$$\text{Gr}(D, p) = \{\text{span}\{V\} : V \in \mathbb{C}^{D \times p}, V^H V = I_p\} \quad (3)$$

对于子空间 $\mathcal{X} = \text{span}(X)$, $\mathcal{Y} = \text{span}(Y) \in \text{Gr}(D, p)$ ，其之间的距离在一般情况下显然不等于欧氏距离，而应该是测地线距离。由于测地线距离的计算复杂度较高，因而一般选择主夹角向量的余弦 $\theta_{\text{vec}} = \{\phi_1, \phi_2, \dots, \phi_p\}$ 取反余弦之后计算 L_2 范数进行近似计算。

参考 Liao 等^[6]的论文，若有广义矩阵 $\dot{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ ，满足 $\dot{A}^H \circ \dot{A} = \dot{I}$ ，其中 $\dot{I} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ ， $I_1 \times I_2 \times \dots \times I_n$ 为广义标量的大小， $D_1 \times D_2$ 为广义矩阵的大小， $\dot{X} \circ \dot{Y}$ 为广义矩阵使用循环卷积乘法代替传统矩阵的乘法。因此，广义矩阵的乘法不同于传统矩阵，满足乘法交换律。所有满足上述条件的 $\mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ 广义矩阵构成的集合称为广义 Stiefel 流形，即：

$$\text{St}_T(D_1, D_2) = \{\dot{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2} : \dot{A}^H \circ \dot{A} = \dot{I}\} \quad (4)$$

式中： St_T 表示广义 Stiefel 流形。

对于所有满足 $\dot{W} \circ \dot{W}^H = \dot{I}$ 的矩阵，其中 $\dot{W} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ 。若有：

$$J(\dot{A} \circ \dot{W}) = \{\dot{A} \circ \dot{W} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2} : \dot{A}^H \circ \dot{A} = \dot{W} \circ \dot{W}^H = \dot{I}\} \quad (5)$$

式中： $\dot{I} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ 。由此得到的集合可以张成同一个广义子空间，进而得到广义 Grassmann 流形的定义：

$$\text{Gr}_T(D_1, D_2) = \{\text{span}\{\dot{V}\} : \dot{V} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}, \dot{V}^H \circ \dot{V} = \dot{I}\} \quad (6)$$

式中： $\dot{I} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ 。

与传统矩阵不同，由于广义 Grassmann 流形是以广义矩阵为数学基础构成的，因此，在计算两个广义子空间距离时，得到的主夹角向量的余弦 $\dot{\theta} = \{\phi_1, \phi_2, \dots, \phi_p\}$ 是一个广义向量， $\phi_1, \phi_2, \dots, \phi_p \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n}$ 。因此，需要计算每一个广义标量的 L_2

范数作为最终得到的主夹角向量的余弦值再进行测地线距离的计算^[7]。

2 广义 Grassmann 流形字典学习

2.1 字典学习

假设在向量空间中存在一个需要从字典中查询的向量 x ，字典中的向量使用 d_i 进行表示。则有 $x = \sum_{i=1}^n y_i d_i$ ，其中 y_i 为各个字典元素的权重，满足 $\sum_{i=1}^n y_i = 1$ 。

给定目标向量 x 和字典，其中 $D = \{d_1, d_2, \dots, d_n\}$ ，字典中的元素 $d_i \in \mathbb{R}^d$ ，字典学习的目标即为最小化一组有限的观测值 $\{d_i\}_{i=1}^m$ 的总成本，对于所有的观测值都有一个最小的编码误差 x_i ，同时产生一个具有期望结构的系数 y_i ，即：

$$\min_{\{y_i\}_{i=1}^m, \mathbb{D}} \left\| x - \sum_{i=1}^N y_i d_i \right\|_2^2 + \lambda \sum_{i=1}^N \|y_i\| \quad (7)$$

2.2 广义 Grassmann 流形字典学习分类算法

基于上述理论，可以得到广义 Grassmann 流形中的字典学习目标函数

$$l_M(\dot{x}, \dot{D}) \triangleq \min_y \left(d_g(\mathcal{X}, C(y, \dot{d}_i))^2 + \lambda f(y) \right) \quad (8)$$

式中： $\dot{x}, \dot{d}_i \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times d \times p}$ ， d_g 是广义 Grassmann 流形中的距离度量， C 是一个编码函数，为每一个系数和字典元素分配一个空间 M 中的元素。

在编码部分，选择使用投影嵌入，将广义 Grassmann 流形嵌入对称矩阵空间中。投影嵌入此前已用多视图信息与图像分类^[8]和多源时间序列聚类^[9]中。

假设 $PSGr_T(d, p)$ 是一个秩为 p 的 $d \times d$ 幂等对称广义矩阵的集合，投影嵌入 $\Pi: Gr_T(d, p) \rightarrow PGr_T(d, p)$ 由 $\Pi(\dot{X}) = \dot{X} \circ \dot{X}^H$ 给出，在后文中，使用 $\hat{X}$ 来代替 $\dot{X} \circ \dot{X}^H$ ，此外， $\langle \cdot, \cdot \rangle$ 表示 Frobenius 内积，如 $\langle \dot{X}, \dot{Y} \rangle = \text{Tr}(\dot{X} \circ \dot{Y})$ 。

字典编码的基本概念是将空间中的某个点表示为空间中其他元素的线性组合，然而在流形中，并没有线性组合的概念。因此，需要找到一种适用于流形的方法来让子空间 $\mathcal{X}$ 可以被其他子空间 $\mathcal{D}_i$ 来表示。参考 Harandi 等的工作，选择使用 Karcher 平均值，其规定黎曼流形 $\mathcal{M}$ 上的点 $\{\mathcal{D}_i\}$ 和权重 $\{y_i\}$ ，能使得 $\sum_{i=1}^n y_i d_g(\mathcal{X}, \mathcal{D}_i)^2$ 最小的点 $\mathcal{X}^*$ 称为权重为 $\{y_i\}$ 的点 $\{\mathcal{D}_i\}$ 的加权 Karcher 平均值。

为了简化计算，提高运算效率，选择使用弦度度量 d_{chord} 来代替式 (8) 中的测地线距离 d_g ，用公式表示为：

$$d_{\text{chord}}(\hat{X}, \hat{Y}) = \|\Pi(\mathcal{X}) - \Pi(\mathcal{Y})\|_F = \|\hat{X} - \hat{Y}\|_F \quad (9)$$

在使用 d_{chord} 代替后 d_g ，Karcher 平均值转化为加权弦度均值，与 Karcher 均值相比，Grassman 流形上的加权弦度均值具有简单的封闭形式。

给定一个包含原子 $[\hat{D}]_i \in PGr_T(d, p)$ 的字典 $\hat{D}$ 和一个目标样

本 $\hat{\mathbf{X}}$, 稀疏编码的问题可以被重新定义为:

$$l(\mathcal{X}, \hat{\mathbf{D}}) \triangleq \min_{\mathbf{y}} \|\hat{\mathbf{X}} - \sum_{j=1}^N [\mathbf{y}]_j \hat{\mathbf{D}}_j\|_F^2 + \lambda \|\mathbf{y}\|_1 \quad (10)$$

展开式 (10) 中的 Frobenius 范数项会得到 $\mathbf{y}$ 的凸函数:

$$\|\hat{\mathbf{X}}\|_F^2 + \|\sum_{i=1}^n [\mathbf{y}]_i \hat{\mathbf{D}}_i\|_F^2 - 2 \langle \sum_{i=1}^n [\mathbf{y}]_i \hat{\mathbf{D}}_i, \hat{\mathbf{X}} \rangle \quad (11)$$

通过将 $[\mathcal{K}(\mathcal{X}, \hat{\mathbf{D}})]_i = \langle \hat{\mathbf{X}}, \hat{\mathbf{D}}_i \rangle$ 定义为目标广义子空间 $\mathcal{X}$ 和字典原子之间相似度的 N 维向量, $[\mathcal{K}(\hat{\mathbf{D}})]_{(i,j)} = \langle \hat{\mathbf{D}}_i, \hat{\mathbf{D}}_j \rangle$ 定义为字典原子之间相似度的 $N \times N$ 对称矩阵可以将式 (11) 修改为:

$$l(\mathcal{X}, \hat{\mathbf{D}}) = \min_{\mathbf{y}} \mathbf{y}^T \mathcal{K}(\hat{\mathbf{D}}) \mathbf{y} - 2 \mathbf{y}^T \mathcal{K}(\mathcal{X}, \hat{\mathbf{D}}) + \lambda \|\mathbf{y}\|_1 \quad (12)$$

又因为, 对于所有的 $\mathbf{v} \in \mathbb{R}^d$ 均满足:

$$\mathbf{v}^T \mathcal{K}(\hat{\mathbf{D}}) \mathbf{v} = \sum_{i=1}^N \sum_{j=1}^N v_i v_j \langle \hat{\mathbf{D}}_i, \hat{\mathbf{D}}_j \rangle = \left\| \sum_{i=1}^N v_i \hat{\mathbf{D}}_i \right\|_F^2 > 0 \quad (13)$$

因此, $\mathcal{K}(\hat{\mathbf{D}})$ 是正定的, 所以问题是凸的, 可以有效求解。

式 (12) 中的问题可以转化为矢量化的稀疏表示编码问题。设 $\mathbf{U} \Sigma \mathbf{V}^T$ 为 $\mathcal{K}(\hat{\mathbf{D}})$ 的高阶奇异值分解 (T-SVD) [6] 的结果, 式 (12) 可以被改写为:

$$l(\mathcal{X}, \hat{\mathbf{D}}) = \min_{\mathbf{y}} \|\mathbf{x}^* - \mathbf{A} \mathbf{y}\|^2 + \lambda \|\mathbf{y}\|_1 \quad (14)$$

式中: $\mathbf{A} = \Sigma^{1/2} \mathbf{U}^T$, $\mathbf{x}^* = \Sigma^{-1/2} \mathbf{U}^T \mathcal{K}(\hat{\mathbf{X}}, \hat{\mathbf{D}})$ 。

算法 1 提供了执行广义 Grassmann 流形字典学习算法 (H-gDL) 的伪代码。

算法 1 H-gDL

输出: 稀疏编码结果 $\mathbf{y}^*$

1. 初始化

2. 对 i, j 从 1 到 n 进行循环

3. 计算 $[\mathcal{K}(\hat{\mathbf{D}})]_{i,j} = \|\hat{\mathbf{D}}_i^T \hat{\mathbf{D}}_j\|_F^2$

4. 结束循环

5. 对 $\mathcal{K}(\hat{\mathbf{D}})$ 进行 T-SVD 分解得到, 得到 $\mathbf{U}, \Sigma, \mathbf{V}^T$

6. $\mathbf{A} = \Sigma^{1/2} \mathbf{U}^T$

7. 对 i 从 1 到 n 进行循环

8. 计算 $[\mathcal{K}(\hat{\mathbf{X}}, \hat{\mathbf{D}})]_i = \|\hat{\mathbf{X}}^T \hat{\mathbf{D}}_i\|_F^2$

9. 结束循环

10. $\mathbf{x}^* = \Sigma^{-1/2} \mathbf{U}^T \mathcal{K}(\hat{\mathbf{X}}, \hat{\mathbf{D}})$

11. $\mathbf{y}^* = \arg \min_{\mathbf{y}} \|\mathbf{x}^* - \mathbf{A} \mathbf{y}\|^2 + \lambda \|\mathbf{y}\|_1$

输出: 分类结果

2.3 优化

虽然在算法中, 使用了各种方法来进行优化。然而, 由于广义矩阵的特殊数学模型, 原矩阵运算中的乘法全部都要使用循环卷积乘法来进行运算。因此, 对于模型的优化也是势在必行。

对于一个广义矩阵 $\hat{\mathbf{A}} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2}$ 对其进行离散余弦变换, 用公式表示为:

$$\mathbf{A}' = \hat{\mathbf{A}} \times_1 \mathbf{T}_1 \times_2 \mathbf{T}_2 \times \dots \times_n \mathbf{T}_n \quad (15)$$

式中: $\mathbf{A}'$ 为经过 DCT 变换后得到的 $\hat{\mathbf{A}}$ 的频域形式, $\mathbf{T}_i$ 为 n

模式乘法, $\mathbf{T}$ 为 DCT 变换矩阵, $\mathbf{T}$ 中的元素用公式表示为:

$$T_{ij} = \begin{cases} \frac{1}{\sqrt{N}}, & i=0 \\ \sqrt{\frac{2}{N}} \cos\left(\frac{(2j+1)i\pi}{2N}\right), & i>0 \end{cases} \quad (16)$$

经过 DCT 变换后, 广义矩阵的每一个广义标量的元素被转换为原广义标量在不同频率下的幅度。因此, 经过 DCT 变换, 广义矩阵可以按频率进行切片, 从原本的 $I_1 \times I_2 \times \dots \times I_n \times D_1 \times D_2$ 维转化为 $I_1 \times I_2 \times \dots \times I_n$ 个相互独立的 $D_1 \times D_2$ 大小的矩阵。

由于广义矩阵的独特机制, 两个广义矩阵的循环卷积乘法可以由经 DCT 变换之后的对应频率矩阵之间的乘法得到, 对算法进行优化。

使用 $\mathbf{A}'_j$ 表示广义矩阵 $\mathbf{A}$ 经 DCT 变换后得到的第 j 个频点的切片, 其中 $j = 1, 2, \dots, I_1 \times I_2 \times \dots \times I_n$ 。所使用的伪代码如算法 2 所示。

算法 2 优化后的 H-gDL 算法

输入: 广义 Grassmann 流形字典 $\{\mathcal{D}_i\}_i^n = \text{span}(\hat{\mathbf{D}}_i)$, 目标流形

点 $\mathcal{X} = \text{span}(\hat{\mathbf{X}})$

输出: 稀疏编码结果 $\mathbf{y}^*$

1. 初始化将广义 Grassmann 流形 $\{\mathcal{D}_i\}_i^n$ 与 $\mathcal{X}$ 经 DCT 变换转化为 $\{\mathcal{D}'_i\}_i^n$ 与 $\mathcal{X}'_j$

2. 对每个切片执行算法 1 得到 $\mathbf{y}'_j$

3. 计算 $\mathbf{y}'_j$ 的众数视为最终分类结果

输出: 分类结果

3 实验

在本文中, 选择了 3 个不同的公开数据集, 与其他五个关于 Grassmann 流形的图像分类算法进行了对比, 即测底线距离的 1-NN 算法、GDA 算法 [10]、RMML 算法 [11]、gSC 算法、MRMML 算法。为了保证公平, 对数据集上的数据进行统一处理。

3.1 数据预处理

由于所选择的数据集均按照相同的时间序列进行排序, 因此在构建 Grassmann 流形和广义 Grassmann 流形时均将同一个时间序列下的所有图片作为一个集合进行流形计算。

假设一个集合含有图像 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, 其中 n 为图片的数量, 将每一张图像 $\mathbf{x}_i$ 向量化得到 $\mathbf{x}_i \in \mathbb{R}^D$, 其中 D 为向量化之后的维度。这样, 集合就变成了 $\mathbf{X} \in \mathbb{R}^{D \times n}$ 的数组。可以通过对数组 $\mathbf{X}$ 进行 SVD 分解选择 $\mathbf{U}$ 的前 p 列, 得到 p 维的线性子空间, 即 $\text{svd}(\mathbf{X}) = \mathbf{U} \Sigma \mathbf{V}^T$ 。

对于广义 Grassmann 流形, 对每一张图片进行零值边界扩展用来保留图像的边界信息。通过使用预设的滑动窗口,

遍历整个数据空间提取局部空间特征块作为广义标量构建基于原始图像数据的广义矩阵。使用高阶奇异值分解得到广义线性子空间。具体过程参考文献 [5]。

3.2 评估设置

3.2.1 数据集

ETH-80 数据集包含了来自 8 个不同类别的视觉对象图像, 这些类别包括苹果、汽车、奶牛、杯子、狗、马、梨和番茄。每个类别包含 10 个不同的对象实例, 且每个对象实例都有从不同视角拍摄的 41 张图像。选择每个种类中的 2 个对象作为探针集, 剩余的对象作为测试集。

RGB-D 数据集中包含 300 个对象, 这些对象来自 51 个种类的物品, 每个对象都包含数百张从不同角度采集的照片。每次实验随机选择其中的 6 个种类的物品用于实验。每个种类中随机选择 25% 的对象作为探针集, 剩下的对象作为测试集。

YOLOv5 手势数据集包含了从 0~9 总共 10 种手势变化, 每个手势约 200 张图片, 总计 1 900 张图片, 所有图片均在单一背景下拍摄。每次实验随机剔除一种手势, 其他手势中每 20 张图片分为一组作为一个对象, 最终选择每种手势的 3 个对象作为探针集, 剩余的对象作为测试集。

在本文的实验中, 所有图片均调整为 $20 \times 20 \times 3$ 。图 1~3 分别展示了三个数据集中的部分图片。

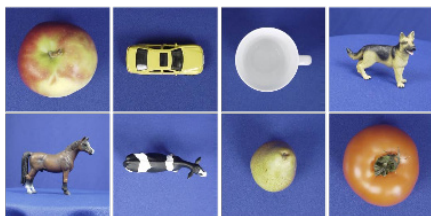


图 1 ETH80 数据集部分图片



图 2 RGB-D 数据集部分图片

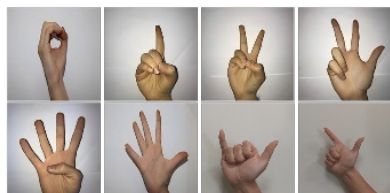


图 3 YOLOv5 数据集部分图片

3.2.2 评价指标

在实验中, 选择分类准确率作为评价指标, 用公式表示为:

$$\text{Accuracy} = \frac{\text{sample}_{\text{correct}}}{\text{sample}_{\text{all}}} \quad (17)$$

式中: $\text{sample}_{\text{all}}$ 代表总样本数, $\text{sample}_{\text{correct}}$ 代表正确分类的样本数。

3.2.3 参数设置

本文提出的算法中, 共包含 $\{t\text{-size}, \lambda, \text{nAtoms}, p\}$ 4 个参数。 $t\text{-size}$ 表示广义矩阵中 t 标量的大小, 设置为 $\mathbb{R}^{3 \times 3}$; λ 为稀疏表示的正则化参数设置为 $1e-1$; nAtoms 表示字典学习中字典原子的数量, 设置为探针集总数的 75%; p 为线性子空间的维度, 统一设置为 12, 即构成的流形中的子空间的维度为 $\mathbb{R}^{D \times 12}$ 。

3.3 结果

表 1 展示了 H-gDL 算法和其他对比算法在 3 个数据集上的分类准确率。从表中的信息可以看出, 提出的算法在每个数据集上均优于其他算法, 尤其是在 RGB-D 这种不规则数据集上, 通过在建模阶段通过广义矩阵提高了对邻域信息的利用率, 最终算法准确率得到了大幅度的提升。

表 1 3 组数据集上不同算法的分类准确率

单位: %				
ID	Methods	ETH80	RGB-D	YOLOv5
1	1-NN	84.296 8	78.555 5	90.870 3
2	GDA	84.484 3	78.527 0	90.574 0
3	RMML	85.109 7	78.583 3	93.980 0
4	gSC	85.484 7	77.083 1	94.814 0
5	MRMML	85.940 0	72.222 0	93.148 0
6	H-gDL	86.703 8	85.055 1	97.224 0

此外, 本文还测试了三个数据集在不同字典规模 l 与正则化参数 λ 条件下的分类准确率。在本项实验中, 选择的字典规模 l 分别为训练集规模的 [15%, 30%, 45%, 60%, 75%, 90%], 正则化参数 λ 为 [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]。最终得到的实验结果通过热图形式的三维折线图进行展示, 如图 4~6 所示。

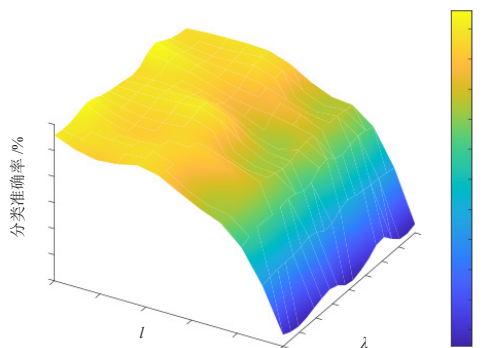


图 4 ETH80 数据集参数测试实验结果

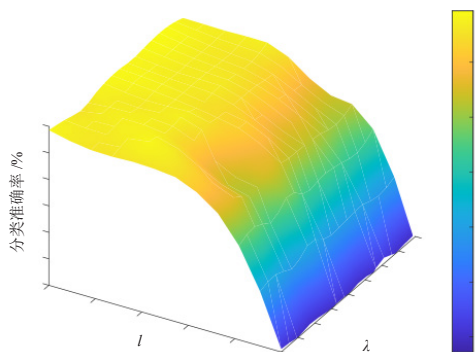


图5 RGB-D数据集参数测试实验结果

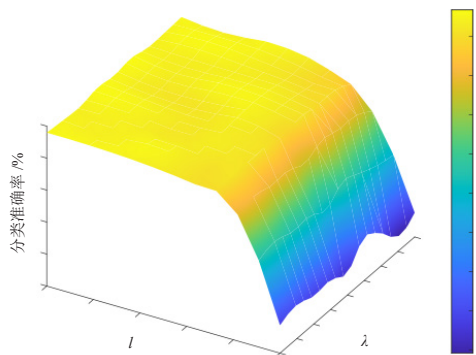


图6 YOLOv5数据集参数测试实验结果

从图中可以看出,算法的分类准确率随着字典原子数 l 的提高和正则化参数的提高 λ 得到了更为优异的表现。另一方面,这种关系并非线性的,根据数据集的不同,每一组数据集分别在不同的参数设置下出现性能拐点,在拐点之前提高参数的数值会使得分类准确率快速提高,过了拐点之后,更大的参数设置仅会对分类性能造成较小的影响。

4 结论

本文围绕图像分类任务中如何有效利用图像的局部结构与邻域信息这一关键问题展开研究,提出了一种基于广义Grassmann流形的字典学习分类算法(H-gDL)。

首先,本文详细阐述了广义Grassmann流形的数学理论基础,通过引入广义矩阵模型和循环卷积运算,成功地将传统Grassmann流形上的线性子空间方法推广至能够处理局部特征的结构化数据形式。在此基础上,构建了广义Grassmann流形上的字典学习目标函数,并利用投影嵌入与弦度量逼近的方法,将复杂的流形稀疏编码问题转化为可高效求解的凸优化问题。

为了提升算法效率,本文进一步提出了一种基于离散余弦变换(DCT)的优化方法。该方法通过将广义矩阵分解为多个独立的频域切片进行并行处理,显著降低了广义矩阵循环卷积乘法的计算复杂度。

在实验部分,本文在ETH-80数据集、RGB-D数据集和YOLOv5手势数据集上,将本文提出的H-gDL算法与多种

基于流形的主流分类算法进行了全面对比。实验结果表明,H-gDL算法在所有数据集上均取得了最高的分类准确率,尤其在背景复杂、分辨率大小不一的RGB-D数据集上性能提升最为显著,充分证明了该算法在获取图像特征、提升模型鲁棒性方面的优越性。此外,对不同大小的字典规模和正则化参数进行对比试验,也为算法的实际应用与参数设置提供了有价值的指导。

参考文献:

- [1] Kim J, Kim T, Kim S, et al. Edge-labeling graph neural network for few-shot learning[PP/OL].(2019-05-04)[2026-05-08].<https://doi.org/10.48550/arXiv.1905.01436>.
- [2] Li Changsheng, Liu Chong, Duan Lixin, et al. Reconstruction regularized deep metric learning for multi-label image classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7):2294-2303.
- [3] Zhang Shuo, Wei Dong, Yan Wenzhu, et al. Probabilistic collaborative representation on Grassmann manifold for image set classification[J]. Neural Computing and Applications, 2021, 33: 2483-2496.
- [4] Wang Rui, Wu Xiaojun, Chen Kaixuan, et al. Multiple riemannian manifold-valued descriptors based image set classification with multi-kernel metric learning[J]. IEEE Transactions on Big Data, 2022, 8(3): 753-769.
- [5] Wei Dong, Shen Xiaobo, Sun Quansen, et al. Sparse representation classifier guided Grassmann reconstruction metric learning with applications to image set analysis[J]. IEEE Transactions on Multimedia, 2023, 25: 4307-4322.
- [6] Liao Liang, Guo Zhuang, Gao Qi, et al. Color image recovery using generalized matrix completion over higher-order finite dimensional algebra[J]. Axioms, 2023, 12(10): 954.
- [7] 王竞宇, 廖亮, 高奇. 基于广义Grassmann流形的图像分类方法[J]. 信息技术与信息化, 2023(3): 112-115.
- [8] Heo S, Teoh A B J, Yu Sunjin, et al. Multivariate time series clustering with state space dynamical modeling and Grassmann manifold learning: a systematic review on human motion data[J]. Applied Sciences, 2025, 15(1): 43.
- [9] Guan Yao, Yao Jiayi, Yan Wenzhu, et al. Robust Grassmann manifold convex hull collaborative representation learning and its kernel extension for image set analysis[J/OL]. Multimedia Systems, 2024[2026-05-08]. <https://link.springer.com/article/10.1007/s00530-024-01522-6>.

面向复杂教学场景的教师行为检测算法优化研究

任甜甜¹

REN Tiantian

摘要

针对“三班联播”复杂教学场景中多视角遮挡、高并发推流与边缘算力受限导致的教师行为漏检问题，文章提出一个轻量化-高鲁棒-可解释三位一体优化框架。设计 Dual-FPN+CBAM 遮挡感知检测头，将板书/手臂多尺度梯度与空间通道双注意力耦合，提升极端姿态召回；构建 TSMini-TCN 分离空洞时序网络，在 RK3399 边缘盒实现 27 ms 延迟、3.8 W 功耗；引入 Transformer 双通道编码器建模教师-学生空间关系，输出五维交互特征。进一步通过边缘实时推流+云端知识蒸馏协同，实现周级云端自更新与月级边缘热更新。实验表明，所提方法 mAP@0.5 达 84.7%，领先次优算法 6.2%，帧率 FPS 保持 36 帧/s，异常报警率低于 0.3%，可用性评分 4.6/5，满足学期无人值守运行需求。

关键词

教师行为检测；轻量化网络；遮挡感知；边缘-云协同；教学场景

doi: 10.3969/j.issn.1672-9528.2026.06.007

0 引言

智慧教室的大规模部署对教师行为实时分析提出迫切需求，然而多摄像机并发、板书遮挡与边缘算力受限导致现有算法在召回率、时延与功耗之间难以兼顾。传统两阶段检测-分类方案依赖高分辨率输入与深度主干，在课堂高遮挡场景下召回率迅速下降，且边缘设备持续运行功耗接近无风扇散热极限^[1]。如何在保证精度的前提下压缩计算与存储开销，成为制约系统落地的关键瓶颈。本文面向“三班联播”复杂教学场景，提出系统化算法优化框架，从遮挡感知检测、轻量化时序建模、空间关系到边缘-云协同部署展开研究，旨

在为智慧教育提供可复制、可扩展的技术路径。

1 复杂教学场景需求与教师行为检测瓶颈

在“三班联播”智慧教室场景中，教师需在 3 间物理教室同步授课，摄像机阵列覆盖讲台、板书区与学生区，形成多视角、高并发、强遮挡的复杂视频流。传统检测模型通常以单视角、无遮挡假设为前提，当教师低头调试实验器材、背身书写板书或半身被讲台遮挡时，可见关键特征面积骤减，导致召回率显著下降^[2]。对 2 000 min 原始录像逐帧标注后发现，低头、背身、板书遮挡三类极端姿态共出现 37% 的帧，对应基线 YOLOv5s 模型的召回率跌至 68% 以下，误检则主要集中在学生人头与投影高光区域，说明空间上下文与遮挡

1. 许昌电气职业学院 河南许昌 461000

[10] Hamm J, Lee D D .Grassmann discriminant analysis: a unifying view on subspace-based learning[C]//Proceedings of the 25th International Conference on Machine Learning (ICML 2008). NewYork: ACM, 2008:376-383.

[11] Zhu Pengfei, Cheng Hao, Hu Qinghua,et al.Towards generalized and efficient metric learning on riemannian manifold[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18). Stockholm: IJCAI, 2018: 3168-3174.

【作者简介】

马鑫(1996—)，男，河南郑州人，硕士研究生，研究方向：

基于广义矩阵的图像分类算法。

李慧勇(1999—)，男，河南商丘人，硕士研究生，研究方向：高维数据与图像处理。

顾强(2000—)，男，江苏扬州人，硕士研究生，研究方向：图像处理与模式识别。

邱枫(2001—)，男，浙江金华人，硕士研究生，研究方向：轴承寿命预测。

廖亮(1975—)，通信作者(email:liaoaliangis@126.com)，男，河南郑州人，博士，研究生导师、中原工学院数学与信息科学学院教授，研究方向：高维数据分析与图像处理通信。

(收稿日期：2025-11-03 修回日期：2026-05-15)