

# 面向无监督域适应的稀疏正则最优传输模型

李怡萱, 马丽涛\*

河北工程大学数理科学与工程学院, 河北 邯郸

收稿日期: 2026年2月11日; 录用日期: 2026年3月9日; 发布日期: 2026年3月18日

## 摘要

无监督域适应旨在利用有标签的源域知识提升模型在无标签目标域上的性能。然而, 现有方法大多依赖全局分布对齐来学习域不变特征, 普遍面临以下挑战: 相似性度量未能充分考虑数据内在的流形结构、域分布不平衡易导致对齐偏差, 以及结果可解释性不强等。针对上述问题, 本文提出一种基于扩散成本的稀疏正则最优传输模型。首先, 基于扩散过程度量跨域样本在流形结构上的相似性, 并结合源域标签信息构建成本矩阵, 从而实现非欧空间中跨分布样本相似性的准确刻画。其次, 在目标函数中引入边际约束惩罚项, 以增强模型对数据不平衡场景的适用性。此外, 引入稀疏正则项来增强对齐样本间的可解释性, 缓解由稠密传输计划引起的噪声误匹配问题。针对模型的非连续性特点, 本文采用简化的SPG算法进行高效求解。最后, 利用模型求解结果训练分类器, 并在三组公开域适应数据集上进行实验评估, 从准确率、精确率、召回率和F1分数四个指标对分类结果进行量化比较, 实验结果表明, 本文所提模型能够有效提升跨域分类精度, 并缓解因域偏移与类别不平衡导致的分类偏差。

## 关键词

无监督域适应, 最优传输, 扩散成本, 边际惩罚, 稀疏正则

# A Sparsity-Regularized Optimal Transport Model for Unsupervised Domain Adaptation

Yixuan Li, Litao Ma\*

School of Mathematics and Physics, Hebei University of Engineering, Handan Hebei

Received: February 11, 2026; accepted: March 9, 2026; published: March 18, 2026

## Abstract

Unsupervised domain adaptation aims to leverage labeled knowledge from a source domain to

\*通讯作者。

improve model performance on an unlabeled target domain. However, most existing methods rely on global distribution alignment to learn domain-invariant features, commonly facing the following challenges: similarity measures fail to adequately account for the intrinsic manifold structure of data; they are sensitive to domain distribution imbalance, which leads to biased alignment; and the interpretability of the alignment results remains limited. To address these issues, this paper proposes a diffusion sparse-regularization optimal transport. Specifically, a diffusion process is employed to measure cross-domain sample similarity on the underlying data manifold firstly. And then source domain label information is incorporated to construct the cost matrix, enabling accurate characterization of cross-distribution similarities in non-Euclidean spaces. Furthermore, a marginal constraint penalty term is introduced into the objective function to enhance robustness under imbalanced data scenarios. In addition, a sparsity regularization term is imposed to improve the interpretability of sample alignment and to mitigate noisy mismatches caused by dense transport plans. To efficiently solve the proposed non-smooth optimization problem, a simplified spectral projected gradient (SPG) algorithm is adopted. Finally, a classifier is trained based on the obtained transport plan, and extensive experiments are conducted on three public domain adaptation datasets. Quantitative evaluations in terms of accuracy, precision, recall, and F1-score demonstrate that the proposed method consistently improves cross-domain classification performance and effectively alleviates classification bias induced by domain shift and class imbalance.

## Keywords

Unsupervised Domain Adaptation, Optimal Transport, Diffusion Cost, Marginal Penalty, Sparsity Regularization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

目前, 机器学习模型在图像分类、目标检测等计算机视觉任务中已取得显著成效[1]-[3]。然而, 尽管这些模型在训练集上表现优异, 在实际应用中仍常面临域间分布差异的挑战。具体而言, 训练阶段所使用数据与目标域真实数据之间常存在明显的特征分布偏移, 导致模型在实际场景中的性能大幅下降。为缓解这一问题, 无监督域适应方法近年来受到了广泛关注[4]。其核心思想是利用带有标注的源域数据, 提升在无标注目标域上的分类器性能, 实现知识的跨域迁移。传统方法[5]-[8]通常将域适应问题转化为缩小域间统计距离的优化问题, 这类方法虽直观有效, 但仅依赖衡量整体分布差异的标量或损失函数来约束全局对齐, 往往难以实现样本或类别层面的细粒度对齐, 从而限制了模型在目标域上的分类精度。

在此背景下, 最优传输理论(Optimal Transport, OT) [9]为跨域分布对齐提供了新的理论框架。该理论利用样本间的传输代价, 建模源域与目标域之间的最优映射关系, 在保持类别结构的同时实现对分布偏移的有效对齐, 并且提供了可解释的传输方案。然而, 传统最优传输模型通常基于样本间成对欧氏距离来定义传输代价, 在高维特征空间中难以准确度量跨域样本间的语义相似性。同时, 传统基于距离的方法往往忽略域内的局部几何结构, 且未能有效利用已有的标签信息。基于此, 文献[10]提出了一种基于扩散几何的传输成本构建方法, 通过“域内-跨域-域内”三阶段扩散过程建模样本关系, 并利用源域标签将语义信息融合到成本中。除此之外, 传统最优传输模型通常要求严格的边际约束, 这在数据质量不平衡或含有噪声样本的实际场景中极大地限制了其应用灵活性。部分研究通过松弛边际约束来提升模型

的适用性, 但为提高求解效率往往会引入熵正则项, 导致生成的传输计划趋于稠密, 不仅削弱了结果的可解释性, 还可能因过度拟合噪声而影响模型的泛化性能。

基于上述分析, 本文拟构建一个面向不平衡域适应的稀疏正则最优传输模型, 采用基于扩散过程的成本函数, 通过图扩散算子显式融合源域与目标域的域内局部几何及跨域关联信息, 并利用源域标签引导扩散方向, 以更全面度量跨域样本间的关系。同时松弛边际约束以提升模型在不平衡数据分布下的性能, 并引入稀疏正则项以增强样本间关联结构的可解释性, 从而进一步强化模型的跨域适应能力。总体而言, 本文的主要工作包括:

1) 通过构建扩散算子同时建模域内几何结构和跨域关系, 并利用标签信息引导扩散方向, 从而更准确量化源域与目标域样本之间的差异, 提高跨域分布对齐的准确性。

2) 引入  $l_2$  惩罚项松弛严格的边际分布约束以实现不平衡传输, 提升模型在类别分布偏移的域适应任务中的性能, 增强其在真实复杂场景下的适用性。

3) 通过在模型中引入  $l_0$  稀疏正则项, 增强传输计划的稀疏性, 从而获得更具可解释性的样本间对应关系, 同时稀疏结构可以在一定程度上抑制噪声样本的干扰, 提升模型的鲁棒性。

4) 针对  $l_0-l_2$  问题的非凸、非光滑及组合优化特性, 设计使用简化的 SPG 算法求解模型, 并利用求解结果训练分类器, 在三个公开数据集上进行实验验证。

## 2. 相关工作

### 2.1. 无监督域适应

无监督域适应(Unsupervised Domain Adaptation, UDA)旨在解决模型从标注充分的源域向未标注且分布不同的目标域迁移时的泛化问题, 其核心在于实现有效的域间分布对齐。早期研究集中于基于统计度量的方法, 其核心是显式定义并最小化域间差异度量来实现分布对齐。Das 等人[11]利用正则化超图匹配建立高阶样本对应关系; Kang 等人[12]提出对比适应网络(CAN)显式建模类内与类间域差异以提升对齐的判别性; Li 等人[13]提出最大密度散度(MDD)以同时最小化域间散度并最大化类内密度。然而, 这类方法仅依赖预定义的统计度量, 难以处理复杂的高维分布及细粒度的结构关系。随着深度学习的发展, 学者们将对抗学习引入域适应问题中, 通过域判别器与特征提取器之间的对抗博弈, 动态学习域不变特征。He 等人[14]提出的贝叶斯双通路网络(BDNet), 通过分离域共享与域特定通路来缓解对抗冲突; 赵恩庭等人[15]引入类别信息作为条件输入进行对抗训练以提升性能; Wang 等人[16]设计动态协作对抗域适应网络(DCADAN)以自适应调整决策边界。尽管对抗方法在实践中表现出较强的适应能力, 但其训练过程往往不稳定, 且难以显式建立跨域样本之间的对应关系。因此, 上述方法在复杂分布的高精度度量与跨域对应关系的显式建立等方面仍面临诸多挑战。

在此背景下, Courty 等人[9]首次将最优传输理论引入域适应任务中, 通过寻找将源域分布变换为目标域分布的最小代价传输计划, 从概率度量角度实现对域间差异的全局刻画, 进一步构建了一个基于正则化最优传输的域适应框架。此外, 通过在模型中引入类别相关正则化以及拉普拉斯正则化, 实现了源域与目标域间的结构化分布对齐, 所得传输计划不仅能建立显式的跨域对应关系, 还可实现结构化的细粒度对齐。随后, 多种基于最优传输的域适应方法[17]-[20]在此框架下引入额外的类别约束或类别正则化项, 以增强跨域对齐过程中的语义一致性, 实现跨域间的全局分布对齐与结构化样本匹配。然而, 这类方法在一定程度上增加了模型优化的复杂性。为了提升求解效率, 部分方法进一步引入熵正则项[21]-[23], 但往往导致所得传输计划趋于稠密, 降低了结果的可解释性。此外, 传统方法通常遵循严格的保质量约束, 在类别不平衡场景下容易因分布偏移而产生匹配偏差。因此, 近年来研究趋势逐渐转向不平衡最优

传输与稀疏最优传输, 以突破上述方法局限。

## 2.2. 不平衡最优传输

最优传输(Optimal Transport, OT)最早由 Monge [24]提出, 其目标是在给定两个概率分布的条件下, 寻找一种传输方案以最小化整体传输成本。随后, Kantorovich [25]通过引入联合概率测度对该问题进行松弛, 将其转化为凸优化形式, 从而能够处理离散分布和复杂的跨空间匹配问题。然而, 无论是 Monge 问题还是 Kantorovich 的松弛形式, 其核心均建立在严格的质量守恒这一前提之上, 即要求源分布与目标分布的总质量完全相等。这一假设在许多现实问题中并不成立, 例如在存在类别不平衡、数据噪声以及部分匹配等场景下, 强制质量守恒会导致无意义的传输或对噪声的过度拟合。为克服其局限性, 不平衡最优传输通过放松严格的边际约束, 从而将最优传输的适用性拓展至质量不相等、需进行模糊匹配等真实场景。为实现不平衡最优传输, 一种主流策略是虚拟点方法。该方法[26]-[28]通过在源域或目标域中引入一个或多个虚拟点, 用于吸收源域的多余质量或为目标域提供不足的质量。这一过程体现为构造一个扩展的代价矩阵, 从而将原始的不平衡传输问题, 重新表述为一个可以求解的、质量守恒的标准平衡最优传输问题。其二为惩罚松弛法。该方法直接修改了 Kantorovich 优化问题, 不再严格要求传输计划的边际分布必须与给定的源、目标分布相等, 而是通过引入一个凸惩罚函数来松弛这一约束。该惩罚项用于度量并惩罚传输计划边际分布与给定分布之间的偏离, 常见的函数包括 KL 散度[29]、平方  $l_2$  距离[30]或总变差[31]等等。此方法的核心优势在于, 它能将不平衡最优传输问题保持为一个结构良好的凸优化问题, 从而允许设计高效、稳定的求解算法。

为高效求解经典最优传输模型, Cuturi 引入熵正则项[29], 将原问题转化为严格凸优化形式, 并利用 Sinkhorn 算法实现快速求解, 极大推动了最优传输在大规模机器学习中的应用。然而, 熵正则项会导致所得传输计划全局稠密, 即几乎所有位置均具有非零传输质量, 这不仅降低了结果的可解释性, 也使得少数噪声样本易与正常样本之间形成虚假连接, 从而扭曲真实匹配关系, 影响模型的鲁棒性与准确性。

## 2.3. 稀疏正则最优传输

在熵正则化最优传输的广泛应用中, 一个公认的局限是其必然产生的全支撑、完全稠密的传输计划。这在高维、大规模或需要明确对应关系(如点云配准、特征匹配)的应用中缺乏可解释性, 且计算和存储成本高昂。为应对此问题, 学者们开始系统性地研究稀疏正则化最优传输模型。其核心思想是通过在标准最优传输问题中引入能诱导稀疏性的结构性约束(如基数约束)或惩罚项(如  $l_1$  范数、Group Lasso 等), 直接优化以得到非零元素显著更少、结构更清晰的传输计划。例如, Tang 等人[32]结合熵正则化与小批量采样及动态自标注机制, 实现隐式稀疏对齐; Courty 等人[9]提出类引导的群稀疏正则( $l_{1,2}$  范数), 约束目标样本仅接收同类源样本的质量, 实现语义结构保持的稀疏匹配; Haasler 等人[33]则引入  $l_{\infty,1}$  群稀疏正则项, 显式促进多变量分布在不同维度上具有一致的稀疏模式。这些方法在保持计算可行性的同时, 显著增强了最优传输在跨域对齐与结构化建模中的表现。为更精确地控制稀疏性, 直接引入  $l_0$  约束在理论层面具有明显优势, 但  $l_0$  范数因其离散性与非凸性, 其优化问题属于 NP 难问题, 现有研究多通过凸松弛(如  $l_1$  范数)或非凸连续惩罚(如 SCAD [34]、MCP [35])进行近似。然而, 凸松弛可能引入估计偏差, 而非凸惩罚的优化算法往往仅能保证子序列收敛, 且依赖较强的几何条件。为在稀疏精确性与算法收敛性之间寻求更优平衡, 可采用 capped- $l_1$  函数作为  $l_0$  惩罚的精确连续松弛。该松弛不仅能保持与原始  $l_0$  问题全局最优解的一致性, 其特有的下界性质还为设计具有全局收敛保证的高效算法奠定了基础, 从而在理论性质与计算效率之间实现了更好的权衡。

### 3. 主要工作

#### 3.1. 问题设置

无监督域适应问题中, 源域和目标域具有定义在相同空间  $X$  上的两个不同的分布  $p_s$  和  $p_t$ 。源域为带有标签信息的样本集  $\{x_i^s\}_{i=1}^{n_s} = X_S \subset (X, p_s)$ , 对应的标签集合为  $\{y_i^s\}_{i=1}^{n_s} = Y_S$ , 目标域为无标签的样本集  $\{x_j^t\}_{j=1}^{n_t} = X_T \subset (X, p_t)$ , 具有与源域相同的类别空间。本文的目标是利用  $X_S$  和  $X_T$  构建模型, 对目标域样本  $X_T$  进行分类。

#### 3.2. 稀疏正则不平衡最优传输模型

##### 3.2.1. 基于扩散过程的成本函数

在最优传输理论中, 成本矩阵通过量化跨域样本间的传输代价, 构成了传输过程的基础约束。这一约束直接决定了样本对之间的耦合强度, 从而影响分布对齐的效果。然而, 传统的最优传输方法往往基于欧式距离直接计算样本间的特征空间距离来构建成本矩阵, 其假设数据处于欧氏空间, 忽视了数据可能存在于低维弯曲流形上的几何特性。此外, 这类方法仅依赖跨域样本对之间的域间距离, 未能有效利用域内结构信息以及源域内可用的标签信息, 因而容易在后续优化过程中产生错误的样本匹配, 降低模型在目标域上的分类性能。为解决上述问题, 受文献[10]启发, 本文首先通过样本之间的亲和度构建源域(或者目标域)内扩散算子, 并通过限制邻域范围实现标签引导, 从而将类别信息融合进扩散过程。具体来说, 对于源样本  $x_i^s$  及其标签  $y_i^s$ , 定义其邻域指标集为  $N_i^s = \{g \mid y_g^s = y_i^s, g \in \{1, \dots, n_s\}\}$ , 将仅在同类源样本中进行扩散; 对于目标样本  $x_j^t$ , 定义其邻域指标集为  $N_j^t = \left\{g \mid g \in \left\| \text{rank} \left( \left\{ \|x_j^t - x_l^t\|^2, l \in \{1, \dots, n_t\} \right\} \right) \right\|_k \right\}$

其中  $\text{rank}(B)$  表示集合  $B$  中元素排序,  $\left\| \text{rank}(B) \right\|_k$  表示取集合  $B$  中前  $k$  个元素的下标, 将在该样本的  $k$  近邻邻域内进行扩散。进而定义源域(或者目标域)内扩散算子如下:

$$D_s(x_i^s, x_g^s) = \begin{cases} \frac{\exp(-\|x_i^s - x_g^s\|_2^2 / \tau)}{\sum_{g' \in N_i^s} \exp(-\|x_i^s - x_{g'}^s\|_2^2 / \tau)}, & \text{if } g \in N_i^s \\ 0, & \text{else} \end{cases} \quad (1)$$

$$D_t(x_j^t, x_g^t) = \begin{cases} \frac{\exp(-\|x_j^t - x_g^t\|_2^2 / \tau)}{\sum_{g' \in N_j^t} \exp(-\|x_j^t - x_{g'}^t\|_2^2 / \tau)}, & \text{if } g \in N_j^t \\ 0, & \text{else} \end{cases}$$

其中,  $\tau$  为高斯核的带宽参数, 较大的  $\tau$  值会使核函数衰减缓慢, 从而允许相对较远的样本点之间仍具有非零相似度; 而较小的  $\tau$  值则使核函数快速衰减, 仅赋予空间距离非常接近的样本点以显著相似度。然后, 构建跨域扩散算子  $Q$ , 衡量源域样本  $x_i^s$  和目标域样本  $x_j^t$  在几何结构上的关联程度。

$$Q(x_i^s, x_j^t) = \frac{\exp(-\|x_i^s - x_j^t\|_2^2 / \tau)}{\sum_{j=1}^{n_t} \exp(-\|x_i^s - x_j^t\|_2^2 / \tau)} \quad (2)$$

最后, 利用上述三个扩散算子构建以下成本矩阵, 充分融合数据的几何结构与标签信息:

$$C = -\log(D_s Q D_t) \quad (3)$$

其中,  $S = D_s Q D_t$  为复合算子, 表示从源点  $x_s^i$  出发, 在  $X_s$  上经过一步扩散到达其邻域, 继而以概率  $Q$  跃迁至  $X_T$  上的一个邻近点, 最后在  $X_T$  上经过一步扩散到达目标点  $x_t^j$  的总概率。因此,  $S$  本质上是一个跨域的、复合的热核, 衡量了通过局部路径连接两个流形上两点的概率流强度。为了得到适用于最优传输问题的成本函数, 对概率矩阵  $S$  应用负对数变换, 将高的扩散概率(表征强的几何连通性)映射为低的传输代价, 从而引导最优传输计划遵循数据内在的流形结构进行质量分配。

### 3.2.2. 模型构建

经典最优传输模型通常要求传输过程满足严格的边际约束, 即源域与目标域分布必须完全对齐, 然而这一假设在实际应用场景中往往存在局限性。一方面, 现实数据常受到噪声、异常值或采样偏差的影响, 过于严格的边际约束会使模型对数据质量过于敏感, 从而降低其稳健性。另一方面, 当源域与目标域仅存在部分对应关系时, 强制要求全局匹配会引入大量无关的传输关联, 进而削弱模型的可解释性与泛化能力。尽管松弛边际约束能够提升模型灵活性, 却容易导致传输矩阵过度稠密、结构模糊, 进而增加过拟合的风险。

为此, 本文提出一种带有稀疏正则项的不平衡最优传输模型, 通过将边际松弛与稀疏约束相结合, 以获取更具有可解释性的样本匹配关系, 同时有助于减少因噪声或异常样本引起的错误对齐, 从而提升模型在不平衡跨域适应任务中的鲁棒性。具体而言, 我们在目标函数中引入  $l_0$  正则项, 并将原边际约束转化为惩罚项加入目标函数。这一处理允许传输计划在源分布与目标分布上存在一定偏离以实现非保质量传输。结合基于扩散过程的成本函数, 得到如下带有稀疏正则项的松弛最优传输问题(Diffusion Sparse-regularization Optimal Transport, DSOT):

$$\min_{T_{ij} \geq 0} \langle C, T \rangle + \rho \left( \|T 1_{n_t} - p_s\|^2 + \|T^T 1_{n_s} - p_t\|^2 \right) + \lambda \|T\|_{0,1} \quad (4)$$

其中,  $\|T\|_{0,1} = \sum_{j=1}^{n_t} \|T_j\|_0$ ,  $\rho$  表示允许偏离惩罚力度的超参数, 当  $\rho \rightarrow \infty$  时, 任何偏离都将带来无穷大的惩罚, 此时模型退化为必须严格满足边际约束的平衡最优传输。 $\lambda > 0$  控制稀疏正则化强度, 通过对传输计划中的非零元素施加惩罚以得到稀疏传输计划。

### 3.2.3. 模型求解

式(4)为矩阵优化, 为方便求解, 将其重新表述为如下带盒约束的向量优化问题,

$$\min_{t \geq 0} c^T t + \frac{\rho}{2} \|At - d\|^2 + \lambda \|t\|_0 \quad (5)$$

其中,  $t = \text{vec}(T) \in \mathbb{R}^{n_s n_t}$ ,  $c = \text{vec}(C) \in \mathbb{R}^{n_s n_t}$ ,  $\text{vec}(\cdot)$  表示对矩阵进行列向量化, 变量  $T$  的边际约束转化为线性约束形式  $A = [I_{n_s} \otimes 1_{n_t}^T; 1_{n_s}^T \otimes I_{n_t}] \in \mathbb{R}^{(n_s+n_t) \times (n_s n_t)}$ ,  $d = [p_s, p_t] \in \mathbb{R}^{n_s+n_t}$ 。由于  $l_0$  正则项具有非凸、不连续等性质, 导致问题(5)优化困难。capped- $l_1$  函数  $\varphi(t) = \frac{|t|}{\nu} - \max\left\{0, \frac{t}{\nu} - 1, -\frac{t}{\nu} - 1\right\}$  通过鼓励小于  $\nu$  的系数被压缩到零, 而对大于  $\nu$  的系数施加恒定惩罚, 可有效逼近  $l_0$  函数, 因此本文将考虑利用 capped- $l_1$  函数, 构造问题(5)的松弛形式。令  $f(t) = c^T t + \frac{\rho}{2} \|At - d\|^2$ , 构造如下连续松弛问题:

$$\min_{t \geq 0} f(t) + \lambda \phi(t) \quad (6)$$

其中,  $\phi(t) = \sum_{i=1}^{n_s n_t} \varphi(t_i)$  具有特殊的非凸结构, 可以表示为一族凸函数  $\phi^d(t)$  的最小值通过迭代求解凸子问

题来逼近原非凸问题的解。具体来说, 对于任意向量  $d = (d_1, \dots, d_{n_s+n_t})^T \in \{1, 2, 3\}^{n_s+n_t}$ , 定义

$$\phi^d(t) = \sum_{i=1}^{n_s+n_t} \left[ \frac{|t_i|}{\nu} - \theta_{d_i}(t_i) \right] \quad (7)$$

其中  $\theta_1(t) = 0$ ,  $\theta_2(t) = \frac{t}{\nu} - 1$ ,  $\theta_3(t) = -\frac{t}{\nu} - 1$ 。文献[36]证明在参数  $\nu$  满足  $\nu < \lambda/L_f$  ( $L_f$  为  $f(t)$  的 Lipschitz 常数)的条件下, capped- $l_1$  连续松弛模型与原始问题具有相同的全局最优解集。由于  $f(t)$  连续可微, 本文可采用简化的 SPG 算法[36]来求解连续松弛问题(6), 得到最优解  $t^*$ , 进而将其矩阵化得到原始问题(4)的最优解  $T^*$ ,

$$T^* = \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} t_{(j-1)m+i}^* \cdot e_i^{(n_s)} \left( e_j^{(n_t)} \right)^T \quad (8)$$

其中,  $e_i^{(n_s)}$  是  $\mathbb{R}^{n_s}$  中的第  $i$  个标准基向量,  $e_j^{(n_t)}$  是  $\mathbb{R}^{n_t}$  中的第  $j$  个标准基向量。

### 3.3. 算法步骤

本节给出了模型 DSOT 的求解算法, 利用 capped- $l_1$  惩罚的凸分解结构和目标函数的连续可微性, SPG [36]算法可以简化为如下算法 1。

#### 算法 1. 简化的 SPG 算法

输入: 成本向量  $c$ , 约束矩阵  $A$ ,  $l_0$  稀疏正则项参数  $\lambda > 0$ ,  $l_2$  边际惩罚项系数  $\rho > 0$ , 常数  $\omega > 1$ , 最大迭代次数  $MaxIter$

输出: 最优传输计划  $T^*$

1、初始化参数:

初始迭代点  $t^{(0)}$ , 初始步长  $\gamma_0$ , 阈值  $\nu > 0$  (满足  $\nu < \lambda/L_f$ )

2、for  $k=0$  to  $MaxIter-1$ :

a、计算梯度  $\nabla f(t^k)$  和函数值  $f(t^k)$

$$\nabla f(t^k) = c + \rho A^T (At^k - d)$$

$$f(t^k) = c^T t^k + \frac{\rho}{2} \|At^k - d\|^2$$

b、确定当前点的分类向量  $d^k$

$$d_i^k = \begin{cases} 1 & t_i < \nu, \\ 2 & t_i \geq \nu, \\ 3 & t_i \leq -\nu \end{cases}$$

c、利用式(7)构建二次近似  $d^k$

$$Q_{d^k, \gamma_k}(t, t^k) = f(t^k) + \langle \nabla f(t^k), t - t^k \rangle + \frac{\gamma_k}{2} \|t - t^k\|^2 + \lambda \phi^{d^k}(t)$$

d、求解近端子问题

$$\hat{t}^{k+1} = \arg \min_{t \in [0, 1]} Q_{d^k, \gamma_k}(t, t^k)$$

e、检查下降条件

$$\text{如果满足 } f(\hat{t}^{k+1}) + \lambda \phi^{d^k}(\hat{t}^{k+1}) \leq Q_{d^k, \gamma_k}(\hat{t}^{k+1}, t^k)$$

$$t^{k+1} = \hat{t}^{k+1}$$

否则:

---

增大步长  $\gamma_k \leftarrow \omega \times \gamma_k$ , 返回(a)重新计算

3、结束循环

4、返回最优解  $T^* = \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} t_{(j-1)m+i}^{k+1} \cdot e_i^{(n_s)} (e_j^{(n_t)})^T$

---

本文采用重心映射方法来建立源域与目标域之间的对应关系。对于源域样本  $x_i^s$  在目标域的重心  $\tilde{x}_i^s$ , 表示为目标域中对应样本的加权组合:

$$\tilde{x}_i^s = \arg \min_x \sum_{j=1}^{n_t} T_{ij}^* \|x - x_j^t\|_2^2 \quad (9)$$

其中,  $T_{ij}^*$  表示从源域样本  $x_i^s$  到目标域样本  $x_j^t$  的传输质量。经该映射重构后的源域样本既服从目标域的空间分布, 又保留了原始标签信息。在此基础上, 可以利用这些与目标域分布一致的样本训练分类器, 进而实现对原始目标域样本的分类任务。

### 3.4. 复杂度分析

本节对简化后的 SPG 算法进行复杂度分析。设问题规模为边际约束个数  $(n_s + n_t)$  和决策变量个数  $(n_s n_t)$ , 算法的计算复杂度可分解如下: 单次线搜索尝试需要计算近端算子、目标函数值和线搜索条件检查, 主要计算量在于矩阵-向量乘法, 复杂度为  $O(n_s n_t (n_s + n_t))$ ; 设平均需要  $L$  次线搜索尝试可以找到满足条件的步长, 所以单轮迭代复杂度为  $O(L n_s n_t (n_s + n_t))$ ; 设收敛所需迭代次数为  $M$ , 则总时间复杂度为  $O(M L n_s n_t (n_s + n_t))$ 。算法的空间复杂度为  $O(n_s n_t)$ , 主要存储传输矩阵、梯度向量及稀疏的约束矩阵, 内存需求较低。

## 4. 实验与结果分析

### 4.1. 数据集介绍

为验证模型的有效性, 本文在三个无监督域适应(UDA)任务中常用的跨域视觉适应数据集上进行了实验验证。通过将数据集中来自不同分布的样本分别设定为源域与目标域, 来构建具有挑战性的跨域迁移任务, 三个数据集的具体介绍如下:

#### (1) COIL20 数据集

COIL-20 数据集包含 20 个物体, 每个物体以  $5^\circ$  间隔采集 72 张  $128 \times 128$  灰度图像, 共计 1440 张。实验中将其划分为两个视角互补的子集: COIL1 (简记为 C1 域) 包含第 1、3 象限角度 ( $[0^\circ, 85^\circ]$  与  $[180^\circ, 265^\circ]$ ) 的图像, COIL2 (简记为 C2 域) 包含第 2、4 象限角度 ( $[90^\circ, 175^\circ]$  与  $[270^\circ, 335^\circ]$ ) 的图像, 二者呈现出相对不同的分布, 可构建 2 个迁移任务: C1  $\rightarrow$  C2 和 C2  $\rightarrow$  C1。

#### (2) Office31 数据集

Office31 是视觉迁移学习领域的常用基准数据集, 包含 31 类办公物品, 共约 4652 张图像。这些图像来自三个不同视觉领域: Amazon (在线商城的商品图像, 简记为 A 域)、DSLR (高分辨率单反相机拍摄图像, 简记为 D 域) 和 Webcam (网络摄像头采集图像, 简记为 W 域), 各领域图像在背景、光照和分辨率等方面存在显著差异。由于经预处理后的 W 域和 D 域已经实现高度对齐, 因此利用其余领域分别作为源域和目标域, 可构建 4 个迁移任务: A  $\rightarrow$  W、A  $\rightarrow$  D、W  $\rightarrow$  A 以及 D  $\rightarrow$  A。

#### (3) RS 遥感数据集

RS 遥感数据集整合了三个采集条件、分辨率、地理分布等特性存在差异的遥感场景图像数据集: RSI-CB (35 类, 约 24,000 张, 分辨率:  $256 \times 256$ , 简记为 R 域)、AID (30 类, 10,000 张, 分辨率:  $600 \times 600$ ,

简记为 A 域)与 UCMerced (21 类, 2100 张, 分辨率:  $256 \times 256$ , 简记为 U 域)。选取不同域间的相同类别构建 6 个迁移任务:  $A \rightarrow R$ 、 $A \rightarrow U$ 、 $R \rightarrow A$ 、 $R \rightarrow U$ 、 $U \rightarrow A$ 、 $U \rightarrow R$ 。

## 4.2. 实验分析

### 4.2.1. 实验相关设置与评价指标

本文实验的操作系统是 Windows 11, CPU 型号为 AMD Ryzen 5 5500U with Radeon Graphics。在实验中, 对于由黑白像素构成的 COIL20 数据集, 将其直接展开为 1024 维向量作为模型输入; 考虑到 Office-31 与 RS 数据集中具有更丰富的视觉信息, 利用 CLIP 预训练模型中的 Vision Transformer 图像编码器[37]提取 512 维通用视觉特征作为模型输入, 利用经重心映射后的源域数据上训练分类器, 在目标域数据上进行评估。为全面衡量模型性能, 本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数作为评价指标。其中, 准确率指正确预测样本数占总样本数的比例, 用于衡量模型的整体分类正确率; 精确率为预测为正类的样本中实际为正类的比例, 反映模型预测正类的准确程度; 召回率是实际为正类的样本中被正确预测的比例, 体现模型对正类别的查全能力; F1 分数为精确率与召回率的调和平均数, 用于综合权衡二者性能, 尤其在数据分布不均衡时提供更稳健的评价。具体公式如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

式(10)~(13)中各指标计算公式中涉及混淆矩阵的基本统计量定义如下: TP (True Positive)指模型正确预测为正类的样本数量; FP (False Positive)为模型错误预测为正类的负类样本数量; FN (False Negative)指模型错误预测为负类的正类样本数量; TN (True Negative)为模型正确预测为负类的样本数量。

### 4.2.2. 对比实验结果与分析

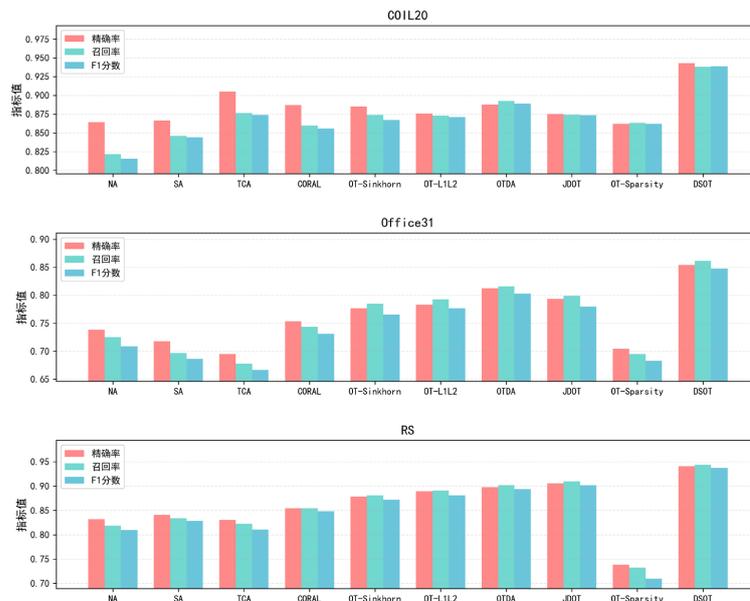
为验证本文提出模型的效果, 将 NA 作为基线方法, 即直接利用源域样本训练的分类器对目标域进行分类。按照算法原理可以将对比方法分为两类: 基于统计度量的方法和基于最优传输的方法。其中, 基于统计度量的方法: SA [38], 通过学习线性变换将源域和目标域的特征子空间进行对齐; TCA [5], 通过最小化源域与目标域在再生核希尔伯特空间中的分布差异, 学习一个公共的降维特征空间; CORAL [6], 通过对齐源域和目标域数据分布的协方差矩阵, 减小域间差异。基于最优传输的方法: Sinkhorn OT [29], 熵正则最优传输方法; OT-L1L2 [9], 基于熵正则和  $l_1 - l_2$  群稀疏正则最优传输; OTDA [9], 基于熵正则和类正则的最优传输方法; JDOT [17], 联合分布最优传输方法; OT-Sparsity [39], 带有稀疏约束的最优传输方法。在对比实验中, 所有方法均在参数空间中进行经验性搜索以获得最优超参数配置, 并报告其最佳结果。为排除分类器性能对模型评估的干扰, 统一采用 1NN 分类器进行测试, 以检验各模型自身的跨域迁移能力。

实验结果如表 1 所示, 在每个任务的性能对比中, NA 的性能表现最差, 表明在源域与目标域存在分布差异时, 直接迁移模型会导致严重的性能下降, 从而凸显了进行域适应的必要性, 最佳结果均已用加粗字体标出。基于统计度量的方法在简单任务上(COIL20 数据集)相较 NA 基线表现出稳定的性能提升,

**Table 1.** The accuracy of the DSOT on various datasets (%)  
**表 1.** DSOT 在各数据集上的准确率(%)

		COIL20		Office31				RS					
方法		C1-C2	C2-C1	A-W	A-D	W-A	D-A	A-R	A-U	U-A	U-R	R-A	R-U
基于统计距离方法	NA	81.67	82.64	77.61	74.90	69.58	69.19	77.19	78.50	88.33	78.22	79.87	84.46
	SA	84.72	84.44	71.45	72.29	66.70	68.05	82.63	80.80	83.50	82.14	81.32	85.31
	TCA	88.06	87.22	65.91	71.89	67.80	67.63	74.83	82.80	84.87	80.57	81.81	83.54
	CORAL	87.92	84.03	78.62	75.90	72.24	72.10	87.48	83.70	86.07	86.63	82.22	84.08
基于最优传输方法	OT-Sinkhorn	87.78	85.97	83.52	82.93	74.58	74.97	86.03	85.80	90.62	89.11	85.61	87.23
	OT-L1L2	88.47	86.11	84.53	83.13	75.54	75.19	86.11	87.80	91.96	91.36	86.77	88.00
	OTDA	90.00	88.47	86.92	84.94	77.07	77.00	87.22	88.50	92.70	92.18	89.07	88.61
基于最优传输方法	JDOT	88.33	86.53	86.16	83.94	75.43	75.36	<b>89.04</b>	91.00	92.52	91.57	89.97	88.31
	OT-Sparsity	86.67	85.97	73.33	68.88	67.84	66.48	65.92	76.60	79.50	75.51	64.99	66.92
	DSOT	<b>93.19</b>	<b>94.44</b>	<b>89.94</b>	<b>86.55</b>	<b>84.13</b>	<b>81.01</b>	88.85	<b>93.70</b>	<b>93.43</b>	<b>98.02</b>	<b>93.64</b>	<b>95.92</b>

验证了分布对齐策略的有效性。然而, 面对复杂的高维数据时, 此类方法提升有限, 在部分任务中甚至性能低于基线方法, 说明其在实际应用中可能引发负迁移现象, 导致模型性能下降。相比之下, 基于最优传输的方法通过几何结构实现对齐, 表现出更强的稳健性, 在多数任务上优于基于统计距离的方法。但在跨域差异较大的任务中, 其性能提升仍较为有限。结合图 1 可以看出, 本文所提模型在各项评估指标上均显著优于其他方法, 表明其在整体分类性能和类别判别能力方面均具有显著优势。同时, 精确率与召回率的同步提升表明该方法在降低误分类的同时有效增强了对真实样本的识别能力。表 2 展示了不同问题规模(不同数量的源域和目标域样本设置)下模型的运行时间, 表明该模型可以处理大规模问题。在不同的域偏移场景下, 各项指标均保持稳定优势, 进一步验证了所提方法的鲁棒性与泛化能力。



**Figure 1.** A comparison chart of the performance metrics of each algorithm on three datasets  
**图 1.** 三个数据集上各算法性能指标的对比图

**Table 2.** Running time of each method under different problem scales (s)**表 2.** 不同问题规模下各方法运行时间(s)

方法	问题规模					
	720 × 720	2817 × 498	2817 × 795	3410 × 1000	9158 × 1300	8101 × 3870
SA	0.5069	1.1534	1.1711	1.4993	2.0354	2.5829
TCA	17.2687	108.8680	111.0380	283.8470	864.9300	1080.7900
CORAL	40.4410	4.0611	5.2820	6.3150	8.0392	9.0265
OT-Sinkhorn	2.6945	10.0408	14.9931	20.0046	62.2495	79.5516
OT-L1L2	19.0983	83.3099	72.7431	117.6000	136.3580	154.0280
OTDA	3.5802	7.7441	11.9951	16.3141	47.1142	78.7379
JDOT	1749.2100	2045.2200	2137.8200	7600.9900	17944.2000	4048.9800
OT- Sparsity	47.9062	89.9665	160.9500	353.1350	752.2630	5342.4800
DSOT	866.5960	1577.8000	1796.6500	4426.2100	5181.9900	6879.6200

#### 4.2.3. 消融实验

为验证本文所提出的基于扩散成本的稀疏正则最优传输模型(DSOT)中各模块的有效性, 本节在 Office31 数据集的 D→A 跨域任务上进行了消融实验。基准模型(Baseline)采用去除所有新增模块后的基于欧式距离的经典最优传输模型。随后, 依次增加基于扩散过程的成本函数、对边际约束的  $l_2$  惩罚项及  $l_0$  稀疏正则项, 以构建消融模型。所有模型均使用相同的参数设置及硬件环境进行训练与评估, 确保比较的公平性。

**Table 3.** Results of the ablation study**表 3.** 消融实验结果

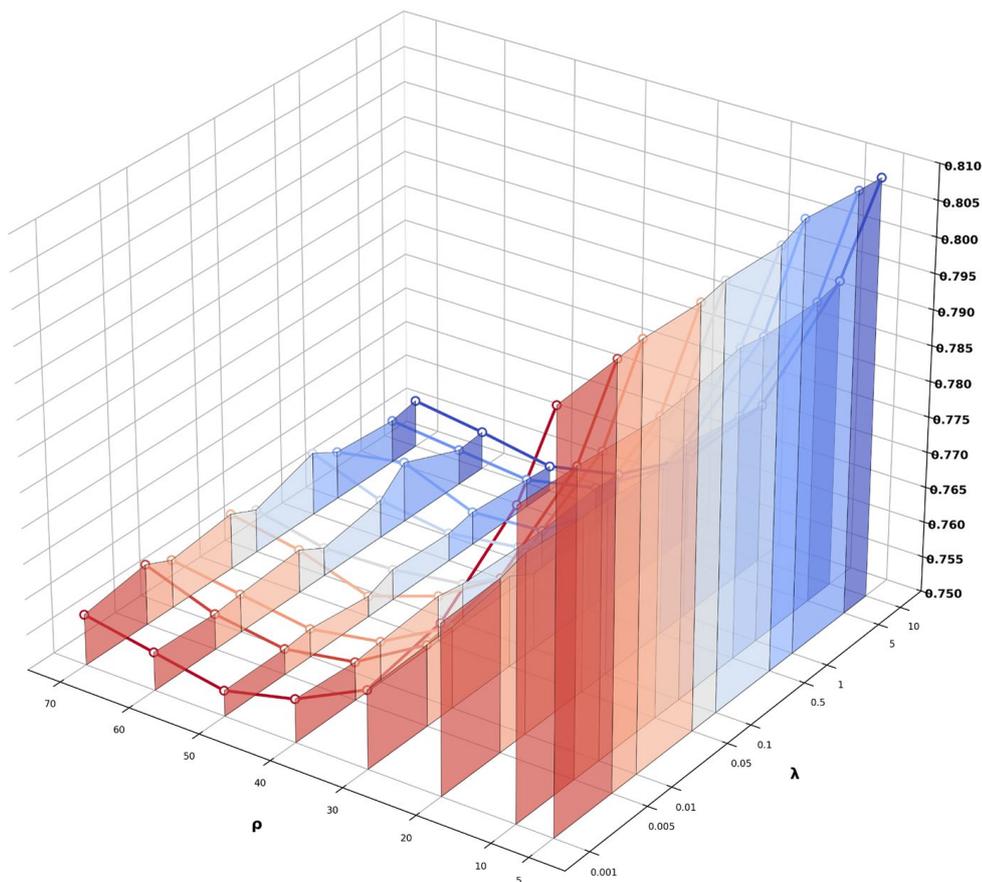
	扩散成本	$l_2$ 边际惩罚项	$l_0$ 稀疏正则项	D→A			
				Accuary	F1	Precision	Recall
(a)	-	-	-	0.6390	0.6354	0.6643	0.6380
(b)	-	√	-	0.6550	0.6577	0.6858	0.6571
(c)	-	-	√	0.6613	0.6243	0.6841	0.6575
(d)	√	-	-	0.6709	0.6232	0.6937	0.6606
(e)	-	√	√	0.6752	0.6772	0.7043	0.6774
(f)	√	-	√	0.7409	0.7339	0.7495	0.7437
(g)	√	√	-	0.7916	0.7915	0.8145	0.7928
(h)	√	√	√	<b>0.8101</b>	<b>0.8089</b>	<b>0.8290</b>	<b>0.8123</b>

实验结果如表 3 所示, 所提方法的各个模块均在提升模型性能方面发挥了重要作用。具体而言, 仅引入  $l_2$  边际惩罚项时, 模型准确率提高了约 2.00%, 且 F1 分数高于准确率, 表明松弛约束能在保持分类精度的同时有效缓解类别不平衡带来的预测偏差。值得注意的是, 在仅加入  $l_0$  稀疏正则项或仅加入扩散成本时, 模型虽然均呈现出分类准确率的提升, 但二者均出现 F1 分数明显偏低的情况, 说明在样本分布不平衡的场景下, 仅依赖稀疏项和扩散成本仍可能导致模型对少数类的识别能力不足。进一步观察发现,

当  $l_0$  稀疏正则项或扩散成本分别与松弛约束结合使用时, F1 值均得到显著提高, 这说明松弛约束在协调各类别间平衡、改善整体分类稳健性方面具有关键作用。综合以上分析, 扩散成本、 $l_2$  边际惩罚项与  $l_0$  稀疏正则项三者的协同机制, 能够有效度量样本间的类别差异, 提升模型在跨域适应中的类别判别能力。该机制不仅保障了不同类别间的预测平衡性, 还增强了模型的泛化能力与匹配关系的可解释性, 从而弥补了单一模块在处理类别不平衡数据时的局限性。实验结果表明, 当三个组件共同作用时, 模型在各项评估指标上均达到最优性能, 充分验证了 DSOT 模型设计的有效性 with 系统性优势。

#### 4.2.4. 参数敏感性分析

为系统评估不同参数组合对模型性能的影响, 本研究在 Office31 的 D→A 任务上进行了全面的参数网格搜索实验。实验中,  $l_0$  稀疏正则项参数  $\lambda$  的取值范围设定为 [0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10],  $l_2$  边际惩罚项参数  $\rho$  的取值范围则为 [5, 10, 20, 30, 40, 50, 60, 70]。针对每一组  $(\lambda, \rho)$  参数对, 记录模型在相应配置下的准确率, 并将所有结果进行可视化, 以便直观比较不同参数设置对性能的影响趋势。当参数  $\lambda$  过大时, 传输计划会被过度压缩, 导致总传输质量趋近于零, 无法实现传输; 而过小则失去稀疏促进作用。参数  $\rho$  过大时, 模型会强制严格匹配边际分布, 从而抑制了  $l_0$  项的稀疏效果, 解趋向于稠密的经典最优传输; 过小则会导致边际严重偏离。因此, 需要通过权衡参数的大小, 在稀疏性和边际匹配精度之间达成平衡。如图 2 所示, 在 D→A 任务中, 当  $\lambda=1, \rho=5$  时, 模型性能达到最优。



**Figure 2.** The performance of the model under different parameter combinations  
**图 2.** 模型在不同参数组合下的表现

#### 4.2.5. 统计显著性检验

为系统评估各模型性能差异的统计显著性,本研究采用非参数统计检验方法,检验结果如图3所示。

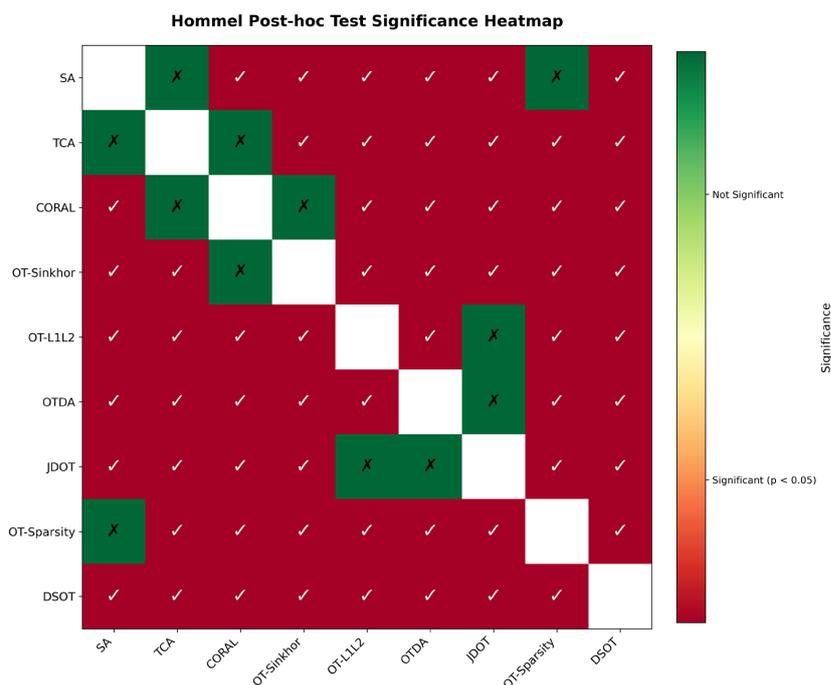


Figure 3. Hommel post hoc test significance heatmap  
图3. Hommel 事后检验显著性热图

首先使用 Friedman 检验[40]对模型在多个数据集上的整体性能差异进行评估。该方法通过对每个数据集内各模型的性能进行排序并比较平均秩次,避免了对数据分布的前提假设。Friedman 检验结果显示,整体存在显著差异( $\chi^2 = 83.7415$ ,  $p < 0.05$ ),表明至少有两个模型在性能上具有统计显著性区别。为进一步识别具体的差异模型,采用 Hommel 校正[41]进行多重比较事后检验。Hommel 方法基于闭合检验原理,通过逐步调整显著性水平,可在严格控制族系错误率( $\text{FWER} \leq \alpha = 0.05$ )的同时保持较高的检验力。结果如图3所示,本文提出的模型与所有对比方法均存在统计显著性差异(所有  $p < 0.05$ ),验证了本文模型在处理跨域任务性能上显著优于现有方法,且此优势具有统计可靠性。

## 5. 结论

本文提出一种基于扩散成本的稀疏正则最优传输模型(DSOT),旨在缓解无监督域适应中源域与目标域之间的分布偏移。该方法在构建成本函数时,充分考虑了数据的几何结构:通过引入扩散算子捕捉数据内在的流形结构,并融合源域标签信息来隐式增强类别约束,从而构建出更具判别力的跨域相似性度量。同时,模型采用松弛边际约束与稀疏正则化相结合的策略,促使传输过程聚焦于跨域间最显著的样本对应关系,这不仅更贴合实际场景中局部匹配的假设,也提升了模型结果的解释性。通过抑制微弱或无关的传输过程,该模型一定程度上能够过滤噪声样本的干扰,进而提升跨域分类的鲁棒性。最后,在三个公开域适应数据集上的实验结果表明,DSOT 在多项评价指标上均取得最优性能,在保持较高分类精度的同时实现了精确率与召回率的最佳平衡,为不平衡场景下的无监督域适应任务提供了一种泛化能力强的有效方法。

## 基金项目

河北省中央引导地方科技专项项目(246Z1825G)。

## 参考文献

- [1] 王静, 刘嘉星, 宋婉莹, 等. 基于空间变换网络和特征分布校准的小样本皮肤图像分类模型[J]. 计算机应用, 2025, 45(8): 2720-2726.
- [2] 周嘉文, 郑小盈, 祝永新, 等. 多头自注意力与双线性池化融合的心肌缺血影像分类[J]. 计算机工程, 2025, 51(11): 246-257.
- [3] 袁志祥, 高永奇. InternDiffuseDet: 结合可变形卷积和扩散模型的目标检测方法[J]. 计算机工程与应用, 2024, 60(12): 203-215.
- [4] Ganin, Y. and Lempitsky, V. (2015) Unsupervised Domain Adaptation by Backpropagation. *International Conference on Machine Learning*, **37**, 1180-1189.
- [5] Pan, S.J., Tsang, I.W., Kwok, J.T. and Yang, Q. (2011) Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, **22**, 199-210. <https://doi.org/10.1109/tnn.2010.2091281>
- [6] Sun, B., Feng, J. and Saenko, K. (2016) Return of Frustratingly Easy Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**, 2058-2065. <https://doi.org/10.1609/aaai.v30i1.10306>
- [7] Chen, C., Fu, Z., Chen, Z., Jin, S., Cheng, Z., Jin, X., et al. (2020) HoMM: Higher-Order Moment Matching for Unsupervised Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 3422-3429. <https://doi.org/10.1609/aaai.v34i04.5745>
- [8] von Seeger, A., Zou, D. and Lerman, G. (2025) Stein Discrepancy for Unsupervised Domain Adaptation. arXiv: 2502.03587.
- [9] Courty, N., Flamary, R., Tuia, D. and Rakotomamonjy, A. (2017) Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1853-1865. <https://doi.org/10.1109/tpami.2016.2615921>
- [10] Maman, G. and Talmon, R. (2025) Geometric Optimal Transport for Unsupervised Domain Adaptation. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=8Nef4vZUzU>
- [11] Das, D. and George Lee, C.S. (2018) Unsupervised Domain Adaptation Using Regularized Hyper-Graph Matching. 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 7-10 October 2018, 3758-3762. <https://doi.org/10.1109/icip.2018.8451152>
- [12] Kang, G., Jiang, L., Yang, Y. and Hauptmann, A.G. (2019) Contrastive Adaptation Network for Unsupervised Domain Adaptation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 15-20 June 2019, 4893-4902. <https://doi.org/10.1109/cvpr.2019.00503>
- [13] Li, J., Chen, E., Ding, Z., Zhu, L., Lu, K. and Shen, H.T. (2021) Maximum Density Divergence for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3918-3930. <https://doi.org/10.1109/tpami.2020.2991050>
- [14] He, Y., Chen, J., Zhang, J., Ke, W. and Gong, Y. (2025) A Bayesian Dual-Pathway Network for Unsupervised Domain Adaptation. *Pattern Recognition*, **164**, Article 111498. <https://doi.org/10.1016/j.patcog.2025.111498>
- [15] 赵恩庭, 张长春, 赵海涛, 等. 基于对抗学习的野生动物图像域适应识别方法[J]. 林业科学, 2025, 61(4): 1-8.
- [16] Wang, X., Jiang, H., Mu, M. and Dong, Y. (2025) A Dynamic Collaborative Adversarial Domain Adaptation Network for Unsupervised Rotating Machinery Fault Diagnosis. *Reliability Engineering & System Safety*, **255**, Article 110662. <https://doi.org/10.1016/j.ress.2024.110662>
- [17] Courty, N., Flamary, R., Habrard, A. and Rakotomamonjy, A. (2017) Joint Distribution Optimal Transportation for Domain Adaptation. *Advances in Neural Information Processing Systems*, **30**, 3733-3742.
- [18] Damodaran, B.B., Kellenberger, B., Flamary, R., Tuia, D. and Courty, N. (2018) Deepjdot: Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y. Eds., *Lecture Notes in Computer Science*. Springer International Publishing, 467-483. [https://doi.org/10.1007/978-3-030-01225-0\\_28](https://doi.org/10.1007/978-3-030-01225-0_28)
- [19] Ida, Y., Kanai, S., Adachi, K., Kumagai, A. and Fujiwara, Y. (2023) Fast Regularized Discrete Optimal Transport with Group-Sparse Regularizers. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 7980-7987. <https://doi.org/10.1609/aaai.v37i7.25965>
- [20] Nguyen, T., Nguyen, V., Le, T., Zhao, H., Tran, Q.H. and Phung, D. (2024) A Class-Aware Optimal Transport Approach with Higher-Order Moment Matching for Unsupervised Domain Adaptation. arXiv:2401.15952.

- [21] Mensch, A. and Peyré, G. (2020) Online Sinkhorn: Optimal Transport Distances from Sample Streams. *Advances in Neural Information Processing Systems*, **33**, 1657-1667.
- [22] Bonafini, M. and Schmitzer, B. (2021) Domain Decomposition for Entropy Regularized Optimal Transport. *Numerische Mathematik*, **149**, 819-870. <https://doi.org/10.1007/s00211-021-01245-0>
- [23] Lin, T., Ho, N. and Jordan, M. I. (2022) On the Efficiency of Entropic Regularized Algorithms for Optimal Transport. *Journal of Machine Learning Research*, **23**, 6143-6184.
- [24] Burkard, R.E., Klinz, B. and Rudolf, R. (1996) Perspectives of Monge Properties in Optimization. *Discrete Applied Mathematics*, **70**, 95-161. [https://doi.org/10.1016/0166-218x\(95\)00103-x](https://doi.org/10.1016/0166-218x(95)00103-x)
- [25] Kantorovich, L.V. (2006) On a Problem of Monge. *Journal of Mathematical Sciences*, **133**, 1383-1383. <https://doi.org/10.1007/s10958-006-0050-9>
- [26] Caffarelli, L. and McCann, R. (2010) Free Boundaries in Optimal Transport and Monge-Ampère Obstacle Problems. *Annals of Mathematics*, **171**, 673-730. <https://doi.org/10.4007/annals.2010.171.673>
- [27] Chapel, L., Alaya, M. Z. and Gasso, G. (2020) Partial Optimal Transport with Applications on Positive-Unlabeled Learning. *Advances in Neural Information Processing Systems*, **33**, 2903-2913.
- [28] Zhang, C., Ren, H. and He, X. (2024) P2OT: Progressive Partial Optimal Transport for Deep Imbalanced Clustering. arXiv:2401.09266.
- [29] Cuturi, M. (2013) Sinkhorn Distances: Lightspeed Computation of Optimal Transport. *Advances in Neural Information Processing Systems*, **26**, 2292-2300.
- [30] Ma, L., Bian, W. and Xue, X. (2024) Point Clouds Matching Based on Discrete Optimal Transport. *IEEE Transactions on Image Processing*, **33**, 5650-5662. <https://doi.org/10.1109/tip.2024.3459594>
- [31] Lacombe, T. (2023) An Homogeneous Unbalanced Regularized Optimal Transport Model with Applications to Optimal Transport with Boundary. *International Conference on Artificial Intelligence and Statistics*, **206**, 7311-7330.
- [32] Tang, W., Ma, Z., Sun, H. and Wang, J. (2023) Learning Sparse Alignments via Optimal Transport for Cross-Domain Fake News Detection. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10095281>
- [33] Haasler, I. and Elvander, F. (2024) Multi-Frequency Tracking via Group-Sparse Optimal Transport. *IEEE Control Systems Letters*, **8**, 1048-1053. <https://doi.org/10.1109/lcsys.2024.3405375>
- [34] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [35] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-aos729>
- [36] Bian, W. and Chen, X. (2020) A Smoothing Proximal Gradient Algorithm for Nonsmooth Convex Regression with Cardinality Penalty. *SIAM Journal on Numerical Analysis*, **58**, 858-883. <https://doi.org/10.1137/18m1186009>
- [37] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I. (2021) Learning Transferable Visual Models from Natural Language Supervision. *International Conference on Machine Learning*, **139**, 8748-8763.
- [38] Fernando, B., Habrard, A., Sebban, M. and Tuytelaars, T. (2013) Unsupervised Visual Domain Adaptation Using Subspace Alignment. 2013 *IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 2960-2967. <https://doi.org/10.1109/iccv.2013.368>
- [39] Liu, T., Puigcerver, J. and Blondel, M. (2022) Sparsity-Constrained Optimal Transport. arXiv:2209.15466.
- [40] Friedman, M. (1937) The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, **32**, 675-701. <https://doi.org/10.1080/01621459.1937.10503522>
- [41] Hommel, G. (1988) A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. *Biometrika*, **75**, 383-386. <https://doi.org/10.1093/biomet/75.2.383>

## 附录

注意到函数  $f(t) = c^T t + \frac{\rho}{2} \|At - d\|^2$  是凸函数, 关于简化的 SPG 算法有如下收敛结论:

**定理**(单调下降性)算法生成的序列  $\{t^k\}$  满足

$$F(t^{k+1}) \leq F(t^k) - \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2$$

其中  $F(t) = f(t) + \lambda\phi(t)$ 。

证明: 因为子问题  $Q_{d^k, \gamma_k}$  是强凸函数, 且满足线搜索下降条件时  $t^{k+1} = \hat{t}^{k+1}$ , 对于任意  $t \in [0, 1]^{n_s n_t}$ , 有

$$\begin{aligned} & f(t^k) + \langle \nabla f(t^k), t^{k+1} - t^k \rangle + \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2 + \lambda\phi^{d^k}(t^{k+1}) \\ & \leq f(t^k) + \langle \nabla f(t^k), t - t^k \rangle + \frac{\gamma_k}{2} \|t - t^k\|^2 + \lambda\phi^{d^k}(t) - \frac{\gamma_k}{2} \|t^{k+1} - t\|^2 \end{aligned} \quad (14)$$

重新整理得到

$$\lambda\phi^{d^k}(t^{k+1}) \leq \lambda\phi^{d^k}(t) + \langle \nabla f(t^k), t - t^{k+1} \rangle + \frac{\gamma_k}{2} \|t - t^k\|^2 - \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2 - \frac{\gamma_k}{2} \|t^{k+1} - t\|^2 \quad (15)$$

下降条件可写为

$$f(t^{k+1}) + \lambda\phi^{d^k}(t^{k+1}) \leq f(t^k) + \langle \nabla f(t^k), t^{k+1} - t^k \rangle + \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2 + \lambda\phi^{d^k}(t^{k+1}) \quad (16)$$

结合(15)和(16)式可以得到

$$f(t^{k+1}) + \lambda\phi^{d^k}(t^{k+1}) \leq f(t^k) + \lambda\phi^{d^k}(t) + \langle \nabla f(t^k), t - t^k \rangle + \frac{\gamma_k}{2} \|t - t^k\|^2 - \frac{\gamma_k}{2} \|t^{k+1} - t\|^2 \quad (17)$$

因为  $f$  是凸函数, 有:

$$f(t^k) + \langle \nabla f(t^k), t - t^k \rangle \leq f(t) \quad (18)$$

由(17)和(18)式得到

$$f(t^{k+1}) + \lambda\phi^{d^k}(t^{k+1}) \leq f(t) + \lambda\phi^{d^k}(t) + \frac{\gamma_k}{2} \|t - t^k\|^2 - \frac{\gamma_k}{2} \|t^{k+1} - t\|^2 \quad (19)$$

令  $t = t^k$ , 我们得到

$$f(t^{k+1}) + \lambda\phi^{d^k}(t^{k+1}) + \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2 \leq f(t^k) + \lambda\phi^{d^k}(t^k) \quad (20)$$

因此

$$F(t^{k+1}) \leq F(t^k) - \frac{\gamma_k}{2} \|t^{k+1} - t^k\|^2 \quad (21)$$

这表明  $\{F(t^k)\}$  单调不增且有下界(因为  $f$  在紧集  $t \in [0, 1]^{n_s n_t}$  上有界), 故收敛。