

Article

Class-Driven Robust Non-Negative Matrix Factorization with Dual-Hypergraph Regularization for Data Clustering

Haiyan Gao ^{1,2,*} and Gaigai Zhou ¹

¹ School of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou 730020, China; zhougg@lzufe.edu.cn

² Key Laboratory of Digital Economy and Social Computing Science of Gansu, Lanzhou 730020, China

* Correspondence: gaohy@lzufe.edu.cn

Abstract

Traditional non-negative matrix factorization (NMF) faces challenges when dealing with complex data, primarily characterized by sensitivity to noise, neglect of data geometric structure, and inability to effectively utilize supervised information. To address these limitations, this paper proposes a class-driven robust non-negative matrix factorization with dual-hypergraph regularization (CRNMFDH) framework. The core contributions of this framework include the following: Firstly, the design of a novel dual-hypergraph regularization term that symmetrically captures and preserves the higher-order geometric structures of both the sample space and feature space, establishing a mutually reinforcing topological relationship between them. Secondly, an introduction of a class-driven mechanism to effectively integrate label information into the decomposition process, significantly enhancing the discriminative capability of the low-dimensional representations. Finally, the employment of a loss function based on correntropy to replace the traditional Euclidean distance, thereby enhancing the model's robustness against noise and outliers. Extensive experiments across nine datasets demonstrate that CRNMFDH significantly outperforms existing state-of-the-art algorithms in multiple clustering evaluation metrics and noise robustness, providing an effective new solution for complex data clustering tasks.

Keywords: non-negative matrix factorization; semi-supervised learning; dual-hypergraph regularization; correntropy; clustering

1. Introduction

As a core technique in unsupervised learning, clustering has become a highly prominent research direction in the fields of pattern recognition and machine learning due to its broad applicability and efficient computational characteristics. In recent years, researchers have focused on developing diverse clustering paradigms to enhance algorithmic performance, such as typical algorithms including K-means clustering [1], hierarchical clustering [2], spectral clustering [3], subspace clustering [4], and non-negative matrix factorization (NMF) [5]. These algorithms have demonstrated significant advantages across multiple application scenarios. Here, NMF achieves dimensionality reduction through non-negative low-rank decomposition. By employing non-negative linear com-

Received: 1 January 2026

Revised: 8 February 2026

Accepted: 9 February 2026

Published: 13 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

binations, it effectively preserves the global linear structure of the data [5]. This characteristic makes it suitable for clustering tasks in non-negative data scenarios such as image processing and text mining.

Although classical NMF demonstrates unique advantages in clustering tasks, its inherent Euclidean distance metric mechanism causes the algorithm to overlook the prevalent local manifold geometric characteristics and underlying structures in high-dimensional spaces. This limitation is particularly pronounced when dealing with datasets with complex topological structures, potentially leading to issues such as feature space distortion and blurred clustering boundaries. To address this, some scholars [6–9] have effectively overcome the aforementioned drawbacks by introducing nearest-neighbor graphs to model local geometric relationships in high-dimensional spaces. However, such models based on simple graphs are constrained by binary adjacency relationships. While they can capture local neighborhood information, they struggle to depict higher-order interaction patterns among data points. In contrast, hypergraphs establish multivariate associations through hyperedges, enabling the construction of a more expressive similarity measurement framework at the topological level. This provides richer and more rigorous mathematical tools for the representation of complex structural data. Hypergraph learning has been successfully applied across multiple domains. For instance, representative works include the dynamic hypergraph neural networks (DHKH) developed by Kang et al. [10] and the hypergraph-regularized NMF (HNMF) by integrating hypergraph learning with NMF, developed by Zeng et al. [11]. Subsequently, Ren et al. [12] introduced adaptive hypergraph regularization within the NMF framework to preserve high-order manifold structures of the data. Yu et al. [13] proposed a robust capped-norm dual-hypergraph regularized tri-factorization NMF (RCHNMTF) by encoding latent multivariate geometric information in both the sample space and the feature space, demonstrating excellent robustness and accuracy in image clustering. To further enhance the model's resistance to noise, Kong et al. [14] employed the $L_{2,1}$ -norm to measure error and proposed a robust NMF ($L_{2,1}$ NMF). By penalizing sparsity in sample dimensions, it handles sample-level outliers and achieves superior and consistent results in clustering tasks. However, its ability to suppress strong noise at the pixel level or non-Gaussian distributions remains limited. Wang et al. [15] designed a logarithmic loss function to measure errors and proposed robust bidirectional random graph regularized matrix factorization (RBSMF). While RBSMF exhibits robustness against heavy-tailed noise distributions, its performance remains dependent on the alignment between the noise distribution and the function parameters. Peng et al. [16] employed the maximum correntropy as the loss function and proposed a correntropy-based graph regularized concept factorization (GCCF), solving the model using the half-quadratic (HQ) optimization technique [17], which effectively suppressed the impact of noise and outliers. As an effective robustness metric, correntropy demonstrates significant advantages in handling non-Gaussian noise and outliers. Subsequently, Yu et al. [18] proposed correntropy-based hypergraph regularized NMF (CHNMF), which not only explores higher-order geometric information among multiple sample points but also further enhances the algorithm's robustness on complex, noisy data.

The aforementioned algorithms are all unsupervised learning methods. When combined with a small amount of labeled data and a large volume of unlabeled data, the accuracy of clustering results can be significantly improved [19–21]. To this end, researchers have integrated label information with manifold information and proposed various semi-supervised NMF algorithms, including discriminative graph-regularized NMF (GDNMF) [22], dual-graph regularized NMF with label discrimination (DCNMFLD) [23], semi-supervised robust hypergraph regularized NMF (SCNMFLD) [24], and semi-supervised exclusive multi-view clustering method (semiEMC) [25], among others. Meanwhile, Li et al.

[26] introduced a class-driven constraint into NMF and proposed class-driven graph regularized NMF (MCDNMF), which forces the learned subsets of basis vectors to represent their respective classes, thereby enhancing the representation consistency of similar data in the low-dimensional space.

1.1. Motivation and Construction

Existing NMF clustering algorithms still face multiple challenges when handling high-dimensional complex data: the objective function based on the Frobenius norm is sensitive to noise and outliers; the sub-structures extracted from the basis matrix have limited capabilities in effectively representing specific classes; and there is insufficient preservation of the inherent local geometric structure of the data. To address these limitations, this paper constructs a hypergraph that integrates the duality relationships between samples and features within a semi-supervised learning framework to preserve the global manifold structure of the data. It introduces a class-driven mechanism to effectively utilize partially labeled information and combines it with the maximum correntropy criterion (MCC) to develop a class-driven robust non-negative matrix factorization with dual-hypergraph regularization for data clustering (CRNMFDH).

As a semi-supervised learning clustering method, the CRNMFDH algorithm effectively utilizes partially known label information to guide the decomposition of the original data matrix into a more discriminative low-dimensional matrix. Subsequently, the K-means algorithm is applied to cluster this low-dimensional representation matrix. Figure 1 illustrates the framework structure of the CRNMFDH algorithm. The core contributions of this study are summarized as follows:

We propose a novel dual-hypergraph regularized NMF framework (CRNMFDH). By constructing a sample hypergraph and a feature hypergraph, this framework symmetrically captures and preserves the higher-order geometric structures of both the sample space and feature space. This design enables the learned low-dimensional representations to simultaneously accommodate the local topology and global discriminative properties of the data.

We have designed an effective semi-supervised learning and robust optimization strategy. By introducing a class-driven discriminative term, we leverage the existing label information to guide the matrix decomposition process, thereby maximizing inter-class discrimination and enhancing the discriminative capability of the basis matrix. Meanwhile, adopting the correntropy criterion as the loss function significantly improves the algorithm's robustness against noise and outliers. We have also designed an efficient iterative solution algorithm for this nonconvex problem based on the half-quadratic (HQ) optimization technique, ensuring the model's convergence and computational efficiency.

We provide comprehensive theoretical analysis and experimental validation. Theoretically, we rigorously prove the convergence of the proposed algorithm and analyze its computational complexity. Experimentally, we conduct a thorough comparison with nine state-of-the-art benchmark algorithms across nine distinct domain datasets. The experimental results strongly demonstrate that the CRNMFDH algorithm achieves significant and consistent performance advantages across multiple key clustering metrics, showcasing its effectiveness and robustness.

1.2. Structure of This Paper

The remainder of this paper is organized as follows. Section 2 provides a brief overview of the fundamental theories of CDNMF and hypergraph learning, laying the theoretical foundation for subsequent algorithm design. Section 3 elaborates on the mathematical model of the CRNMFDH algorithm, explains its optimization process, and provides a theoretical analysis of its convergence and time complexity. Section 4 systematically

evaluates the effectiveness of this algorithm through multiple experiments and provides an in-depth discussion of the results. Section 5 summarizes the entire paper and proposes several feasible directions for future research.

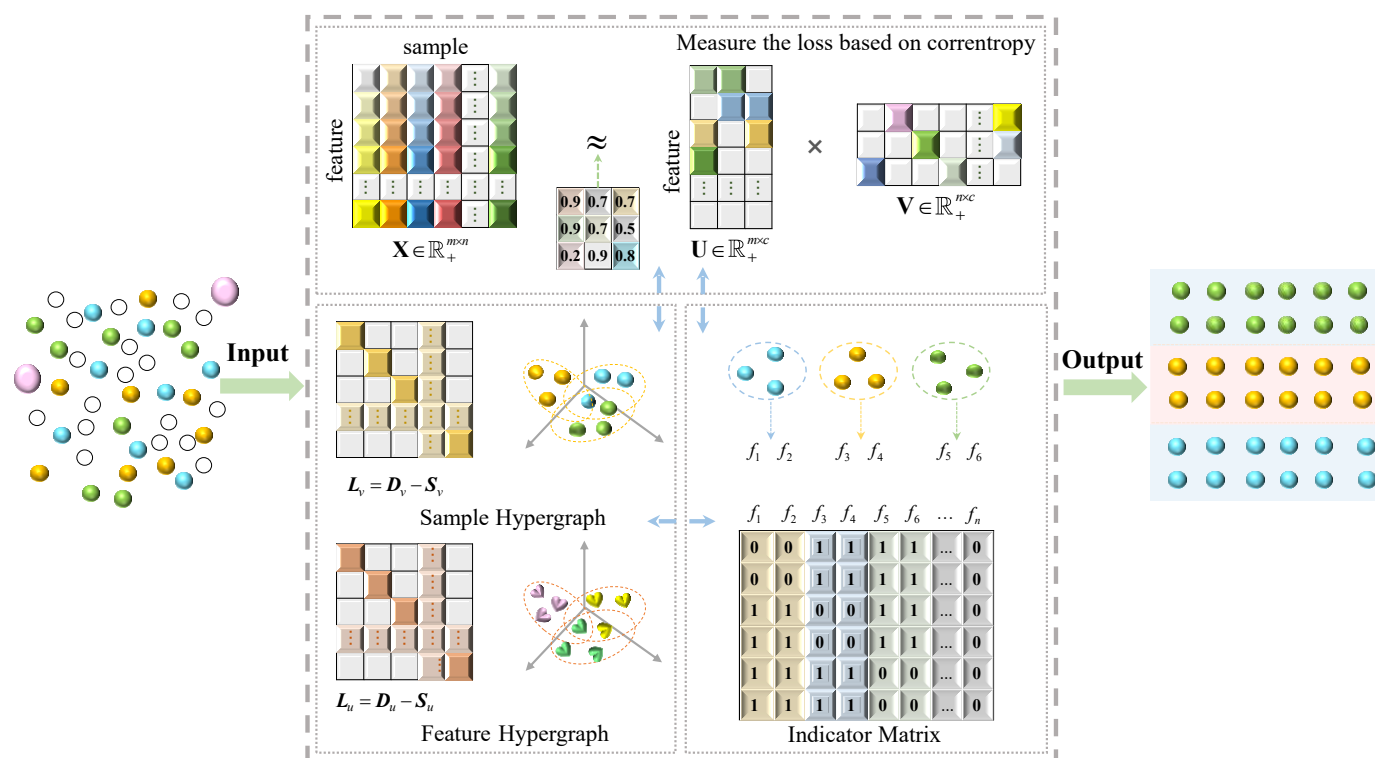


Figure 1. Schematic diagram of the proposed CRNMFHDH algorithm.

2. Related Work

2.1. Class-Driven NMF (CDNMF)

Standard NMF, as a non-negative constrained linear data representation and dimensionality reduction tool, aims to decompose the original data matrix $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}_+^{m \times n}$ into an approximate form of the product of the basis matrix $U = (u_1, u_2, \dots, u_c) \in \mathbb{R}_+^{m \times c}$ and the coefficient matrix $V = (v_1, v_2, \dots, v_c) \in \mathbb{R}_+^{c \times n}$, i.e., $X \approx UV^T$. NMF employs the Frobenius norm metric UV^T to measure the approximation degree of X , expressed as the following optimization problem [5]:

$$\min_{U, V \geq 0} \|X - UV^T\|_F^2 \quad (1)$$

Although NMF considers the global structural features of data, the absence of labeled information to guide the clustering process results in low-dimensional representations lacking sufficient discriminative power, limiting its application in clustering tasks. To enhance inter-class separability and clustering accuracy, Xiao et al. [27] introduced a class-driven constraint into the NMF framework, proposing class-driven NMF (CDNMF).

Assume the first l data points in the data matrix X are labeled $b_j \in \{1, 2, \dots, z\}$, where z is the total number of classes, and the remaining data points are unlabeled. To learn a more discriminative basis matrix $U = (u_1, u_2, \dots, u_c) \in \mathbb{R}_+^{m \times c}$, where u_i effectively represents the data of class i , while weakly representing other classes. Let each

class correspond to r basis vectors satisfying $c = r \times z$. Based on this, the indicator matrix $Y = (y_{b_1}, y_{b_2}, \dots, y_{b_n})^T \in \mathbb{R}^{n \times c}$ for any labeled sample x_j with label b_j and low-dimensional representation v_j must satisfy the following conditions:

$$y_{b_j}^T v_j = 0 \quad (2)$$

Here, the coefficient represents the selection of v_j with respect to y_{b_j} , i.e., the coefficient corresponding to the basis vectors not belonging to u_b . Equation (2) represents the class-driven constraint. The CDNMF algorithm corresponds to the optimization problem [27]:

$$\min_{U, V \geq 0} \|X - UV^T\|_F^2 + \lambda \text{Tr}(YV^T) \quad (3)$$

The multiplication update rule for U and V is as follows:

$$U_{ik} \leftarrow U_{ik} \frac{(XV)_{ik}}{(UV^T V)_{ik}} \quad (4)$$

$$V_{jk} \leftarrow V_{jk} \frac{(X^T U)_{jk}}{(VU^T U + (\lambda Y) / 2)_{jk}} \quad (5)$$

2.2. Hypergraph Learning

Traditional NMF minimizes the global reconstruction error as its optimization objective, failing to effectively uncover and preserve the local manifold structure of the data. Based on the manifold hypothesis, adjacent sample points in high-dimensional Euclidean space should retain their local neighborhood relationships after embedding into a low-dimensional manifold space. Graph-regularized NMF (GNMF) [6] pioneered the characterization of local geometric structures within data by constructing sample similarity graphs. Its optimization objective is

$$\min_{U, V \geq 0} \|X - UV^T\|_F^2 + \alpha \text{Tr}(V^T L V) \quad (6)$$

Here, $S \in \mathbb{R}^{n \times n}$ denotes the weight matrix, $L = D - S$ represents the Laplacian matrix, and D denotes the degree matrix. α serves as a non-negative regularization parameter, whose magnitude directly determines the balance between reconstruction error and manifold structure preservation.

However, real-world data typically exhibits more complex higher-order correlation patterns. GNMF can only characterize local pairwise similarity relationships and struggles to capture the multimodal, multi-type higher-order semantic correlations prevalent in high-dimensional data. This limitation restricts its ability to model complex manifold structures. Research indicates that preserving higher-order geometric relationships among data points can further enhance the clustering performance of algorithms [11,28,29]. To this end, hypergraph learning has been introduced into the non-negative matrix factorization framework. A hyperedge in a hypergraph can connect multiple vertices, typically constructed using the KNN method centered on each vertex. In this case, the number of vertices in the hypergraph equals the number of hyperedges. Figure 2 shows an example model of a graph and a hypergraph.

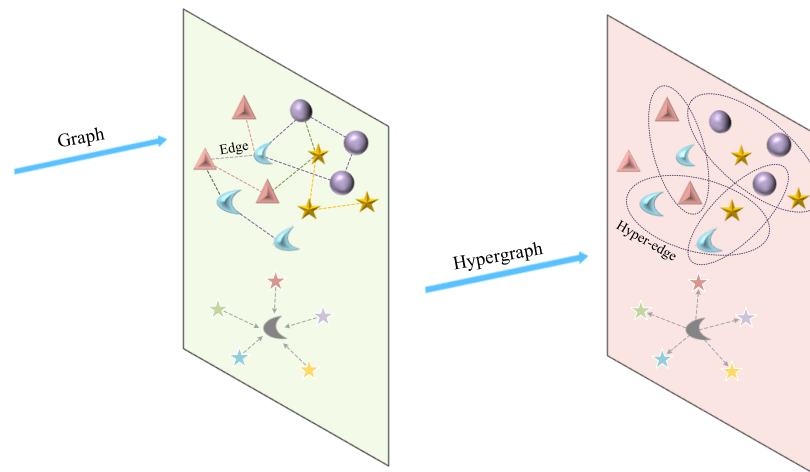


Figure 2. Example of a graph and a hypergraph.

Define the weight of the i -th hyperedge in the hypergraph as

$$w(e) = W_{ii} = \sum_{v_j \in e_i} \exp\left(-\|v_i - v_j\|^2 / \eta^2\right) \quad (7)$$

where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ denotes the vertex set, $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$ denotes the hyperedge set, $\eta = 1/k \sum_{v_j \in e_i} \|v_i - v_j\|$, and k represents the number of nearest neighbors for each vertex. The non-normalized hypergraph Laplacian matrix L_v is defined as

$$L_v = D_v - S_v \quad (8)$$

Here, $S_v = HWD_e^{-1}H^T$. The incidence matrix H reflects the correspondence between vertices and hyperedges, $D_v = \text{diag}(HW\mathbf{1})$ is the diagonal matrix of vertex degrees, $D_e = \text{diag}(\mathbf{1}^T H)$ denotes the diagonal matrix of hyperedge degrees, and $\mathbf{1} = (1, 1, \dots, 1)^T$. Then, hypergraph regularization can be transformed into the following minimization problem:

$$\ell_{\text{data}} = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{i, j \in e} \frac{w(e)}{\delta(e)} \|\mathbf{v}_i^T - \mathbf{v}_j^T\|^2 = \text{Tr}(\mathbf{V}^T (D_v - S_v) \mathbf{V}) = \text{Tr}(\mathbf{V}^T L_v \mathbf{V}) \quad (9)$$

where $\mathbf{v}_i^T$ and $\mathbf{v}_j^T$ represent the low-dimensional representations of the original data $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively.

Given that high-dimensional data exhibits not only local geometric correlations in the sample space but also embodies manifold structures in the feature space [9], a feature space-based hypergraph regularization term is introduced to more comprehensively model the intrinsic structural relationships within the data. The non-normalized Laplacian matrix of the feature hypergraph is defined as $L_u = D_u - S_u$. The objective function for the dual-hypergraph regularized non-negative matrix factorization model can then be expressed as

$$\min_{U, V \geq 0} \|\mathbf{X} - UV^T\|_F^2 + \alpha \text{Tr}(\mathbf{V}^T L_v \mathbf{V}) + \gamma \text{Tr}(\mathbf{U}^T L_u \mathbf{U}) \quad (10)$$

Here, α and β are non-negative hyperparameters that respectively constrain the weights of the regularization terms in the sample hypergraph and feature hypergraph.

Both L_v and L_u are symmetric positive definite matrices, which respectively smooth the low-dimensional embeddings of samples and features.

3. The Proposed Method

3.1. Objective Function

Standard NMF employs the Frobenius norm to measure reconstruction error. The squared error term significantly amplifies the influence of outlier samples, causing the learned basis matrix and coefficient matrix to deviate from the true data structure. This reduces the model's performance in feature extraction and clustering tasks. To reduce the model's sensitivity to noise, we introduce correntropy as the loss function. Correntropy, as a nonlinear function of local similarity metrics, has demonstrated robustness [30]. Its core idea involves applying a nonlinear mapping to the error through a Gaussian kernel function:

$$V(X, Y) = E \left[\exp \left(-(X - Y)^2 / 2\sigma^2 \right) \right] = \int \exp \left(-(x - y)^2 / 2\sigma^2 \right) dF_{XY}(x, y) \quad (11)$$

where X and Y are two similar random variables, σ is the kernel bandwidth parameter, and $E[\cdot]$ is the expectation operator. However, in practical applications, the joint distribution function $F_{XY}(x, y)$ is often unknown. Typically, correntropy is represented using a finite number of data samples $\{(x_i, y_i)\}_{i=1}^n$:

$$V_{N, \sigma}(X, Y) = \frac{1}{n} \sum_{i=1}^N g(x_i - y_i) = \frac{1}{n} \sum_{i=1}^N \exp \left(-(x_i - y_i)^2 / 2\sigma^2 \right) \quad (12)$$

Here, $V_{n, \sigma}$ represents the estimator of correntropy. As shown in Equation (12), when outliers cause excessive reconstruction errors (i.e., $|x_i - y_i|$ is too large), the corresponding $\exp \left(-(x_i - y_i)^2 / 2\sigma^2 \right)$ value becomes smaller. The loss function for NMF based on correntropy [31] is

$$\ell = \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) \quad (13)$$

Unlike the L_2 loss, which is sensitive to outliers, the correntropy criterion serves as a nonlinear measure of local similarity. From an optimization standpoint, for samples with large errors (i.e., noise or outliers), the gradient of the loss function based on correntropy rapidly diminishes towards zero. This implies that during the optimization of the objective function, the loss function based on the MCC metric automatically reduces the effective weight assigned to such samples. As a result, noise and outliers contribute minimally to the updates of model parameters, ensuring stability in the decomposition process and robustness in the final results.

Based on this, we incorporate the dual-hypergraph regularization term and class-driven constraints into the objective function (13) to enhance the discriminative capability of the basis matrices and the structured representation of the data, thereby improving the overall performance of semi-supervised clustering tasks on complex, noisy data. We propose a class-driven robust dual-hypergraph regularized NMF algorithm (CRNMFHDH), with an objective function defined as follows:

$$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \alpha \text{Tr}(V^T L_v V) - \gamma \text{Tr}(U^T L_u U) - \beta \text{Tr}(YV^T) \quad (14)$$

Here, α , β , and γ are all non-negative regularization parameters. The first term $\sum_{i=1}^m g\left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2}\right)$ is a loss function based on correntropy, which mitigates the algorithm's robustness to noisy data by assigning smaller weights to anomalous samples. The second approach $\text{Tr}(\mathbf{V}^T \mathbf{L}_v \mathbf{V})$ explores higher-order geometric structures embedded within the sample space by constructing a sample hypergraph of the data, thereby enhancing the discriminative power and accuracy of low-dimensional matrices. The third approach $\text{Tr}(\mathbf{U}^T \mathbf{L}_u \mathbf{U})$ extracts latent higher-order information among multiple features by constructing a feature hypergraph of the data, thereby fully preserving both the local geometric properties and global discriminative structure of the data. This symmetrical regularization design ensures that the manifold structures of both the sample and feature spaces are preserved simultaneously. The fourth term $\text{Tr}(\mathbf{YV}^T)$ is a class-driven constraint that associates class label information with each basis vector. This forces the learning of a more discriminative set of basis vectors, thereby enhancing the discriminative power of the basis matrix. Addressing the shortcomings of existing algorithms that rely solely on simple graph structures or simple pairwise constraints while neglecting higher-order geometric structures and the integration of supervisory information, the CRN-MFDH algorithm demonstrates significant advantages. Especially in noisy environments where label information is scarce and outliers exist, existing algorithms face challenges in maintaining both information integrity and discriminative power. CRNMFDH provides an effective solution to this problem.

3.2. Optimization

Since the objective function in Equation (14) is both nonlinear and nonconvex, in order to better solve this nonconvex objective function, we employ the half-quadratic (HQ) optimization technique. This allows us to transform the objective function into a convex one with a multiplicative form and convert the maximization objective function into an easily solvable minimization optimization problem. First, according to Proposition 1 [31], the first term of Equation (14) can be rewritten as

$$\begin{aligned} \max_{U, V \geq 0} \sum_{i=1}^m g\left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2}\right) &= \min_{U, V \geq 0} (-\ell) = \text{Tr}((X^T - \mathbf{V}U^T)\mathbf{A}(X - UV^T)) \\ &= \text{Tr}((X^T \mathbf{A}X - 2\mathbf{V}U^T \mathbf{A}X + \mathbf{V}U^T \mathbf{A}UV^T)) \end{aligned} \quad (15)$$

where $\mathbf{A} \in \mathbb{R}_+^{m \times m}$ is a positive diagonal matrix with diagonal elements equal to

$$\mathbf{A}_{ii} = -\exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2\right) / \sigma^2 \quad (16)$$

The value of σ is determined empirically [16], and its calculation formula is as follows:

$$\sigma = \sqrt{\frac{1}{2m} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \quad (17)$$

Therefore, the objective function (14) is equivalent to optimizing the following problem:

$$f = \text{Tr}(X^T \mathbf{A}X - 2\mathbf{V}U^T \mathbf{A}X + \mathbf{V}U^T \mathbf{A}UV^T) + \alpha \text{Tr}(\mathbf{V}^T \mathbf{L}_v \mathbf{V}) + \gamma \text{Tr}(\mathbf{U}^T \mathbf{L}_u \mathbf{U}) + \beta \text{Tr}(\mathbf{YV}^T) \quad (18)$$

The augmented Lagrangian of Equation (18) is

$$\begin{aligned}\mathfrak{L}(\mathbf{U}, \mathbf{V}) = & \text{Tr}(\mathbf{X}^T \mathbf{A} \mathbf{X} - 2\mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{X} + \mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{U} \mathbf{V}^T) + \alpha \text{Tr}(\mathbf{V}^T \mathbf{L}_v \mathbf{V}) \\ & + \gamma \text{Tr}(\mathbf{U}^T \mathbf{L}_u \mathbf{U}) + \beta \text{Tr}(\mathbf{Y} \mathbf{V}^T) - \text{Tr}(\boldsymbol{\Phi} \mathbf{U}^T) - \text{Tr}(\boldsymbol{\Psi} \mathbf{V}^T)\end{aligned}\quad (19)$$

Here, $\boldsymbol{\Phi} = (\Phi_{ij})$ and $\boldsymbol{\Psi} = (\Psi_{ij})$ are the Lagrange multipliers for $\mathbf{U} \geq 0$ and $\mathbf{V} \geq 0$. The first-order partial derivatives of Equation (19) with respect to $\mathbf{U}$ and $\mathbf{V}$ are respectively

$$\frac{\partial \mathfrak{L}}{\partial \mathbf{U}} = -2\mathbf{A} \mathbf{X} \mathbf{V} + 2\mathbf{A} \mathbf{U} \mathbf{V}^T \mathbf{V} + 2\gamma \mathbf{L}_u \mathbf{U} - \boldsymbol{\Phi} \quad (20)$$

$$\frac{\partial \mathfrak{L}}{\partial \mathbf{V}} = -2\mathbf{X}^T \mathbf{A} \mathbf{U} + 2\mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{U} + 2\alpha \mathbf{L}_v \mathbf{V} - \beta \mathbf{Y} - \boldsymbol{\Psi} \quad (21)$$

Given that $\Phi_{ij} U_{ij} = 0$ and $\Psi_{ij} V_{ij} = 0$, according to the condition [32], solving for $\frac{\partial \mathfrak{L}}{\partial \mathbf{U}} = 0$ and $\frac{\partial \mathfrak{L}}{\partial \mathbf{V}} = 0$ yields

$$(-2\mathbf{A} \mathbf{X} \mathbf{V} + 2\mathbf{A} \mathbf{U} \mathbf{V}^T \mathbf{V} + 2\gamma \mathbf{L}_u \mathbf{U})_{ij} U_{ij} = 0 \quad (22)$$

$$(-2\mathbf{X}^T \mathbf{A} \mathbf{U} + 2\mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{U} + 2\alpha \mathbf{L}_v \mathbf{V} + \beta \mathbf{Y})_{ij} V_{ij} = 0 \quad (23)$$

Thus, the update rules for $\mathbf{U}$ and $\mathbf{V}$ are

$$U_{ik} \leftarrow U_{ik} \frac{(\mathbf{A} \mathbf{X} \mathbf{V} + \gamma \mathbf{S}_u \mathbf{U})_{ik}}{(\mathbf{A} \mathbf{U} \mathbf{V}^T \mathbf{V} + \gamma \mathbf{D}_u \mathbf{U})_{ik}} \quad (24)$$

$$V_{jk} \leftarrow V_{jk} \frac{(\mathbf{X}^T \mathbf{A} \mathbf{U} + \alpha \mathbf{S}_v \mathbf{V})_{jk}}{(\mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{U} + \alpha \mathbf{D}_v \mathbf{V} + \beta \mathbf{Y} / 2)_{jk}} \quad (25)$$

3.3. Convergence Analysis

We employ the auxiliary function method [5,33] to prove the convergence of the CRNMFHD algorithm. Since the lower bound of the objective function in Equation (18) is non-negative, it suffices to demonstrate that the objective function is non-increasing under the update rules defined by Equations (24) and (25).

Let the first derivative of function f be denoted as f' . In the objective function (18), the parts involving U_{ik} and V_{jk} are denoted as $f_{U_{ik}}(\mathbf{U})$ and $f_{V_{jk}}(\mathbf{V})$, respectively. Prove that the auxiliary functions of $f_{U_{ik}}(\mathbf{U})$ and $f_{V_{jk}}(\mathbf{V})$ are respectively

$$g(\mathbf{U}, \mathbf{U}_{ik}^t) = f_{U_{ik}}(\mathbf{U}_{ik}^t) + f'_{U_{ik}}(\mathbf{U}_{ik}^t)(\mathbf{U} - \mathbf{U}_{ik}^t) + \frac{(\mathbf{A} \mathbf{U} \mathbf{V}^T \mathbf{V} + \gamma \mathbf{D}_u \mathbf{U})_{ik}}{\mathbf{U}_{ik}^t} (\mathbf{U} - \mathbf{U}_{ik}^t)^2 \quad (26)$$

$$g(\mathbf{V}, \mathbf{V}_{jk}^t) = f_{V_{jk}}(\mathbf{V}_{jk}^t) + f'_{V_{jk}}(\mathbf{V}_{jk}^t)(\mathbf{V} - \mathbf{V}_{jk}^t) + \frac{(\mathbf{V} \mathbf{U}^T \mathbf{A} \mathbf{U} + \alpha \mathbf{D}_v \mathbf{V} + \beta \mathbf{Y} / 2)_{jk}}{\mathbf{V}_{jk}^t} (\mathbf{V} - \mathbf{V}_{jk}^t)^2 \quad (27)$$

From Equations (26) and (27), it follows readily that $g(\mathbf{U}, \mathbf{U}) = f_{U_{ik}}(\mathbf{U})$ and $g(\mathbf{V}, \mathbf{V}) = f_{V_{jk}}(\mathbf{V})$. Let f'' denote the second derivative of f . Then the Taylor expansions of $f_{U_{ik}}(\mathbf{U})$ and $f_{V_{jk}}(\mathbf{V})$ at $\mathbf{U} = \mathbf{U}_{ik}^t$ and $\mathbf{V} = \mathbf{V}_{jk}^t$, respectively, are given by

$$f_{U_{ik}}(\mathbf{U}) = f_{U_{ik}}(\mathbf{U}_{ik}^t) + f'_{U_{ik}}(\mathbf{U}_{ik}^t)(\mathbf{U} - \mathbf{U}_{ik}^t) + \frac{f''_{U_{ik}}(\mathbf{U}_{ik}^t)}{2}(\mathbf{U} - \mathbf{U}_{ik}^t)^2 \quad (28)$$

$$f_{V_{jk}}(\mathbf{V}) = f_{V_{jk}}(\mathbf{V}_{jk}^t) + f'_{V_{jk}}(\mathbf{V}_{jk}^t)(\mathbf{V} - \mathbf{V}_{jk}^t) + \frac{f''_{V_{jk}}(\mathbf{V}_{jk}^t)}{2}(\mathbf{V} - \mathbf{V}_{jk}^t)^2 \quad (29)$$

Then proving $g(\mathbf{U}, \mathbf{U}_{ik}^t) \geq f_{U_{ik}}(\mathbf{U})$ and $g(\mathbf{V}, \mathbf{V}_{jk}^t) \geq f_{V_{jk}}(\mathbf{V})$ is equivalent to proving

$$\frac{(\mathbf{A}\mathbf{U}\mathbf{V}^T\mathbf{V} + \gamma\mathbf{D}_u\mathbf{U})_{ik}}{\mathbf{U}_{ik}^t} \geq \frac{f''_{U_{ik}}(\mathbf{U}_{ik}^t)}{2} = \mathbf{A}_{ii}(\mathbf{V}^T\mathbf{V})_{kk} + \gamma(\mathbf{L}_u)_{ii} \quad (30)$$

$$\frac{(\mathbf{V}\mathbf{U}^T\mathbf{A}\mathbf{U} + \alpha\mathbf{D}_v\mathbf{V} + \beta\mathbf{Y}/2)_{jk}}{\mathbf{V}_{jk}^t} \geq \frac{f''_{V_{jk}}(\mathbf{V}_{jk}^t)}{2} = (\mathbf{U}^T\mathbf{A}\mathbf{U})_{kk} + \alpha(\mathbf{L}_v)_{jj} \quad (31)$$

Equate Equation (30) to

$$(\mathbf{A}\mathbf{U}\mathbf{V}^T\mathbf{V})_{ik} = \sum_{l=1}^n \mathbf{U}_{il}^t \mathbf{A}_{li} (\mathbf{V}^T\mathbf{V})_{lk} \geq \mathbf{U}_{ik}^t \mathbf{A}_{ii} (\mathbf{V}^T\mathbf{V})_{kk} \quad (32)$$

$$(\mathbf{D}_u\mathbf{U})_{jk} = \sum_{q=1}^n (\mathbf{D}_u)_{jq} \mathbf{U}_{qk}^t \geq (\mathbf{D}_u)_{ii} \mathbf{U}_{ik}^t \geq (\mathbf{D}_u - \mathbf{S}_u)_{ii} \mathbf{U}_{ik}^t \quad (33)$$

The following can be obtained:

$$(\mathbf{A}\mathbf{U}\mathbf{V}^T\mathbf{V} + \gamma\mathbf{D}_u\mathbf{U})_{ik} \geq \mathbf{U}_{ik}^t [\mathbf{A}_{ii}(\mathbf{V}^T\mathbf{V})_{kk} + \gamma(\mathbf{L}_u)_{ii}] \quad (34)$$

Therefore, $g(\mathbf{U}, \mathbf{U}_{ik}^t) \geq f_{U_{ik}}(\mathbf{U})$, meaning $g(\mathbf{U}, \mathbf{U}_{ik}^t)$ is an auxiliary function of function $f_{U_{ik}}(\mathbf{U})$.

Equate Equation (31) to

$$(\mathbf{V}\mathbf{U}^T\mathbf{A}\mathbf{U})_{jk} = \sum_{r=1}^n \mathbf{V}_{jr}^t (\mathbf{U}^T\mathbf{A}\mathbf{U})_{rk} \geq \mathbf{V}_{jk}^t (\mathbf{U}^T\mathbf{A}\mathbf{U})_{kk} \quad (35)$$

$$(\alpha\mathbf{D}_v\mathbf{V})_{jk} = \alpha \sum_{w=1}^n (\mathbf{D}_v)_{jw} \mathbf{V}_{wk}^t \geq \alpha(\mathbf{D}_v)_{jj} \mathbf{V}_{jk}^t \geq \alpha(\mathbf{D}_v - \mathbf{S}_v)_{jj} \mathbf{V}_{jk}^t \quad (36)$$

Since $\beta\mathbf{Y} \geq 0$, it follows that

$$(\mathbf{V}\mathbf{U}^T\mathbf{A}\mathbf{U} + \alpha\mathbf{D}_v\mathbf{V} + \beta\mathbf{Y}/2)_{jk} \geq \mathbf{V}_{jk}^t [(\mathbf{U}^T\mathbf{A}\mathbf{U})_{kk} + \alpha(\mathbf{L}_v)_{jj}] \quad (37)$$

It is proven that $g(\mathbf{V}, \mathbf{V}_{jk}^t) \geq f_{V_{jk}}(\mathbf{V})$, i.e., $g(\mathbf{V}, \mathbf{V}_{jk}^t)$ is an auxiliary function of function $f_{V_{jk}}(\mathbf{V})$.

In summary, Equations (26) and (27) serve as auxiliary functions for functions $f_{U_{ik}}(\mathbf{U})$ and $f_{V_{jk}}(\mathbf{V})$, respectively. Thus, the objective function ultimately converges to a local optimum through the iterative Equations (24) and (25). Algorithm 1 provides the specific implementation steps of the CRNMFDH algorithm.

Algorithm 1 CRNMFDH

Input: Data matrix X , number of nearest neighbors k_1 and k_2 in sample hypergraph and feature hypergraph, label ratio, indicator matrix Y , parameters α , β , and γ , maximum iteration count T , $\varepsilon = 10^{-10}$.

Output: Basic matrix U , coefficient matrix V , and clustering results.

Steps:

1. Randomly initialize the values of U^0 and V^0 ;
2. If the convergence condition $\sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) \leq \varepsilon$ holds and the iteration count reaches T , proceed to step 7; otherwise, proceed to step 3;
3. Compute the Laplacian matrices L_v and L_u ;
4. Fix U and V , update σ and A ;
5. Fix V , update U according to Equation (24), namely:

$$U_{ik}^{t+1} \leftarrow U_{ik}^t \frac{(A^T X V^t + \gamma S_u U^t)_{ik}}{(A^T U^t (V^t)^T V^t + \gamma D_u V^t)_{ik}}, \quad \forall i \in \{1, 2, \dots, m\}, \quad \forall k \in \{1, 2, \dots, c\}$$

6. Fix U , update V according to Equation (25), namely:

$$V_{jk}^{t+1} \leftarrow V_{jk}^t \frac{(X^T A^T U^t + \alpha S_v V^t)_{jk}}{(V^t (U^t)^T A^T U^t + \alpha D_v V^t + \beta Y / 2)_{jk}}, \quad \forall j \in \{1, 2, \dots, n\}, \quad \forall k \in \{1, 2, \dots, c\}$$

Here $t = T$

7. Perform clustering on V using the K-means algorithm;

3.4. Time Complexity Analysis

This section analyzes the time complexity of the CRNMFDH algorithm, denoted by O . The computational complexity required for constructing the sample hypergraph and feature hypergraph in CRNMFDH is $O(mn^2 + nm^2)$, where m is the feature dimension and n is the number of samples. The computational complexity for updating σ , A , U , and V is $O(mnc)$, where c is the embedding dimension. Assuming the CRNMFDH algorithm terminates after the T -th iteration, the overall computational complexity of the algorithm is denoted as $O(Tmnc + mn^2 + nm^2)$. Additionally, Table 1 presents a comparison of the time complexities between the proposed algorithm and CRNMFDH. It can be observed that the complexity of the CRNMFDH algorithm is higher than that of NMF, CDNMF, and MCDNMF. However, this increased complexity significantly enhances the accuracy of matrix decomposition and the performance of clustering.

Table 1. Time complexity comparison of eight benchmark algorithms.

Algorithm	Objective Function	Time Complexity
NMF [5]	$\min_{U, V \geq 0} \ X - UV^T\ _F^2$	$O(Tmnc)$
GNMF [6]	$\min_{U, V \geq 0} \ X - UV^T\ _F^2 + \alpha \text{Tr}(V^T L V)$	$O(Tmnc + mn^2)$
HNMF [11]	$\min_{U, V \geq 0} \ X - UV^T\ _F^2 + \alpha \text{Tr}(V^T L_v V)$	$O(Tmnc + mn^2)$
DHCF [34]	$\min_{U, V \geq 0} \ X - XUV^T\ _F^2 + \alpha \text{Tr}(V^T L_v V) + \beta \text{Tr}(U^T L_u U)$	$O(Tn^2 c + mn^2 + nm^2)$
MCCNMF [31]	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right)$	$O(Tmnc)$
CNMF [35]	$\min_{U, Z \geq 0} \ X - UZ^T A^T\ _F^2$	$O(Tmnc + mn^2)$

CDNMF [27]	$\min_{U, V \geq 0} \ X - UV^T\ _F^2 + \alpha \text{Tr}(YV^T)$	$O(Tmnc)$
MCDNMF [26]	$\min_{U, V \geq 0} \ X - UV^T\ _F^2 + \alpha \text{Tr}(V^T L V) + \beta \text{Tr}(YV^T)$	$O(Tmnc + mn^2)$
CRNMFHDH	$\max_{U, V \geq 0} \sum_{j=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \alpha \text{Tr}(V^T L V) - \beta (U^T L U) - \gamma \text{Tr}(YV^T)$	$O(Tmnc + mn^2 + nm^2)$

4. Experiments and Analysis

This section conducts a comparative analysis between the proposed algorithm and nine state-of-the-art algorithms on nine publicly available datasets. Given that this study presents an improved NMF algorithm, we selected eight relevant and advanced NMF variants, along with the K-means benchmark algorithm, for comparison. For semi-supervised algorithms, 20% of samples in each class were randomly selected as labeled data, and the indicator matrix Y was constructed accordingly. During the experiments, we set the embedding dimension c of the new data space to be equal to the total number of categories z . The success of clustering algorithms depends not only on their consistency with labels but also on their ability to preserve the internal structure of the data—namely, the compactness of clusters and the distances between them. We employ four commonly used external evaluation metrics: Accuracy (ACC), Normalized Mutual Information (NMI), Purity (PUR), and Adjusted Rand Index (ARI). Additionally, the Silhouette Coefficient (SC) is utilized to measure the compactness and separability of clusters [36,37]. The values of ACC, NMI, and PUR range from $[0, 1]$, while ARI ranges from $[-1, 1]$. The closer their values approach 1, the higher the precision of the clustering results [38,39]. The SC value ranges from $[-1, 1]$, with values closer to 1 indicating better cohesion and separation within clusters. The experiment was implemented using Python 3.9.13. The computer environment configuration for the experiment is as follows: 11th Gen Intel (R) Core (TM) i5-1135G7 processor @ 2.40 GHz–2.42 GHz, 16 GB of memory, and a 64-bit Windows 10 operating system.

4.1. Datasets

The nine datasets used in the experiment are Semeion, UMIST, COIL20, warpPIE10P, Jaffe, USPS, att1, wdbc, and Vehicle. Semeion, wdbc, and vehicle are available at (<http://archive.ics.uci.edu/ml>), while COIL20 and USPS datasets can be downloaded from (<http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html>). For the USPS dataset, 200 images were randomly selected from each category to form the experimental subset. The Jaffe and UMIST datasets can be accessed at (<http://www.kasrl.org/jaffe.html> and <http://images.ee.umist.ac.uk/danny/database.html>), respectively. The warpPIE10P dataset [40] is a subset of the PIE dataset, comprising 210 samples across 10 categories. The att1 dataset, created by the AT&T Laboratories at Cambridge University, contains 400 facial images from 40 individuals and can be downloaded from (<https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>). Table 2 presents statistical information for the nine datasets.

Table 2. Basic information about datasets.

Dataset	Samples (n)	Features (m)	Classes (z)	Data Type	Image Size
Semeion	1593	256	10	digital images	16×16
UMIST	574	10,304	20	facial images	112×92
COIL20	1440	1024	20	object images	32×32
warpPIE10P	210	2420	10	facial images	32×32
Jaffe	213	1024	10	facial images	32×32
USPS	9298	256	10	digital images	16×16
att1	400	644	40	facial images	92×112
wdbc	569	30	2	medical images	— —
Vehicle	846	18	4	vehicle contour data	— —

4.2. Comparative Algorithms and Parameter Settings

A brief introduction to the nine comparison algorithms is as follows:

- (1) K-means [1] (unsupervised) achieves hard clustering of data through an iterative optimization process by minimizing the sum of squared Euclidean distances from samples within each cluster to their respective cluster centroids.
- (2) NMF [5] (unsupervised) aims to decompose a non-negative matrix into the product of two low-rank factor matrices, which partially represent the whole, and approximates the factor matrices by calculating reconstruction errors.
- (3) GNMF [6] (unsupervised) effectively learns latent local manifold information in data by introducing a graph regularization term.
- (4) HNMF [11] (unsupervised) encodes geometric information within the data space by constructing a data hypergraph, thereby preserving high-order manifold information across multiple samples.
- (5) DHCF [34] (unsupervised) constructs hypergraphs in both the sample space and feature space based on the dual graph regularization decomposition algorithm, thereby preserving their respective higher-order geometric structures.
- (6) MCCNMF [31] (unsupervised) reduces the algorithm's sensitivity to noise and outliers by reconstructing the correntropy error.
- (7) CNMF [35] (semi-supervised) incorporates labeled information as hard constraints into NMF, thereby enhancing clustering accuracy.
- (8) CDNMF [27] (semi-supervised) introduces a class-driven constraint within the NMF framework, associating label information with each basis vector. This ensures that features of data belonging to the same class become more similar, while features of data from different classes become more distinguishable.
- (9) MCDNMF [26] (semi-supervised) builds upon CDNMF by introducing a graph regularization term to preserve latent local geometric structures within the data space.

To ensure the fairness of the experiments, the number of nearest neighbors for all graph construction algorithms and those preserving higher-order geometric structures is uniformly set to $k = 5$. For each baseline algorithm, the parameter configurations are determined within the recommended ranges specified in their original literature, employing a unified grid search strategy. This ensures that the performance comparison of each algorithm is conducted under its optimal settings. The optimal hyperparameter combinations for each algorithm across different datasets are detailed in Table 3.

Table 3. Parameter settings of CRNMFHDH and comparative algorithms across the 9 datasets.

Dataset	CDNMF (α)	MCDNMF (α, β)	GNMF (α)	HNMF (α)	DHCF (α, β)	CRNMFHDH (α, β, γ)
Semeion	100	0.01, 100	10	0.001	0.01, 0.01	0.01, 0.1, 100
UMIST	100	10, 100	10	10	1, 1	0.1, 10, 100
COIL20	100	1, 100	10	0.01	0.1, 0.1	0.01, 10, 1000
warPIE10P	100	0.1, 100	0.1	0.01	0.01, 0.01	0.1, 10, 100
Jaffe	100	0.1, 100	1	0.01	0.01, 0.01	0.1, 0.001, 100
USPS	100	10, 100	10	0.01	0.01, 0.001	0.01, 1, 100
att1	10	0.01, 100	1	0.001	0.001, 0.001	0.001, 0.001, 100
wdbc	100	0.01, 100	10	0.001	0.001, 0.01	0.001, 1, 100
Vehicle	100	10, 100	10	0.001	0.001, 0.01	0.01, 0.001, 100

4.3. Results and Analysis

To minimize the impact of experimental randomness on results, each algorithm was independently run 20 times across nine datasets. The final clustering performance was determined by the mean $\pm$ standard deviation of each evaluation metric. Tables 4–8 summarize the specific values of the ACC, NMI, PUR, ARI, and SC evaluation metrics for 10 algorithms across nine datasets, with bolded numbers indicating the optimal result for the corresponding dataset under that metric.

Table 4. ACC of 10 algorithms on 9 datasets (%).

Algorithm Dataset	CRNMFHDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCCNMF	DHCF	MCDNMF
Semeion	85.62 $\pm$ 0.10	59.79 $\pm$ 4.10	51.07 $\pm$ 3.32	68.05 $\pm$ 4.47	53.08 $\pm$ 3.99	74.39 $\pm$ 0.44	55.19 $\pm$ 4.41	51.59 $\pm$ 3.75	61.11 $\pm$ 8.50	74.54 $\pm$ 0.82
UMIST	89.79 $\pm$ 1.99	43.57 $\pm$ 2.15	39.27 $\pm$ 2.24	45.30 $\pm$ 2.21	53.13 $\pm$ 3.39	59.26 $\pm$ 2.25	62.62 $\pm$ 1.16	34.90 $\pm$ 1.91	64.99 $\pm$ 2.72	64.91 $\pm$ 3.07
COIL20	91.22 $\pm$ 1.86	68.40 $\pm$ 2.66	64.45 $\pm$ 2.43	71.69 $\pm$ 2.29	73.84 $\pm$ 1.83	75.14 $\pm$ 0.57	75.73 $\pm$ 1.54	63.66 $\pm$ 2.70	82.66 $\pm$ 1.87	77.81 $\pm$ 1.78
warPIE10P	75.17 $\pm$ 2.97	26.07 $\pm$ 1.16	47.38 $\pm$ 3.19	59.74 $\pm$ 2.41	48.31 $\pm$ 3.71	65.05 $\pm$ 3.32	46.79 $\pm$ 2.68	45.36 $\pm$ 2.37	49.60 $\pm$ 2.06	63.67 $\pm$ 2.21
Jaffe	98.85 $\pm$ 0.23	86.46 $\pm$ 5.58	82.82 $\pm$ 4.32	88.52 $\pm$ 4.22	93.90 $\pm$ 4.53	94.34 $\pm$ 1.14	95.68 $\pm$ 2.50	90.19 $\pm$ 5.68	96.88 $\pm$ 3.56	97.56 $\pm$ 0.95
USPS200	92.40 $\pm$ 0.03	65.44 $\pm$ 0.56	66.30 $\pm$ 3.42	78.54 $\pm$ 2.17	80.14 $\pm$ 4.12	82.00 $\pm$ 0.11	70.14 $\pm$ 7.35	56.47 $\pm$ 2.79	72.30 $\pm$ 5.85	83.91 $\pm$ 0.28
att1	84.55 $\pm$ 1.49	45.12 $\pm$ 0.38	70.66 $\pm$ 2.23	75.49 $\pm$ 2.70	79.59 $\pm$ 1.45	75.20 $\pm$ 2.37	81.36 $\pm$ 1.75	69.42 $\pm$ 2.25	79.00 $\pm$ 2.10	83.31 $\pm$ 1.45
wdbc	93.48 $\pm$ 0.05	85.41 $\pm$ 0.00	72.62 $\pm$ 12.50	82.17 $\pm$ 12.13	89.82 $\pm$ 0.82	80.62 $\pm$ 8.94	90.01 $\pm$ 0.64	77.50 $\pm$ 13.24	90.03 $\pm$ 0.63	92.82 $\pm$ 0.31
Vehicle	65.75 $\pm$ 0.57	45.12 $\pm$ 0.38	38.50 $\pm$ 1.81	42.23 $\pm$ 3.32	45.09 $\pm$ 0.34	47.50 $\pm$ 4.28	45.22 $\pm$ 0.21	55.44 $\pm$ 2.84	45.53 $\pm$ 0.21	59.05 $\pm$ 0.47

Table 5. NMI of 10 algorithms on 9 datasets (%).

Algorithm Dataset	CRNMFHDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCCNMF	DHCF	MCDNMF
Semeion	77.30 $\pm$ 0.18	63.27 $\pm$ 3.09	43.74 $\pm$ 2.26	56.32 $\pm$ 2.91	52.09 $\pm$ 4.35	60.80 $\pm$ 0.46	54.36 $\pm$ 2.90	44.00 $\pm$ 2.80	56.11 $\pm$ 6.13	61.63 $\pm$ 1.10
UMIST	91.56 $\pm$ 0.34	51.03 $\pm$ 2.11	58.80 $\pm$ 1.82	45.30 $\pm$ 2.21	71.36 $\pm$ 1.86	66.53 $\pm$ 1.91	82.98 $\pm$ 0.47	54.05 $\pm$ 1.90	80.63 $\pm$ 0.82	78.12 $\pm$ 2.09

COIL20	94.99 ± 0.53	74.50 ± 1.73	74.83 ± 1.53	79.52 ± 1.33	81.96 ± 1.17	80.14 ± 0.51	84.35 ± 0.83	74.66 ± 1.52	89.98 ± 0.64	83.98 ± 0.78
warpPIE10P	75.88 ± 2.34	29.67 ± 1.20	53.59 ± 2.82	62.43 ± 1.93	53.12 ± 2.25	62.11 ± 2.19	51.65 ± 2.21	52.07 ± 2.79	55.79 ± 2.32	64.07 ± 2.10
Jaffe	98.25 ± 0.41	89.53 ± 3.65	84.28 ± 2.69	87.44 ± 2.33	92.86 ± 3.38	92.67 ± 0.95	94.73 ± 2.17	92.12 ± 3.25	96.44 ± 2.31	96.64 ± 1.12
USPS200	84.50 ± 0.06	69.15 ± 0.44	58.43 ± 2.03	67.33 ± 1.89	75.34 ± 2.31	70.88 ± 0.18	67.39 ± 4.23	51.73 ± 2.01	68.87 ± 3.35	73.95 ± 0.32
att1	91.15 ± 0.70	48.43 ± 2.96	82.60 ± 1.03	87.66 ± 1.35	89.67 ± 0.74	83.39 ± 1.40	90.16 ± 0.83	83.78 ± 1.43	90.55 ± 0.63	90.76 ± 0.77
wdbc	65.95 ± 0.30	85.41 ± 0.00	21.45 ± 18.43	38.61 ± 20.50	52.07 ± 2.64	35.18 ± 12.12	53.31 ± 2.23	28.18 ± 16.08	53.69 ± 2.24	61.29 ± 1.19
Vehicle	41.70 ± 0.85	48.43 ± 2.96	13.88 ± 2.22	16.96 ± 2.51	20.52 ± 0.67	23.39 ± 2.32	20.63 ± 0.21	50.57 ± 2.05	20.79 ± 0.21	30.07 ± 0.52

Table 6. PUR of 10 algorithms on 9 datasets (%).

Algorithm Dataset	CRNMF	FDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCCNMF	DHCF	MCDNMF
Semeion	85.62 ± 0.10		55.06 ± 2.01	53.01 ± 2.84	68.71 ± 3.85	56.73 ± 3.73	74.39 ± 0.44	57.58 ± 3.42	53.33 ± 3.75	62.25 ± 8.49	74.54 ± 0.82
UMIST	89.75 ± 1.04		65.89 ± 1.68	45.73 ± 2.16	51.37 ± 2.31	60.65 ± 3.13	60.41 ± 1.98	70.91 ± 0.27	41.22 ± 2.15	70.79 ± 1.05	70.34 ± 2.71
COIL20	91.85 ± 1.23		78.97 ± 1.27	67.34 ± 2.43	73.80 ± 1.85	75.51 ± 1.73	75.14 ± 0.57	77.27 ± 1.31	66.60 ± 2.26	84.38 ± 1.41	79.90 ± 1.31
warpPIE10P	76.05 ± 2.70		25.93 ± 2.05	49.62 ± 2.70	61.05 ± 2.30	49.69 ± 2.00	65.48 ± 2.94	47.76 ± 2.47	47.29 ± 2.16	52.88 ± 2.14	65.05 ± 2.07
Jaffe	98.85 ± 0.23		88.95 ± 3.61	84.25 ± 3.20	89.01 ± 3.16	94.13 ± 3.78	94.34 ± 1.14	95.70 ± 2.44	90.89 ± 4.89	97.09 ± 2.66	97.56 ± 0.95
USPS 200	92.40 ± 0.03		57.05 ± 0.35	68.20 ± 2.76	78.74 ± 1.94	81.32 ± 3.21	82.00 ± 0.11	71.25 ± 7.61	58.28 ± 2.15	73.56 ± 6.44	83.91 ± 0.28
att1	85.21 ± 1.29		18.61 ± 0.09	71.56 ± 1.76	78.30 ± 2.54	81.45 ± 1.24	75.67 ± 2.20	81.36 ± 1.75	72.28 ± 2.09	81.81 ± 1.53	83.79 ± 1.24
wdbc	93.48 ± 0.05		46.48 ± 0.00	74.24 ± 10.36	83.02 ± 10.34	89.82 ± 0.82	81.07 ± 7.72	90.01 ± 0.64	79.48 ± 9.78	90.03 ± 0.63	92.82 ± 0.31
Vehicle	65.75 ± 0.57		18.61 ± 0.09	40.72 ± 1.41	44.63 ± 2.89	45.23 ± 0.42	49.26 ± 3.37	45.22 ± 0.20	57.55 ± 2.59	45.53 ± 0.21	59.05 ± 0.47

Table 7. ARI of 10 algorithms on 9 datasets (%).

Algorithm Dataset	CRNMF	FDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCCNMF	DHCF	MCDNMF
Semeion	71.75 ± 0.22		40.96 ± 3.20	30.31 ± 2.71	47.17 ± 3.79	35.69 ± 3.89	53.03 ± 0.60	34.67 ± 3.22	31.10 ± 3.55	42.11 ± 6.76	53.29 ± 1.05
UMIST	80.15 ± 1.68		34.54 ± 2.09	28.22 ± 2.47	33.66 ± 2.32	43.25 ± 3.75	38.97 ± 2.85	50.30 ± 1.79	22.90 ± 2.19	55.43 ± 2.55	42.02 ± 5.46
COIL20	86.03 ± 1.95		62.05 ± 2.50	53.34 ± 2.43	63.56 ± 1.87	66.75 ± 2.27	64.08 ± 0.85	70.29 ± 1.53	55.30 ± 2.36	78.10 ± 1.65	68.41 ± 2.04
warpPIE10P	64.05 ± 2.70	6.49 ± 1.10	26.80 ± 2.67	40.57 ± 2.53	26.84 ± 1.88	30.42 ± 3.74	24.88 ± 2.03	25.31 ± 2.12	30.87 ± 1.99	37.00 ± 3.33	
Jaffe	97.44 ± 0.52		79.20 ± 6.57	73.09 ± 4.86	79.13 ± 4.76	88.64 ± 6.28	88.37 ± 1.91	91.56 ± 3.92	84.97 ± 6.63	94.13 ± 4.36	94.83 ± 1.90
USPS 200	83.85 ± 0.06		47.12 ± 0.53	48.04 ± 2.95	60.60 ± 2.79	68.40 ± 3.41	65.00 ± 0.20	57.15 ± 6.28	38.67 ± 2.43	59.19 ± 5.31	68.15 ± 0.55

att1	72.59 ± 3.28	12.09 ± 0.08	55.46 ± 2.37	66.18 ± 3.49	71.61 ± 1.80	43.24 ± 5.05	71.60 ± 2.28	56.93 ± 2.01	72.38 ± 1.85	70.90 ± 3.47
wdbc	75.39 ± 0.18	49.14 ± 0.00	26.24 ± 22.85	46.99 ± 25.04	63.10 ± 2.54	40.09 ± 17.78	63.65 ± 2.02	36.90 ± 21.16	63.69 ± 1.99	73.19 ± 1.07
Vehicle	36.17 ± 0.90	12.09 ± 0.08	10.71 ± 1.36	13.48 ± 2.84	13.75 ± 0.37	15.82 ± 2.35	13.79 ± 0.23	35.92 ± 2.65	14.49 ± 0.24	27.49 ± 0.64

Table 8. SC of 10 algorithms on 9 datasets (%).

Algorithm Dataset	CRNMF	FDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCCNMF	DHCF	MCDNMF
Semeion	67.58 ± 0.07		6.35 ± 0.25	20.77 ± 1.27	24.98 ± 1.34	32.83 ± 4.02	40.91 ± 0.29	45.30 ± 0.97	23.07 ± 0.60	33.97 ± 0.74	45.35 ± 1.12
UMIST	72.84 ± 0.65		17.55 ± 0.34	24.42 ± 0.98	25.56 ± 1.10	29.24 ± 1.02	45.15 ± 0.89	55.15 ± 2.34	25.25 ± 0.65	62.42 ± 0.46	68.24 ± 1.48
COIL20	76.95 ± 0.89		24.33 ± 0.57	37.69 ± 1.13	40.95 ± 1.03	45.23 ± 0.91	60.11 ± 0.99	60.11 ± 0.99	37.68 ± 1.19	68.82 ± 0.64	69.05 ± 0.59
warpPIE10P	63.03 ± 0.94		26.40 ± 0.78	38.05 ± 1.49	42.14 ± 2.19	42.54 ± 1.45	54.23 ± 2.62	55.07 ± 2.61	38.11 ± 1.64	37.21 ± 1.24	41.23 ± 0.87
Jaffe	87.99 ± 0.12		25.67 ± 0.35	44.96 ± 1.63	49.30 ± 1.38	72.62 ± 1.72	67.74 ± 1.04	77.07 ± 0.63	45.88 ± 1.85	68.04 ± 1.08	71.55 ± 0.42
USPS 200	60.15 ± 0.02		15.76 ± 0.08	27.58 ± 1.30	34.15 ± 1.50	49.10 ± 3.72	49.93 ± 0.11	53.58 ± 0.11	27.12 ± 0.68	49.34 ± 3.67	50.92 ± 5.05
att1	66.31 ± 0.73		18.15 ± 0.34	24.08 ± 0.78	26.22 ± 0.70	60.82 ± 0.69	42.03 ± 0.97	48.37 ± 1.05	22.12 ± 0.68	33.32 ± 0.50	36.26 ± 0.38
wdbc	97.59 ± 0.37		50.65 ± 0.03	60.31 ± 2.83	64.88 ± 4.18	72.87 ± 2.43	63.09 ± 4.88	71.57 ± 0.80	51.51 ± 3.07	84.12 ± 11.61	84.22 ± 11.71
Vehicle	96.60 ± 0.86		46.12 ± 0.05	36.74 ± 3.98	36.32 ± 3.89	61.71 ± 0.98	50.91 ± 2.71	62.94 ± 4.66	36.86 ± 4.03	86.51 ± 12.15	89.39 ± 9.63

As shown in Tables 4–8, the CRNMF_{FDH} algorithm achieved the highest clustering metrics across all nine datasets. This fully demonstrates the algorithm's outstanding clustering performance and robustness in clustering tasks. Specifically, when evaluating the quality metric SC for clustering structures, the CRNMF_{FDH} algorithm achieved an average SC value of 76.56%, significantly outperforming the other nine comparison algorithms. This demonstrates that its learned low-dimensional representations exhibit outstanding performance in terms of intra-class compactness and inter-class separability. The primary reasons for the proposed algorithm's superiority over other algorithms are as follows:

- (1) Compared to unsupervised algorithms, such as K-means, NMF, and GNMF, CRNMF_{FDH} demonstrates superior clustering performance. Specifically, on the UMIST dataset, its ACC, NMI, PUR, ARI, and SC metrics show average improvements of 46.22%, 40.53%, 23.86%, 45.61 and 55.29% over K-means, compared to NMF, it achieved average improvements of 50.52%, 32.76%, 44.02%, 51.93%, and 48.42%, respectively. Compared to GNMF, it achieved average improvements of 36.66%, 20.20%, 29.10%, 36.90%, and 43.6%, respectively. This is because CRNMF_{FDH} constructs Laplacian hypergraph regularization terms in both the data space and feature space, effectively preserving the multi-dimensional geometric structure information of both the data manifold and feature manifold. This overcomes the limitation of traditional graph methods, which can only model pairwise relationships, thereby enhancing the discriminative power of the low-dimensional representations.

- (2) Compared with hypergraph algorithms, such as HNMF and DHCF, CRNMFDH achieves superior average values for the ACC, NMI, PUR, ARI, and SC metrics. Specifically, it outperforms HNMF by 0.1724, 0.1352, 0.1577, 0.2106, and 0.1776, respectively, and surpasses DHCF by 0.1509, 0.1205, 0.1340, 0.1745, and 0.1837, respectively. The remarkable performance of CRNMFDH stems from its approach of associating label information with basis vectors through the construction of an indicator matrix. This constraint compels the learned basis vectors to represent instances of the same class more effectively while representing those of different classes less effectively, thereby enhancing clustering performance.
- (3) Compared to semi-supervised algorithms, such as CNMF, CDNMF, and MCDNMF, CRNMFDH achieves higher average ACC values by 18.35%, 13.82%, and 8.92%, respectively. The average NMI improved by 19.97%, 16.24%, and 2.31%, respectively. The average PUR increased by 16.70%, 13.47%, and 8.00%, respectively. The average ARI rose by 24.01%, 25.38%, and 14.68%, respectively. The average improvements in SC values were 0.3828, 0.2388, and 0.1476, respectively. This is because CRNMFDH employs correntropy to reconstruct errors. Within the objective function, correntropy assigns smaller weights to dissimilar sample points, thereby reducing the impact of outliers while enhancing the method's robustness.

4.4. Statistical Significance Test

This section employs the Friedman test, augmented by the Iman–Davenport correction, to verify whether the performance advantage of the CRNMFDH algorithm over the comparison methods is statistically significant. Following a significant result in the Friedman test, we proceeded with pairwise post hoc comparisons using the Wilcoxon signed-rank test (one-tailed, right-sided) to determine whether the CRNMFDH algorithm was significantly superior to each of the competitor algorithms. To control the family-wise error rate (FWER) arising from multiple comparisons, the p -values from all pairwise tests were adjusted using Holm's correction. The detailed results of these statistical analyses are summarized in Tables 9 and 10. The bolded numbers indicate the optimal results achieved by the corresponding algorithm under that metric.

Table 9. Average rankings from Friedman test.

Algorithm Dataset	CRNMFDH	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCC-NMF	DHCF	MCDNMF
ACC	1.00	8.11	9.00	6.33	5.89	4.67	5.22	8.44	4.00	2.33
NMI	1.33	6.33	8.78	7.11	5.67	5.67	5.33	7.57	4.22	3.00
PUR	1.00	8.44	8.89	6.33	5.56	5.00	5.33	8.11	3.89	2.44
ARI	1.00	8.11	9.11	6.22	4.89	5.67	5.33	7.89	3.56	3.22
SC	1.00	9.68	8.44	7.00	4.67	4.83	4.67	8.33	3.11	3.28

Table 9 results indicate that CRNMFDH achieved the lowest average ranking across all five clustering evaluation metrics, with corrected p -values below 0.001 for every metric. This strongly rejects the null hypothesis that “all algorithms exhibit no significant performance differences,” demonstrating that the proposed CRNMFDH algorithm possesses statistically significant performance advantages over the comparison methods. The clustering performance of MCDNMF and DHCF is inferior to that of the CRNMFDH algorithm, while traditional K-means clustering and basic NMF methods rank relatively lower. The statistical results further confirm that CRNMFDH achieves significant and consistent performance improvements across multiple evaluation metrics by integrating dual-hypergraph regularization, label information guidance, and a correntropy-based robustness

mechanism. This approach not only enhances the structural preservation capability of feature representations but also improves the robustness of the clustering process against noise and outliers, thereby demonstrating outstanding discriminative power and stability in complex high-dimensional data clustering tasks.

Table 10. Results of Wilcoxon signed-rank right-tailed test with Holm correction.

Algorithm Dataset	K-Means	NMF	CNMF	GNMF	CDNMF	HNMF	MCC-NMF	DHCF	MCDNMF
ACC	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.01
NMI	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.02
PUR	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.01
ARI	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.01
SC	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.01

The statistically corrected test results are presented in Table 10. At a significance level of $\alpha = 0.05$, CRNMFHDH demonstrated a significant superiority in all 45 pairwise comparisons, with all adjusted p -values falling below 0.05. Specifically, CRNMFHDH achieved statistically significant improvements over all competing algorithms across the ACC, NMI, PUR, ARI, and SC metrics. The results indicate that even after strictly controlling the family-wise error rate (FWER) to ensure more conservative statistical inference, CRNMFHDH maintains a statistically significant advantage across all comparisons. These findings further substantiate that—through the integration of dual-hypergraph regularization, label-information guidance, and the correntropy-based robust mechanism—CRNMFHDH delivers robust and reliable performance gains across multiple clustering evaluation dimensions, rather than merely reflecting random statistical fluctuations. Consequently, the results provide strong statistical support for the robustness and discriminative power of CRNMFHDH in practical applications.

4.5. Ablation Experiments

To systematically evaluate the effectiveness of each regularization term in the CRNMFHDH algorithm, ablation experiments were conducted in this section. The names of the six degenerate variants of CRNMFHDH and their objective functions are listed in Table 11. The experiments also included MCCNMF without regularization as the baseline comparison algorithm, whose ACC, NMI, PUR, and ARI metrics are presented in Tables 4–7. All degenerate variants were evaluated on nine public datasets. Their clustering performance metrics are summarized in Tables 12–15, with optimal results highlighted in bold.

Table 11. Degraded objective functions of CRNMFHDH.

No.	Algorithm Name	Objective Function After CRNMFHDH Degradation
1	CRNMFHDH-1	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \alpha \text{Tr}(V^T L_v V)$
2	CRNMFHDH-2	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \beta (U^T L_u U)$
3	CRNMFHDH-3	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \gamma \text{Tr}(YV^T)$
4	CRNMFHDH-4	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \alpha \text{Tr}(V^T L_v V) - \beta (U^T L_u U)$

5	CRNMFHDH-5	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \alpha \text{Tr}(V^T L_v V) - \gamma \text{Tr}(YV^T)$
6	CRNMFHDH-6	$\max_{U, V \geq 0} \sum_{i=1}^m g \left(\sqrt{\sum_{j=1}^n (X_{ij} - (UV^T)_{ij})^2} \right) - \beta (U^T L_u U) - \gamma \text{Tr}(YV^T)$

Table 12. ACC values in the CRNMFHDH ablation experiment (%).

Algorithm Dataset	CRNMFHDH-1	CRNMFHDH-2	CRNMFHDH-3	CRNMFHDH-4	CRNMFHDH-5	CRNMFHDH-6	CRNMFHDH
Semeion	57.11 ± 3.11	52.32 ± 4.84	74.81 ± 1.88	67.19 ± 4.79	75.83 ± 0.63	75.18 ± 0.42	85.62 ± 0.10
UMIST	63.28 ± 3.14	35.55 ± 1.85	59.30 ± 1.92	65.34 ± 3.45	70.90 ± 1.02	68.92 ± 1.34	89.79 ± 1.99
COIL20	77.63 ± 1.24	64.11 ± 2.77	79.15 ± 1.39	82.31 ± 1.69	80.19 ± 1.64	79.20 ± 1.18	91.22 ± 1.86
warpPIE10P	48.43 ± 2.30	47.33 ± 2.53	66.48 ± 2.43	50.76 ± 2.00	69.50 ± 2.62	68.74 ± 2.97	75.17 ± 2.97
Jaffe	96.78 ± 0.31	92.53 ± 4.77	95.49 ± 1.92	97.09 ± 0.19	98.52 ± 0.17	96.00 ± 2.64	98.85 ± 0.23
USPS 200	71.14 ± 4.83	59.70 ± 4.98	80.52 ± 0.17	72.83 ± 1.20	87.29 ± 0.04	81.43 ± 0.06	92.40 ± 0.03
att1	81.25 ± 1.20	72.91 ± 1.84	75.05 ± 1.91	82.69 ± 1.69	83.52 ± 1.47	77.17 ± 2.01	84.55 ± 1.49
wdbc	90.22 ± 10.17	84.46 ± 8.95	82.89 ± 3.97	89.75 ± 0.93	92.98 ± 0.30	88.38 ± 7.24	93.48 ± 0.05
Vehicle	44.94 ± 0.13	38.81 ± 1.87	45.51 ± 2.63	46.98 ± 0.10	56.28 ± 0.10	46.25 ± 3.91	65.75 ± 0.57

Table 13. NMI values in the CRNMFHDH ablation experiment (%).

Algorithm Dataset	CRNMFHDH-1	CRNMFHDH-2	CRNMFHDH-3	CRNMFHDH-4	CRNMFHDH-5	CRNMFHDH-6	CRNMFHDH
Semeion	49.14 ± 1.74	44.63 ± 2.75	61.68 ± 1.59	61.87 ± 2.92	63.73 ± 0.54	62.14 ± 0.38	77.30 ± 0.18
UMIST	81.69 ± 1.56	54.10 ± 2.06	66.36 ± 1.41	82.82 ± 2.50	84.34 ± 0.87	80.25 ± 1.00	91.56 ± 0.34
COIL20	87.99 ± 1.25	74.09 ± 1.82	79.48 ± 1.02	88.90 ± 1.13	87.50 ± 0.54	79.90 ± 1.04	94.99 ± 0.53
warpPIE10P	52.43 ± 1.91	51.56 ± 2.42	57.83 ± 2.34	55.91 ± 4.22	61.64 ± 2.41	59.43 ± 3.03	75.88 ± 2.34
Jaffe	96.52 ± 0.54	93.36 ± 3.64	93.77 ± 2.00	97.06 ± 0.33	98.06 ± 0.29	94.01 ± 2.28	98.25 ± 0.41
USPS 200	66.14 ± 4.83	54.49 ± 8.23	67.38 ± 0.19	71.37 ± 1.26	76.65 ± 0.07	73.97 ± 0.07	84.50 ± 0.06
att1	90.82 ± 0.62	85.05 ± 0.93	87.44 ± 0.95	90.98 ± 0.71	91.04 ± 0.97	89.23 ± 1.14	91.15 ± 0.70
wdbc	51.40 ± 20.65	41.96 ± 11.09	33.99 ± 6.53	52.65 ± 3.34	63.47 ± 0.54	47.63 ± 10.56	65.95 ± 0.30
Vehicle	19.87 ± 0.07	14.86 ± 2.07	18.82 ± 2.24	21.93 ± 0.04	28.92 ± 0.34	23.31 ± 3.73	41.70 ± 0.85

Table 14. PUR values in the CRNMFHDH ablation experiment (%).

Algorithm Dataset	CRNMFHDH-1	CRNMFHDH-2	CRNMFHDH-3	CRNMFHDH-4	CRNMFHDH-5	CRNMFHDH-6	CRNMFHDH
Semeion	58.41 ± 2.33	53.98 ± 4.07	74.81 ± 1.88	69.09 ± 3.86	75.83 ± 0.63	75.18 ± 0.42	85.62 ± 0.10
UMIST	70.70 ± 1.63	41.76 ± 2.13	59.43 ± 1.78	71.02 ± 2.56	73.95 ± 0.89	70.45 ± 1.30	89.75 ± 1.04
COIL20	80.80 ± 1.53	64.97 ± 2.63	74.28 ± 1.26	84.12 ± 1.51	81.23 ± 1.13	74.31 ± 1.13	91.85 ± 1.23
warpPIE10P	49.19 ± 2.04	48.38 ± 2.19	53.52 ± 2.38	50.55 ± 2.16	59.71 ± 2.66	55.50 ± 2.65	76.05 ± 2.70
Jaffe	96.78 ± 0.31	92.81 ± 4.39	95.49 ± 1.92	97.09 ± 0.19	98.52 ± 0.17	95.00 ± 2.64	98.85 ± 0.23
USPS 200	73.55 ± 3.22	59.79 ± 4.98	81.52 ± 0.17	77.09 ± 1.20	87.29 ± 0.04	82.43 ± 0.06	92.40 ± 0.03
att1	82.26 ± 1.27	75.14 ± 1.62	80.01 ± 1.74	83.17 ± 1.38	84.45 ± 1.26	81.65 ± 1.90	85.21 ± 1.29
wdbc	90.22 ± 10.17	83.27 ± 6.21	82.89 ± 3.97	89.75 ± 0.93	92.98 ± 0.30	85.38 ± 7.24	93.48 ± 0.05
Vehicle	44.94 ± 0.13	41.34 ± 1.29	45.87 ± 2.43	44.98 ± 0.10	59.28 ± 0.10	49.70 ± 3.77	65.75 ± 0.57

Table 15. ARI values in the CRNMFHDH ablation experiment (%).

Algorithm Dataset	CRNMFHDH-1	CRNMFHDH-2	CRNMFHDH-3	CRNMFHDH-4	CRNMFHDH-5	CRNMFHDH-6	CRNMFHDH
Semeion	36.23 ± 2.45	31.98 ± 3.56	53.82 ± 2.11	48.06 ± 4.13	55.57 ± 0.80	54.33 ± 0.65	71.75 ± 0.22
UMIST	54.36 ± 4.46	23.42 ± 1.93	39.40 ± 2.33	56.30 ± 3.89	59.31 ± 1.76	49.01 ± 1.91	80.15 ± 1.68

COIL20	74.42 ± 1.52	54.47 ± 2.51	60.49 ± 1.71	77.63 ± 1.77	72.30 ± 2.16	60.95 ± 1.62	86.03 ± 1.95
warpPIE10P	26.80 ± 1.56	25.60 ± 2.00	30.45 ± 1.82	28.68 ± 2.29	35.77 ± 3.10	33.78 ± 2.34	64.05 ± 2.70
Jaffe	93.50 ± 0.68	88.99 ± 6.75	90.55 ± 3.56	94.16 ± 0.41	96.79 ± 0.37	91.07 ± 4.36	97.44 ± 0.52
USPS 200	55.40 ± 3.15	40.92 ± 5.67	61.90 ± 0.25	60.81 ± 1.70	67.73 ± 0.07	62.08 ± 0.08	83.85 ± 0.06
att1	71.56 ± 1.70	63.56 ± 2.15	69.47 ± 4.06	72.58 ± 1.82	72.21 ± 3.68	71.61 ± 5.20	72.59 ± 3.28
wdbc	51.20 ± 23.40	41.12 ± 12.69	43.67 ± 8.52	62.83 ± 2.93	73.65 ± 1.05	44.08 ± 15.51	75.39 ± 0.18
Vehicle	13.19 ± 0.07	11.27 ± 1.40	10.95 ± 2.29	14.98 ± 0.10	25.56 ± 0.13	19.10 ± 3.23	36.17 ± 0.90

The results presented in Tables 12–15 validate the effectiveness of the CRNMFDH algorithm, which achieves optimal values for ACC, NMI, PUR, and ARI metrics across all nine datasets.

- (1) Both CRNMFDH-1 and CRNMFDH-3 exhibit significant improvements in clustering performance compared to MCCNMF. Specifically, they achieve average enhancements of 9.59% and 12.74% in the ACC metric, 7.20% and 3.95% in the NMI metric, 8.88% and 8.99% in the PUR metric, and 9.85% and 8.08% in the ARI metric, respectively. Although CRNMFDH-2 and MCCNMF demonstrate comparable performance on certain datasets due to limited feature discrimination between categories, the CRNMFDH-2 model, which is further optimized by introducing a class-driven constraint, still shows superior overall average performance compared to CRNMFDH-3. This indicates that the combined introduction of sample hypergraph regularization, feature hypergraph regularization, and class-driven constraint can effectively enhance the model's clustering discrimination capability and structural representation consistency, thereby systematically improving clustering performance.
- (2) Compared to CRNMFDH-4, CRNMFDH achieves average improvements of 0.1354, 0.1086, 0.1245, and 0.1682 in ACC, NMI, PUR, and ARI metrics, respectively, across the 9 datasets. When Compared to CRNMFDH-5, the average improvements in these four metrics are 0.0686, 0.0732, 0.0708, and 0.1206, respectively. In comparison with CRNMFDH-6, significant average increases of 0.1061, 0.1238, 0.1215, and 0.2016 are observed in the ACC, NMI, PUR, and ARI metrics, respectively. These consistent and significant results fully demonstrate that integrating correntropy metrics with feature hypergraph regularization, sample hypergraph regularization, and a class-driven constraint into a unified NMF framework can effectively synergistically enhance the model's structural representation capability and clustering performance.

4.6. Clustering Visualization Analysis

To systematically evaluate the clustering performance of the CRNMFDH algorithm, a comparative visualization analysis was conducted using the t-SNE method. Taking NMF as the baseline algorithm, six improved NMF clustering algorithms with relatively superior performance are selected for performance validation on the warpPIE10P, Se-meion, and USPS datasets respectively. In the experimental design, t-SNE was first employed to project the obtained low-dimensional data onto a two-dimensional plane. Subsequently, to eliminate random errors, each algorithm underwent 20 independent replicate experiments under optimal parameter configurations. The mean matrix was ultimately used as the visualization input data. The results are shown in Figures 3–5, where the “+” represents a sample point and the square represents the centroid. Figures 3–5 demonstrate that during the clustering process, CRNMFDH exhibits outstanding performance in terms of intra-class compactness and inter-class separation. This indicates that the CRNMFDH algorithm can more effectively separate data of different categories in the feature subspace, thereby enhancing the data's distinguishability. Notably, even on the USPS dataset, which contains noise, the cluster structure of the clustering results obtained by the CRNMFDH algorithm does not exhibit significant mixing or outliers. This visually

validates the robustness of the proposed algorithm and its effectiveness in maintaining intra-cluster compactness. Therefore, for high-dimensional data clustering problems, CRNMFHDH can be employed for dimensionality reduction preprocessing, followed by K-means clustering on the resulting low-dimensional representations, thereby significantly improving the final clustering accuracy.

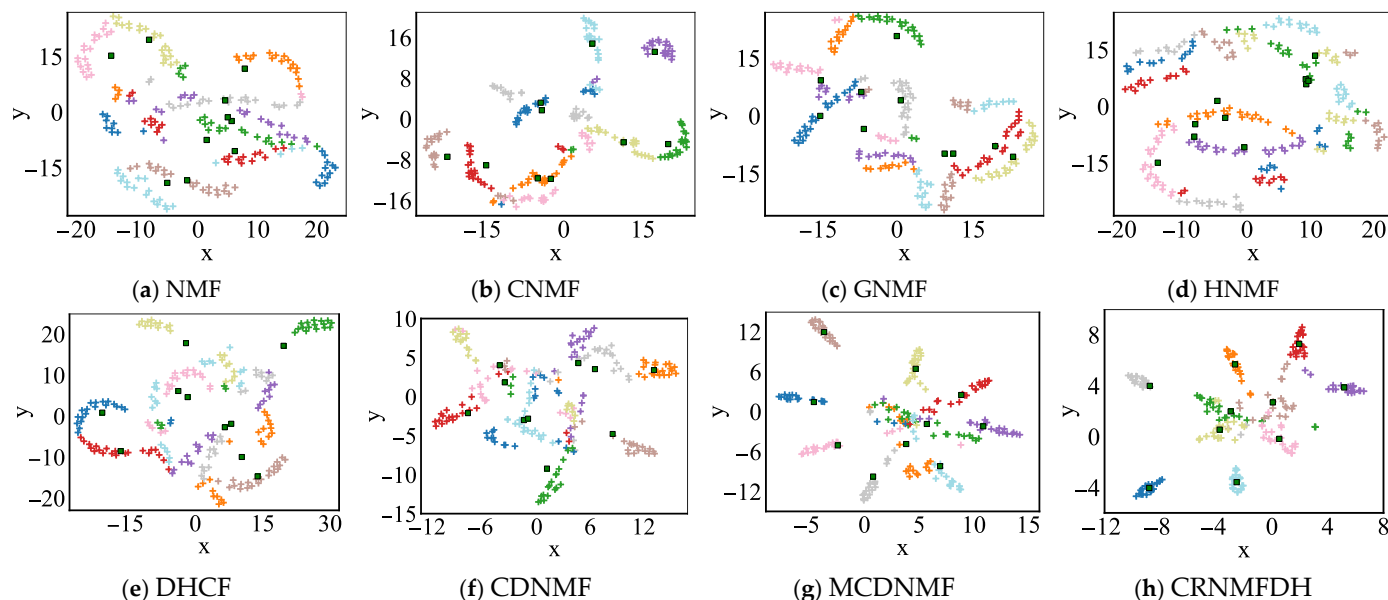


Figure 3. Low-rank representations of warpPIE10P learned by 9 algorithms.

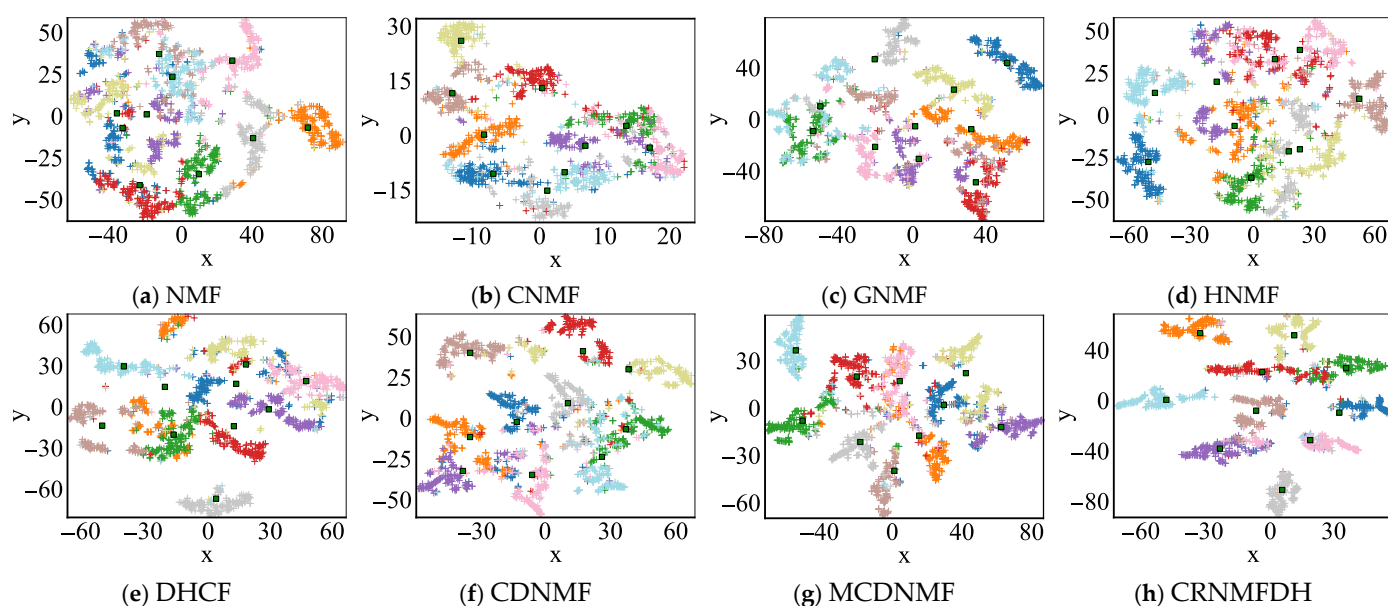
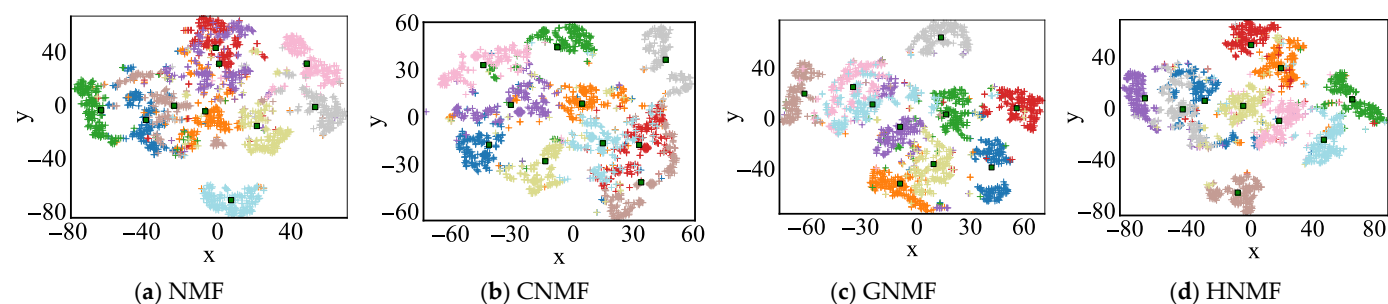


Figure 4. Low-rank representations of Semeion learned by 9 algorithms.



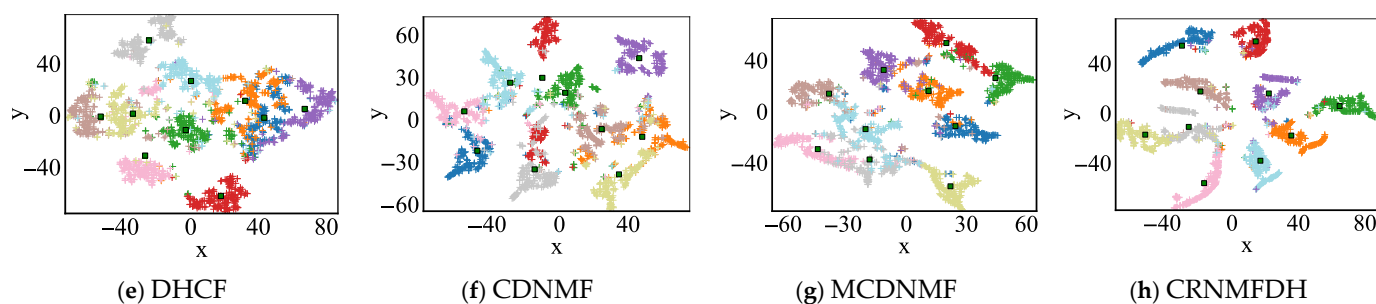


Figure 5. Low-rank representations of USPS learned by 9 algorithms.

4.7. Impact of Label Information

The proposed CRNMFHDH is a semi-supervised learning algorithm whose performance is closely tied to the volume of labeled data. To evaluate the impact of labeled data volume on algorithm performance, we conducted comparative experiments across nine datasets: for each category within each dataset, we randomly sampled labeled data at ratios of 10%, 20%, 30%, 40%, and 50%, with the remaining data serving as unlabeled data. To ensure statistical significance of the experimental results, each algorithm was independently run 20 times under each label ratio setting. The final average performance metric values and their standard deviations were aggregated to produce the results shown in Figures 6–14.

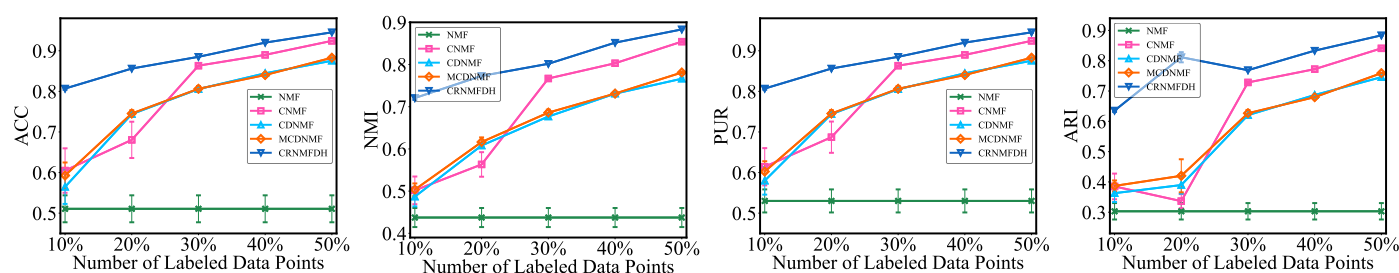


Figure 6. Semi-supervised clustering performance versus label ratio on Semeion.

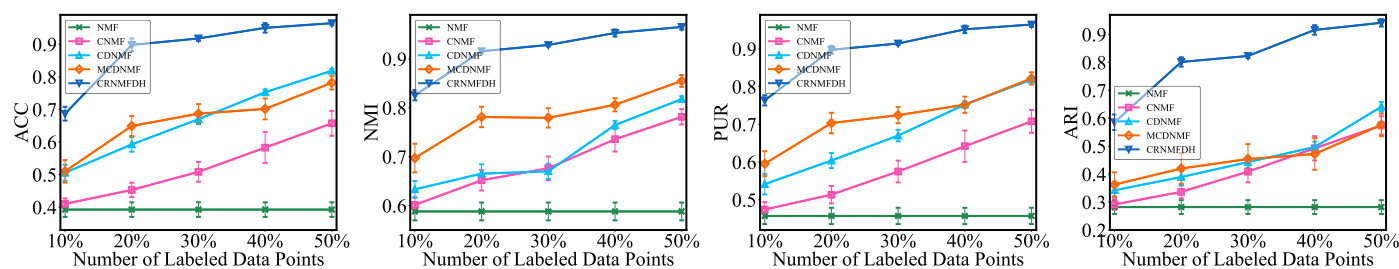


Figure 7. Semi-supervised clustering performance versus label ratio on UMIST.

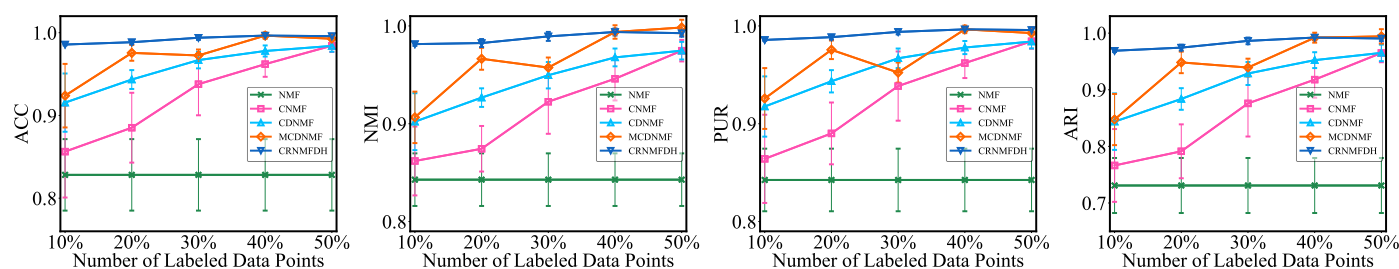


Figure 8. Semi-supervised clustering performance versus label ratio on Jaffe.

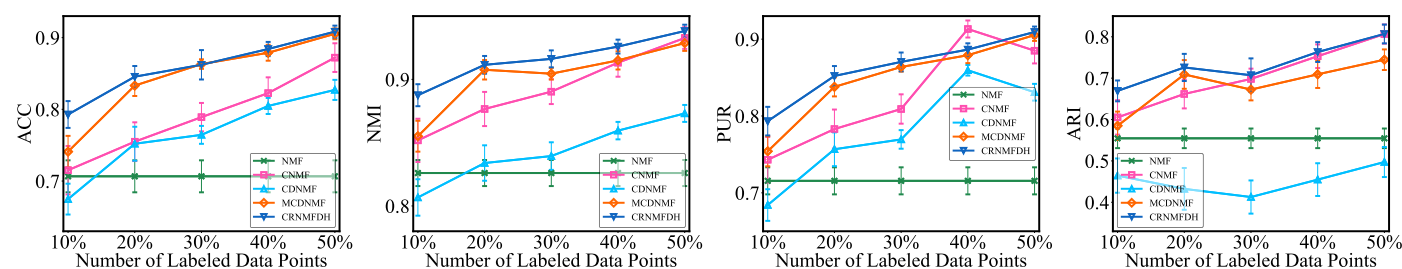


Figure 9. Semi-supervised clustering performance versus label ratio on att1.

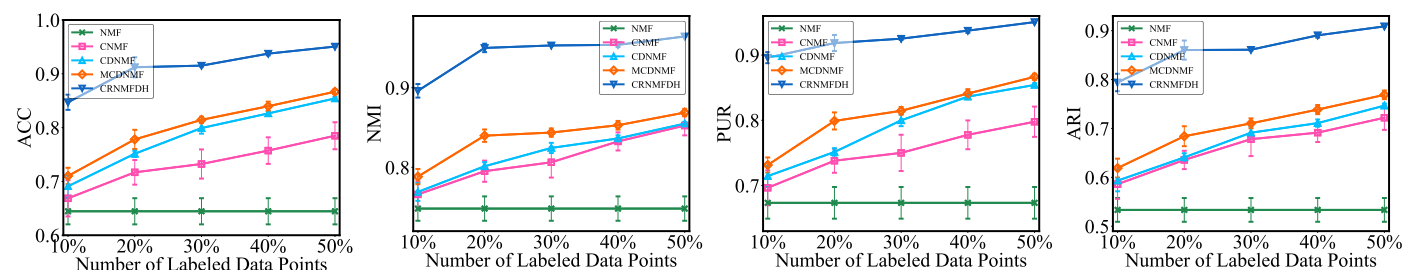


Figure 10. Semi-supervised clustering performance versus label ratio on COIL20.

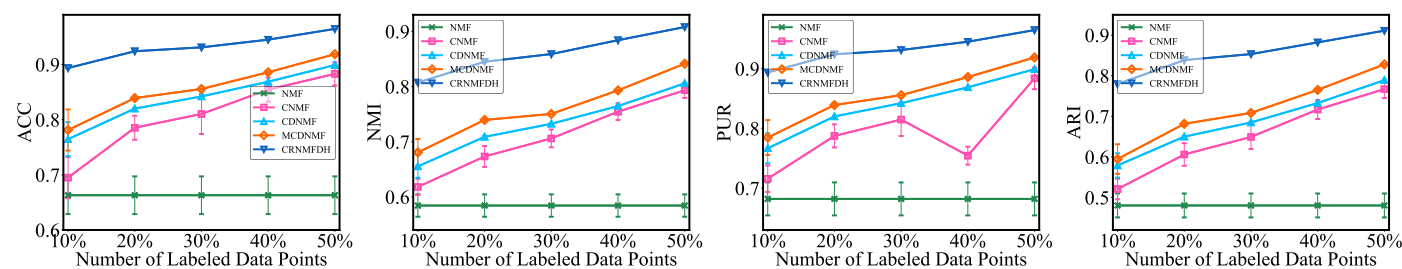


Figure 11. Semi-supervised clustering performance versus label ratio on USPS.

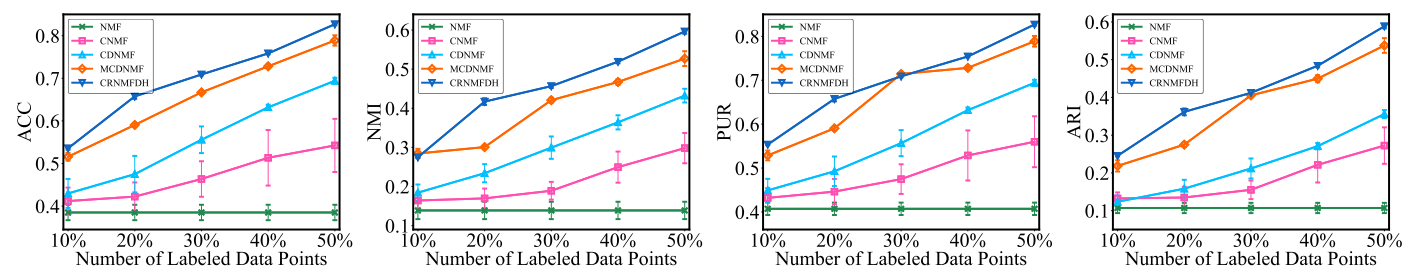


Figure 12. Semi-supervised clustering performance versus label ratio on Vehicle.

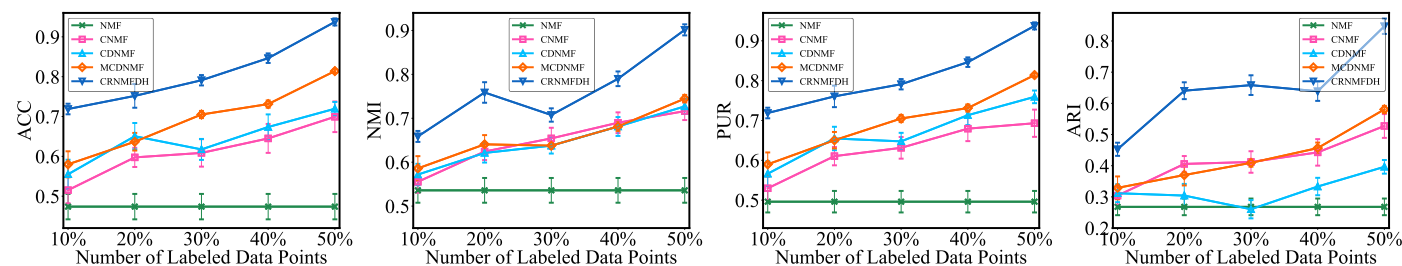


Figure 13. Semi-supervised clustering performance versus label ratio on warpPIE10P.

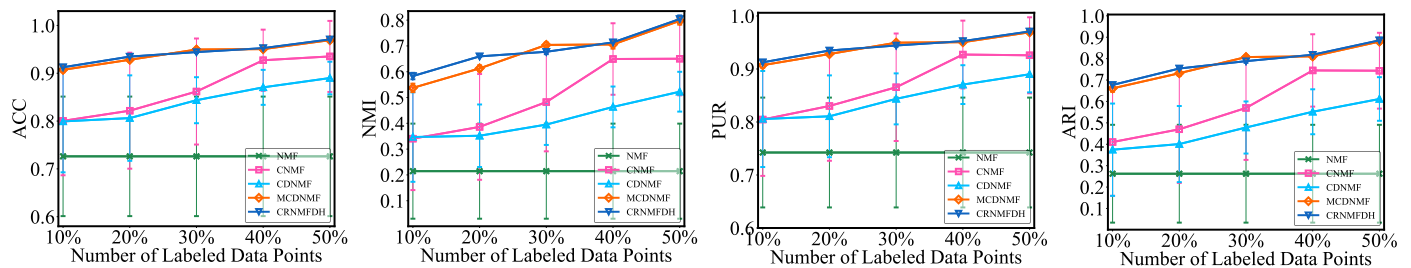


Figure 14. Semi-supervised clustering performance versus label ratio on wdbc.

As illustrated in Figures 6–14, with the gradual increase in the proportion of labeled data, the clustering performance of the four semi-supervised algorithms (CNMF, CDNMF, MCDNMF, and CRNMFHD) systematically improved across the nine datasets. In contrast, the clustering performance of unsupervised NMF methods does not change with an increase in the number of labeled samples. Notably, even with limited labeled samples, CRNMFHD consistently demonstrated significantly superior clustering performance compared to other methods across all datasets. The results indicate that incorporating appropriate structured label prior information into the model can effectively enhance the discriminative performance and generalization capability of clustering tasks.

4.8. Robustness Analysis

To validate the robustness of the CRNMFHD algorithm against noise, we conducted experiments using the Jaffe facial image dataset. We added Gaussian noise (0, 0.02, 0.04, 0.06, 0.08, 0.1) and salt-and-pepper noise (0, 0.05, 0.1, 0.15, 0.2, 0.25) of varying intensities to the images to simulate pixel damage. The noisy images are shown in Figures 15 and 16. Using NMF as the baseline algorithm, we selected six algorithms with superior clustering performance: CNMF, CDNMF, MCCNMF, DHCF, MCDNMF, and CRNMFHD. Each algorithm was run independently 20 times, with results shown in Figures 17 and 18.

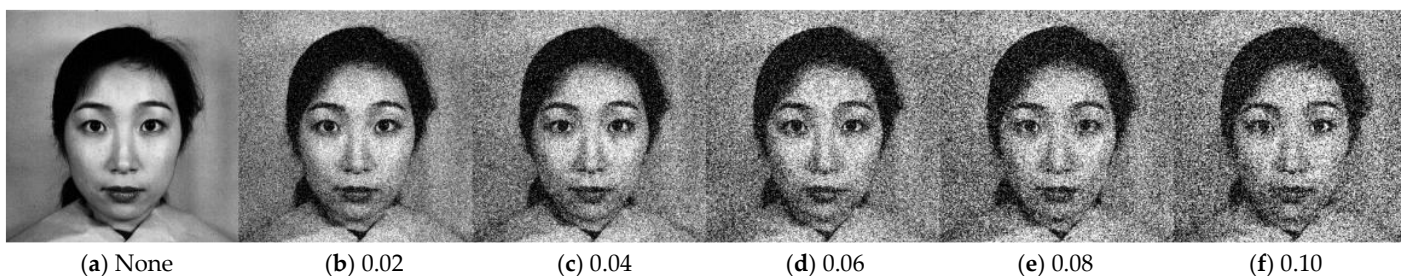


Figure 15. Partial Jaffe images corrupted by Gaussian noise at different variances.

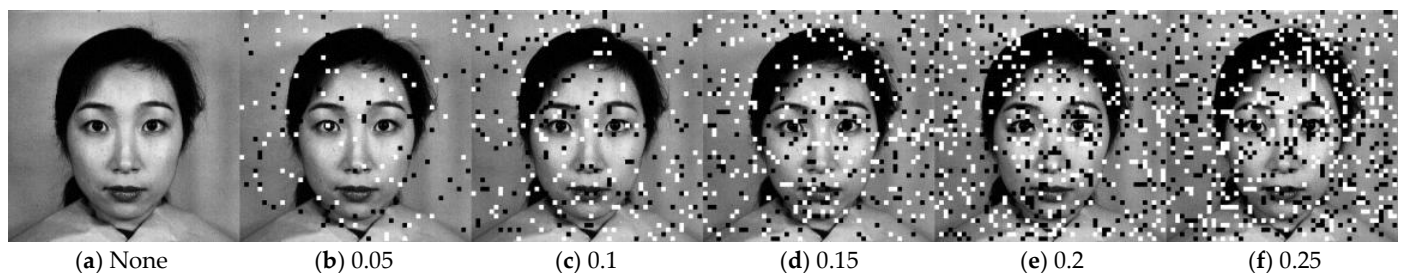


Figure 16. Partial Jaffe images corrupted by salt-and-pepper noise at different variances.

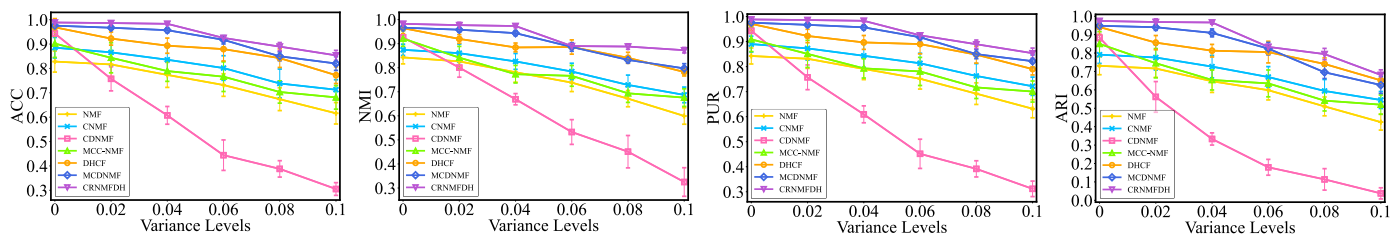


Figure 17. Clustering performance vs. Gaussian-noise level on Jaffe for seven algorithms.

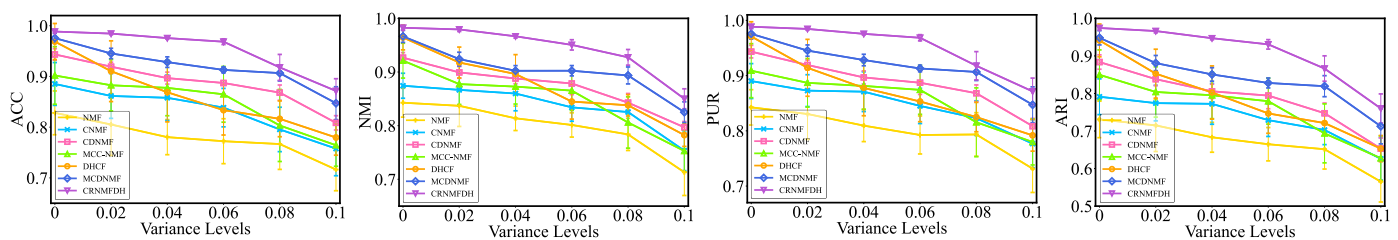


Figure 18. Clustering performance vs. salt-and-pepper noise level on Jaffe for seven algorithms.

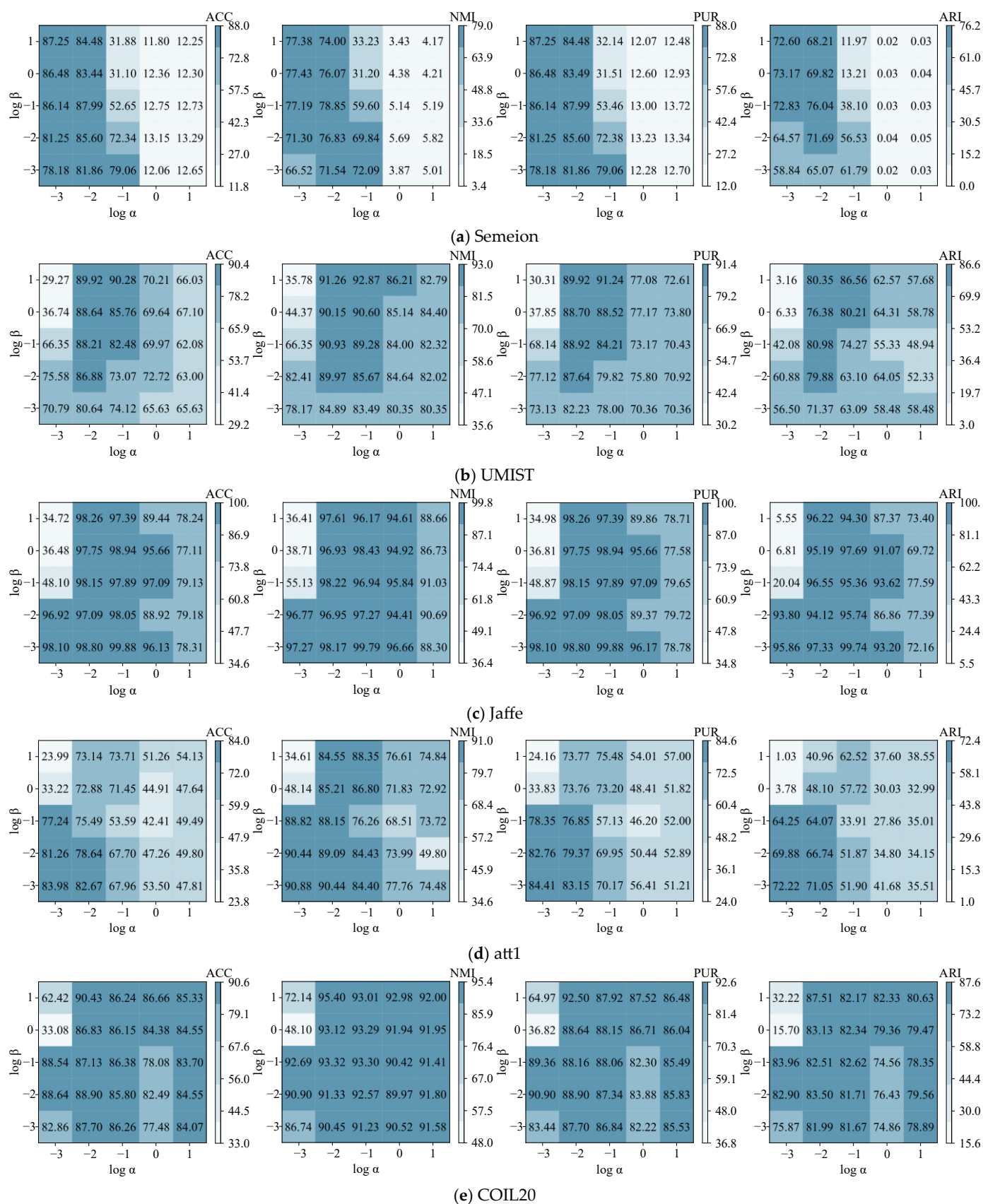
Experimental results indicate that under controlled noise interference conditions, the average clustering performance of all seven algorithms exhibited a significant decline. Noise interference degrades the accuracy of matrix decomposition in these algorithms, thereby affecting clustering performance. Compared to the baseline algorithm, the CRNMFDH algorithm consistently produces superior clustering results across all noise scenarios. This is because correntropy reconstructs errors through a Gaussian kernel function, making the model insensitive to noise and outliers. Furthermore, CRNMFDH effectively leverages existing label information in the data, making the resulting basis matrix and coefficient matrix more discriminative. Figures 17 and 18 demonstrate that although the clustering performance of the CRNMFDH algorithm declines noticeably with increasing noise levels, it remains superior to other algorithms. In summary, CRNMFDH exhibits strong robustness and discriminative power, effectively mitigating the interference of Gaussian and non-Gaussian noise on clustering results.

4.9. Parameter Sensitivity Analysis

This section discusses the impact of parameters α , β , and γ on clustering performance to evaluate their influence on the CRNMFDH algorithm.

4.9.1. Parameters α and β Sensitivity Analysis

To analyze the impact of parameters α and β on the CRNMFDH algorithm, we randomly sampled 20% of data from each category as labeled data. The number of nearest neighbors k_1 and k_2 for both the sample hypergraph and feature hypergraph were uniformly set to 5, while the class-driven term parameter γ was empirically set to 100. The values for the sample hypergraph regularization parameter α and the feature hypergraph regularization parameter β were both set within the range of $\{10^f : f \in \{-3, -2, -1, 0, 1\}\}$. Figure 19 presents the average clustering results of the CRNMFDH algorithm across nine datasets, obtained from twenty independent repetitions of experiments under different parameter combinations $\log \alpha$ and $\log \beta$. Darker colors in the figure indicate superior clustering performance of CRNMFDH under that parameter combination.



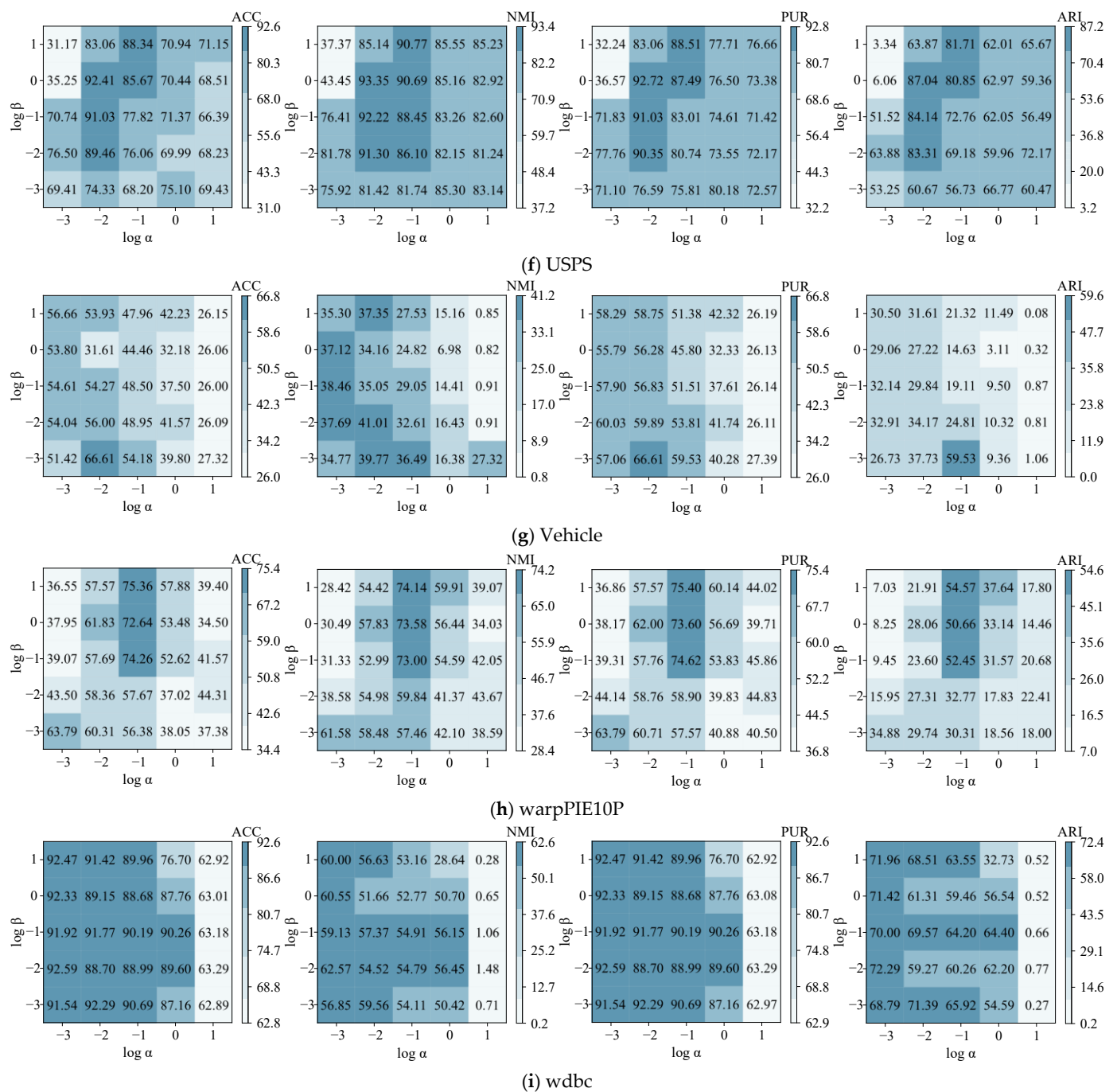


Figure 19. Parameter sensitivity analysis of CRNMFHD w.r.t. α and β on various datasets.

Figure 19 shows that, with other parameters held constant, the clustering performance of the CRNMFHD algorithm on the Semeion, UMIST, Jaffe, att1, COIL20, and wdbc datasets is insensitive to variations in parameter combinations α and β . The algorithm achieves good performance with smaller parameter combinations α and β . However, on the USPS, Vehicle, and warpPIE10P datasets, it is relatively sensitive to changes in parameters α and β . Experiments reveal that when $\alpha = 0.01$ and $\beta = 1$, the CRNMFHD algorithm exhibits good average clustering performance across all 9 datasets. Therefore, when processing new datasets, the parameter combination $\alpha = 0.01$, $\beta = 1$, $\gamma = 100$ should be prioritized.

4.9.2. Parameter γ Sensitivity Analysis

To determine the optimal parameter γ for the CRNMFDH algorithm, other parameters were fixed with $\alpha = 0.01$ and $\beta = 1$, and the hypergraph neighbor count was uniformly set to 5. Parameter γ was varied within the range $\{10^f : f \in \{-2, -1, 0, 1, 2, 3\}\}$. For each parameter value, each algorithm was run independently 20 times, and the mean and standard deviation of its clustering metrics were calculated. The results are shown in Figure 20, where the horizontal axis represents the corresponding $\log \gamma$ value.

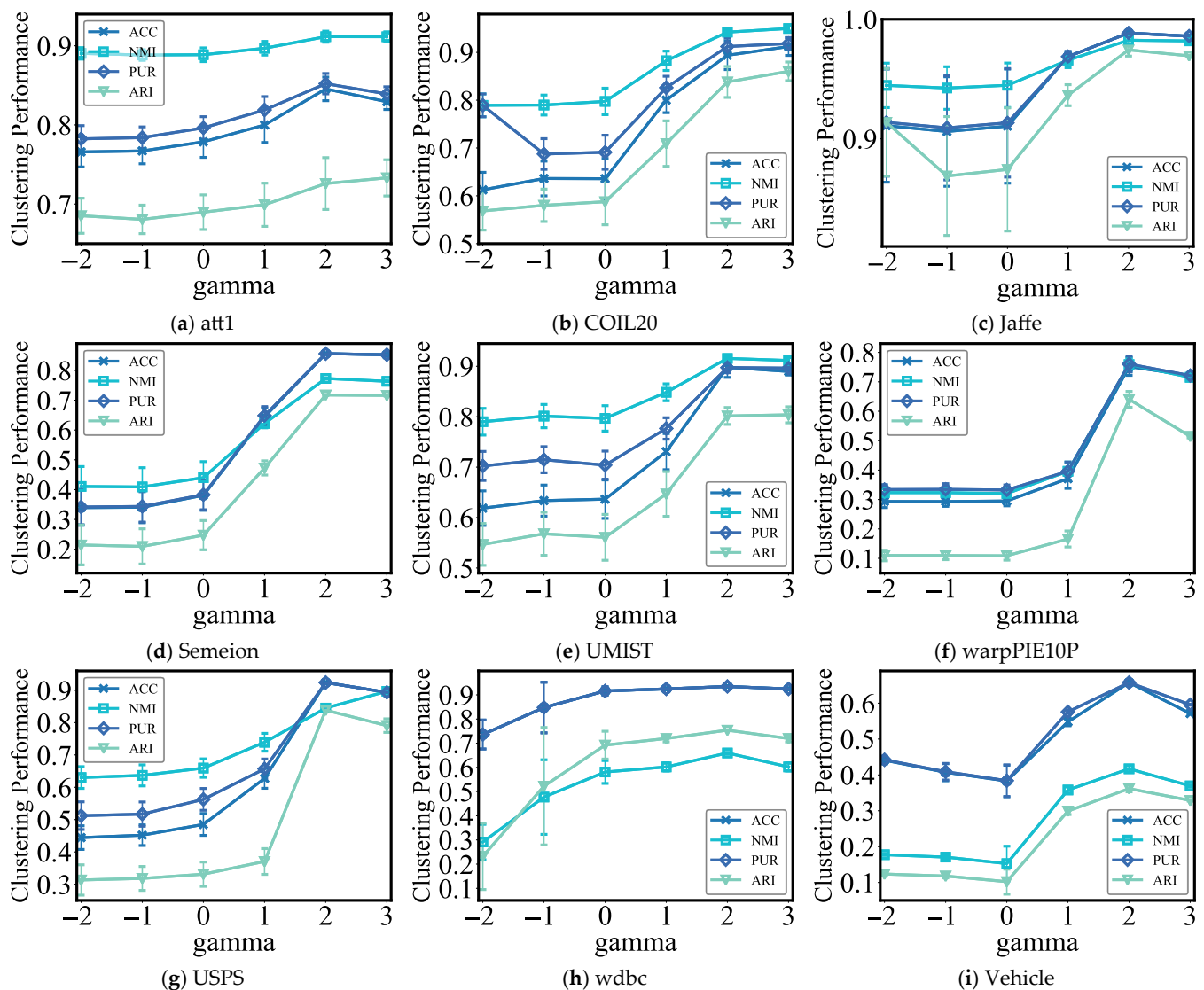


Figure 20. Parameter sensitivity analysis of CRNMFDH w.r.t. γ on various datasets.

Figure 20 illustrates the clustering performance of the CRNMFDH algorithm across different datasets as a function of parameter γ . It can be observed that when γ is set to larger values, the clustering accuracy of the CRNMFDH algorithm improves, while smaller values of γ yield poorer clustering accuracy. Therefore, the algorithm should select larger values of γ . When $\gamma = 100$, CRNMFDH achieves good clustering performance across all nine datasets. Consequently, when applying the CRNMFDH algorithm to new data, the parameter combinations $\alpha = 0.01$, $\beta = 1$, and $\gamma = 100$ should be prioritized.

4.10. Empirical Convergence Analysis

To evaluate the convergence performance of the CRNMFDH algorithm, we conducted experiments using optimal parameter settings for each dataset. Figure 21 illustrates the convergence process of this algorithm across nine datasets. The horizontal axis represents the iteration count, while the vertical axis corresponds to the normalized value of the objective function. Observation shows that the objective function value of CRNMFDH continuously decreases across all datasets as iterations increase, reaching a steady state within 150 iterations and ultimately converging to a similar level. This phenomenon aligns with the theoretical conclusions presented in Section 3.3, fully validating the effectiveness of the CRNMFDH update rule in terms of convergence efficiency and stability.

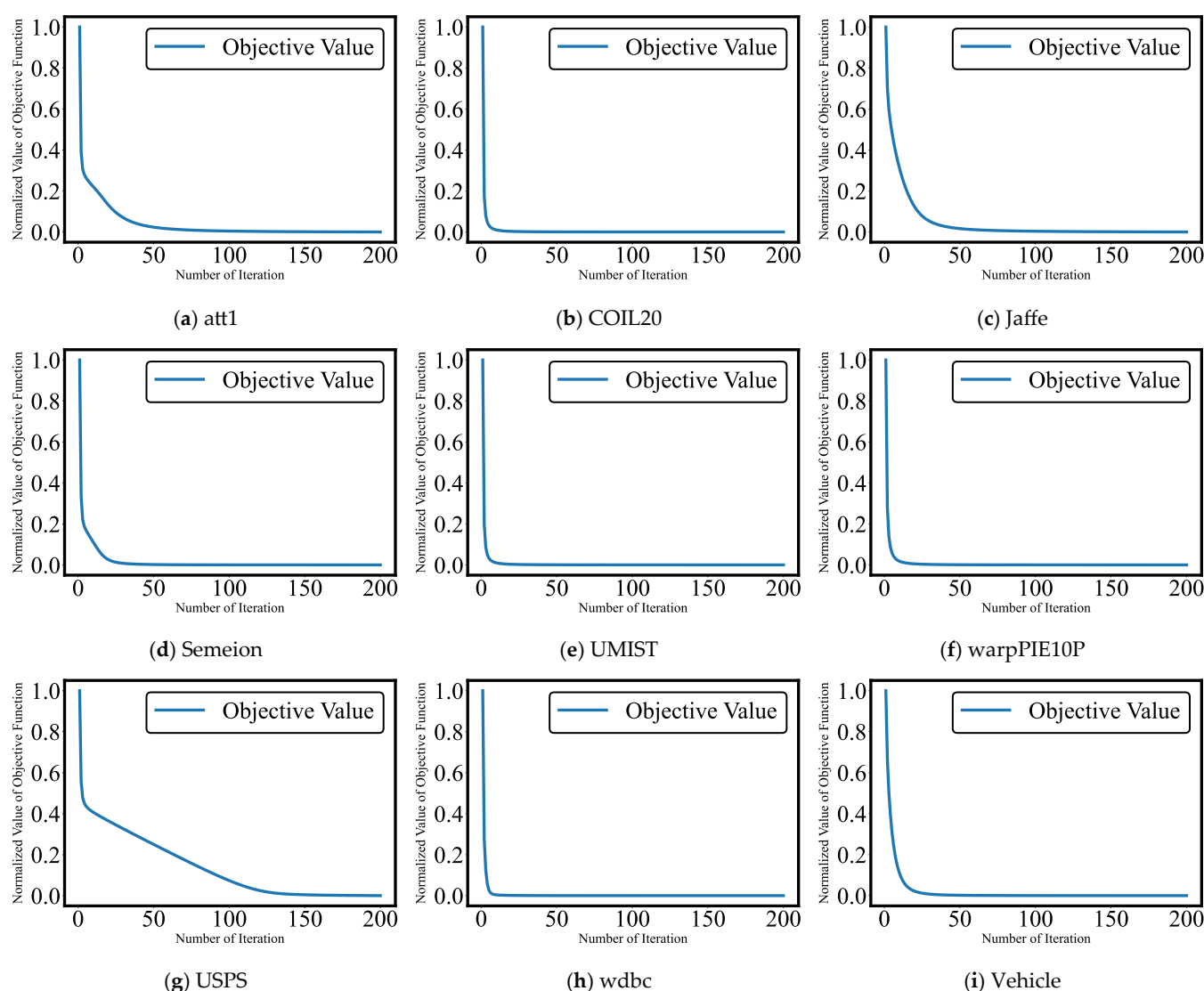


Figure 21. Convergence curves of CRNMFDH on various datasets.

5. Conclusions

In this paper, we propose a novel semi-supervised clustering algorithm, the class-driven robust NMF with dual-hypergraph regularization (CRNMFDH). To enhance model robustness, we adopt the MCC reconstruction error metric, effectively suppressing the impact of noise and outliers on the decomposition process. To fully capture the data's high-order geometric structure and local manifold relationships, we construct hypergraph

regularization terms in both the sample space and feature space. This design simultaneously preserves high-order neighbor relationships across both spaces, enhancing the model's ability to retain local geometric structures. The symmetry between sample and feature hypergraph regularization terms helps balance the influence of samples and features during matrix decomposition. To enhance representational discriminability, we introduce a class-driven heterogeneity representation penalty term. This constraint leverages limited labeled information to ensure that representations of samples within the same class are as compact as possible, while representations of samples from different classes are as separated as possible. This approach strengthens both the compactness of intra-class representations and the distinctiveness of inter-class representations. Furthermore, we elaborate on the objective function and iterative update rules of the CRNMFDH algorithm, providing a theoretical proof of its convergence. Extensive experimental results demonstrate that the CRNMFDH algorithm significantly outperforms competing algorithms in both clustering evaluation metrics and noise robustness, validating its effectiveness and robustness.

Although the CRNMFDH algorithm has demonstrated outstanding performance in clustering tasks, there are still two aspects that warrant further optimization and exploration: Firstly, the algorithm implementation involves a relatively large number of parameters that need to be set, and the process of parameter tuning is rather cumbersome. For instance, the hypergraph neighborhood size, as a critical parameter influencing the algorithm's performance, still relies on manual setting. In the future, it is necessary to develop a more efficient mechanism for adaptive parameter selection. Secondly, the current model is based on the NMF framework, and its information representation capability is constrained by the single-view data analysis paradigm. To enhance the algorithm's ability to handle complex-structured data, subsequent research could explore extending this algorithm framework to a multi-view learning paradigm, thereby enabling the information fusion and joint analysis of multi-source heterogeneous data.

Author Contributions: Conceptualization, methodology, resources, writing—review and editing: H.G. and G.Z.; validation, supervision, project administration, funding acquisition: H.G.; writing—original draft, software, formal analysis, data curation, visualization: G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Social Science Fund of China (NO. 21BTJ042) and the National Key Statistical Science Research Project (NO. 2025LZ007); as well as by the Gansu Provincial Natural Science Foundation (NO. 23JRRA1186) and the Gansu Provincial Universities' Young Doctor Support Program (NO. 2025QB-058).

Data Availability Statement: This study used nine datasets, among which the Semeion, wdbc, and vehicle datasets are publicly available at (<http://archive.ics.uci.edu/ml>), accessed on 23 April 2025; the COIL20 and USPS datasets are publicly available for download at (<http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html>), accessed on 25 April 2025; Jaffe is available at (<http://www.kasrl.org/jaffe.html>), accessed on 25 April 2025; UMIST is available at (<http://www.kasrl.org/jaffe.html>), accessed on 25 April 2025; the warpPIE10P dataset is a subset of the PIE dataset, containing 210 samples from 10 categories, available at (<https://www.ri.cmu.edu/project/pie-database/>), accessed on 29 April 2025; the att1 dataset was created by AT&T Laboratory at the University of Cambridge, containing 400 facial images of 40 individuals, publicly available for download at (<https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>), accessed on 28 April 2025.

Acknowledgments: The authors wish to thank the editors and reviewers for their careful review and constructive suggestions, which have significantly improved the quality of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. society. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. <https://doi.org/10.2307/2346830>.
- Karypis, G.; Han, E.-H.; Kumar, V. Chameleon: Hierarchical clustering using dynamic modeling. *Computer* **1999**, *32*, 68–75. <https://doi.org/10.1109/2.781637>.
- Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. <https://doi.org/10.1007/s11222-007-9033-z>.
- Parsons, L.; Haque, E.; Liu, H. Subspace clustering for high dimensional data: A review. *ACM Sigkdd Explor. Newsl.* **2004**, *6*, 90–105. <https://doi.org/10.1145/1007730.1007731>.
- Lee, D.-D.; Seung, H.-S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. <https://doi.org/10.1038/44565>.
- Cai, D.; He, X.; Han, J.; Huang, T.S. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 1548–1560. <https://doi.org/10.1109/tpami.2010.231>.
- Chen, Y.; Qu, G.; Zhao, J. Orthogonal graph regularized non-negative matrix factorization under sparse constraints for clustering. *Expert Syst. Appl.* **2024**, *249*, 123797. <https://doi.org/10.1016/j.eswa.2024.123797>.
- Wan, M.; Cai, M.; Yang, G. Robust exponential graph regularization non-negative matrix factorization technology for feature extraction. *Mathematics* **2023**, *11*, 1716. <https://doi.org/10.3390/math11071716>.
- Shang, F.; Jiao, L.; Wang, F. Graph dual regularization non-negative matrix factorization for co-clustering. *Pattern Recognit.* **2012**, *45*, 2237–2250. <https://doi.org/10.1016/j.patcog.2011.12.015>.
- Kang, X.; Li, X.; Yao, H.; Li, D.; Jiang, B.; Peng, X.; Wu, T.; Qi, S.; Dong, L. Dynamic hypergraph neural networks based on key hyperedges. *Inf. Sci.* **2022**, *616*, 37–51. <https://doi.org/10.1016/j.ins.2022.10.006>.
- Zeng, K.; Yu, J.; Li, C.; You, J.; Jin, T. Image clustering by hyper-graph regularized non-negative matrix factorization. *Neurocomputing* **2014**, *138*, 209–217. <https://doi.org/10.1016/j.neucom.2014.01.043>.
- Ren, X.; Yang, Y. Adaptive hypergraph structure regularized semi-supervised non-negative matrix factorization for image clustering. *Neurocomputing* **2025**, *651*, 130895. <https://doi.org/10.1016/j.neucom.2025.130895>.
- Yu, J.; Pan, B.; Yu, S.; Leung, M.-F. Robust capped norm dual hyper-graph regularized non-negative matrix tri-factorization. *Math. Biosci. Eng.* **2023**, *20*, 12486–12509. <https://doi.org/10.3934/mbe.2023556>.
- Kong, D.; Ding, C.; Huang, H. Robust nonnegative matrix factorization using $L_{2,1}$ -norm. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; pp. 673–682. <https://doi.org/10.1145/2063576.2063676>.
- Wang, Q.; He, X.; Jiang, X.; Li, X. Robust bi-stochastic graph regularized matrix factorization for data clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 390–403. <https://doi.org/10.1109/tpami.2020.3007673>.
- Peng, S.; Ser, W.; Chen, B.; Sun, L.; Lin, Z. Correntropy based graph regularized concept factorization for clustering. *Neurocomputing* **2018**, *316*, 34–48. <https://doi.org/10.1016/j.neucom.2018.07.049>.
- Nikolova, M.; Ng, M.K. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.* **2005**, *27*, 937–966. <https://doi.org/10.1137/030600862>.
- Yu, N.; Wu, M.-J.; Liu, J.-X.; Zheng, C.-H.; Xu, Y. Correntropy-based hypergraph regularized NMF for clustering and feature selection on multi-cancer integrated data. *IEEE Trans. Cybern.* **2020**, *51*, 3952–3963. <https://doi.org/10.1109/tcyb.2020.3000799>.
- Shi, L.; Ma, X.; Xi, L.; Duan, Q.; Zhao, J. Rough set and ensemble learning based semi-supervised algorithm for text classification. *Expert Syst. Appl.* **2011**, *38*, 6300–6306. <https://doi.org/10.1016/j.eswa.2010.11.069>.
- Lu, Z.; Wang, L. Noise-robust semi-supervised learning via fast sparse coding. *Pattern Recognit.* **2015**, *48*, 605–612. <https://doi.org/10.1016/j.patcog.2014.08.019>.
- Belkin, M.; Niyogi, P.; Sindhiani, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.
- Li, H.; Zhang, J.; Shi, G.; Liu, J. Graph-based discriminative nonnegative matrix factorization with label information. *Neurocomputing* **2017**, *266*, 91–100. <https://doi.org/10.1016/j.neucom.2017.04.067>.
- Li, J.; Li, Y.; Li, C. Dual-graph-regularization constrained nonnegative matrix factorization with label discrimination for data clustering. *Mathematics* **2023**, *12*, 96. <https://doi.org/10.3390/math12010096>.
- Luo, M.; Li, S.; Tao, J.; Vladimirovich, P.P. Semi-supervised correntropy-based non-negative matrix factorization with hypergraph regularization. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 10037–10056. <https://doi.org/10.1007/s13042-024-02523-7>.

25. Zhou, H.; Yin, H.; Li, Y.; Chai, Y. Multiview clustering via exclusive non-negative subspace learning and constraint propagation. *Inf. Sci.* **2021**, *552*, 102–117. <https://doi.org/10.1016/j.ins.2020.11.037>.
26. Li, H.; Zhou, Y.; Zhao, P.; Wang, L.; Yu, C. Class-driven nonnegative matrix factorization with manifold regularization for data clustering. *Neurocomputing* **2024**, *592*, 127751. <https://doi.org/10.1016/j.neucom.2024.127751>.
27. Xiao, Y.-H.; Zhu, Z.-F.; Zhao, Y.; Wei, Y.-C. Class-driven non-negative matrix factorization for image representation. *J. Comput. Sci. Technol.* **2013**, *28*, 751–761. <https://doi.org/10.1007/s11390-013-1374-9>.
28. Luo, W.; Wu, Z.; Zhou, N. Hypergraph-based convex semi-supervised unconstrained symmetric matrix factorization for image clustering. *Inf. Sci.* **2024**, *680*, 121138. <https://doi.org/10.1016/j.ins.2024.121138>.
29. Xu, Y.; Lu, L.; Liu, Q.; Chen, Z. Hypergraph-Regularized L_p Smooth Nonnegative Matrix Factorization for Data Representation. *Mathematics* **2023**, *11*, 2821. <https://doi.org/10.3390/math11132821>.
30. Yang, B.; Wu, J.; Zhou, Y.; Zhang, X.; Lin, Z.; Nie, F.; Chen, B. Robust spectral embedded bilateral orthogonal concept factorization for clustering. *Pattern Recognit.* **2024**, *150*, 110308. <https://doi.org/10.1016/j.patcog.2024.110308>.
31. Wang, J.-Y.; Wang, X.; Gao, X. Non-negative matrix factorization by maximizing correntropy for cancer clustering. *BMC Bioinform.* **2013**, *14*, 107. <https://doi.org/10.1186/1471-2105-14-107>.
32. Dreves, A.; Facchinei, F.; Kanzow, C.; Sagratella, S. On the solution of the KKT conditions of generalized Nash equilibrium problems. *SIAM J. Optim.* **2011**, *21*, 1082–1108. <https://doi.org/10.1137/100817000>.
33. Cai, D.; He, X.; Wu, X.; Han, J. Non-negative matrix factorization on manifold. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 63–72. <https://doi.org/10.1109/icdm.2008.57>.
34. Ye, J.; Jin, Z. Dual Hypergraph Regularized Concept Factorization Algorithm for Data Representation. *Comput. Sci.* **2017**, *44*, 309–314.
35. Liu, H.; Wu, Z.; Li, X.; Cai, D.; Huang, T.S. Constrained nonnegative matrix factorization for image representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1299–1311. <https://doi.org/10.1109/tpami.2011.217>.
36. Zhu, L.J.; Ma, B.X.; Zhao, X.Q. Analysis of clustering validity based on silhouette coefficient. *J. Comput. Appl.* **2010**, *30*, 139–141+198.
37. Shahapure, K.-R.; Nicholas, C. Cluster quality analysis using silhouette score. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; pp. 747–748. <https://doi.org/10.1109/dsaa49011.2020.00096>.
38. Cai, H.; Liu, B.; Xiao, Y.; Lin, L. Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization. *Knowl.-Based Syst.* **2019**, *182*, 104798. <https://doi.org/10.1016/j.knosys.2019.06.006>.
39. Mirzal, A. A convergent algorithm for orthogonal nonnegative matrix factorization. *J. Comput. Appl. Math.* **2014**, *260*, 149–166. <https://doi.org/10.1016/j.cam.2013.09.022>.
40. Zhang, Y.; Lu, Z.; Wang, S. Unsupervised feature selection via transformed auto-encoder. *Knowl.-Based Syst.* **2021**, *215*, 106748. <https://doi.org/10.1016/j.knosys.2021.106748>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.