



LJMU Research Online

Shi, Y, Wan, Y, Wang, X and Li, H

Incorporation of Histogram Intersection and Semantic Information into Non-Negative Local Laplacian Sparse Coding for Image Classification

<http://researchonline.ljmu.ac.uk/id/eprint/25558/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Shi, Y, Wan, Y, Wang, X and Li, H (2025) Incorporation of Histogram Intersection and Semantic Information into Non-Negative Local Laplacian Sparse Coding for Image Classification. Mathematics, 13 (2).

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

Article

Incorporation of Histogram Intersection and Semantic Information into Non-Negative Local Laplacian Sparse Coding for Image Classification

Ying Shi ¹, Yuan Wan ^{2,*}, Xinjian Wang ³  and Huanhuan Li ^{4,*} ¹ Department of Basic Courses, Suzhou City University, Suzhou 215104, China; wzj105@suda.edu.cn² School of Science, Wuhan University of Technology, Wuhan 430070, China³ Navigation College, Dalian Maritime University, Dalian 116026, China; wangxinjian@dlnu.edu.cn⁴ Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK

* Correspondence: wanyuan@whut.edu.cn (Y.W.); h.li2@ljmu.ac.uk (H.L.)

Abstract: Traditional sparse coding has proven to be an effective method for image feature representation in recent years, yielding promising results in image classification. However, it faces several challenges, such as sensitivity to feature variations, code instability, and inadequate distance measures. Additionally, image representation and classification often operate independently, potentially resulting in the loss of semantic relationships. To address these issues, a new method is proposed, called Histogram intersection and Semantic information-based Non-negativity Local Laplacian Sparse Coding (HS-NLLSC) for image classification. This method integrates Non-negativity and Locality into Laplacian Sparse Coding (NLLSC) optimisation, enhancing coding stability and ensuring that similar features are encoded into similar codewords. In addition, histogram intersection is introduced to redefine the distance between feature vectors and codebooks, effectively preserving their similarity. By comprehensively considering both the processes of image representation and classification, more semantic information is retained, thereby leading to a more effective image representation. Finally, a multi-class linear Support Vector Machine (SVM) is employed for image classification. Experimental results on four standard and three maritime image datasets demonstrate superior performance compared to the previous six algorithms. Specifically, the classification accuracy of our approach improved by 5% to 19% compared to the previous six methods. This research provides valuable insights for various stakeholders in selecting the most suitable method for specific circumstances.

Keywords: semantic information; image representation; image classification; sparse coding; support vector machine

MSC: 90C30



Academic Editors: Ming Yang and Liqun Shan

Received: 18 December 2024

Revised: 6 January 2025

Accepted: 8 January 2025

Published: 10 January 2025

Citation: Shi, Y.; Wan, Y.; Wang, X.; Li, H. Incorporation of Histogram Intersection and Semantic Information into Non-Negative Local Laplacian Sparse Coding for Image Classification. *Mathematics* **2025**, *13*, 219. <https://doi.org/10.3390/math13020219>

Copyright: © 2025 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image classification is a pivotal component within the realm of computer vision [1–3], a field dedicated to extracting meaningful information from images, transforming this information into features, and encoding it for computer processing. This process allows computers to train, learn, and categorise images into various groups. In the era of the internet, characterised by widespread image accessibility, the diversity and volume of image categories have seen substantial growth. Consequently, the efficient organisation, analysis, and accurate classification and prediction of large volumes of image data have

become essential research pursuits in computer vision. The technology of image classification serves as an interdisciplinary field with diverse applications in sectors such as medical engineering [4–6], environmental monitoring [7,8], industrial manufacturing [9,10], and autonomous driving [11,12]. For instance, in the field of medical engineering, it can be employed to identify and categorise cells and tissue structures in biomedical images, thereby contributing to medical research and treatment. In the context of environmental monitoring, image classification aids in the monitoring of image data related to the atmosphere, water bodies, land, and other environmental elements. In industrial manufacturing, it is used for product quality control and defect detection on production lines, enhancing the level of automation in manufacturing processes. In the domain of autonomous driving, image classification is applied to recognise and understand elements like traffic signs, pedestrians, and vehicles on roads, facilitating vehicle decision-making and operations. Furthermore, within maritime traffic [13–15], image classification can be used for vessel identification, buoy recognition, maritime event monitoring, maritime boundary surveillance, and meteorological monitoring.

In recent years, deep learning technologies [16–18], particularly Convolutional Neural Network (CNN) [19], have achieved remarkable advancements in image classification tasks. Deep learning models, by learning to extract high-level features from raw pixel data, have led to groundbreaking progress in the accuracy and performance of image classification. Researchers such as Yu et al. took a comprehensive approach by integrating spectral–spatial features and extracting valuable information independently through two separate dense Convolutional Neural Networks (CNNs) [20]. They introduced a spatial–spectral dense CNN framework with a feedback attention mechanism, specially tailored for hyperspectral image classification. Ozkaraca et al. developed a new modular deep learning model to preserve the existing advantages of established transfer learning methods, including DenseNet, VGG16, and basic CNN architectures, while eliminating their limitations in the classification of Magnetic Resonance (MR) images [21]. Shamshad et al. investigated the applications of transformers in various medical image tasks such as segmentation, detection, classification, restoration, synthesis, registration, and clinical report generation [22]. They have developed taxonomies for each application, identified challenges specific to each, provided insights into solutions, and highlighted emerging trends. Building upon the attention mechanism of the transformer, Roy et al. introduced a new morphological transformer (morphFormer) [23]. This innovative approach integrates learnable spectral and spatial morphological networks, enhancing the interaction between structural and shape information in the hyperspectral image token and the CLS token. Zhou et al. proposed a novel Feature Learning network based on Transformer (FL-Tran), aiming to learn salient features and excavate potential useful features [24]. Overall, these advancements in deep learning and attention mechanisms have significantly improved image classification methods and their applications across various domains.

From this, it can be seen that deep learning has achieved significant success in image classification tasks, but there are also some drawbacks. For example, deep learning models often require a large amount of annotated data for training, as well as substantial computational resources during the training process. This limitation hinders their application in certain industries [25]. Secondly, the performance of deep learning models is often highly sensitive to the choice of hyperparameters and model tuning. Adjusting these parameters requires some level of expertise and computational resources, sometimes involving extensive experimentation [26]. Finally, deep learning models are often considered black-box models, making it challenging to interpret their internal decision-making processes. This may pose a problem in applications where interpretability and explainability are crucial [27].

Based on these drawbacks, traditional sparse coding models [28–30] have some advantages in image classification. Sparse coding involves representing images sparsely, emphasising important local features in the images. This helps extract crucial information from the images and reduces redundancy. Secondly, sparse coding performs relatively well on small-sample data because it learns features by encoding training samples, demonstrating robustness with relatively fewer samples. Additionally, sparse coding can be used to reduce the dimensionality of images, extracting essential information and thereby reducing the complexity of the feature space. Finally, the sparse representations generated by sparse coding are relatively easy to interpret. The sparse coding for each image can be viewed as the weight allocation to a set of bases, aiding in understanding how the model learns discriminative features for images. Therefore, many scholars have conducted various studies on sparse coding models. Yang et al. proposed a sparse coding (SC) algorithm based on Spatial Pyramid Matching (SPM), which effectively reduces quantization error [31]. Nevertheless, traditional SC exhibits instability in the encoding process, where similar features might be mapped to various codewords. Thus, Gao et al. introduced the Laplacian matrix to preserve the consistency of encoding similar local features, proposing the Laplacian Sparse Coding (LSC) algorithm to extract spatial geometric information from images, making the encoding process no longer independent [32]. Considering the locality between features, Wang et al. proposed the Locality-constrained Linear Coding (LLC) image classification algorithm, ensuring that similar features receive similar encodings [33]. Min et al. introduced the Laplacian matrix into LLC to maintain the consistency of encoding similar features [34]. While LLC utilises K-nearest neighbour encoding, the absolute difference between certain positive and negative elements in the encoding increases with the increase in K. To address this issue, Liu et al. introduced non-negativity constraints and proposed the non-negative LLC image classification algorithm [35]. In addition, since combinatorial optimisation problems involve a mixture of addition and subtraction operations, the application of subtraction might cancel out features from each other. To solve this problem, Lee et al. introduced non-negativity and employed Non-negative Matrix Factorization (NMF) to learn partial representations of objects, proposing a corresponding model [36]. To improve the robustness of NMF, a novel algorithm named Robust NMF (RNMF) was proposed in [37]. Hoyer combined NMF with SC to propose non-negative sparse coding [38]. Cai et al. proposed a graph-regularised NMF based on data representation [39]. Furthermore, Han et al. presented the SC method based on non-negativity and dependency constraints (Lap-NMF-SPM), which utilises NMF and Laplacian operators to preserve the relationships between local features [40].

Among the mentioned encoding methods, Euclidean distance is commonly used to measure the similarity between features and the dictionary. However, the local features of images are based on the histogram of statistical variables. Therefore, Euclidean distance may not effectively measure the relationship between them. Wu et al. proposed a method that goes beyond Euclidean distance, called the Histogram Intersection Kernel (HIK), which more effectively measures the similarity between features and codebooks [41]. Chen et al. introduced a histogram intersection-based LLC for scene image classification algorithm based on LLC [42]. Wan et al. incorporated histogram intersection and the Elastic Net model into the optimisation problem, resulting in an Elastic Net and Histogram Intersection-based Non-negative Local Sparse Coding (EH-NLSC) method [43].

In addition, the image feature representation and classification in these methods are two relatively independent processes. The feature quantization methods involved in encoding ignore potential semantic information, which can affect the effectiveness of image classification. To overcome these issues, the concept of semantic representation based on image representation [44] has been introduced. Based on generative models, Rasiwasia et al.

utilised a low-dimensional semantic space generated by Gaussian mixture models for scene classification and image retrieval [45,46]. On the other hand, based on discriminative models, Zhang et al. constructed the semantic space of images using a discriminative model to retain more semantic information [47]. They combined the SC model to propose a joint image representation and classification algorithm in Random Semantic Space (RSS). Shen et al. used global image features and considered the semantic information of labels to propose a method that combines image segmentation and classification in a joint framework [48].

Although previous studies have demonstrated the effectiveness of sparse coding in image classification under standard conditions, these methods often face three challenges, as follows:

- (1) Sparse coding is highly sensitive to feature variations, leading to coding instability where similar features are encoded into different codewords. The previous studies only take into account two of the main three features in the optimisation problem: non-negativity, locality, and Laplacian regularisation.
- (2) In addition, Euclidean distance could not effectively measure the relationship between feature vectors and codebooks.
- (3) The processes of image representation and classification are relatively independent. The feature quantization methods involved in coding neglect the potential contextual information in local regions, resulting in the loss of visual and semantic information in images, thus impeding the effectiveness of image classification.

To enhance the extraction of comprehensive and effective information from images, and subsequently improve image classification accuracy, this paper integrates histogram intersection and semantic information. The specific research contributions are outlined as follows:

- (1) Incorporate non-negativity and locality into the LSC model, constructing the NLLSC method. This method preserves local information among features and spatial geometric information, significantly improving the instability of encoding.
- (2) Introduce histogram intersection to redefine the distance between feature vectors and the dictionary in the locality constraint of the sparse coding model. This redefinition provides a more accurate measurement of their similarity, ensuring that similar features can share their local bases.
- (3) After obtaining the fused locality and non-negativity in Laplacian Sparse Coding, integrate image representation and classification, which incorporates semantic information to preserve the contextual relationship between image features. This approach more comprehensively and effectively captures the essence of the image.
- (4) Conduct comparative experiments with the other six state-of-the-art methods in four standard image datasets and three maritime datasets to validate the performance of the proposed methodology.

The remainder of this paper is structured as follows: Section 2 provides a review of related work on SC, LSC, and LLC; Section 3 introduces the proposed coding method, Histogram intersection, and Semantic information-based Non-negativity Local Laplacian Sparse Coding (HS-NLLSC); Section 4 presents experimental results on several datasets; and Section 5 offers the conclusions.

2. Preliminary Methods

The proper encoding of local features is crucial for image classification, as it not only faithfully represents images but also improves the accuracy of image classification. Recently, numerous scholars have proposed various encoding methods, and these methods have demonstrated promising classification results. This section primarily introduces three typical encoding models: sparse coding, Laplacian Sparse Coding, and Locality-constrained Linear Coding.

Let the feature matrix of an image be denoted as $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$, the dictionary be denoted as $U = [u_1, u_2, \dots, u_M] \in R^{D \times M}$, and the corresponding sparse coding be represented as $V = [v_1, v_2, \dots, v_N] \in R^{M \times N}$.

2.1. Sparse Coding

In light of the quantization errors arising from vector quantization methods and the potential lack of semantic information in the C-means method, the SC method has been introduced. The central challenge it addresses is the learning of an over-complete dictionary U in an M -dimensional space (i.e., $M \gg D$; namely, the number of base vectors significantly exceeds its dimension). The objective is to choose as few base vectors as possible to represent the feature vector. The particular optimisation problem is articulated below:

$$\begin{cases} \min_{U, V} \sum_{i=1}^N (\|x_i - Uv_i\|_2^2 + \lambda \|v_i\|_1) \\ \text{s.t. } \|u_j\|_2^2 \leq 1, \forall j = 1, 2, \dots, M \end{cases} \quad (1)$$

Here, λ represents the regularisation parameter, which balances the trade-off between reconstruction error and the sparsity of the coding. And u_j represents the j -th column vector of dictionary U .

The general solution for Equation (1) is to alternately fix U (or V) and optimise V (or U) until the value of the objective function achieves the specified extreme value.

2.2. Laplacian Sparse Coding

To address the encoding instability in traditional SC, where similar features might be encoded into different codewords, LSC was introduced in [32]. LSC incorporates the Laplacian matrix to maintain the stability of encoding similar local features, thus eliminating the independence of the encoding process. The specific optimisation problem is presented as follows:

$$\begin{cases} \min_{U, V} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \beta \text{tr}(VLV^T) \\ \text{s.t. } \|u_j\|_2^2 \leq 1, \forall j = 1, 2, \dots, M \end{cases} \quad (2)$$

where $\beta \text{tr}(VLV^T)$ is used to extract the spatial geometric information of the image and reduce quantization errors, and L represents the Laplacian matrix.

2.3. Locality-Constrained Linear Coding

The LLC method was introduced in [33], highlighting that local non-zero coefficients are frequently assigned to bases in proximity to the coding feature data. This approach utilises multiple codewords from the codebook to enhance the accuracy of representing a feature descriptor. Additionally, similar features utilise similar coding patterns by sharing their local codewords, effectively addressing the instability issue present in SC. The specific optimisation problem is presented as follows:

$$\begin{cases} \min_V \sum_{i=1}^N (\|x_i - Uv_i\|_2^2 + \lambda \|d_i \odot v_i\|_2^2) \\ \text{s.t. } 1^T v_i = 1, \forall i = 1, 2, \dots, N \end{cases} \quad (3)$$

where \odot represents element-wise multiplication (for column vectors) and λ denotes a regularisation parameter. $d_i \in R^M$ denotes a local adaptor, defined as follows:

$$\begin{aligned} d_i &= \exp\left(\frac{\text{dist}(x_i, U)}{\sigma}\right) \\ \text{dist}(x_i, U) &= [\text{dist}(x_i, u_1), \dots, \text{dist}(x_i, u_M)]^T \end{aligned} \quad (4)$$

where $dist(x_i, u_j)$ is the Euclidean distance between x_i and u_j , and σ is a parameter used to adjust weight decay. The constraint condition $1^T v_i = 1$ indicates the translation invariance of the LLC method.

To facilitate a more intuitive comparison of these three methods, Table 1 provides an overview of their advantages, disadvantages, and applications.

Table 1. Comparison of three different methods.

Methods	Advantages	Disadvantages	Applications
SC	<ol style="list-style-type: none"> 1. Compact data representation; 2. Feature selection and dimensionality reduction; 3. Robustness to noise and redundant data; 4. Nonlinear mapping. 	<ol style="list-style-type: none"> 1. Higher computational complexity; 2. Sensitivity to parameters; 3. Risk of overfitting. 	<ol style="list-style-type: none"> 1. Image processing: Image compression, denoising, and feature extraction; 2. Signal processing; 3. Machine learning; 4. Neuroscience.
LSC	<ol style="list-style-type: none"> 1. Preservation of local structural information; 2. Emphasis on the non-negativity of data; 3. Addressing the instability of encoding. 	<ol style="list-style-type: none"> 1. Higher computational complexity; 2. Dependency on parameters. 	<ol style="list-style-type: none"> 1. Image processing: Image compression, denoising, and feature extraction; 2. Signal processing; 3. Bioinformatics.
LLC	<ol style="list-style-type: none"> 1. Emphasis on local features; 2. Feature selection and dimensionality reduction; 3. Higher computational efficiency. 	<ol style="list-style-type: none"> 1. Dependency on parameters; 2. Sensitivity to data distribution; 3. Limited global information. 	Image processing: image feature extraction, image classification, object recognition, and image compression.

3. Methodology

3.1. The Proposed Framework

The proposed framework comprises four primary components, as depicted in Figure 1. The first involves the extraction of common SIFT features from images, while the second involves image representation using NLLSC based on Histogram Intersection (HI-NLLSC). In the third part, semantic information is integrated between image representation and image classification based on HI-NLLSC to acquire the final HS-NLLSC with its updated features. Finally, the SVM classifier is utilised to classify these images within the semantic spaces of the third part.

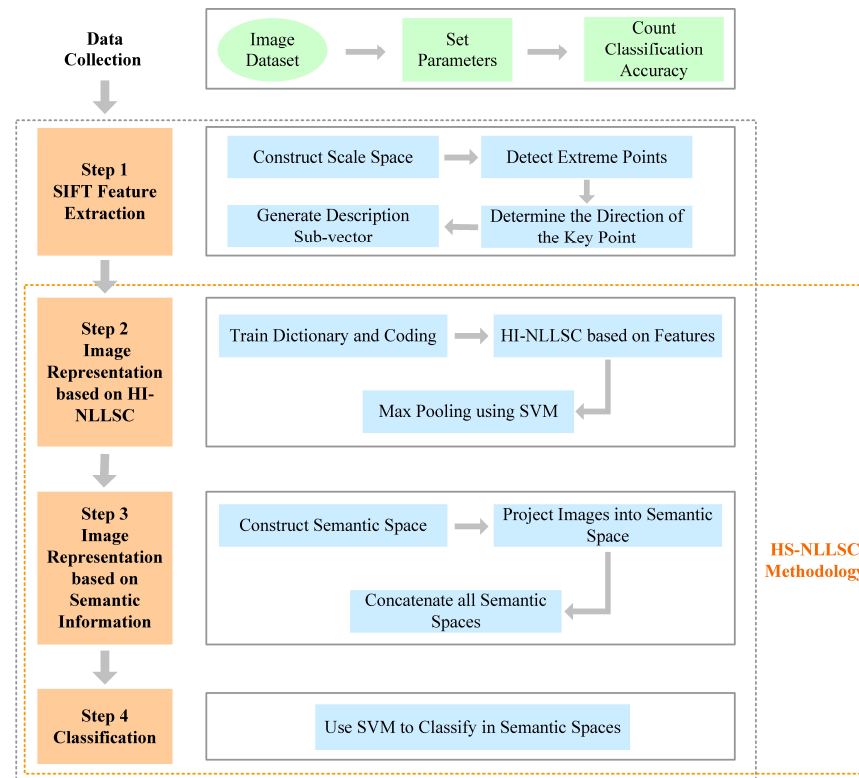


Figure 1. The framework of the HS-NLLSC algorithm.

3.2. The Proposed HS-NLLSC Algorithm

In this paper, the NLLSC method is introduced, which incorporates non-negativity and locality constraints based on the LSC model. Additionally, histogram intersection is integrated into the locality constraint of the optimisation problem. Moreover, the HS-NLLSC method is proposed by considering both image representation and classification.

Firstly, the HI-NLLSC method is employed to encode the local features of the images, utilising Max Pooling (MP) to derive the original image representations.

Secondly, a subset of image representations is randomly selected to construct a semantic information-based space. Within this space, all training images are projected using a trained classifier, resulting in projected image feature representations that serve as the final image representations.

Finally, an SVM classifier is utilised for both training and classification, with the output providing class information. This comprehensive framework integrates non-negativity and locality constraints, histogram intersection, and semantic information to enhance image representation and classification within the HS-NLLSC method.

3.3. Image Representation Using HI-NLLSC

This section primarily outlines the process of deriving the original image representations through the HI-NLLSC method. By integrating histogram intersection into the optimisation problem of NLLSC, the distance between features and the dictionary is redefined, effectively quantifying their similarity.

3.3.1. Train Dictionary and Corresponding Coding

Due to the extensive number of extracted local features, constructing the local adaptors [30] and the Laplacian matrix incurs high computation complexity. Thus, template features are employed to train the dictionary and corresponding coding, randomly selected from all local features. Firstly, the initial formulation of the HI-NLLSC method is presented. Given X as the input non-negative feature matrix, B as the non-negative dictionary, and S as the corresponding non-negative sparse coding, where $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$, $B = [b_1, b_2, \dots, b_M] \in R^{D \times M}$, $S = [s_1, s_2, \dots, s_N] \in R^{M \times N}$, by incorporating locality and non-negativity into LSC, the optimisation problem is provided as follows:

$$\begin{cases} \min_{B,S} \sum_{i=1}^N (\|x_i - Bs_i\|_2^2 + \lambda \|d_i \odot s_i\|_2^2) + \beta \text{tr}(SLS^T) \\ \text{s.t.} \quad \text{sparseness}(b_j) = S_b \\ \|b_j\|_2^2 \leq 1, B \geq 0, S \geq 0, \forall j \end{cases} \quad (5)$$

where λ , β , and S_b represent specified constants, and the sparseness S_b is defined based on the relationship between the l_1 -norm and l_2 -norm, which is represented as follows:

$$\text{sparseness}(b_j) = \frac{\sqrt{D} - \|b_j\|_1 / \|b_j\|_2}{\sqrt{D} - 1} \quad (6)$$

where D is the dimensionality of b_j , i.e., $b_j \in R^{D \times 1}$.

In this paper, an improvement is made to calculate the Euclidean distance between features and the dictionary in the LLC model. A similarity measurement method based on histogram intersection is proposed. $d_i \in R^M$ is a local adaptor, defined as follows:

$$\begin{aligned} d_i &= \sigma / \text{dist}(x_i, B) \\ \text{dist}(x_i, B) &= [\text{dist}(x_i, b_1), \dots, \text{dist}(x_i, b_M)]^T \end{aligned} \quad (7)$$

where σ represents a parameter used to adjust weight decay. $dist(x_i, b_j)$ represents the distance between x_i and b_j , which is measured using histogram intersection. The calculation method is defined as follows:

$$dist(x_i, b_j) = \sum_{k=1}^M \min(x_{ik}, b_{jk}) \tag{8}$$

where M is the dimensionality of the two histograms (size of the dictionary), and x_{ik} and b_{jk} , respectively, represent the k -th elements of the features x_i and b_j .

The method of using alternating fixed B (or S) to optimise S (or B) is employed to solve Equation (5). Firstly, X and B are fixed, S is optimised, and the following optimisation problem is obtained:

$$\begin{cases} \min_S \|X - BS\|_F^2 + \lambda \|d \odot S\|_F^2 + \beta tr(SLS^T) \\ s.t. S \geq 0 \end{cases} \tag{9}$$

where $d = [d_1, d_2, \dots, d_N] \in R^{M \times N}$, and \odot represents element-wise multiplication (for matrices).

For Equation (9), the objective function is first transformed into a trace form of matrices. Then, utilising the Lagrange Multiplier Method (LMM) and Karush–Kuhn–Tucker (KKT) conditions, the update rule for S can be obtained as follows:

$$s_{ij} = s_{ij} \cdot \frac{(B^T X + \beta S W)_{ij}}{(B^T B S + \beta S D + \lambda \text{diag}(g_i) \bar{S})_{ij}}, \tag{10}$$

$$\forall i = 1, 2, \dots, M, j = 1, 2, \dots, N$$

where $\text{diag}(g_i)$ is a diagonal matrix with g_i as its diagonal elements, $g_i = [d_{i1}^2, d_{i2}^2, \dots, d_{iN}^2]$ is an N -dimensional row vector, $d_{ij}^2 = (\sigma / dist(x_j, b_i))^2$, and $\bar{S} = [\text{diag}(s_1), \text{diag}(s_2), \dots, \text{diag}(s_N)]^T$.

Next, B is optimised by fixing X and S . The optimisation problem is as follows:

$$\begin{cases} \min_B \|X - BS\|_F^2 \\ s.t. sparseness(b_j) = S_b, B \geq 0, \|b_j\|_2^2 \leq 1, \forall j = 1, 2, \dots, M \end{cases} \tag{11}$$

For Equation (11), the diagonal matrix Λ can be obtained using the Lagrange dual problem and conjugate gradient method. After solving for Λ , it is substituted into the following equation to obtain B , namely:

$$B = (XS^T)(SS^T + \Lambda)^{-1} \tag{12}$$

3.3.2. HI-NLLSC Based on New Features

Some template features X are randomly selected to train B and S in Section 3.3.1. When a new feature matrix H of local features appears, the HI-NLLSC method proceeds by using B and S . So the optimisation problem can be written as follows:

$$\begin{cases} \min_V \|H - BV\|_F^2 + \lambda \|d \odot V\|_F^2 + \frac{\beta}{2} \sum_{ji} \|v_j - s_i\|_2^2 w_{ji} \\ s.t. v_{ij} \geq 0, \forall i, j \end{cases} \tag{13}$$

where s_i represents the i -th column vector of S , and V represents the sparse coding of the updated feature matrix H . The elements w_{ji} in the similarity matrix W are obtained by calculating the K-nearest neighbour relationship between the new feature and the template feature, where the K-nearest neighbour relationship is measured using histogram

intersection. If the template feature x_i and the new feature h_j have a K-nearest neighbour relationship, $w_{ji} = 1$; otherwise, $w_{ji} = 0$. The metric function is the same as in Equation (8).

The update rule for V can be obtained using the Lagrange Multiplier Method (LMM) and KKT conditions as follows:

$$v_{ij} = v_{ij} \frac{(B^T X + \beta S W^T)_{ij}}{(B^T B V + 0.5 * \beta V A + \lambda \text{diag}(g_i) \bar{V})_{ij}} \tag{14}$$

where A is the diagonal weight matrix with its diagonal elements being $a_{jj} = \sum_i w_{ji}$, $\text{diag}(g_i)$ is the same as the definition in Equation (10), and $\bar{V} = [\text{diag}(v_1), \text{diag}(v_2), \dots, \text{diag}(v_N)]^T$. After obtaining B and S from the template features, Equation (14) can be utilised to perform HI-NLLSC on the new features.

For the feature fusion stage, this paper employs the MP method. The specific approach is as follows:

$$z_l = \max\{|v_{1l}|, |v_{2l}|, \dots, |v_{Nl}|\}, l = 1, 2, \dots, M \tag{15}$$

where v_{Nl} is the l -th element of the sparse coding v_N , and z_l is the l -th element of the vector z . Thus, the image of a single spatial pyramid region can be described by an M -dimensional vector z , as shown below: $z = [z_1, z_2, \dots, z_M]^T$.

After obtaining the image representation for each region, the final image representation is obtained using the SPM method. In the image classification stage, this paper utilises a multi-class linear SVM.

3.4. Image Representation Based on Semantic Information

To overcome the challenges posed by the relatively independent nature of the HI-NLLSC method and the semantic gap between visual features and human understanding, as well as the oversight of semantic information in local regions during feature quantization, we propose the HS-NLLSC image classification algorithm. This algorithm seeks to comprehensively integrate visual and semantic information in images, capturing the relationships between semantic objects and their surrounding environments. Initially, the HI-NLLSC method is utilised to encode the local features of images, producing the original image representation. Subsequently, a semantic space is constructed to generate the final representation of images. Finally, SVM is employed to classify the obtained image representations within the semantic space.

3.4.1. Construct Semantic Space

The semantic space is defined as the collective space obtained from all image representations during classifier training, serving the purpose of image representation. Each distinct semantic space is created by training the classifier on randomly selected images.

Assuming that the original image representations of P training images using the HI-NLLSC method are obtained, denoted as r^1, r^2, \dots, r^P , with a total of C classes and their corresponding class labels y^1, y^2, \dots, y^P , from these representations, L ($L \leq P$) images are randomly selected to construct the semantic space and this selection process is repeated T times. The corresponding results are denoted as $\{(r^{1,1}, y^{1,1}), \dots, (r^{L,1}, y^{L,1})\}, \dots, \{(r^{1,T}, y^{1,T}), \dots, (r^{L,T}, y^{L,T})\}$. For the t -th random selection of images $\{(r^{1,t}, y^{1,t}), \dots, (r^{L,t}, y^{L,t})\}$, the SVM classifier is utilised to construct the corresponding semantic space, namely:

$$f_c^t(r^{l,t}) = \bar{y}_c^{l,t} = w_c^t r^{l,t} + b_c^t, \quad l = 1, 2, \dots, L, \quad c = 1, 2, \dots, C \tag{16}$$

Then, the corresponding optimisation problem using the Hinge loss function is constructed, namely:

$$\min_{w_c^t} \|w_c^t\|^2 + \alpha \sum_{l=1}^L l(\bar{y}_c^{l,t}, y^{l,t}) \tag{17}$$

By solving Equation (17), the corresponding w_c^t and b_c^t are obtained.

Each dimension of the semantic space corresponds to a classifier trained using randomly selected samples. As there are C classes of images, the generated semantic space is C -dimensional.

3.4.2. Project Images into Semantic Space

After training the SVM classifier, all the training images are projected into the aforementioned semantic space, namely:

$$r_{i,k}^{ss} = (f_1^t(r^p); f_2^t(r^p); \dots; f_C^t(r^p)), t = 1, 2, \dots, T; p = 1, 2, \dots, P \tag{18}$$

where the superscript ‘ss’ represents ‘semantic space’.

3.4.3. Concatenate All Semantic Spaces

Upon acquiring knowledge of all joint spaces, the training images are projected into all the T generated joint spaces. The connection of all image features in these spaces forms the final image representation, namely:

$$r_k^{ss} = (r_{1,p}^{ss}; r_{2,p}^{ss}; \dots; r_{T,p}^{ss}), p = 1, 2, \dots, P \tag{19}$$

Following the acquisition of the final image representation using Equation (19), a multi-class SVM is employed for image classification.

3.5. Description of Three Algorithms

After obtaining expressions for matrices B and S , the dictionary and associated sparse codes for template features are acquired through the systematic use of the following algorithms.

Algorithm 1 is designed for the iterative update of the Lagrangian diagonal matrix to obtain the representation of B . Subsequently, Algorithm 2 is employed to identify the optimal approximation B^* that ensures the appropriate sparsity of B . In this process, B is replaced by B^* .

Algorithm 1 (Iteratively update Λ to calculate B)

Input: non-negative matrix X ; sparse coding V ; precision ϵ ;

Output: diagonal matrix Λ ; dictionary B

1. **Initiate** Λ^0 and let $k = 0$;
 2. **while** convergence is not achieved, **do**
 3. Let $g_0 = \nabla f(\Lambda^0)$, **if** $\|g_0\|_2 \leq \epsilon$ **then**
 4. Λ^0 is the desired extremum;
 5. **else**
 6. $d^0 = -g_0$;
 7. **end if**
 8. Set $g_{k+1} = \nabla f(\Lambda^{k+1})$, **if** $\|g_{k+1}\|_2 \leq \epsilon$ or $\|\Lambda^{k+1} - \Lambda^k\|_F \leq \epsilon \|\Lambda^1 - \Lambda^0\|_F$ **then**
 9. Λ^{k+1} is the required extreme value;
 10. **else**
 11. $d^{k+1} = -g_{k+1} + \beta_k d^k$, $\beta_k = \|g_{k+1}\|_2^2 / \|g_k\|_2^2$;
 12. **end if**
 13. Determine the optimal step ζ_k size using an approximate one-dimensional search, namely: $f(\Lambda^k + \zeta_k d^k) = \min_{\zeta} f(\Lambda^k + \zeta d^k)$;
 14. $\Lambda^{k+1} = \Lambda^k + \zeta_k d^k$;
 15. Set $k = k + 1$; return Step 8;
 16. **end while**
 17. Return Λ , and obtain dictionary B according to Equation (12).
-

Algorithm 2 (The optimal approximation B^* for proper sparseness of B)

Input: a random column vector b of matrix B

Output: the nearest non-negative vector b^* of B^*

1. Compute the sparseness S_{b^*} of the column vector b^* with Equation $sparseness(b^*) = \frac{\sqrt{D} - \|b^*\|_1 / \|b^*\|_2}{\sqrt{D} - 1}$, and $k_1 = \|b^*\|_1 / \|b^*\|_2 = \sqrt{D} - (\sqrt{D} - 1)S_{b^*}$, where D represents the dimensionality of vector u or b^* ;
2. Map vector b into the k_1 constraint space, $b_i^* = b_i + \frac{k_1 - \|b\|_1}{D}$ for $\forall i$, namely make $\|b^*\|_1 = k_1$;
3. Let $Z = \{\}$ to be an initial negative element set;
4. **while not** iteratively finding the closest non-negative vector b^* and meeting $\|b^*\|_2^2 = \|b\|_2^2$ **do**;
5. Set midpoint $m_i = \begin{cases} \frac{k_1}{D - length(Z)}, i \notin Z \\ 0, i \in Z \end{cases}$ in k_1 constraint space;
6. Get the non-negative solution α by solving a quadratic equation $\|b^*\|_2^2 = \|m + \alpha(b^* - m)\|_2^2$, and replace $b^* = m + \alpha(b^* - m)$ with α to update the vector b^* ;
7. **if** all elements of b^* are non-negative **then**
8. Return b^* ;
9. **else**
10. **for each** $i \in Z$ **do**
11. Let all negative elements be zero by $Z = Z \cup \{i : b_i^* < 0\}$ and set $b_i^* = 0, \forall i \in Z$;
12. Recompute the projection, keep b^* invariant in k_1 constraint space, namely:
 $b_i^* = b_i^* + (k_1 - \|b^*\|_1) / (D - length(Z))$;
13. Go to Step 5;
14. **end for**
15. **end if**
16. **end while**

Following this, Algorithm 3 is applied to learn the dictionary of HS-NLLSC and its corresponding coding. This entails iterative updates of both B and S until the established stop criterion is satisfied. The implementation of Algorithm 3 integrates the functionalities of both Algorithms 1 and 2.

In summary, the algorithmic process can be outlined as follows:

Algorithm 3 (HS-NLLSC)

Input: non-negative feature matrix X , original dictionary B , original sparse coding S , Laplacian matrix L , parameter $\lambda, \beta, \sigma, S_b$, number of training images P , number of training iterations T

Input: dictionary B , sparse coding S , class labels

1. **Preprocessing:** $X = X / \max(X(\cdot))$, $S = S / \|S\|_1$;
2. **While** convergence is not achieved, **do**
3. Update sparse coding S with Equation (10);
4. Normalize B and S according to the following equations: $s_{ij} = s_{ij} / \sqrt{\sum_i s_{ij}}$, $b_{ij} = b_{ij} / \sqrt{\sum_i s_{ij}}$;
5. Update Lagrange dual matrix Λ using **Algorithm 1**;
6. Project each column vector of matrix B using **Algorithm 2** to obtain B^* , and let $B = B^*$, thereby obtaining the optimal dictionary B and corresponding sparse coding S ;
7. Set $k = k + 1$;
8. **If** Step 8 in **Algorithm 1** is satisfied **then**
9. Return B and S ;
10. **else**
11. Return Step 3;
12. **end if**
13. **end while**
14. After obtaining B and S for the template features, calculate the sparse coding V for the new features according to Equation (14);
15. Use SPM with Equation (15) to perform MP on the obtained coding and obtain the original image representation;
16. After obtaining the image representation using HI-NLLSC, construct the semantic space according to Equation (16), compute w_c^t and b_c^t according to Equation (17);
17. Project all training images into the semantic space according to Equation (18);
18. Connect all semantic spaces and generate the final image representation according to Equation (19);
19. Use multi-class linear SVM to classify the image in the semantic space.

4. Experiments

This part mainly presents three experiments to validate the feasibility of the HS-NLLSC algorithm. The first subsection provides information about the datasets used in the experiments. The second subsection describes the parameter settings of the experiments. The third subsection mainly introduces the design and results analysis of the three experiments. Following this, the fourth subsection provides an analysis of algorithm stability. Finally, the last subsection discusses the complexity analysis of the algorithm.

4.1. Experimental Datasets

In this section, detailed descriptions of four standard datasets are provided, namely, Corel-10, Scene-15, Caltech-101, and Caltech-256 datasets. The specific information is presented in Table 2. Additionally, partial images from the Caltech-101 dataset are displayed in Figure 2.

Table 2. Four standard image datasets.

Image Datasets	Number of Classes	Number of Images Per Class	Total Number of Images
Corel-10	10	100	1000
Scene-15	15	200~400	4485
Caltech-101	101	31~800	9144
Caltech-256	256	≥ 80	29,780

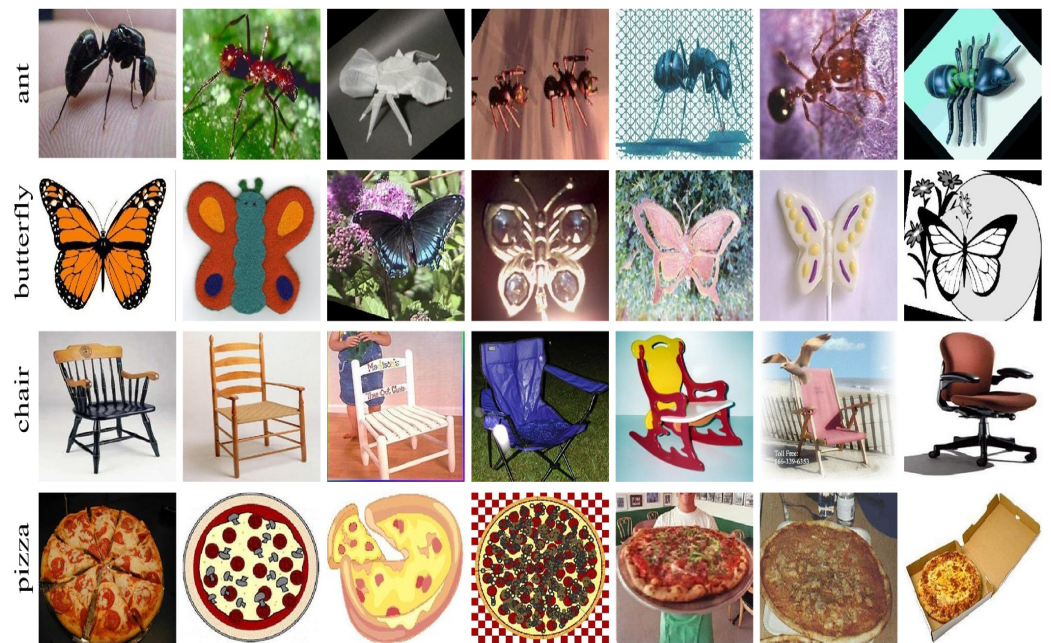


Figure 2. Some pictures of the Caltech-101 dataset.

Furthermore, three maritime datasets are discussed, namely, the Singapore Maritime Dataset (SMD), the Open Seaship dataset, and the Marine Image Dataset (MID). The SMD is divided into three parts, comprising on-shore videos, on-board videos, and near-infrared (NIR) videos. The distribution of the SMD is outlined in Table 3. As for the Open Seaship dataset, it currently contains 31,455 images covering seven common ship types (i.e., ore carriers, bulk carriers, general cargo ships, container ships, fishing vessels, passenger ships, and mixed types). The specific information is detailed in Table 4.

Moreover, the MID consists of eight video sequences for marine obstacle detection. It comprises 2655 labelled images with a resolution of 640×480 pixels captured from our Jinghai VIII USV. Partial images from the MID are shown in Figure 3.

Table 3. The SMD.

Subdataset	Videos (Annotated)	Labelled Frames	Number of Labels
VIS on-shore	40 (36)	17,967	154,495
VIS on-board	11 (4)	2400	3173
NIR	30 (23)	11,286	83,174
Total	81 (63)	31,653	240,842

Table 4. The Open Seaship dataset.

Ship Category	Images	Percentage
Ore carrier	5126	0.1630
Bulk cargo carrier	5067	0.1610
Container ship	3657	0.1163
General cargo ship	5342	0.1698
Fishing boat	5652	0.1797
Passenger ship	3171	0.1008
Mixed type	3440	0.1094
Total	31455	1

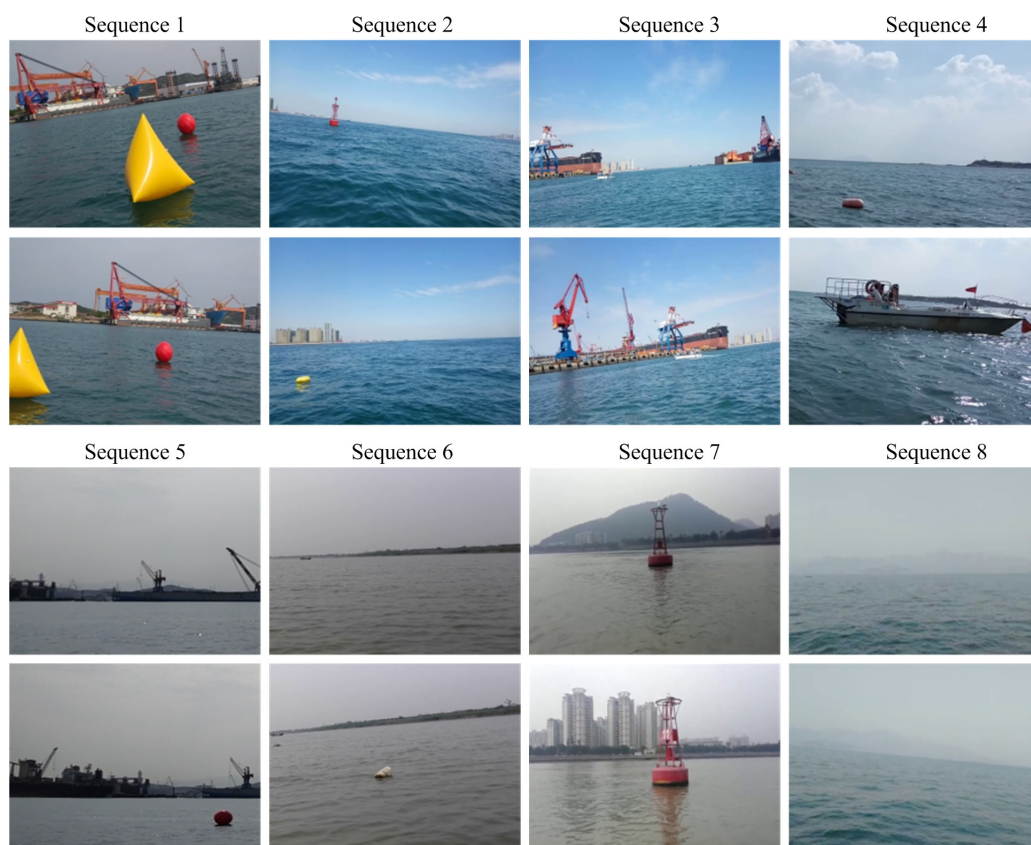


Figure 3. Some pictures of the MID.

4.2. Experimental Settings

For the four standard datasets, different training and testing samples are selected for the experiments. Specifically, for the Corel-10 and Scene-15 datasets, 50 and 100 images from each category are randomly selected as training samples, while the remaining images in each category are regarded as testing samples.

Regarding the Caltech-101 dataset, 15 or 30 images from each category are randomly chosen as training samples, and the remaining images in each category are used as test samples. For the Caltech-256 dataset, 15, 30, 45, or 60 images are randomly selected from

each category to be used as training images, while the remaining in each category are taken as test images.

For the three maritime datasets, 50, 100, 150, 200, or 250 images are randomly chosen from each category to be used as training images, and the remaining images in each category are taken as test images.

During the feature extraction stage, a step size of 8 and a window of 16×16 are used to extract SIFT features for each image. Each local feature descriptor is 128-dimensional, namely, $D = 128$. Regarding the process of dictionary learning, the dictionary size is set to 1024. In the optimisation problem, there are four key parameters, namely, λ , β , σ , and S_b . As for λ and β , in the SC algorithm, $\lambda \in [0.1, 0.3]$. For instance, in the LSC algorithm, $\lambda = 0.4$ and $\beta = 0.2$ are set for the Corel-10 and Scene-15 datasets, while for the Caltech-101 and Caltech-256 datasets, $\lambda = 0.3$ and $\beta = 0.1$ are adopted. Then, it can be concluded that $\lambda \in [0.1, 0.4]$ and $\beta \in [0.1, 0.4]$. In the proposed method, after comparing several different values, $\lambda = 0.4$ and $\beta = 0.2$ are ultimately set, as presented in Section 4.3.3. Additionally, $\sigma = 100$ and $S_b = 0.4$ are determined according to References [33,40]. In the generation of the semantic space, $L = 0.3P$ and $T = 30$ are configured. Detailed information can be found in Table 5.

Table 5. The experimental parameter setting for each stage.

Image Datasets	Corel-10	Scene-15	Caltech-101	Caltech-256	MID	SMD	Seaship
The feature extraction stage							
Methods	SIFT	SIFT	SIFT	SIFT	SIFT	SIFT	SIFT
Step size	8	8	8	8	8	8	8
Window	16×16	16×16	16×16	16×16	16×16	16×16	16×16
The HS-NLLSC coding stage							
Training images	50	100	15, 30	15, 30, 45, 60	50, 100, 150, 200, 250	50, 100, 150, 200, 250	50, 100, 150, 200, 250
Test images	the rest	the rest	the rest	the rest	the rest	the rest	the rest
Dictionary size	1024	1024	1024	1024	1024	1024	1024
λ	0.4	0.4	0.4	0.4	0.4	0.4	0.4
β	0.2	0.2	0.2	0.2	0.2	0.2	0.2
σ	100	100	100	100	100	100	100
Sparseness S_b	0.4	0.4	0.4	0.4	0.4	0.4	0.4
The generation of the semantic space stage							
The number of images used to construct the semantic space (L)	0.3P	0.3P	0.3P	0.3P	0.3P	0.3P	0.3P
The times of repetitions selected (T)	30	30	30	30	30	30	30

4.3. Experimental Design and Result Analysis

This section comprises three experimental design components. Experiment 1 involves the visualisation of learned dictionaries for SC, LSC, and the proposed HS-NLLSC method. In Experiment 2, each dataset is randomly divided into 10 subsets, and a 10-fold cross-validation approach is utilised to determine the average classification accuracy and standard deviation of the proposed HS-NLLSC method. Experiment 3 investigates the influence of two parameters, λ and β , on the classification performance across the four standard datasets.

4.3.1. Visualisation of Learned Dictionaries

In this subsection, Figure 4 illustrates the dictionaries learned using the SC, LSC, and HS-NLLSC methods. These images are displayed in grayscale format to effectively highlight the original features' attributes, specifically non-negativity, locality, bandpass characteristics, and directionality.

(1) Non-negativity

Non-negativity ensures that the pixel values in the dictionary images are non-negative. Dictionaries with strong non-negativity typically appear brighter, with minimal dark regions. This characteristic is essential for ensuring that the dictionaries accurately represent features in a physically interpretable manner.

(2) Locality

Locality refers to the concentration of dictionary atoms in specific regions, appearing as localised patches rather than being distributed across the entire image. Dictionaries with good locality effectively capture local patterns, which are critical for tasks such as object recognition and texture analysis.

(3) Bandpass characteristics

Bandpass characteristics represent a balance between high-frequency and low-frequency features in the dictionaries. High-frequency features, such as fine textures, coexist with low-frequency features, such as smooth regions. This balance ensures that the dictionaries can capture both detailed and broader structural elements in the data.

(4) Directionality

Directionality reflects the ability of dictionary atoms to capture specific directional patterns, such as horizontal, vertical, or diagonal edges. Dictionaries with strong directionality exhibit clear streaks or gradients, indicating their sensitivity to directional features in the input data. This characteristic is particularly valuable for applications involving edge detection and orientation analysis.

These four characteristics are critical for evaluating the quality of dictionaries learned by different methods. As shown in Figure 4, the dictionaries obtained by the HS-NLLSC method demonstrate a superior representation of these characteristics compared to the SC and LSC methods, highlighting its ability to effectively capture complex patterns in the data.

From Figure 4, it can be seen that the dictionaries generated by the SC and LSC methods (as shown in Figure 4a,b) exhibit common attributes such as locality, bandpass characteristics, and directionality. However, due to the differential operations used in their optimisation processes, negative bases may exist in these dictionaries, leading to a lack of non-negativity. In contrast, as shown in Figure 4c, the dictionary obtained using the HS-NLLSC method exhibits more discernible characteristics, which encompasses locality, non-negativity, bandpass characteristics, and directionality. Furthermore, considering the sparseness of the NLLSC dictionary, it can be concluded that the sparser the dictionary, the weaker its directionality and bandpass characteristics (see [40]). Therefore, in Figure 4c, the appropriate sparseness of the dictionary S_b is 0.4, which results in better performance in characterising the features of such images compared to other methods.

Additionally, to showcase the performance of HS-NLLSC, the image of code V obtained by the non-negative dictionary in Scene-15 is generated using different methods. Subsequently, V is visualised, as depicted in Figure 5. The representation of HS-NLLSC is presented in Figure 5c, while SC and EH-NLSC are illustrated in Figures 5a and 5b, respectively.

In Figure 5, the non-zero elements in V are depicted as white pixels. The distribution of V depicted in Figure 5c demonstrates a more uniform pattern, indicating the incorporation

of locality, sparsity, and semantic information. This reflects the consideration of both group effect and topology information.

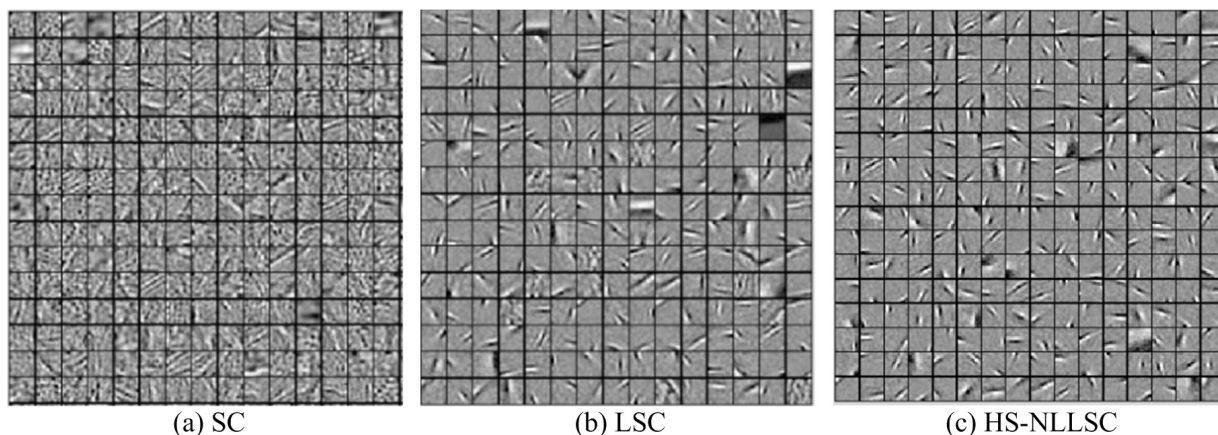


Figure 4. The obtained dictionaries with non-negativity, locality, bandpass characteristics, and directionality for the three methods.

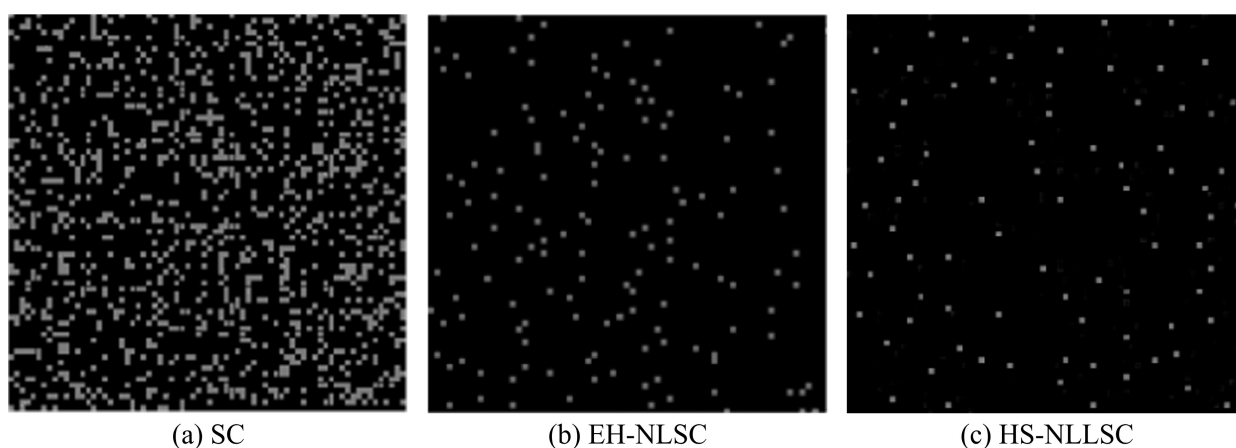


Figure 5. Visualisation of code V learned from (a) SC, (b) EH-NLSC, and (c) HS-NLLSC in Scene-15.

4.3.2. Comparison of Average Classification Accuracy

Table 6 presents a comparison of the HS-NLLSC method to six state-of-the-art sparse coding methods, including SC, LSC, LLC, Lap-NMF-SPM, RSS, and EH-NLSC, across four standard datasets: Corel-10, Scene-15, Caltech-101, and Caltech-256. The comparative results clearly indicate that the HS-NLLSC approach exceeds the performance of state-of-the-art methods on four datasets. This superiority can be attributed to several factors. Firstly, the HS-NLLSC method integrates non-negativity, locality, Laplacian regularisation, and histogram intersection, ensuring the accurate encoding of similar local features and the precise measurement of feature-dictionary similarity, thereby reducing coding instability. Secondly, in generating the semantic space, the proposed method comprehensively addresses both image representation and classification, maximising the utilisation of semantic information for a more effective representation. In contrast, previous methods such as SC, LSC, and LLC may suffer from feature cancellation due to the use of addition and subtraction in their optimisation problems, while Lap-NMF-SPM may lack local information, leading to inaccurate representations. Lastly, EH-NLSC, despite using histogram intersection, lacks Laplacian regularisation, impacting its ability to extract spatial geometric information effectively and rendering the encoding process independent.

Table 6. Classification accuracy (average value \pm standard deviation) on four standard image datasets (%).

Methods	Corel-10	Scene-15	Caltech-101 (15)	Caltech-101 (30)	Caltech-256 (15)	Caltech-256 (30)	Caltech-256 (45)	Caltech-256 (45)
SC [31]	86.76 \pm 1.18	81.12 \pm 0.45	66.87 \pm 0.45	72.10 \pm 1.14	27.53 \pm 0.42	33.86 \pm 0.55	37.35 \pm 1.64	40.08 \pm 0.79
LSC [32]	88.43 \pm 0.75	89.65 \pm 0.41	70.32 \pm 1.35	74.86 \pm 0.53	29.88 \pm 0.15	35.67 \pm 0.33	38.37 \pm 0.46	40.35 \pm 0.24
LLC [33]	87.83 \pm 1.03	81.53 \pm 0.87	68.57 \pm 0.88	72.54 \pm 0.71	31.27 \pm 0.85	34.17 \pm 0.33	35.93 \pm 0.51	37.58 \pm 0.49
Lap-NMF-SPM [40]	91.24 \pm 0.95	90.46 \pm 0.87	74.35 \pm 0.94	76.81 \pm 0.49	35.24 \pm 0.83	37.46 \pm 0.32	39.87 \pm 0.75	41.35 \pm 0.72
RSS [47]	95.72 \pm 0.78	92.45 \pm 0.93	77.63 \pm 0.89	82.91 \pm 0.22	40.16 \pm 0.53	44.96 \pm 0.85	48.25 \pm 0.47	51.32 \pm 0.41
EH-NLSC [43]	93.64 \pm 0.78	91.82 \pm 0.67	73.34 \pm 0.62	78.89 \pm 0.39	35.89 \pm 0.56	38.87 \pm 0.59	41.65 \pm 0.53	43.61 \pm 0.47
HS-NLLSC	98.86 \pm 0.23	97.56 \pm 0.64	81.54 \pm 0.48	87.73 \pm 0.67	46.34 \pm 0.45	49.86 \pm 0.65	53.60 \pm 0.78	57.68 \pm 0.12

Moreover, these five methods operate relatively independently regarding image representation and classification, neglecting the semantic information of the images. This oversight can lead to a lack of semantic details. Although the RSS method utilises semantic information to construct the semantic space during the image representation process, it still relies on the traditional SC method in the encoding stage. This reliance leads to instability and a lack of discriminative original image representation. In contrast, the proposed method incorporates non-negativity, locality, Laplacian regularisation, histogram intersection, and semantic information. This comprehensive approach preserves more features and ensures consistency in encoding among similar features, ultimately enhancing the performance of image classification. Overall, the HS-NLLSC method significantly improves the accuracy of image classification.

As shown in Table 6, the average classification accuracy of the HS-NLLSC method is higher than that of the other methods, indicating that this algorithm performs well overall on these four datasets. Furthermore, except for slightly higher variances on the Caltech-256 (45) and Caltech-101 (30) datasets, the variances of classification accuracy are generally low, demonstrating that our algorithm is relatively stable and exhibits strong robustness.

For a more intuitive comparison, the classification results have been transformed into Figure 6. Figure 6 displays the classification results, including the average value \pm standard deviation, for seven different methods across four standard image datasets. It is evident that the HS-NLLSC method has significantly enhanced the classification accuracy, ranging from about 5% to 19% compared to several other methods.

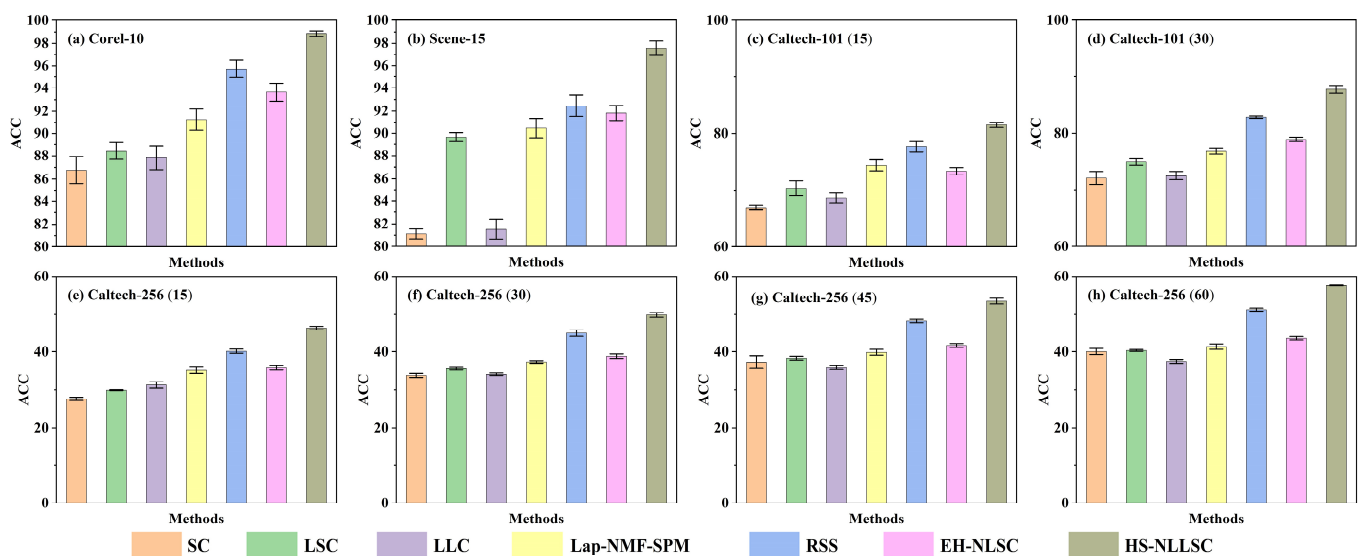


Figure 6. Classification accuracy (average value \pm standard deviation) for seven different methods in four standard image datasets.

Focused on ship classification, the analysis is centred on the Seaship and Singapore Maritime datasets, as outlined in Table 7, which illustrates the average classification accuracy for these maritime datasets. To offer a clearer visual representation of the classification outcomes, the data from Table 6 have been transformed into a visual depiction, as shown in Figure 7.

Table 7. Classification accuracy (average value \pm standard deviation) on the three maritime datasets (%).

Training Images	50	100	150	200	250
MID	100	100	100	100	100
Seaship	78.62 \pm 0.72	89.28 \pm 0.26	92.99 \pm 0.13	94.62 \pm 0.23	94.61 \pm 0.43
Seaship-trimming	87.59 \pm 0.67	92.16 \pm 0.62	95.33 \pm 0.23	96.64 \pm 0.16	97.63 \pm 0.23
SMD-trimming	93.22 \pm 0.36	96.41 \pm 0.41	97.73 \pm 0.29	98.97 \pm 0.15	99.28 \pm 0.08

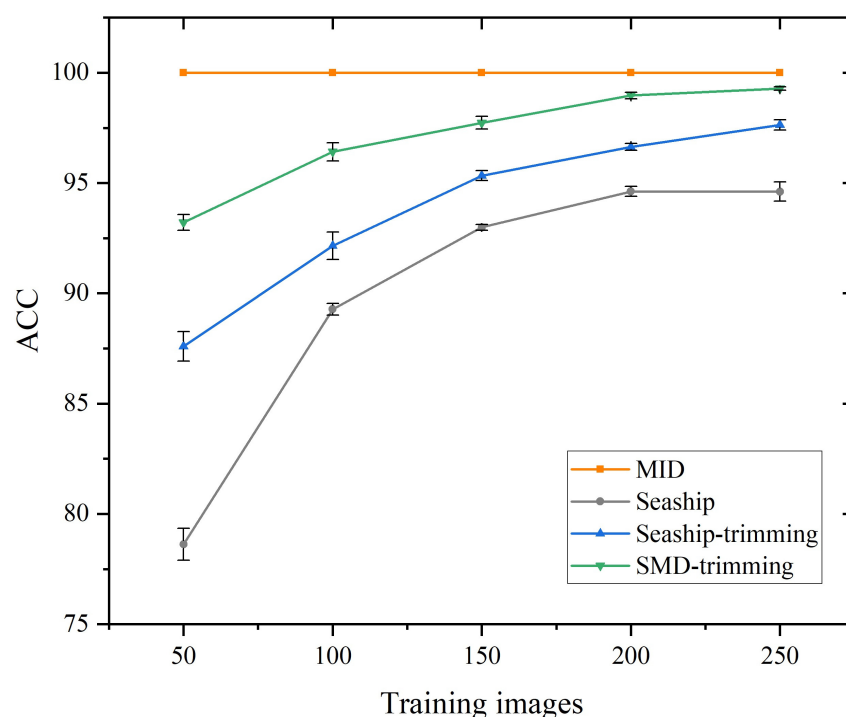


Figure 7. Classification accuracy (average value \pm standard deviation) for the three maritime datasets.

From Table 7 and Figure 7, it becomes apparent that the HS-NLLSC method yields favourable classification results across the three maritime datasets. Generally, with an increase in the number of training images, there is an observable improvement in classification accuracy, typically ranging approximately from 1% to 16%. Notably, for the MID, the classification accuracy remains at 100%, regardless of the number of training images.

4.3.3. Sensitivity Analysis of Different Parameters

In this experiment, the influence of various parameter settings on the classification accuracy across four standard datasets is examined. Specifically, different values are assigned to the parameters λ and β , namely, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, and 0.4, and the corresponding classification accuracies are illustrated in Figure 8. From Figure 8, we can see that the classification accuracy is highest when $\lambda = 0.4$ and $\beta = 0.2$.

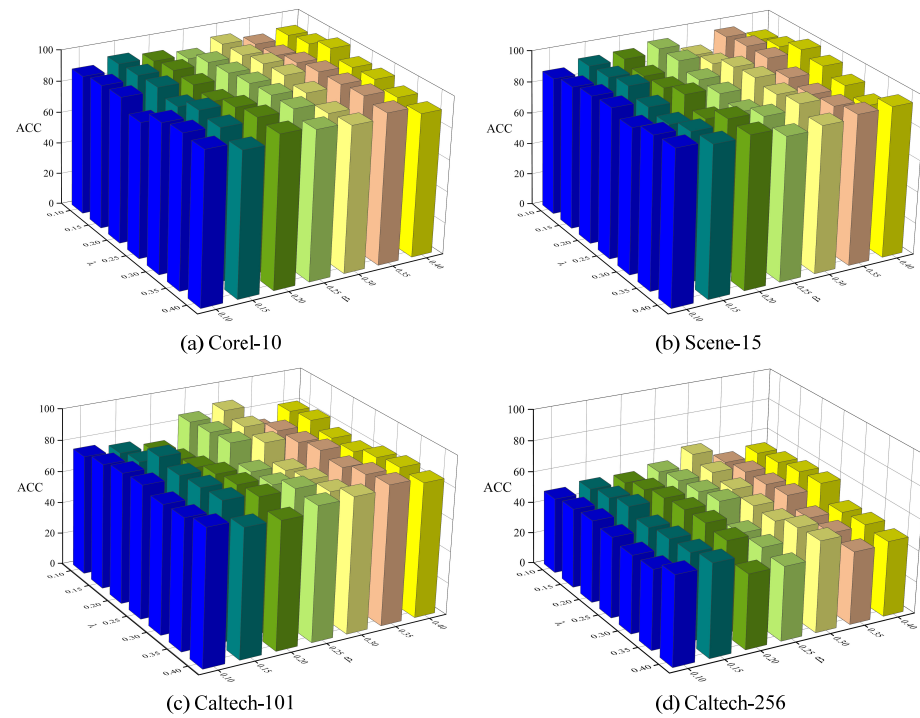


Figure 8. The impact of λ and β on the classification results.

4.4. Algorithm Stability Analysis

The image representations in Caltech-256 of the NMF, SC, and LLC methods are compared with HS-NLLSC, as shown in Figure 9.

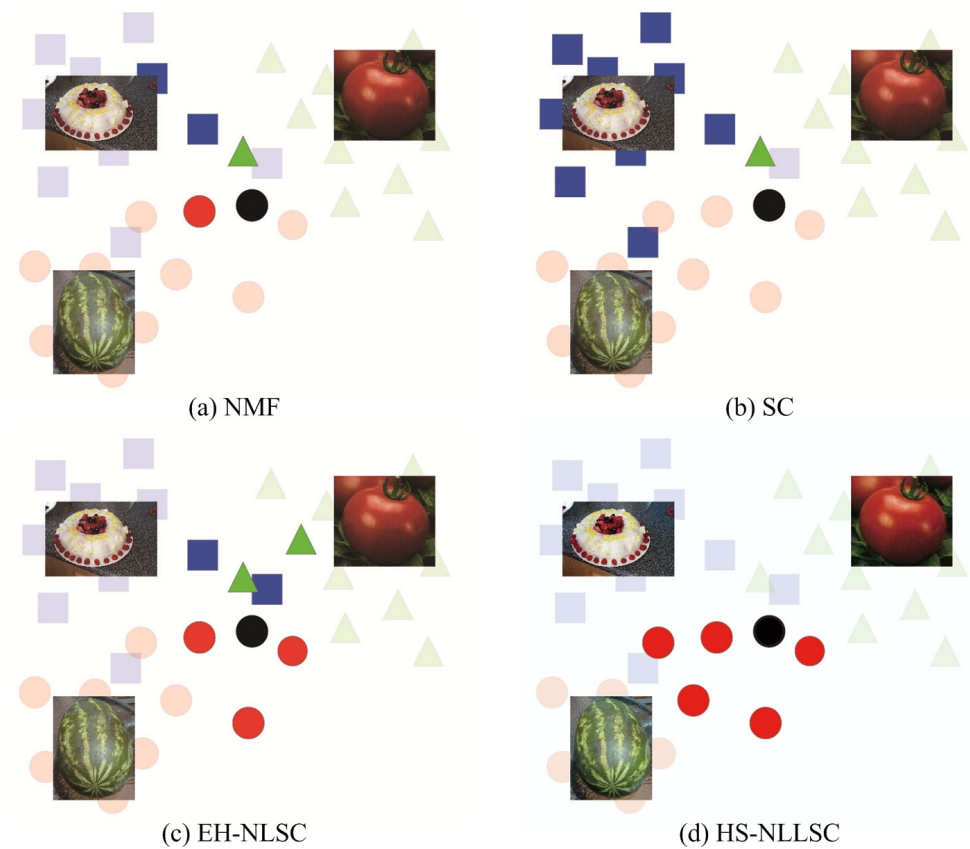


Figure 9. Image representations of four different methods in Caltech-256.

As depicted in Figure 9, the target data are represented by black circles, while data from three distinct image categories (watermelon, cake, and tomato) in the Caltech-256 dataset are denoted by red circles, blue squares, and green triangles, respectively. Figure 9a illustrates the effect of non-negative constraints applied to both the dictionary and encoding in NMF, resulting in non-zero coefficients appearing only in specific regions. However, these coefficients may lack sparsity within the same region. In Figure 9b, the image representation of the SC method is displayed, where only a few coefficients are non-zero for a given target data, leading to sparsity in the coefficient vector. In contrast, Figure 9c depicts the representation generated by the EH-NLSC method, which lacks semantic information and tends to select codewords near the input feature matrix for encoding. Finally, Figure 9d showcases the image representation produced by the HS-NLLSC method, which incorporates non-negativity, locality, and semantic information. This approach ensures similarity between the input data and neighbouring codewords, enhancing the stability and consistency of the encoding process. By addressing the limitations observed in other methods, the HS-NLLSC method offers a more robust and comprehensive image representation.

4.5. Algorithm Complexity Analysis

Given the number of local features in an image (N), the number of template features ($N1$), and the size of the dictionary (M), the total complexity of similarity calculation between all local features and template features is $o(N \times N1)$. The complexity of the feature sign search algorithm is $o(N \times M)$. Therefore, the overall complexity of the coding stage in LSC is $o(N \times N1 + N \times M) = o(N \times (N1 + M))$.

After adding histogram intersection and local constraints, the complexity of the HI-NLLSC coding stage is $o(M^2)$. The computational complexity of the non-negativity constraint is $o(M + N)$. Thus, the total complexity of the HI-NLLSC encoding stage is $o(N \times (N1 + M) + M^2)$. In the MP stage, since the SPM process involves the number of pyramid levels ($pLevels$) and the number of histogram bins ($nBins$), the complexity of this process is $o(N + pLevels \times nBins)$. Hence, the total complexity of the HI-NLLSC stage is $o(N \times (N1 + M) + M^2 + pLevels \times nBins)$.

Regarding the semantic information stage, given the number of cross-validation folds ($nRounds$) and the number of categories in the image dataset (C) for SVM classification, the complexity of this stage is $o(nRounds \times C)$. In summary, the overall computational complexity of the proposed HS-NLLSC algorithm is $o(N \times (N1 + M) + M^2 + pLevels \times nBins + nRounds \times C)$.

The complexity of the above-mentioned different stages is listed in Table 8.

Table 8. Complexity analysis of different stages.

Different Stages	Complexity
Coding stage in LSC	$o(N \times N1 + N \times M) = o(N \times (N1 + M))$
HI-NLLSC coding stage	$o(N \times (N1 + M) + M^2)$
MP stage	$o(N + pLevels \times nBins)$
Overall HI-NLLSC stage	$o(N \times (N1 + M) + M^2 + pLevels \times nBins)$
Semantic information stage	$o(nRounds \times C)$
HS-NLLSC algorithm	$o(N \times (N1 + M) + M^2 + pLevels \times nBins + nRounds \times C)$

5. Conclusions

This study presents HS-NLLSC, an innovative approach to image classification that addresses key limitations of traditional sparse coding methods, such as their inability to effectively link image representation with classification and to fully capture the rela-

tionships between features and dictionaries. By integrating non-negativity, locality, and Laplacian regularisation, HS-NLLSC improves feature retention and ensures the coherence and interdependence of the coding process. A major advancement in HS-NLLSC is its use of histogram intersection to accurately measure the similarity between feature vectors and codebooks, enabling it to construct a semantic space that bridges the gap between image representation and classification. This comprehensive strategy allows for a contextual and semantic representation of images aligned closely with classification objectives.

The key findings of this study demonstrate that HS-NLLSC provides a more precise and comprehensive depiction of original images. By leveraging the similarity and interdependence of local features, the method enhances classification accuracy. Its effectiveness is validated across four benchmark image datasets, where it outperforms existing methods. Additionally, its robust classification capabilities are confirmed through its application to three maritime datasets, underscoring its versatility and practical utility.

Despite its promising results, HS-NLLSC has some limitations. It can be sensitive to noise and outliers, which may affect stability and performance. Additionally, its computational demands and storage requirements pose challenges for large datasets. To address these challenges and expand its applicability, future research could focus on the following areas: (1) Develop strategies to improve the robustness of sparse coding, enabling it to effectively handle noise and outliers, thus enhancing stability and performance. (2) Further explore methods to integrate richer semantic information into the classification process, enabling more accurate and meaningful image representation. These future directions aim to overcome current limitations, broaden the application scope of HS-NLLSC, and enhance its performance in various domains. By addressing these areas, HS-NLLSC has the potential to become a cornerstone in the field of sparse coding and image classification, contributing to advancements in both academic research and practical applications.

Author Contributions: Conceptualization, Y.S. and H.L.; methodology, Y.S., Y.W. and H.L.; software, Y.S. and H.L.; validation, Y.S., X.W. and H.L.; formal analysis, Y.S., Y.W. and X.W.; investigation, Y.S. and Y.W.; resources, H.L.; data curation, Y.S., Y.W. and H.L.; writing—original draft preparation, Y.S. and H.L.; writing—review and editing, Y.S., Y.W., X.W. and H.L.; visualization, Y.S., X.W. and H.L.; project administration, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 52101399).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

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

References

1. Liu, S.; Zhang, Q.; Huang, L. Graphic Image Classification Method Based on an Attention Mechanism and Fusion of Multilevel and Multiscale Deep Features. *Comput. Commun.* **2023**, *209*, 230–238. [[CrossRef](#)]
2. Gao, X.; Xiao, Z.; Deng, Z. High Accuracy Food Image Classification via Vision Transformer with Data Augmentation and Feature Augmentation. *J. Food Eng.* **2024**, *365*, 111833. [[CrossRef](#)]
3. Hu, J.; Zhong, W.; Zhang, M.; Kang, S.; Yan, O. EiGAN: An Explicitly and Implicitly Feature-Aligned GAN for Degraded Image Classification. *Pattern Recognit. Lett.* **2024**, *178*, 195–201. [[CrossRef](#)]
4. Li, Y.; Huang, Y.; He, N.; Ma, K.; Zheng, Y. Improving Vision Transformer for Medical Image Classification via Token-Wise Perturbation. *J. Vis. Commun. Image Represent.* **2024**, *98*, 104022. [[CrossRef](#)]

5. Yan, F.; Yan, B.; Liang, W.; Pei, M. Token Labeling-Guided Multi-Scale Medical Image Classification. *Pattern Recognit. Lett.* **2024**, *178*, 28–34. [[CrossRef](#)]
6. Mahmood, M.J.; Raj, P.; Agarwal, D.; Kumari, S.; Singh, P. SPLAL: Similarity-Based Pseudo-Labeling with Alignment Loss for Semi-Supervised Medical Image Classification. *Biomed. Signal Process. Control* **2024**, *89*, 105665. [[CrossRef](#)]
7. Lv, Z.; Guo, H.; Zhang, L.; Liang, D.; Zhu, Q.; Liu, X.; Zhou, H.; Liu, Y.; Gou, Y.; Dou, X.; et al. Urban Public Lighting Classification Method and Analysis of Energy and Environmental Effects Based on SDGSAT-1 Glimmer Imager Data. *Appl. Energy* **2024**, *355*, 122355. [[CrossRef](#)]
8. Mushtaq, Z.; Su, S.-F.; Tran, Q.-V. Spectral Images Based Environmental Sound Classification Using CNN with Meaningful Data Augmentation. *Appl. Acoust.* **2021**, *172*, 107581. [[CrossRef](#)]
9. Koulali, R.; Zaidani, H.; Zaim, M. Image Classification Approach Using Machine Learning and an Industrial Hadoop Based Data Pipeline. *Big Data Res.* **2021**, *24*, 100184. [[CrossRef](#)]
10. Sajitha, P.; Andrushia, A.D.; Anand, N.; Naser, M.Z. A Review on Machine Learning and Deep Learning Image-Based Plant Disease Classification for Industrial Farming Systems. *J. Ind. Inf. Integr.* **2024**, *38*, 100572. [[CrossRef](#)]
11. Cui, Y.; Chen, R.; Chu, W.; Chen, L.; Tian, D.; Li, Y.; Cao, D. Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 722–739. [[CrossRef](#)]
12. Li, G.; Ji, Z.; Chang, Y.; Li, S.; Qu, X.; Cao, D. ML-ANet: A Transfer Learning Approach Using Adaptation Network for Multi-Label Image Classification in Autonomous Driving. *Chin. J. Mech. Eng.* **2021**, *34*, 78. [[CrossRef](#)]
13. Mishra, N.; Kumar, A.; Choudhury, K. Deep Convolutional Neural Network Based Ship Images Classification. *Def. Sci. J.* **2021**, *71*, 200–208. [[CrossRef](#)]
14. Leonidas, L.A.; Jie, Y. Ship Classification Based on Improved Convolutional Neural Network Architecture for Intelligent Transport Systems. *Information* **2021**, *12*, 302. [[CrossRef](#)]
15. Petković, M.; Vujović, I.; Kaštelan, N.; Šoda, J. Every Vessel Counts: Neural Network Based Maritime Traffic Counting System. *Sensors* **2023**, *23*, 6777. [[CrossRef](#)] [[PubMed](#)]
16. Ji, X.; Wang, J.; Yan, Z. A Stock Price Prediction Method Based on Deep Learning Technology. *Int. J. Crowd Sci.* **2021**, *5*, 55–72. [[CrossRef](#)]
17. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the State-of-the-Art Technologies of Semantic Segmentation Based on Deep Learning. *Neurocomputing* **2022**, *493*, 626–646. [[CrossRef](#)]
18. Khalil, I.; Mehmood, A.; Kim, H.; Kim, J. OCTNet: A Modified Multi-Scale Attention Feature Fusion Network with InceptionV3 for Retinal OCT Image Classification. *Mathematics* **2024**, *12*, 3003. [[CrossRef](#)]
19. Ketkar, N.; Moolayil, J. Convolutional Neural Networks. In *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*; Ketkar, N., Moolayil, J., Eds.; Apress: Berkeley, CA, USA, 2021; pp. 197–242. ISBN 978-1-4842-5364-9.
20. Yu, C.; Han, R.; Song, M.; Liu, C.; Chang, C.-I. Feedback Attention-Based Dense CNN for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
21. Özkaraça, O.; Bağrıaçık, O.İ.; Gürüler, H.; Khan, F.; Hussain, J.; Khan, J.; Laila, U. Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images. *Life* **2023**, *13*, 349. [[CrossRef](#)]
22. Shamshad, F.; Khan, S.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in Medical Imaging: A Survey. *Med. Image Anal.* **2023**, *88*, 102802. [[CrossRef](#)] [[PubMed](#)]
23. Roy, S.K.; Deria, A.; Shah, C.; Haut, J.M.; Du, Q.; Plaza, A. Spectral-Spatial Morphological Attention Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [[CrossRef](#)]
24. Zhou, W.; Dou, P.; Su, T.; Hu, H.; Zheng, Z. Feature Learning Network with Transformer for Multi-Label Image Classification. *Pattern Recognit.* **2023**, *136*, 109203. [[CrossRef](#)]
25. Sahoo, S.; Kumar, S.; Abedin, M.Z.; Lim, W.M.; Jakhar, S.K. Deep Learning Applications in Manufacturing Operations: A Review of Trends and Ways Forward. *J. Enterp. Inf. Manag.* **2022**, *36*, 221–251. [[CrossRef](#)]
26. Fang, W.; Zhuo, W.; Song, Y.; Yan, J.; Zhou, T.; Qin, J. Δ free-LSTM: An Error Distribution Free Deep Learning for Short-Term Traffic Flow Forecasting. *Neurocomputing* **2023**, *526*, 180–190. [[CrossRef](#)]
27. Dhar, T.; Dey, N.; Borra, S.; Sherratt, R.S. Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust. *IEEE Trans. Technol. Soc.* **2023**, *4*, 68–75. [[CrossRef](#)]
28. Luo, F.; Huang, Y.; Tu, W.; Liu, J. Local Manifold Sparse Model for Image Classification. *Neurocomputing* **2020**, *382*, 162–173. [[CrossRef](#)]
29. Zhang, B.; Liu, J. Discriminative Convolutional Sparse Coding of ECG Signals for Automated Recognition of Cardiac Arrhythmias. *Mathematics* **2022**, *10*, 2874. [[CrossRef](#)]
30. Duan, Y.; Wang, N.; Zhang, Y.; Song, C. Tensor-Based Sparse Representation for Hyperspectral Image Reconstruction Using RGB Inputs. *Mathematics* **2024**, *12*, 708. [[CrossRef](#)]

31. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20 June 2009; pp. 1794–1801.
32. Gao, S.; Tsang, I.W.-H.; Chia, L.-T.; Zhao, P. Local Features Are Not Lonely—Laplacian Sparse Coding for Image Classification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3555–3561.
33. Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-Constrained Linear Coding for Image Classification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
34. Min, H.; Liang, M.; Luo, R.; Zhu, J. Laplacian Regularized Locality-Constrained Coding for Image Classification. *Neurocomputing* **2016**, *171*, 1486–1495. [[CrossRef](#)]
35. Liu, G.; Liu, Y.; Guo, M.; Liu, P.; Wang, C. Non-Negative Locality-Constrained Linear Coding for Image Classification. In *Intelligence Science and Big Data Engineering. Image and Video Data Engineering, Proceedings of the ISClDE 2015, Suzhou, China, 14–16 June 2015*; He, X., Gao, X., Zhang, Y., Zhou, Z.-H., Liu, Z.-Y., Fu, B., Hu, F., Zhang, Z., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 462–471.
36. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)] [[PubMed](#)]
37. Zhang, L.; Chen, Z.; Zheng, M.; He, X. Robust Non-Negative Matrix Factorization. *Front. Electr. Electron. Eng. China* **2011**, *6*, 192–200. [[CrossRef](#)]
38. Hoyer, P.O. Non-Negative Sparse Coding. In Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, Martigny, Switzerland, 6 September 2002; pp. 557–565.
39. Cai, D.; He, X.; Han, J.; Huang, T.S. Graph Regularized Non-negative Matrix Factorization for Data Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1548–1560. [[CrossRef](#)]
40. Han, H.; Liu, S.; Gan, L. Non-Negativity and Dependence Constrained Sparse Coding for Image Classification. *J. Vis. Commun. Image Represent.* **2015**, *26*, 247–254. [[CrossRef](#)]
41. Wu, J.; Rehg, J.M. Beyond the Euclidean Distance: Creating Effective Visual Codebooks Using the Histogram Intersection Kernel. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 630–637.
42. Chen, H.; Xie, K.; Wang, H.; Zhao, C. Scene Image Classification Using Locality-Constrained Linear Coding Based on Histogram Intersection. *Multimed. Tools Appl.* **2018**, *77*, 4081–4092. [[CrossRef](#)]
43. Yuan, W.; Jinghui, Z.; Zhiping, C.; Xiaojing, M. Non-Negative Local Sparse Coding Algorithm Based on Elastic Net and Histogram Intersection. *J. Comput. Appl.* **2019**, *39*, 706. [[CrossRef](#)]
44. Zhang, C.; Liu, J.; Tian, Q.; Liang, C.; Huang, Q. Beyond Visual Features: A Weak Semantic Image Representation Using Exemplar Classifiers for Classification. *Neurocomputing* **2013**, *120*, 318–324. [[CrossRef](#)]
45. Rasiwasia, N.; Vasconcelos, N. Scene Classification with Low-Dimensional Semantic Spaces and Weak Supervision. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–6.
46. Rasiwasia, N.; Vasconcelos, N. Image Retrieval Using Query by Contextual Example. In Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, Vancouver, BC, Canada, 30 October 2008; Association for Computing Machinery: New York, NY, USA; pp. 164–171.
47. Zhang, C.; Zhu, X.; Li, L.; Zhang, Y.; Liu, J.; Huang, Q.; Tian, Q. Joint Image Representation and Classification in Random Semantic Spaces. *Neurocomputing* **2015**, *156*, 79–85. [[CrossRef](#)]
48. Shen, F.; Zeng, G. Semantic Image Segmentation via Guidance of Image Classification. *Neurocomputing* **2019**, *330*, 259–266. [[CrossRef](#)]

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.