Computer Science and Application 计算机科学与应用, 2026, 16(7), 196-211
Published Online July 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.167252

Hans汉斯

基于特权信息学习的多视图支持向量机

孙　伟[1*]，陈平华[2]

[1]广州理工学院计算机科学与工程学院，广东　广州
[2]广东工业大学计算机学院，广东　广州

收稿日期：2026年6月22日；录用日期：2026年7月20日；发布日期：2026年7月29日

摘　要

多视图支持向量机是处理多视图特征数据的典型学习方法，在小样本、高可解释性需求场景下仍具有不可替代的优势。现有多视图支持向量机方法直接在原始高维数据上构建分类器，未考虑高维数据中的冗余与噪声，且多数将“降维”与“分类”视为独立过程，导致分类性能受限。针对上述问题，本文提出一种基于特权信息学习的多视图支持向量机(MVDRP)，首次将高维数据降维与多视图SVM分类器训练融合为统一凸优化模型并同步优化，同时引入特权信息学习实现视图间互补信息的双向传递，确保模型同时满足多视图学习的一致性与互补性原则。在图像、文本、多媒体三类真实多视图数据集上的实验表明，本文模型不仅在分类精度与F1分数上优于传统“降维 + 分类”串联方法及现有多视图SVM方法，在小样本、低计算资源场景下相比主流深度学习多视图方法也具有显著优势。

关键词

多视图学习，降维，特权信息学习，支持向量机，凸优化

Multi-View Support Vector Machine Based on Privileged Information Learning

Wei Sun[1*], Pinghua Chen[2]

[1]School of Computer Science and Engineering, Guangzhou Institute of Science and Technology, Guangzhou Guangdong
[2]School of Computer Science and Technology, Guangdong University of Technology, Guangzhou Guangdong

Received: June 22, 2026; accepted: July 20, 2026; published: July 29, 2026

Abstract

Multi-view Support Vector Machine is a typical learning method for processing multi-view feature

*通讯作者。

文章引用: 孙伟, 陈平华. 基于特权信息学习的多视图支持向量机[J]. 计算机科学与应用, 2026, 16(7): 196-211.
DOI: 10.12677/csa.2026.167252

data, which still has irreplaceable advantages in scenarios requiring small samples and high interpretability. Existing MvSVM methods construct classifiers directly on original high-dimensional data without considering the redundancy and noise, and most treat "dimensionality reduction" and "classification" as independent processes, leading to limited classification performance. To address these issues, this paper proposes a Privileged Information Learning-based Multi-view Support Vector Machine. For the first time, this model integrates high-dimensional data dimensionality reduction and MvSVM classifier training into a unified convex optimization model for simultaneous optimization. Meanwhile, privileged information learning is introduced to realize bidirectional transmission of complementary information between views, ensuring that the model conforms to both the consistency and complementarity principles of multi-view learning. Experiments on three types of real-world multi-view datasets (image, text and multimedia) show that the proposed model outperforms traditional "dimensionality reduction + classification" tandem methods and existing MvSVM methods in both classification accuracy and F1-score, and also has significant advantages over mainstream deep learning multi-view methods in small sample and low computational resource scenarios.

Keywords

Multi-View Learning, Dimensionality Reduction, Privileged Information Learning, Support Vector Machine, Convex Optimization

1. 引言

支持向量机(Support vector machine, SVM)是机器学习领域的经典研究方向，在非线性数据分类任务中表现出优异的泛化能力和可解释性，其核心是通过最大化结构风险与最小化经验风险找到最优分类超平面[1]。针对二分类问题，SVM 通过最大化类间间隔实现样本分离，并基于学习到的超平面完成测试样本预测，近年来已广泛应用于模式识别[2]、时间序列预测[3]和故障诊断[4]等领域。多视图学习[5]通过在联合优化函数中建模各视图特征，提升模型泛化能力[6]。多视图学习遵循两大核心原则：互补性原则和一致性原则。

多视图支持向量机[7]在处理多视图数据领域取得实质性进展，例如，MTSVM [8]将协同训练与特征空间内的随机采样方法相结合，通过提供额外的信息标签来训练子分类器；多视图最小二乘支持向量机[9]通过最小化所有视图误差来实现一致性；Yu 等[10]提出一种基于实例的多视图支持向量机算法，该算法研究每个实例的独特特征；Tang 等[11]提出一种基于非并行 SVM 的视图学习模型，通过最大间隔机制和一致性原则将非并行 SVM 扩展到多视图学习；Tang 等[12]利用特权信息整合学习过程中的所有信息，并保留不同视图上的特征以实现互补原则。上述方法要么遵循一致性原则，要么遵循互补原则。与这些方法不同，Tang 等[13]提出一种多视图特权 SVM (PSVM-2V)，它同时体现这两个原则，根据特权信息学习(Learning using privileged information, LUPI)范式利用不同视图作为特权信息，实现互补原则。此外，引入正则化约束来弥合这些不同视图之间的差异，从而确保遵守一致性原则。

近年来，多视图学习的主流研究方向逐渐转向深度学习框架，如 MVCNN [14]、MvDNN [15]、SCA-PVNet [16]等方法通过深度神经网络自动提取多视图特征，在大规模标注数据场景下取得优异性能。然而，深度学习方法存在以下缺陷：1) 依赖大量标注数据，在小样本场景下泛化能力急剧下降；2) 模型可解释性差，无法明确各视图特征对分类结果的贡献；3) 计算资源需求高，难以部署在边缘设备上。相比之下，

传统 SVM 方法在小样本、高可解释性、低计算资源场景下仍具有显著优势，因此对多视图 SVM 的优化研究仍具有重要的理论与应用价值。

多视图数据常具有高维性，如图像分类的 SIFT 特征、文本分类的词向量[17] [18]，直接在高维数据上训练多视图支持向量机会因冗余与噪声导致分类精度下降[19]。传统解决方案分为两类：(1) 无监督降维+多视图 SVM，如 PCA [20]、因子分析[21]等无监督降维方法，虽能降低维度，但无法保留标签信息，导致后续多视图 SVM 性能受限[22]；(2) 监督降维 + 多视图 SVM，如 LDA [23]等监督降维方法，先将数据降维至低维空间，再训练多视图 SVM 分类器，但此类方案将“降维”与“多视图 SVM 分类”视为独立过程，忽略二者的内在联系。降维与分类是多视图 SVM 学习的两个核心环节，二者需协同而非独立。尽管 Tang 等[13]在 2017 年提出了结合特权信息的多视图 SVM 方法，但该方法仍未解决高维数据的冗余与噪声问题。为此，本文提出基于特权信息学习的多视图支持向量机(MVDRP)，将“降维”与“多视图 SVM 分类”融合为统一优化模型，实现投影矩阵与分类参数的同步优化，确保降维过程保留多视图 SVM 所需的判别信息；同时引入特权信息学习，以一个视图作为另一个视图的特权知识，强化视图间的互补性；最后采用交替优化框架求解模型，图 1 展示 MVDRP 方法的模型框架图。本文的核心创新如下：

(1) 提出“降维–分类”联合优化的多视图 SVM 框架，同步优化降维投影与分类参数，解决传统串联方案中降维过程与分类需求脱节导致的信息丢失问题；

(2) 模型同时满足多视图学习的一致性与互补性原则，实现了视图间信息的双向传递；

(3) 严格证明目标函数的凸性与算法的收敛性，为模型的理论可靠性提供保障。

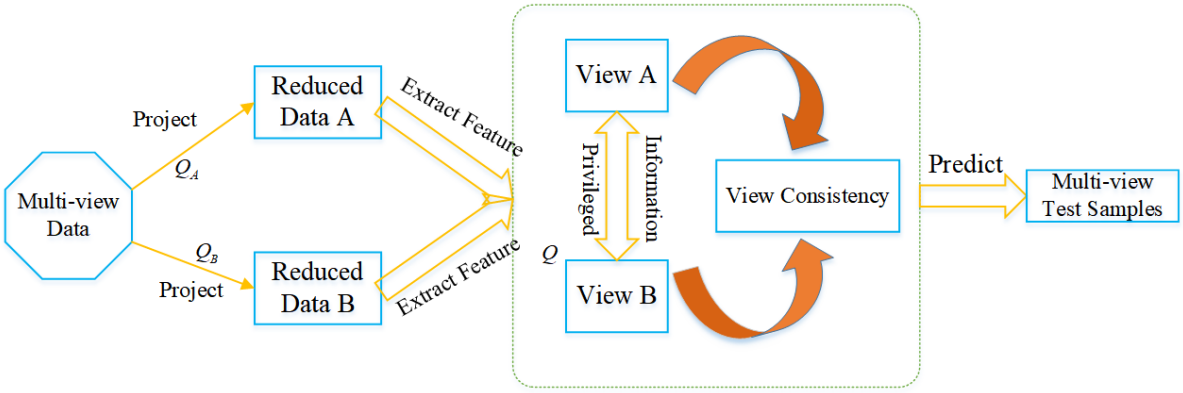

Figure 1. Model framework figure
图 1. 模型框架图

2. 相关工作

2.1. 多视图学习

Xu 等人[24]采用一种多视图算法来促进全面的空间学习，融合来自各个视图的互补信息，形成数据集的整体表示；Li 等人[25]提出 MTSVM 方法，该方法将协同训练与采样模型相结合；MV3MR [26]将信息量更大的视图分配给具有更高权重的特定核，并在利用标签相关性学习视图权重的过程中探索不同视图之间的互补性；MVLapSVM [27]是一种在 SVM 框架内结合流形正则化和多视图正则化的技术，利用标记数据和未标记数据来缩小不同视图之间的差距。文献[7]对 SVM-2K 模型进行改进，提出一种基于核典型相关分析的 SVM-2K 模型。现有的多视图 SVM 方法直接在原始数据上学习分类器，没有考虑高维数据的特性[28]，当在高维数据上学习分类器，其分类精度可能会受到限制。

2.2. 降维

降维是一种将高维数据转换为低维表示，同时保留原始数据分布基本特征的技术[29]。现有的降维研究[30] [31]可分为三类，第一类是无监督方法，不考虑标签信息，例如，主成分分析是无监督降维的代表性方法之一。第二类是监督方法，它在训练过程中考虑标签信息，例如，线性判别分析是一种典型的监督方法。第三类是半监督方法，训练数据集包含未标记数据和标记数据，例如，拉普拉斯正则化最小二乘法[32]和半监督判别分析[33]是两种广泛使用的半监督降维方法。现有的降维方法大多首先将原始数据降维为低维数据[34]，然后将低维数据作为输入训练多视图分类器，将降维和多视图分类器训练视为两个独立的过程，忽略它们之间的联系。

2.3. 深度学习多视图分类方法

近年来，深度学习在多视图学习领域取得了显著进展。MVCNN [14]通过卷积神经网络分别提取每个视图的特征，然后通过最大池化融合多视图特征进行分类；MvDNN [15]采用多层感知机分别处理每个视图，通过共享层实现视图间的信息交互；SCA-PVNet [16]引入交叉注意力机制，动态捕捉不同视图之间的关联关系。这些方法在大规模标注数据场景下表现优异，但存在小样本泛化能力差、可解释性弱、计算成本高等问题。本文提出 MVDRP 方法结合传统 SVM 的小样本优势与多视图学习的互补性优势，为小样本高维多视图分类任务提供一种新的解决方案。

3. 模型整体设计

3.1. 目标函数

3.1.1. 符号定义说明

具体来说，假设训练集有 l 个训练样本 $\left\{\left(x_i^A, x_i^B, y_i\right)\right\}_{i=1}^{l}$，其中 $x_i^A \in R^{D_A}$ 是视图 A 的第 i 个样本；$x_i^B \in R^{D_B}$ 是视图 B 的第 i 个样本；$y_i \in \{-1,1\}$ 是第 i 个样本的标签，原始数据 x_i^A 和 x_i^B 分别通过投影矩阵 $\boldsymbol{Q}_A \in R^{D_A \times d_A}$ 和 $\boldsymbol{Q}_A \in R^{D_A \times d_A}$ 投影到两个低维空间中，其中 d_A 和 d_B 是低维空间的维数。对于 Q_A 和 Q_B，对其施加正交矩阵约束 $Q_A^{\mathrm{T}} Q_A = I_A$ 和 $Q_B^{\mathrm{T}} Q_B = I_B$，加速特征值求解与模型收敛。多视图 SVM 的分类边界定义为：视图 A 的分类器 $f_A(x) = Q_A W_A \cdot \phi\left(x_i^A\right)$，其中 $W_A \in R^X$，X 为视图 A 的范数向量，$\phi\left(x_i^A\right)$ 为样本 x_i^A 在特征空间的映射。同理，视图 B 的分类器 $f_B(x) = Q_B W_B \cdot \phi\left(x_i^B\right)$。

3.1.2. 联合优化目标函数

MVDRP 的目标函数满足在降低多视图数据维度的同时，确保低维特征满足多视图 SVM 的最大间隔约束，且视图间兼具一致性与互补性。其具体形式如公式(1)所示：

$$\min_{\hat{W}_A, \hat{W}_B, Q_A, Q_B} \frac{1}{2}\left(\left\|\hat{W}_A\right\|^2 + \gamma\left\|\hat{W}_B\right\|^2\right) + C^A \sum_{i=1}^{l} \xi_i^{A^*} + C^B \sum_{i=1}^{l} \xi_i^{B^*} + C \sum_{i=1}^{l} \eta_i \tag{1}$$

$$\text{s.t.} \quad \left|\hat{W}_A^{\mathrm{T}} Q_A^{\mathrm{T}} \cdot \phi_A\left(x_i^A\right) - \hat{W}_B^{\mathrm{T}} Q_B^{\mathrm{T}} \cdot \phi_B\left(x_i^B\right)\right| \le \varepsilon + \eta_i, \quad \eta_i \ge 0$$

$$y_i\left(\hat{W}_A^{\mathrm{T}} Q_A^{\mathrm{T}} \cdot \phi_A\left(x_i^A\right)\right) \ge 1 - \xi_i^{A^*}$$

$$y_i\left(\hat{W}_B^{\mathrm{T}} Q_B^{\mathrm{T}} \cdot \phi_B\left(x_i^B\right)\right) \ge 1 - \xi_i^{B^*}$$

$$\xi_i^{A^*} \ge y_i\left(\hat{W}_B^{\mathrm{T}} Q_B^{\mathrm{T}} \cdot \phi_B\left(x_i^B\right)\right), \quad \xi_i^{A^*} \ge 0$$

$$\xi_i^{B^*} \ge y_i\left(\hat{W}_A^{\mathrm{T}} Q_A^{\mathrm{T}} \cdot \phi_A\left(x_i^A\right)\right), \quad \xi_i^{B^*} \ge 0$$

$$Q_A^{\mathrm{T}} Q_A = I_A, \quad Q_B^{\mathrm{T}} Q_B = I_B$$

1) $\gamma \geq 0, C_A \geq 0, C_B \geq 0$ 和 $C \geq 0$ 是权衡参数，$\eta_i, \xi_i^{A^*}$ 和 $\xi_i^{B^*}$ 是松弛变量，$\frac{1}{\left\|\hat{W}_A\right\|^2}$ 和 $\frac{1}{\left\|\hat{W}_B\right\|^2}$ 分别是视图 A 和视图 B 的分类边界。通过最小化 $\left\|\hat{W}_A\right\|^2$ 和 $\left\|\hat{W}_B\right\|^2$ 来最大化这两个边界，γ 是权衡视图 A 和视图 B 中分类器关系的参数。

2) 参数 $\varepsilon \geq 0$ 是非负视图一致性参数，η_i 控制视图 A 和视图 B 中预测值的差异，如果预测值的差异不大于阈值，则 $\eta_i = 0$，否则，$\eta_i > 0$，确保视图 A 和视图 B 的标签一致性。

3) 第四个和第五个约束分别将 $\xi_i^{A^*}$ 和 $\xi_i^{B^*}$ 视为视图 A 和视图 B 的松弛变量，这两个松弛变量由两个视图相互评估，体现多视图学习的互补原理。

4) $Q_A^{\mathrm{T}} Q_A = I_A$ 和 $Q_B^{\mathrm{T}} Q_B = I_B$ 确保投影矩阵正交，简化优化并保留特征分布信息，降维过程通过投影矩阵与分类器参数的同步优化，确保低维特征服务于多视图 SVM 的最大间隔需求。

3.2. 优化过程

由于约束条件 $Q_A^{\mathrm{T}} Q_A = I_A$ 和 $Q_B^{\mathrm{T}} Q_B = I_B$ 导致非凸性，需通过变量替换与交替优化求解。

3.2.1. 变量重定义与目标函数转化

首先定义 $W_A \in R^{D_A}$ 和 $W_B \in R^{D_B}$，如下所示：

$$W_A = Q_A \hat{W}_A, \quad W_B = Q_B \hat{W}_B \tag{2}$$

将公式(2)代入公式(1)中，结合正交约束的性质，可转化为凸优化问题(证明见定理 1)，公式(1)重写为公式(3)：

$$\min_{W_A, W_B, Q} \frac{1}{2}\left(\|W_A\|^2 + \gamma \|W_B\|^2\right) + C^A \sum_{i=1}^{l} \xi_i^{A^*} + C^B \sum_{i=1}^{l} \xi_i^{B^*} + C \sum_{i=1}^{l} \eta_i \tag{3}$$

$$\begin{aligned}
\text{s.t. } & \left|W_A \cdot \phi_A\left(x_i^A\right) - W_B \cdot \phi_B\left(x_i^B\right)\right| \leq \varepsilon + \eta_i, \quad \eta_i \geq 0 \\
& y_i\left(W_A \cdot \phi_A\left(x_i^A\right)\right) \geq 1 - \xi_i^{A^*} \\
& y_i\left(W_B \cdot \phi_B\left(x_i^B\right)\right) \geq 1 - \xi_i^{B^*} \\
& \xi_i^{A^*} \geq y_i\left(W_B \cdot \phi_B\left(x_i^B\right)\right), \quad \xi_i^{A^*} \geq 0 \\
& \xi_i^{B^*} \geq y_i\left(W_A \cdot \phi_A\left(x_i^A\right)\right), \quad \xi_i^{B^*} \geq 0 \\
& Q_A^{\mathrm{T}} Q_A = I_A, \quad Q_B^{\mathrm{T}} Q_B = I_B \\
& W_A = Q_A Q_A^{\mathrm{T}} W_A, \quad W_B = Q_B Q_B^{\mathrm{T}} W_B
\end{aligned}$$

在介绍公式(3)的解决方法之前，先给出公式转化的证明如下：

定理 1. 考虑有 $W_A = Q_A \hat{W}_A, W_B = Q_B \hat{W}_B, Q_A^{\mathrm{T}} Q_A = I_A$ 和 $Q_B^{\mathrm{T}} Q_B = I_B$，可得到如下公式。

1) $\hat{W}_A = Q_A^{\mathrm{T}} W_A$； $\hat{W}_B = Q_B^{\mathrm{T}} W_B$。

2) $W_A = Q_A Q_A^{\mathrm{T}} W_A$； $W_B = Q_B Q_B^{\mathrm{T}} W_B$。

3) $\hat{W}_A^{\mathrm{T}} \hat{W}_A = W_A^{\mathrm{T}} W_A$； $\hat{W}_B^{\mathrm{T}} \hat{W}_B = W_B^{\mathrm{T}} W_B$。

证明 1. 在定理 1 中，有 $W_A \in R^{d_A}, W_B \in R^{d_B}, \hat{W}_A \in R^{D_A}, \hat{W}_B \in R^{D_B}, Q_A \in R^{D_A \times d_A}$ 和 $Q_B \in R^{D_B \times d_B}$，基于此，可以得到以下偏差。

1) $\hat{W}_A = Q_A^{\mathrm{T}} Q_A \hat{W}_A = Q_A^{\mathrm{T}} W_A$。

2) $\hat{W}_B = Q_B^{\mathrm{T}} Q_B \hat{W}_B = Q_B^{\mathrm{T}} W_B$。

3) $W_A = Q_A \hat{W}_A = Q_A\left(Q_A^{\mathrm{T}} W_A\right) = Q_A Q_A^{\mathrm{T}} W_A$。

4) $W_B = Q_B\hat{W}_B = Q_B\left(Q_B^{\mathrm{T}}W_B\right) = Q_BQ_B^{\mathrm{T}}W_B$。

5) $\hat{W}_A^{\mathrm{T}}\hat{W}_A = \left(Q_A^{\mathrm{T}}W_A\right)^{\mathrm{T}}Q_A^{\mathrm{T}}W_A = W_A^{\mathrm{T}}Q_AQ_A^{\mathrm{T}}W_A = W_A^{\mathrm{T}}W_A$。

6) $\hat{W}_B^{\mathrm{T}}\hat{W}_B = \left(Q_B^{\mathrm{T}}W_B\right)^{\mathrm{T}}Q_B^{\mathrm{T}}W_B = W_B^{\mathrm{T}}Q_BQ_B^{\mathrm{T}}W_B = W_B^{\mathrm{T}}W_B$。

解 Q_A 和 Q_B 与约束 $Q_A^{\mathrm{T}}Q_A = I_A$ 和 $Q_B^{\mathrm{T}}Q_B = I_B$ 相关，利用定理 1 将公式(1)转化为公式(3)，一方面，根据定理 1，有 $\hat{W}_A^{\mathrm{T}}\hat{W}_A = W_A^{\mathrm{T}}W_A$ 且 $\hat{W}_B^{\mathrm{T}}\hat{W}_B = W_B^{\mathrm{T}}W_B$，公式(1)中的项 $\hat{W}_B^{\mathrm{T}}\hat{W}_B = W_B^{\mathrm{T}}W_B$ 和 $\left\|\hat{W}_B\right\|^2$ 可以分别替换为公式(3)中的 $\|W_A\|^2$ 和 $\|W_B\|^2$，另一方面，有 $W_A = Q_A\hat{W}_A$ 和 $W_B = Q_B\hat{W}_B$，因此，分别将公式(1)中的 $\hat{W}_A^{\mathrm{T}}Q_A^{\mathrm{T}}$ 和 $\hat{W}_B^{\mathrm{T}}Q_B^{\mathrm{T}}$ 替换为 W_A 和 W_B，即可得到公式(3)。

3.2.2. 优化 Q_A 和 Q_B

本节将介绍如何求解 Q_A 和 Q_B。为了优化 Q_A 和 Q_B，首先去除对 Q_A 和 Q_B 的约束，即 $Q_iQ_i = I_i$ 和 $\hat{W}_i = Q_iQ_i\hat{W}_i\left(i = A,B\right)$，求解公式(3)得到 W_A 和 W_B 的初始值，然后，固定 W_A 和 W_B 的值，对 Q_A 和 Q_B 进行优化，具体步骤如下：

求解 W_A 和 W_B：去除对 Q_A 和 Q_B 的约束，将目标函数转化为满足多视角学习中的一致性原理和互补性原理的二次规划，如下所示。

$$\min_{W_A,W_B}\frac{1}{2}\left(\|W_A\|^2 + \gamma\|W_B\|^2\right) + C^A\sum_{i=1}^{l}\xi_i^{A^*} + C^B\sum_{i=1}^{l}\xi_i^{B^*} + C\sum_{i=1}^{l}\eta_i \tag{4}$$

$$\text{s.t.}\quad \left|W_A\cdot\phi_A\left(x_i^A\right) - W_B\cdot\phi_B\left(x_i^B\right)\right| \le \varepsilon + \eta_i,\quad \eta_i \ge 0$$

$$y_i\left(W_A\cdot\phi_A\left(x_i^A\right)\right) \ge 1 - \xi_i^{A^*}$$

$$y_i\left(W_B\cdot\phi_B\left(x_i^B\right)\right) \ge 1 - \xi_i^{B^*}$$

$$\xi_i^{A^*} \ge y_i\left(W_B\cdot\phi_B\left(x_i^B\right)\right),\quad \xi_i^{A^*} \ge 0$$

$$\xi_i^{B^*} \ge y_i\left(W_A\cdot\phi_A\left(x_i^A\right)\right),\quad \xi_i^{B^*} \ge 0$$

求解 Q_A 和 Q_B：设 W_A^* 和 W_B^* 分别为公式(4)得到的最优值，则 $W_A^* \in R^{D_A}, W_B^* \in R^{D_B}, Q_A \in R^{D_A\times d_A}, Q_B \in R^{D_B\times d_B}$，有如下公式：

$$Q_A^{\mathrm{T}}Q_A = I_A, W_A^* = Q_AQ_A^{\mathrm{T}}W_A^* \tag{5}$$

$$Q_B^{\mathrm{T}}Q_B = I_B, W_B^* = Q_BQ_B^{\mathrm{T}}W_B^* \tag{6}$$

公式(5)包含 $D_A\times d_A$ 个变量和 $\left(d_A\times d_A + D_A\right)$ 个方程，公式(6)包含 $D_B\times d_B$ 个变量和 $\left(d_B\times d_B + D_B\right)$ 个方程。当方程数量大于变量数量，即 $D_A\times d_A \le d_A\times d_A + D_A$ 或 $D_B\times d_B \le d_B\times d_B + D_B$，则可能无解。为了避免这种情况，$Q_AQ_A^{\mathrm{T}}$ 和 $Q_BQ_B^{\mathrm{T}}$ 通过 W_A^* 和 W_B^* 以使尽可能接近 $Q_AQ_A^{\mathrm{T}}W_A^*$ 和 $Q_BQ_B^{\mathrm{T}}W_B^*$，因此，可以通过以下优化公式求解 Q_A 和 Q_B：

$$\min_{Q_A^{\mathrm{T}}Q_A = I_A}\sum_{i=1}^{l}\left\|Q_AQ_A^{\mathrm{T}}W_A^* - W_A^*\right\|^2, \tag{7}$$

$$\min_{Q_B^{\mathrm{T}}Q_B = I_B}\sum_{i=1}^{l}\left\|Q_BQ_B^{\mathrm{T}}W_B^* - W_B^*\right\|^2, \tag{8}$$

基于线性数学中的规范化形式，将公式(7)和(8)重写如下：

$$\max_{Q_A^{\mathrm{T}}Q_A = I_A} tr\left(Q_A^{\mathrm{T}}\boldsymbol{W}_A^*\boldsymbol{W}_A^{*\mathrm{T}}Q_A\right), \tag{9}$$

$$\max_{Q_B^{\mathrm{T}}Q_B = I_B} tr\left(Q_B^{\mathrm{T}}\boldsymbol{W}_B^*\boldsymbol{W}_B^{*\mathrm{T}}Q_B\right), \tag{10}$$

公式(9)和(10)是带正交矩阵约束的迹最大化优化问题，其中最优解 Q_A 和 Q_B 分别是矩阵 $\boldsymbol{W}_A^*$ 和 $\boldsymbol{W}_B^*$ 的 d_A 和 d_B 最大特征值对应的特征向量，同时可以将 Q_A 和 Q_B 计算为 $\boldsymbol{W}_A^*$ 和 $\boldsymbol{W}_B^*$ 的左奇异向量。

3.2.3. 通过固定 Q_A 和 Q_B 来优化 $\hat{W}_A$ 和 $\hat{W}_B$

公式(7)和(8)得到 Q_A 和 Q_B，且有 $Q_A = Q_A^*$ 和 $Q_A = Q_A^*$，通过以下公式计算(1)中的 $\hat{W}_A$ 和 $\hat{W}_B$，

$$\min_{\hat{W}_A,\hat{W}_B} \frac{1}{2}\left(\left\|\hat{W}_A\right\|^2 + \gamma\left\|\hat{W}_B\right\|^2\right) + C^A\sum_{i=1}^{l}\xi_i^{A^*} + C^B\sum_{i=1}^{l}\xi_i^{B^*} + C\sum_{i=1}^{l}\eta_i \tag{11}$$

$$\text{s.t. } \left|\hat{W}_A^{\mathrm{T}}\cdot\hat{\phi}_A\left(x_i^A\right) - \hat{W}_B^{\mathrm{T}}\cdot\hat{\phi}_B\left(x_i^B\right)\right| \le \varepsilon + \eta_i,\ \ \eta_i \ge 0$$

$$y_i\left(\hat{W}_A^{\mathrm{T}}\cdot\hat{\phi}_A\left(x_i^A\right)\right) \ge 1 - \xi_i^{A^*}$$

$$y_i\left(\hat{W}_B^{\mathrm{T}}\cdot\hat{\phi}_B\left(x_i^B\right)\right) \ge 1 - \xi_i^{B^*}$$

$$\xi_i^{A^*} \ge y_i\left(\hat{W}_B^{\mathrm{T}}\cdot\hat{\phi}_B\left(x_i^B\right)\right),\ \ \xi_i^{A^*} \ge 0$$

$$\xi_i^{B^*} \ge y_i\left(\hat{W}_A^{\mathrm{T}}\cdot\hat{\phi}_A\left(x_i^A\right)\right),\ \ \xi_i^{B^*} \ge 0$$

上述公式是 QP 问题，可以通过其对偶形式求解。引入非负拉格朗日函数 $\alpha_i^A,\alpha_i^B,\beta_i^+,\beta_i^-,\beta_i^A,\beta_i^B$ 和 τ_i，可得到公式(11)的拉格朗日函数 L。

根据对偶理论，最小化原目标函数(11)对应于最大化关于 $\alpha_i^A,\alpha_i^B,\beta_i^+,\beta_i^-,\beta_i^A,\beta_i^B$ 和 τ_i 的拉格朗日函数。因此，对变量 $\hat{W}_A,\hat{W}_B,\xi_i^{A^*},\xi_i^{B^*}$ 和 η_i 取导数，并使其为零，可得到如下方程：

$$\partial_{\hat{W}_A} L = \hat{W}_A - \sum_{i=1}^{l}\left(\alpha_i^A y_i - \beta_i^+ + \beta_i^- - \lambda_i^B y_i\right)\hat{\phi}_A\left(x_i^A\right) = 0, \tag{12}$$

$$\partial_{\hat{W}_B} L = \gamma\hat{W}_B - \sum_{i=1}^{l}\left(\alpha_i^B y_i - \beta_i^+ + \beta_i^- - \lambda_i^A y_i\right)\hat{\phi}_B\left(x_i^B\right) = 0, \tag{13}$$

$$\partial_{\xi_i^{A^*}} L = C^A - \left(\alpha_i^A + \lambda_i^A + \beta_i^A\right) = 0 \tag{14}$$

$$\partial_{\xi_i^{B^*}} L = C^B - \left(\alpha_i^B + \lambda_i^B + \beta_i^B\right) = 0 \tag{15}$$

$$\partial_{\eta_i} L = C - \left(\beta_i^+ + \beta_i^- + \tau_i\right) = 0 \tag{16}$$

由于 $\beta_i^A \ge 0, \beta_i^B \ge 0$ 和 τ_i 成立，根据公式(14)~(16)，可得 $\alpha_i^A + \lambda_i^A \le C^A, \alpha_i^B + \lambda_i^B \le C^B$ 和 $\beta_i^+ + \beta_i^- \le C$，将这些约束代入拉格朗日函数，可得到对偶形式(17)。

$$\begin{aligned}\min \frac{1}{2}\sum_{i,j=1}^{l}\Bigg[&\left(\alpha_i^A y_i - \beta_i^+ + \beta_i^- - \lambda_i^B y_i\right)\left(\alpha_j^A y_i - \beta_j^+ + \beta_j^- - \lambda_i^B y_i\right)\kappa_A\left(x_i^A, x_j^A\right)\\ &+ \frac{1}{\gamma}\left(\alpha_i^B y_i + \beta_i^+ - \beta_i^- - \lambda_i^A y_i\right)\kappa_B\left(x_i^B, x_j^B\right)\Bigg] - \sum_{i=1}^{l}\left(\alpha_i^A + \alpha_i^B\right) + \varepsilon\sum_{i=1}^{l}\left(\beta_i^+ + \beta_i^-\right)\\ \text{s.t. } &0 \le \alpha_i^A + \lambda_i^A \le C^A, 0 \le \alpha_i^B + \lambda_i^B \le C^B, 0 \le \beta_i^+ + \beta_i^- \le C\\ &\alpha_i^A,\alpha_i^B,\beta_i^+,\beta_i^-,\lambda_i^A,\lambda_i^B \ge 0\end{aligned} \tag{17}$$

求解对偶形式(17)后，可得到 $\alpha^A,\alpha^B,\beta^+,\beta^-,\lambda^A,\lambda^B$ 的值，由此进一步可得到 $\hat{W}_A$ 和 $\hat{W}_B$ 的解为：

$$\hat{W}_A = \sum_{i=1}^{l}\left(\alpha_i^A y_i - \beta_i^+ + \beta_i^- - \lambda_i^B y_i\right)\hat{\phi}_A\left(x_i^A\right) \tag{18}$$

$$\hat{W}_B = \frac{1}{\gamma}\sum_{i=1}^{l}\left(\alpha_i^B y_i + \beta_i^+ - \beta_i^- - \lambda_i^A y_i\right)\hat{\phi}_B\left(x_i^B\right) \tag{19}$$

一旦获得最优 $\hat{W}_A^*$ 和 $\hat{W}_B^*$，使用以下公式分别从视图 A 和 B 预测新样本 $x_t=\left(x_t^A,x_t^B\right)$ 的标签：

$$f_A\left(x_t^A\right)=\hat{W}_A^{*\mathrm{T}}\hat{\phi}_A\left(x_t^A\right) \tag{20}$$

$$f_B\left(x_t^B\right)=\hat{W}_B^{*\mathrm{T}}\hat{\phi}_B\left(x_t^B\right) \tag{21}$$

视图 A 和视图 B 的特征可以互补，最终的预测结果由各个视图的结果决定：

$$y_t=sign\left[\alpha f_A\left(x_t^A\right)+(1-\alpha)f_B\left(x_t^B\right)\right] \tag{22}$$

其中 α 是权重参数，表示每个视图对最终预测的重要性，在本实验中，将 α 设置为 0.5，MVDRP 的概述如算法 1 所示。

算法 1 MVDRP 概述

输入：

维度 d，训练数据 $\left(x_i^A,x_i^B,y_i\right)_{i=1}^l$。

输出：

投影矩阵 $\boldsymbol{Q}_A$ 和 $\boldsymbol{Q}_B$，分类参数 $\hat{W}_A$ 和 $\hat{W}_B$。

1：$\left(W_A^*,W_B^*\right)\leftarrow$ 使用 $\left(x_i^A,x_i^B,y_i\right)$ 解决问题(7)和(8)；

2：$\left(\boldsymbol{Q}_A,\boldsymbol{Q}_B\right)\leftarrow$ 关于 d 个最大奇异值的 W_A^* 和 W_B^* 的左奇异向量；

3：$\hat{\phi}_A\left(x_i^A\right)\leftarrow Q_A\cdot\phi\left(x_i^A\right)$ 和 $\hat{\phi}_B\left(x_i^B\right)\leftarrow Q_B\cdot\phi\left(x_i^B\right)$；

4：$\left(\hat{W}_A^*,\hat{W}_B^*\right)\leftarrow$ 解决问题(17)；

5：对于一个测试样本，$x_t=\left(x_t^A,x_t^B\right)$，根据公式(22)预测它的标签 y_t。

3.2.4. 收敛性分析

本文采用交替优化策略求解模型，其收敛性可通过目标函数的单调性与有界性证明如下：

定理 1. MVDRP 交替优化，其目标函数值序列 $\left\{J^{(t)}\right\}_{t=1}^{\infty}$ 单调不增，且存在下界，算法必然收敛。

(1) 单调性证明：设第 t 轮迭代中，先固定投影矩阵 $W_A^{(t)},W_B^{(t)}$，求解分类参数 $w_A^{(t)},w_B^{(t)},b_A^{(t)},b_B^{(t)}$。该子问题为凸二次规划，求得的全局最优解必然满足：

$$J\left(W_A^{(t)},W_B^{(t)},w_A^{(t)},w_B^{(t)}\right)\le J\left(W_A^{(t)},W_B^{(t)},w_A^{(t-1)},w_B^{(t-1)}\right)$$

随后固定分类参数，求解投影矩阵 $W_A^{(t+1)},W_B^{(t+1)}$，该子问题的最优解满足：

$$J\left(W_A^{(t+1)},W_B^{(t+1)},w_A^{(t)},w_B^{(t)}\right)\le J\left(W_A^{(t)},W_B^{(t)},w_A^{(t)},w_B^{(t)}\right)$$

联立可得 $J^{(t+1)}\le J^{(t)}$，即目标函数值随迭代单调不增。

(2) 有界性证明：目标函数中 $\frac{1}{2}\|w_A\|^2+\frac{1}{2}\|w_B\|^2\ge 0$，松弛变量项 $C_1\sum\xi_A+C_2\sum\xi_B\ge 0$，一致性约束项与特权信息项均为非负正则项，因此目标函数存在下界 0。

综上，单调递减且有下界的序列必然收敛，证毕。

4. 实验结果与分析

4.1. 数据集

选择三类具有本质差异的多视图数据集，如表 1 所示，验证 MVDRP 的通用性与鲁棒性。

Table 1. Dataset introduction
表 1. 数据集介绍

数据集	类型	视图特征(维度)	类别数	样本量
Caltech	图像	视图 A：LBP (928)；视图 B：HOG (1984)	7	1474
WebKB	文本网页	视图 A：内容词向量(1703)；视图 B：引用链接(569)	5	195
Movies	多媒体	视图 A：关键词(1878)；视图 B：演员(1398)	17	617

4.2. 基线方法

为了客观地评价本模型的性能，分为四类对比方法，明确“降维”与“分类”的逻辑关系：

(1) 传统“降维 + 多视图 SVM”串联：PCA + PSVM-2V (PCA 降维后用 PSVM-2V 分类)、LDA + PSVM-2V (LDA 降维后用 PSVM-2V 分类)；

(2) 传统多视图 SVM 方法：SVM-2K (仅满足一致性原则的多视图 SVM)、PSVM-2V (兼顾一致性与互补性的多视图特权 SVM)、MvSVM-2C (融合共识与互补信息的多视图 SVM)、MPWTSVM (特权加权孪生多视图 SVM)、GMIGSVMs (广义特征值多视图 SVM)；

(3) 特权信息学习方法：SVM+ (基于特权信息调整松弛变量的 SVM)；

(4) 深度学习多视图分类方法：MVCNN (多视图卷积神经网络)、MvDNN (多视图深度神经网络)、SCA-PVNet (交叉注意力多视图网络)。

4.3. 实验设置

(1) 所有方法采用径向基函数(RBF)核，核参数 σ 的取值范围为{0.001, 0.01, 0.1, 1, 10, 100}。

(2) PCA + PSVM-2V、LDA + PSVM-2V、SVM+和 MvSVM-2C 中的参数 C 在集合{0.01, 0.1, 1, 10, 100, 1000}范围内进行调整。

(3) 在 MPWTSVM 中，令 $C^A = C^B = C$，C 和 D 的值在{0.01, 0.1, 1, 10, 100, 1000}中搜索。

(4) 在 GMIGSVM 中，设置 c1 = c2，c1 和 λ 的值在{0.01, 0.1, 1, 10, 100, 1000}中选择。

(5) 在 SVM-2K、SVM+和 MVDRP 中，设 γ 并从{0.01, 0.1, 1, 10, 100, 1000}范围中选择，参数 γ 设置为 0.01。

(6) SVM+、PSVM-2V 和 MVDRP 的权衡参数 γ 在范围{0.01, 0.1, 1, 10, 100, 1000}内调整。

(7) 深度学习方法采用 Adam 优化器，学习率为 0.001，批大小为 32，训练轮数为 100，采用早停策略防止过拟合。

对于每种基线方法，分别得到两种不同视图的性能，并将每种视图的输入特征链接起来以形成混合视图，得到混合视图的结果。除了准确率之外，还得到 F1 分数，以评估多视图分类的性能。

4.4. 实验结果

4.4.1. Caltech 数据集(高维视觉数据)

表 2 显示，MVDRP 在单视图与混合视图上均优于所有对比方法。混合视图下，MVDRP 的 Acc (92.85%)比串联方案 PCA + PSVM-2V (80.26%)、LDA + PSVM-2V (82.08%)分别提升 12.59%、10.77%，同时，从表 2 可以看出，MVDRP 的 F1 分数(95.31%)也明显高于 PCA (83.62%)和 LDA (85.47%)。PCA + PSVM-2V、LDA + PSVM-2V 的降维过程未考虑多视图 SVM 的分类需求(如最大间隔)，导致降维后特征丢失判别信息；而 MVDRP 通过联合优化，使降维过程服务于多视图 SVM 的间隔最大化目标，保留关键分类特征。与深度学习方法相比，MVDRP 在 Caltech 数据集上的 Acc 比 MVCNN (90.12%)高 2.73%，比 MvDNN

(88.76%)高 4.09%，比 SCA-PVNet (91.23%)高 1.62%。这是因为 Caltech 数据集样本量较小(1474 个样本)，深度学习方法容易过拟合，而 SVM 方法在小样本场景下具有更好的泛化能力。

Table 2. Classification accuracy and F1-score on Caltech dataset
表 2. Caltech 数据集的分类准确率和 F1 分数

模型	View A		View B		Hybrid View	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
PCA	77.43 ± 1.21	81.26 ± 1.35	79.14 ± 1.18	82.86 ± 1.27	80.26 ± 1.09	83.62 ± 1.15
LDA	78.62 ± 1.15	82.47 ± 1.22	80.37 ± 1.09	84.31 ± 1.16	82.08 ± 1.03	85.47 ± 1.08
PSVM-2V	85.74 ± 0.98	87.94 ± 1.05	88.02 ± 0.92	91.28 ± 0.97	89.14 ± 0.86	91.15 ± 0.91
SVM-2K	84.19 ± 1.03	86.81 ± 1.09	86.43 ± 0.98	89.57 ± 1.03	87.52 ± 0.92	90.26 ± 0.97
SVM+	82.96 ± 1.12	85.39 ± 1.18	84.59 ± 1.06	88.16 ± 1.12	85.36 ± 1.01	89.39 ± 1.06
MvSVM-2C	85.37 ± 0.95	87.32 ± 1.01	87.28 ± 0.89	90.15 ± 0.94	88.71 ± 0.83	90.78 ± 0.88
MPWTSVM	86.11 ± 0.91	88.39 ± 0.97	88.93 ± 0.85	92.19 ± 0.90	90.15 ± 0.79	92.14 ± 0.84
GMIGSVMs	85.93 ± 0.94	88.12 ± 1.00	88.42 ± 0.88	91.96 ± 0.93	89.71 ± 0.82	91.86 ± 0.87
MVCNN	86.72 ± 1.35	89.25 ± 1.42	88.56 ± 1.28	91.73 ± 1.33	90.12 ± 1.19	92.37 ± 1.25
MvDNN	85.39 ± 1.47	87.86 ± 1.54	87.12 ± 1.41	90.25 ± 1.46	88.76 ± 1.32	91.02 ± 1.38
SCA-PVNet	87.85 ± 1.26	90.37 ± 1.33	**89.64 ± 1.19**	**92.81 ± 1.24**	91.23 ± 1.10	93.56 ± 1.16
MVDRP	**88.57 ± 0.76**	**91.18 ± 0.82**	89.26 ± 0.71	92.52 ± 0.76	**92.85 ± 0.65**	**95.31 ± 0.70**

4.4.2. WebKB 数据集(异构文本数据)

表 3 显示，MVDRP 在混合视图上的 Acc (88.75%)与 F1 (92.39%)均为最优。比传统多视图 SVM 提升显著，MVDRP 的 Acc 比 SVM-2K (83.12%)提升 5.63%，比 PSVM-2V (84.97%)提升 3.78%，比 MPWTSVM (85.48)提升 3.27%；WebKB 的视图 A 与视图 B 维度差异大，传统多视图 SVM 直接处理高维文本特征时，冗余与噪声导致分类偏差；MVDRP 通过联合优化，在降维过程中消除异构视图的维度差异与冗余，同时保留视图互补信息，提升分类精度。与深度学习方法相比，MVDRP 在 WebKB 数据集上的 Acc 比 MVCNN (82.35%)高 6.40%，比 MvDNN (80.78%)高 7.97%，比 SCA-PVNet (84.12%)高 4.63%。WebKB 数据集样本量极小，深度学习方法严重过拟合，SVM 方法表现出极强的小样本优势。

Table 3. Classification accuracy and F1-score on WebKB dataset
表 3. WebKB 数据集的分类准确率和 F1 分数

模型	View A		View B		Hybrid View	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
PCA	74.83 ± 2.15	78.26 ± 2.32	76.53 ± 2.08	79.82 ± 2.25	77.64 ± 1.97	80.37 ± 2.14
LDA	76.21 ± 2.07	79.53 ± 2.23	77.82 ± 2.01	80.49 ± 2.17	78.95 ± 1.90	81.85 ± 2.06
PSVM-2V	83.09 ± 1.62	85.76 ± 1.75	84.79 ± 1.56	87.06 ± 1.69	84.97 ± 1.48	88.22 ± 1.61
SVM-2K	81.16 ± 1.75	83.94 ± 1.88	82.36 ± 1.69	84.71 ± 1.82	83.12 ± 1.61	85.28 ± 1.74
SVM+	80.04 ± 1.83	82.75 ± 1.96	81.25 ± 1.77	83.36 ± 1.90	82.53 ± 1.69	84.46 ± 1.82
MvSVM-2C	82.87 ± 1.65	85.19 ± 1.78	83.64 ± 1.59	86.57 ± 1.72	84.66 ± 1.51	87.71 ± 1.64

续表

MPWTSVM	83.51 ± 1.58	86.04 ± 1.71	85.18 ± 1.52	87.95 ± 1.65	85.48 ± 1.44	88.93 ± 1.57
GMIGSVMs	83.26 ± 1.61	85.92 ± 1.74	84.92 ± 1.55	87.52 ± 1.68	85.12 ± 1.47	88.57 ± 1.60
MVCNN	79.26 ± 2.53	81.73 ± 2.71	80.58 ± 2.46	83.25 ± 2.63	82.35 ± 2.34	85.68 ± 2.51
MvDNN	77.89 ± 2.67	80.25 ± 2.85	79.14 ± 2.60	81.86 ± 2.77	80.78 ± 2.48	84.12 ± 2.65
SCA-PVNet	81.37 ± 2.38	83.86 ± 2.56	82.65 ± 2.31	85.19 ± 2.48	84.12 ± 2.20	87.35 ± 2.37
MVDRP	**86.78 ± 1.32**	**89.18 ± 1.45**	**88.47 ± 1.26**	**91.24 ± 1.39**	**88.75 ± 1.18**	**92.39 ± 1.31**

4.4.3. Movies 数据集(非平衡多媒体数据)

表 4 显示，MVDRP 在混合视图上的 Acc (74.36%)与 F1 (71.49%)最优。MVDRP 的 Acc 比特权信息学习方法 SVM+ (65.93%)提升 8.43%，比传统多视图 SVM 中最优的 MPWTSVM (71.02%)提升 3.34%；Movies 数据集类别不平衡，传统方法易偏向多数类；MVDRP-SVM 通过联合优化中的一致性约束平衡不同视图的分类偏差，同时特权信息学习修正少数类的分类误差，提升鲁棒性。与深度学习方法相比，MVDRP 在 Movies 数据集上的 Acc 比 MVCNN (70.25%)高 4.11%，比 MvDNN (68.79%)高 5.57%，比 SCA-PVNet (72.14%)高 2.22%。Movies 数据集类别不平衡且样本量有限，深度学习方法难以学习到少数类的特征，而 SVM 方法通过最大间隔机制对不平衡数据具有更好的鲁棒性。

Table 4. Classification accuracy and F1-score on Movies dataset
表 4. Movies 数据集的分类准确率和 F1 分数

模型	View A		View B		Hybrid View	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
PCA	57.52 ± 2.34	53.94 ± 2.51	58.47 ± 2.27	55.16 ± 2.44	59.28 ± 2.16	54.92 ± 2.33
LDA	59.85 ± 2.26	56.22 ± 2.43	61.13 ± 2.19	57.83 ± 2.36	61.85 ± 2.08	58.35 ± 2.25
PSVM-2V	68.93 ± 1.87	65.37 ± 2.04	70.61 ± 1.80	67.82 ± 1.97	70.52 ± 1.69	67.64 ± 1.86
SVM-2K	65.37 ± 2.01	61.63 ± 2.18	66.89 ± 1.94	63.59 ± 2.11	67.27 ± 1.83	64.28 ± 2.00
SVM+	63.74 ± 2.12	60.18 ± 2.29	64.36 ± 2.05	62.41 ± 2.22	65.93 ± 1.94	63.11 ± 2.11
MvSVM-2C	68.26 ± 1.92	64.82 ± 2.09	68.94 ± 1.85	65.95 ± 2.02	69.41 ± 1.74	66.73 ± 1.91
MPWTSVM	69.83 ± 1.78	68.05 ± 1.95	71.62 ± 1.71	68.79 ± 1.88	71.02 ± 1.60	68.16 ± 1.77
GMIGSVMs	69.52 ± 1.81	67.19 ± 1.98	71.86 ± 1.74	68.94 ± 1.91	70.85 ± 1.63	67.92 ± 1.80
MVCNN	66.28 ± 2.45	63.12 ± 2.62	67.56 ± 2.38	65.37 ± 2.55	70.25 ± 2.27	66.89 ± 2.44
MvDNN	64.75 ± 2.58	61.59 ± 2.75	66.03 ± 2.51	63.86 ± 2.68	68.79 ± 2.40	65.42 ± 2.57
SCA-PVNet	68.14 ± 2.29	65.76 ± 2.46	69.87 ± 2.22	67.53 ± 2.39	72.14 ± 2.11	69.27 ± 2.28
MVDRP	**70.69 ± 1.65**	**68.36 ± 1.82**	**72.14 ± 1.58**	**69.11 ± 1.75**	**74.36 ± 1.47**	**71.49 ± 1.64**

4.4.4. 消融实验

为验证每个核心模块的贡献，设计三个变体模型：(1) MVDRP-NoDR：去掉联合降维模块，直接在原始高维数据上训练。(2) MVDRP-NoCons：去掉视图一致性约束模块。(3) MVDRP-NoPriv：去掉特权信息互补性模块。

Table 5. Ablation study results (Accuracy, %)
表 5. 消融实验结果(准确率%)

模型	Caltech	WebKB	Movies
MVDRP-NoDR	87.23	82.56	68.79
MVDRP-NoCons	89.57	85.34	70.25
MVDRP-NoPriv	88.91	84.67	69.83
MVDRP	**92.85**	**88.75**	**74.36**

从表 5 可以看出，去掉联合降维模块后，模型性能下降最明显，在三个数据集上分别下降 5.62%、6.19%、5.57%，这表明联合降维是本文方法最重要的创新点，有效解决高维数据的冗余与噪声问题；去掉视图一致性约束模块后，性能分别下降 3.28%、3.41%、4.11%，这表明一致性约束可以有效平衡不同视图的分类偏差；去掉特权信息互补性模块后，性能分别下降 3.94%、4.08%、4.53%，这表明特权信息学习可以有效利用视图间的互补信息，提升分类精度。三个模块的协同作用使得完整模型取得最优性能。

4.4.5. 不同维度数下的性能变化

图 2 显示在 Caltech、WebKB 和 Movies 数据集上本文所提方法在不同维数的错误率。从图 2 中可以看出，维度并非越高越好，随着维数 d 从 1 增加到某个转折点，错误率明显下降，之后错误率曲线趋于平稳，表明原始特征空间中的主要信息可以用最大的特征向量来表示。从公式(7)和(8)中投影矩阵 Q_A 和 Q_B 的计算可以看出，Q_A 和 Q_B 包含了 W_A 和 W_B 对应的特征向量，因此最大的特征向量具有最多的判别信息。如果将维数降低到一个较小的数，例如 $d=1$，可能会遗漏一些有价值的信息，导致错误率相对较大。此外最小误差出现在维度数相对较小而非较大的情况下，在 WebKB 和 Movies 数据集上，当维度数大于 200 时，错误率会略有增加，这种现象的一个可能解释是，额外的维度可能不会为分类模型提供有价值的信息，相反，它们可能会引入冗余甚至噪声，从而降低整体的分类性能。

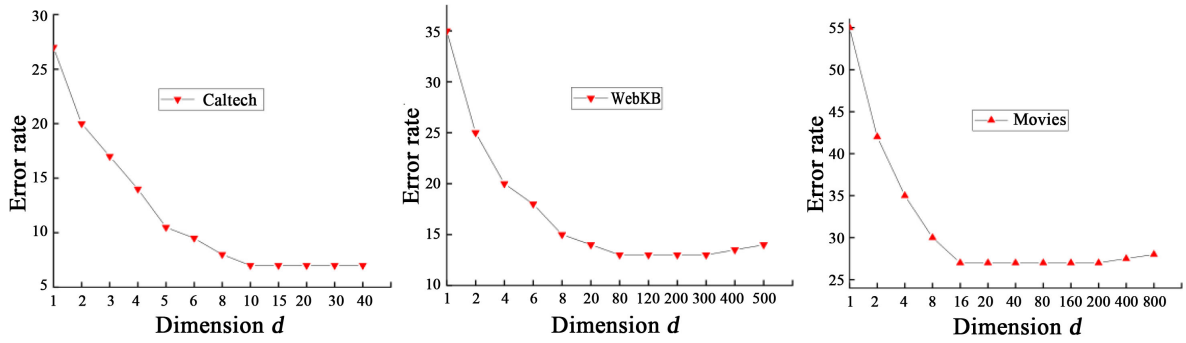

Figure 2. Performance variation of MVDRP with different dimension numbers
图 2. 不同维数的 MVDRP 性能变化

4.4.6. 参数敏感性分析

在本小节中，探讨 MVDRP 对不同参数的敏感性，通过设置不同的参数 γ 值，调整对不同视角的偏好。此外，参数 C^A、C^B 和 C 是正则化参数，与学习误差相关。设置 C，让参数 γ 和 γ 的值从 0.01 到 1000 变化，得到在三个数据集上的准确率，通过观察图 3 中的结果，得到以下结论，参数 γ 与项 $\left\|\hat{W}_B\right\|^2$ 相关，当 $\gamma>1$ 时，视图 B 对整体学习模型的影响比视图 A 更大；否则，视图 A 更具决定性，例如，在 Caltech 数据集上，当 $\gamma=10$ 时，分类结果最好，说明在此数据集上，视图 B 比视图 A 更具决定性。此外，在

WebKB 和 Movies 数据集上，视图 A 对整体学习模型的影响比视图 B 更大，因为当 $\gamma=0.1$ 时，分类准确率最高。此外，当参数 C 取值达到 10 或 100 时，分类性能最佳。参数 C 与学习误差相关，当参数 C 过大时，模型倾向于最小化学习误差，而不是最大化分类器边距，这可能导致过拟合。当 C 过小时，学习模型倾向于最大化分类器边距，而不是最小化学习误差，这可能导致相对较高的分类误差。

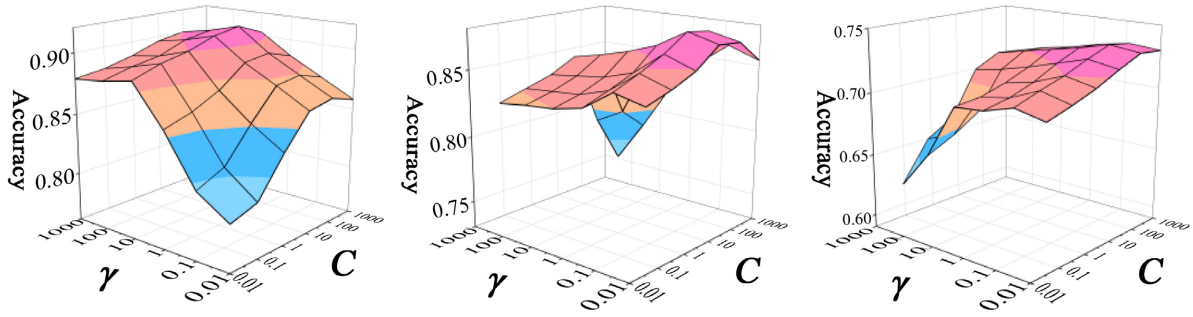

Figure 3. Performance variation with different values of parameters γ and C
图 3. 参数 γ 和 C 取不同值时的性能变化

4.4.7. 噪声鲁棒性实验

为了验证模型对特权信息噪声的鲁棒性，在三个数据集的视图 B 中添加不同比例的高斯噪声 ($\sigma=0.1,0.2,0.3,0.4,0.5$)，测试 MVDRP 与 PSVM-2V 的分类准确率变化，从图 4 可以看出：随着噪声强度的增加，所有方法的准确率均有所下降，但 MVDRP 的下降速度明显慢于 PSVM-2V；当高斯噪声 ($\sigma=0.5$)时，MVDRP 在 Caltech、WebKB、Movies 数据集上的准确率分别下降 4.23%、5.17%、6.35%，而 PSVM-2V 分别下降了 8.76%、10.23%、12.45%；这表明 MVDRP 通过联合降维与特权信息学习，对噪声具有更强的鲁棒性。降维过程可以过滤掉大部分噪声，而特权信息约束通过松弛变量进一步吸收剩余噪声的影响。

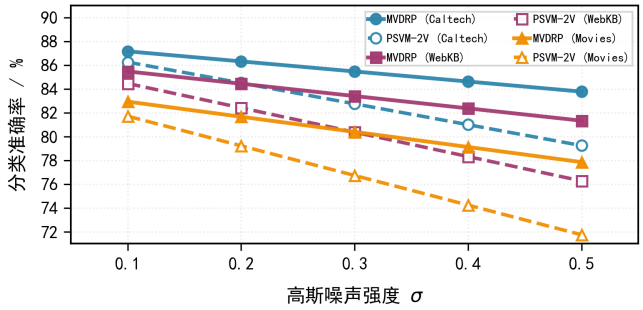

Figure 4. Experiment on robustness against gaussian noise
图 4. 高斯噪声鲁棒性实验

4.4.8. 计算效率对比实验

对比本文方法与基线方法在三个数据集上的平均训练时间和内存占用。从表 6 可以看出，本文方法的训练时间略长于传统 SVM 方法，但远低于深度学习方法，仅为 MVCNN 的 19.0%，SCA-PVNet 的 14.3%；本文方法的内存占用与传统 SVM 方法相当，远低于深度学习方法，仅为 MVCNN 的 13.5%，SCA-PVNet 的 9.0%；随着数据集规模的增加，本文方法的计算效率优势更加明显。这表明本文方法在保证分类精度的同时，具有较低的计算资源需求，适合部署在边缘设备和资源受限的场景中。

Table 6. Comparison of computational efficiency
表 6. 计算效率对比

模型	Caltech		WebKB		Movies	
	训练时间 (s)	内存占用 (MB)	训练时间 (s)	内存占用 (MB)	训练时间 (s)	内存占用 (MB)
PCA + PSVM-2V	12.35	85.62	2.17	32.45	5.89	56.78
LDA + PSVM-2V	15.72	92.37	2.89	38.62	7.23	63.54
PSVM-2V	28.64	125.89	4.56	56.37	12.78	89.25
MVCNN	185.32	1024.56	32.78	512.34	98.65	768.92
MvDNN	126.78	768.23	25.43	384.67	72.34	576.45
SCA-PVNet	245.67	1536.89	45.21	768.45	135.79	1024.67
MVDRP	35.26	138.54	5.72	62.89	15.34	98.76

5. 结论

针对多视图支持向量机的高维数据学习问题，本文提出基于特权信息学习的多视图支持向量机，将降维与多视图 SVM 分类融合为统一凸优化模型，同步优化投影矩阵与分类参数，避免传统串联方案的信息丢失；通过同一样本视图分类的一致性约束与不同视图互为特权信息的互补性约束，同时满足多视图学习的两大核心原则。本文严格证明目标函数的凸性与算法的收敛性，为模型的理论可靠性提供保障。在三类异构数据集上的实验表明，MVDRP 的分类精度与 F1 分数均优于传统“降维 + 分类”串联方案及现有多视图 SVM 方法，在小样本、低计算资源场景下相比主流深度学习多视图方法也具有显著优势。

本文模型仍存在以下潜在局限：其一，降维模块采用线性投影，对特征空间存在复杂非线性结构的数据集适配性不足；其二，交替优化的求解方式虽保证收敛，但每轮迭代需求解二次规划问题，在大规模数据集上的训练效率仍有提升空间。针对上述局限，后续研究将从两个方向展开：第一，引入核化投影与流形嵌入，将线性降维拓展至非线性降维框架，提升模型对复杂特征的适配能力；第二，结合傅里叶特征与随机梯度下降策略，替代传统二次规划求解，进一步提升模型在大规模数据场景下的可扩展性。

基金项目

广东省重点领域研发计划项目(2023B1111050010)。

参考文献

[1] Sain, S.R. (1996) The Nature of Statistical Learning Theory. *Technometrics*, **38**, 409-409. https://doi.org/10.1080/00401706.1996.10484565

[2] Gong, F.M. and Li, H.J. (2011) Traffic Sign Detection and Pattern Recognition Based on Binary Tree Support Vector Machines. *Advanced Materials Research*, **204**, 1394-1398. https://doi.org/10.4028/www.scientific.net/amr.204-210.1394

[3] Jaramillo, J., Velasquez, J.D. and Franco, C.J. (2017) Research in Financial Time Series Forecasting with SVM: Contributions from Literature. *IEEE Latin America Transactions*, **15**, 145-153. https://doi.org/10.1109/tla.2017.7827918

[4] Wang, T., Qi, J., Xu, H., Wang, Y., Liu, L. and Gao, D. (2016) Fault Diagnosis Method Based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter. *ISA Transactions*, **60**, 156-163. https://doi.org/10.1016/j.isatra.2015.11.018

[5] Xie, X. and Xiong, Y. (2022) Generalized Multi-View Learning Based on Generalized Eigenvalues Proximal Support Vector Machines. *Expert Systems with Applications*, **194**, Article ID: 116491. https://doi.org/10.1016/j.eswa.2021.116491

[6] He, Y., Tian, Y. and Liu, D. (2019) Multi-View Transfer Learning with Privileged Learning Framework. *Neurocomputing*,

335, 131-142. https://doi.org/10.1016/j.neucom.2019.01.019

[7] Farquhar, J., Hardoon, D., Meng, H.Y., Shawe-Taylor, J. and Szedmak, S. (2005) Two View Learning: SVM-2K, Theory and Practice. *Advances in Neural Information Processing Systems*, **18**.

[8] Sun, S. and Jin, F. (2011) Robust Co-Training. *International Journal of Pattern Recognition and Artificial Intelligence*, **25**, 1113-1126. https://doi.org/10.1142/s0218001411008981

[9] Houthuys, L., Langone, R. and Suykens, J.A.K. (2018) Multi-View Least Squares Support Vector Machines Classification. *Neurocomputing*, **282**, 78-88. https://doi.org/10.1016/j.neucom.2017.12.029

[10] Yu, S., Li, X., Sun, S., Wang, H., Zhang, X. and Chen, S. (2022) IBMvSVM: An Instance Based Multi-View SVM Algorithm for Classification. *Applied Intelligence*, **52**, 14739-14755.

[11] Tang, J., Li, D., Tian, Y. and Liu, D. (2018) Multi-View Learning Based on Nonparallel Support Vector Machine. *Knowledge-Based Systems*, **158**, 94-108. https://doi.org/10.1016/j.knosys.2018.05.036

[12] Tang, J., Tian, Y., Liu, D. and Kou, G. (2019) Coupling Privileged Kernel Method for Multi-View Learning. *Information Sciences*, **481**, 110-127. https://doi.org/10.1016/j.ins.2018.12.058

[13] Tang, J., Tian, Y., Zhang, P. and Liu, X. (2018) Multiview Privileged Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, **29**, 3463-3477. https://doi.org/10.1109/tnnls.2017.2728139

[14] He, X., Bai, S., Chu, J. and Bai, X. (2020) An Improved Multi-View Convolutional Neural Network for 3D Object Retrieval. *IEEE Transactions on Image Processing*, **29**, 7917-7930. https://doi.org/10.1109/tip.2020.3008970

[15] Huang, Z., Zhou, J.T., Zhu, H., Zhang, C., Lv, J. and Peng, X. (2021) Deep Spectral Representation Learning from Multi-View Data. *IEEE Transactions on Image Processing*, **30**, 5352-5362. https://doi.org/10.1109/tip.2021.3083072

[16] Lin, D., Cheng, Y., Guo, A., Mao, S. and Li, Y. (2024) SCA-PVNet: Self-and-Cross Attention Based Aggregation of Point Cloud and Multi-View for 3D Object Retrieval. *Knowledge-Based Systems*, **296**, Article ID: 111920. https://doi.org/10.1016/j.knosys.2024.111920

[17] Guo, W., Wang, Z., Yang, H. and Du, W. (2023) Multi-View Dimensionality Reduction Learning with Hierarchical Sparse Feature Selection. *Applied Intelligence*, **53**, 12774-12791. https://doi.org/10.1007/s10489-022-04161-4

[18] Feng, L., Meng, X. and Wang, H. (2020) Multi-View Locality Low-Rank Embedding for Dimension Reduction. *Knowledge-Based Systems*, **191**, Article ID: 105172. https://doi.org/10.1016/j.knosys.2019.105172

[19] Li, H., Gao, M., Wang, H. and Jeon, G. (2023) Multi-View Projection Learning via Adaptive Graph Embedding for Dimensionality Reduction. *Electronics*, **12**, Article 2934. https://doi.org/10.3390/electronics12132934

[20] Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. *WIREs Computational Statistics*, **2**, 433-459. https://doi.org/10.1002/wics.101

[21] Kalia, V., Walker, D.I., Krasnodemski, K.M., Jones, D.P., Miller, G.W. and Kioumourtzoglou, M. (2020) Unsupervised Dimensionality Reduction for Exposome Research. *Current Opinion in Environmental Science & Health*, **15**, 32-38. https://doi.org/10.1016/j.coesh.2020.05.001

[22] Hyvärinen, A. and Oja, E. (2000) Independent Component Analysis: Algorithms and Applications. *Neural Networks*, **13**, 411-430. https://doi.org/10.1016/s0893-6080(00)00026-5

[23] Izenman, A.J. (2013) Linear Discriminant Analysis. In: Izenman, A.J., Ed., *Modern Multivariate Statistical Techniques*, Springer, 237-280. https://doi.org/10.1007/978-0-387-78189-1_8

[24] Xu, C., Tao, D. and Xu, C. (2014) Large-Margin Multi-Viewinformation Bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 1559-1572. https://doi.org/10.1109/tpami.2013.2296528

[25] Li, J., Allinson, N., Tao, D. and Li, X. (2006) Multitraining Support Vector Machine for Image Retrieval. *IEEE Transactions on Image Processing*, **15**, 3597-3601. https://doi.org/10.1109/tip.2006.881938

[26] Luo, Y., Tao, D., Xu, C., Xu, C., Liu, H. and Wen, Y.G. (2013) Multiview Vector-Valued Manifold Regularization for Multilabel Image Classification. *IEEE Transactions on Neural Networks and Learning Systems*, **24**, 709-722. https://doi.org/10.1109/tnnls.2013.2238682

[27] Sun, S. (2013) Multi-View Laplacian Support Vector Machines. arXiv: 1307.7024.

[28] Wang, H., Peng, J. and Fu, X. (2019) Co-Regularized Multi-View Sparse Reconstruction Embedding for Dimension Reduction. *Neurocomputing*, **347**, 191-199. https://doi.org/10.1016/j.neucom.2019.03.080

[29] Buathong, W. (2016) TLiSVM (Triple Linear SVM Weight) for Dimensionality Reduction. 2016 11*th International Conference on Computer Science & Education* (*ICCSE*), Nagoya, 23-25 August 2016, 273-278. https://doi.org/10.1109/iccse.2016.7581593

[30] Xu, X., Yang, Y., Deng, C. and Nie, F. (2019) Adaptive Graph Weighting for Multi-View Dimensionality Reduction. *Signal Processing*, **165**, 186-196. https://doi.org/10.1016/j.sigpro.2019.06.026

[31] Apasiba Abeo, T., Shen, X., Gou, J., Mao, Q., Bao, B. and Li, S. (2019) Dictionary-Induced Least Squares Framework

for Multi-View Dimensionality Reduction with Multi-Manifold Embeddings. *IET Computer Vision*, **13**, 97-108. https://doi.org/10.1049/iet-cvi.2018.5135

[32] Zhang, X., Yang, C. and Jiao, L. (2010) A Semi-Supervised SAR Target Recognition Based on Laplacian Regularized Least Squares Classification. *Journal of Software*, **21**, 586-596. https://doi.org/10.3724/sp.j.1001.2010.03538

[33] Cai, D., He, X. and Han, J. (2007) Semi-Supervised Discriminant Analysis. 2007 *IEEE* 11*th International Conference on Computer Vision*, Rio de Janeiro, 14-21 October 2007, 1-7. https://doi.org/10.1109/iccv.2007.4408856

[34] Bousahba, N., Adjoudj, R., Belhia, S. and Chachou, L. (2022) Dimensionality Reduction Method Apply for Multi-View Multimodal Person Identification. *International Journal of Computing and Digital Systems*, **12**, 675-685. https://doi.org/10.12785/ijcds/120155