

Detection of Iranian Traffic Signs Using HOG Features and Sparse Classification Based on Fisher Criterion and Locally Linear Embedding (LLE)

Amir bahador bayat

Master of Electronics, Iran, Tehran,

amirbahador87@ymail.com

Abstract

The detection and recognition of traffic signs is a fundamental component of advanced driver assistance systems (ADAS), contributing significantly to road safety, driving comfort, and the development of autonomous driving technologies. Despite extensive research in this domain, accurate and real-time traffic sign recognition remains challenging due to varying illumination conditions, occlusions, scale variations, complex backgrounds, and the diversity of traffic sign appearances in real-world environments. To address these challenges, this study proposes a novel two-stage approach for efficient and real-time traffic sign detection and recognition under actual traffic conditions. In the first stage, traffic sign detection and initial classification are performed using Histogram of Oriented Gradients (HOG) features combined with a sparse classification framework, enabling robust feature representation and efficient discrimination between sign and non-sign regions. In the second stage, a more fine-grained recognition process is applied, where traffic signs are further categorized into subclasses using sparse classification enhanced by the Fisher criterion and Locally Linear Embedding (LLE), which improves intra-class compactness and inter-class separability. The proposed method is evaluated using the Iranian Traffic Sign Detection Benchmark Dataset, which contains 2,182 annotated traffic sign images spanning multiple categories. Experimental results demonstrate that the proposed framework achieves a high recognition accuracy of 99.2%, while maintaining significantly lower computational complexity compared to existing approaches in the literature. Moreover, the system demonstrates strong real-time performance, with an average processing time of 138 milliseconds per frame, corresponding to approximately 7.2 frames per second. Overall, the results indicate that the proposed two-stage method provides an effective balance between accuracy and computational efficiency, making it suitable for real-time ADAS applications and potential deployment in intelligent transportation systems.

Keywords: Traffic Sign Detection, HOG Features, Sparse Representation, Fisher Criterion, Locally Linear Embedding

1. Introduction

In modern road infrastructure, traffic signs are essential components placed along road corridors to guide transportation activities. Designed with specific colors, shapes, characters, and symbols, traffic signs convey critical signals to road users, manage driving behaviors, and ensure efficient navigation. By providing accurate and timely information, traffic signs enable road users to reach their destinations quickly, facilitate smoother traffic flow, and enhance driving safety. Furthermore, traffic signs support law enforcement efforts by traffic management authorities [1]. The development and expansion of intelligent transportation systems, particularly in traffic sign recognition, have been significant challenges in recent years. Traffic sign detection, a cornerstone of intelligent transportation systems, aims to improve road safety and efficiency by interpreting signs captured in images or videos by cameras [2].

Vehicle-mounted or roadside tools transmit information, such as speed limits, warning signs, and directional signs, to drivers, improving navigation efficiency and safety [3]. Traffic Sign Recognition (TSR) involves locating traffic signs in digital images or video frames and classifying them into specific categories [4]. TSR methods primarily rely on visual cues, such as the shape and color of signs. However, conventional TSR algorithms face limitations in real-time testing due to factors such as lighting conditions, camera angles, occlusions, driving speed, and other environmental variables. Achieving multi-target detection is particularly challenging, as visual objects can be missed due to slow detection processes [5].

Given the importance and widespread use of traffic signs, significant attention has been devoted to measuring traffic signs using advanced techniques. Initially, traffic sign measurement relied on manual on-site operations by trained workers, which were inefficient and labor-intensive, especially for large road networks. In recent decades, advancements in imaging sensors—offering improved resolution, quality, and affordability—have enabled the effective use of image data for traffic sign measurement tasks. Rapidly capturing traffic sign images along road corridors significantly reduces manpower requirements and enhances operational efficiency. Consequently, numerous strategies for intelligently interpreting road scene images for high-performance traffic sign measurement have been proposed in the literature. Traffic sign detection remains a trending topic and serves as a prerequisite for a wide range of applications. Specifically, it involves localizing traffic sign instances to identify their presence and determining their types to interpret their signals.

To date, various algorithms for traffic sign detection with high accuracy and efficiency have been developed [5-6], some of which have been applied in real-world scenarios. However, challenges persist due to the diversity of imaging sensors, variations in illumination, weather conditions, and the physical state of traffic signs (e.g., size, shape, type, deformation, perspective, and occlusion). The level of automation and accuracy in traffic sign detection still lags behind human-level performance, creating a noticeable gap. Achieving higher detection

accuracy and fewer misidentification errors with fully automated processing schemes, compared to human manual detection, remains a significant challenge under varying conditions. Thus, reliability and practicality must be improved through more advanced solutions to expand the integration and value of traffic sign detection modules in real-world transportation applications.

Traffic signs vary in shape and color and are generally categorized into five main classes: warning signs (typically red triangles), prohibition signs (typically red circles), reservation signs (typically blue rectangles), mandatory signs (blue circles), and temporary signs (yellow triangles). The primary distinction of Persian traffic signs lies in the use of Persian text or English numerals within the signs, with some signs featuring both Persian instructions and their English translations. Certain Persian letters include dots above or below, which, if omitted, alter the letter and its meaning. These dots are integral to the sign's meaning and must not be overlooked or removed by TSR systems.

The cost-saving and life-saving benefits of developing a Persian TSR system are evident, particularly for applications in autonomous vehicles, where traffic sign detection is a key component. While autonomous vehicle technology has advanced recently, it still lacks optimized universal detection models for both traffic signs and the broader vehicle environment. This necessitates ongoing development of traffic sign detection models suitable for practical applications. Beyond autonomous vehicles, traffic sign detection has other applications, such as assisting visually impaired pedestrians. Individuals with visual impairments may use advanced walking canes or guide dogs [1-3], but they still require verbal notifications of traffic signs. This technology can also be adapted to detect other signs for visually impaired individuals. The motivation for this work, as presented in this paper, is clear: it offers utility across multiple applications, providing significant benefits for both autonomous vehicles and accessibility for the visually impaired. Researchers from various countries and languages must focus on traffic signs in different languages, as applications are not geographically restricted. Such systems can also alert drivers to overlooked signs, potentially preventing accidents.

TSR research not only helps prevent road accidents and protect drivers but also facilitates efficient and accurate inspection of traffic signs, reducing unnecessary manpower and resources. Additionally, it provides technical support for unmanned and assisted driving. Thus, research on automated detection is of paramount importance and holds significant value for daily life [1].

This paper primarily investigates achieving an accurate and real-time TSR model based on sparse representation. The main objective is to develop an improved model for classifying and detecting Persian traffic sign images, with the ultimate goal of creating a model applicable to practical scenarios for societal benefit. Our contributions are threefold. First, we collect and augment sample images to create a new dataset for Persian traffic signs, comprising 2,182

images across 26 classes. Second, we extract image features using the latest HOG technique and evaluate TSR performance on our dataset using sparse classification based on the Fisher criterion and LLE. Finally, we conduct a detailed comparison of TSR performance with other algorithms and report the results.

The proposed flowchart of our study is presented in Figure (1), with its components detailed in Sections 3–5.

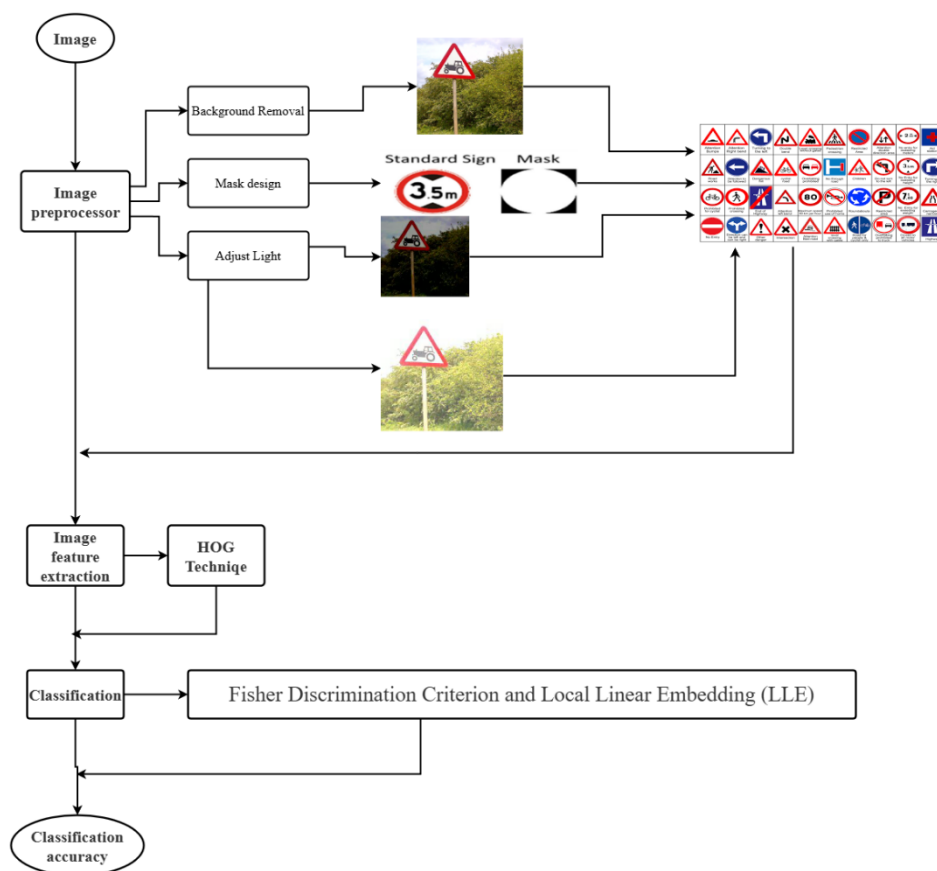


Figure (1): Proposed Flowchart

The paper is organized as follows: Section 2 reviews the literature. Section 3 describes the data. Section 4 presents the feature extraction method, and Section 5 outlines the proposed classification method. Section 6 provides the conclusion.

2. Literature Review

In recent years, road safety has become increasingly critical, particularly in developing countries. According to the World Health Organization's 2018 Global Status Report on Road

Safety, over 1.35 million people die annually in road accidents worldwide [1], with 20 to 50 million others injured or disabled. Most road accidents stem from ignorance of traffic rules, failure to recognize traffic signs, or driver distraction. Adhering to road signs can prevent the majority of these accidents. Consequently, various initiatives, such as developing automated road sign detection systems to enhance driving safety and ease, have been adopted to improve road safety [2].

Computer vision methods for traffic sign detection can be categorized into three groups: color- and shape-based, traditional machine learning, and deep learning. These methods are introduced sequentially for a comprehensive review of related work [3].

Traffic signs are typically red, yellow, or blue, leading to several color-based detection approaches. Deshmukh et al. [9] used color thresholding to separate road sign regions from their backgrounds in RGB space. However, RGB color space is highly sensitive to varying conditions, such as vehicle lighting or natural light effects. To address this, other color spaces have been employed. Hechri and Mitbaa [9] proposed a method for traffic sign detection in YCbCr color space, converting input images from RGB to YCbCr and identifying bright red based on the Cr component range. The YUV model was used in [11] to detect blue rectangular signs. Shaposhnikov et al. [12] utilized the HSI color space, while Vitabile et al. [13] employed it to identify all possible road sign candidates. Additionally, Fleyeh [14] used an improved color space model based on brightness and saturation to extract red from images.

Given that traffic signs have regular geometric shapes, shape is another critical feature for detection, offering greater robustness to deformation and illumination changes. Several shape-based detection methods have been proposed. The most common approach involves the Hough Transform, as seen in [15], where authors used the classic Hough algorithm to detect straight lines for triangular traffic signs. Yakimov and Fursov [16] developed a generalized Hough Transform-based method for detecting triangular traffic signs in real road scenes. The main drawbacks of these techniques are high memory requirements and computationally intensive processing. To address this, Barnes and Zelinsky [17] proposed a radial symmetry detector, similar to the Hough Transform, to reduce processing time. Loy and Zelinsky [18] used fast radial symmetry detection for each region of interest (ROI) to identify circular shapes, an approach extended by Loy and Barnes [19] for regular polygons with successful implementation. However, these methods' robustness is often affected by noise and occlusions, complicating the detection process. Another notable shape classification method involves corner detection. In [20], authors identified triangular and square shapes by detecting corners in each ROI using the Harris corner detection algorithm. Additionally, pattern-matching methods for shape detection have been proposed, which are well-suited for real-time road sign detection. The core idea is to match a pattern against all images in a database to find the optimal match [21].

Rajesh et al. [22] employed a convolutional neural network (CNN) based on an end-to-end learning approach, Modified Mask R-CNN (RM R-CNN). Evaluation results showed an error rate below 3%. The performance of RM R-CNN was compared with conventional deep neural network architectures like Fast R-CNN and Mask R-CNN, achieving a precision of 97.08%, surpassing Mask R-CNN and Faster R-CNN.

Chen et al. [23] categorized traffic detection methods into three main techniques: conventional methods, deep learning, and hybrid approaches. Results indicated that hybrid detection algorithms outperformed others in true positive rates, false positive rates, number of test images, accuracy, and processing time. These findings highlight the potential of hybrid methods in developing accurate and effective TSD systems, paving the way for further research in this area.

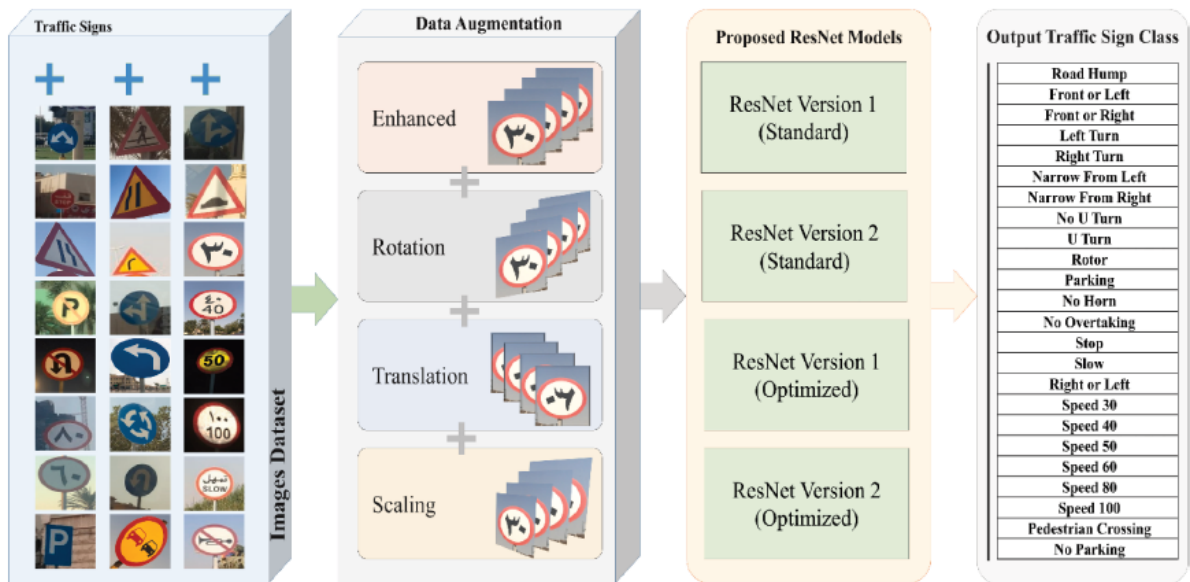


Figure (2): Workflow of the Traffic Sign Recognition Method

By understanding the various categories of traffic signs and their variations across countries, researchers can develop robust detection algorithms for recognizing and classifying different sign types. This comprehensive understanding also aids in developing and evaluating TSR systems and identifying their benchmarks. The types of traffic signs in Iran are illustrated in Figure (3):



Figure (3): Types of Traffic Signs in Iran

3. Data Collection

For this experiment, we selected eight classes of high-visibility and safety-critical traffic signs. Due to the dispersed nature of traffic signs on roads, we collected images of traffic signs instead of driving videos. These images were captured using our mobile phone cameras in the streets of our city and divided into two groups.

Our dataset comprises 2,182 traffic sign images labeled as “No U-Turn” (271 images), “Speed Bump” (329 images), “Road Work” (294 images), “Watch for Children” (176 images), “Crossing Ahead” (313 images), “Yield” (317 images), “Stop” (286 images), and “No Entry” (196 images), as shown in Table 1.

Table (1): Types of Traffic Signs

ClassId	Name
0	Speed limit (20 km/h)
1	Speed limit (30 km/h)

2	Speed limit (50 km/h)
3	Speed limit (60 km/h)
4	Speed limit (70 km/h)
5	Speed limit (80 km/h)
6	End of speed limit (80 km/h)
7	Speed limit (100 km/h)
8	Speed limit (120 km/h)
9	No passing
10	No passing for vehicles over 3.5 metric tons
11	Right-of-way at the next intersection
12	Priority road
13	Yield
14	Stop
15	No vehicles
16	Vehicles over 3.5 metric tons prohibited
17	No entry
18	General caution
19	Dangerous curve to the left
20	Dangerous curve to the right
21	Double curve
22	Bumpy road
23	Slippery road
24	Road narrows on the right
25	Road work
26	Traffic signals

3.1 Dataset Augmentation

For the raw data in our dataset, some images were captured in landscape orientation. We first used JPEG Autorotate software to rotate images to portrait orientation. Due to the high resolution of images leading to excessively long training times, we resized the images while maintaining the same aspect ratio between width and height. Thus, all images in our dataset were normalized to 1128×2016 and 1536×2048 pixels. Each label consists of five parameters: classification index, center point coordinates (x, y), width (w, $w \geq x \geq 1$), and height (h, $h \geq y \geq 1$). After completing all labeling tasks, we grouped the images. The dataset was split into training and testing sets with an 8:2 ratio. Images and their corresponding annotation files were organized into respective folders.

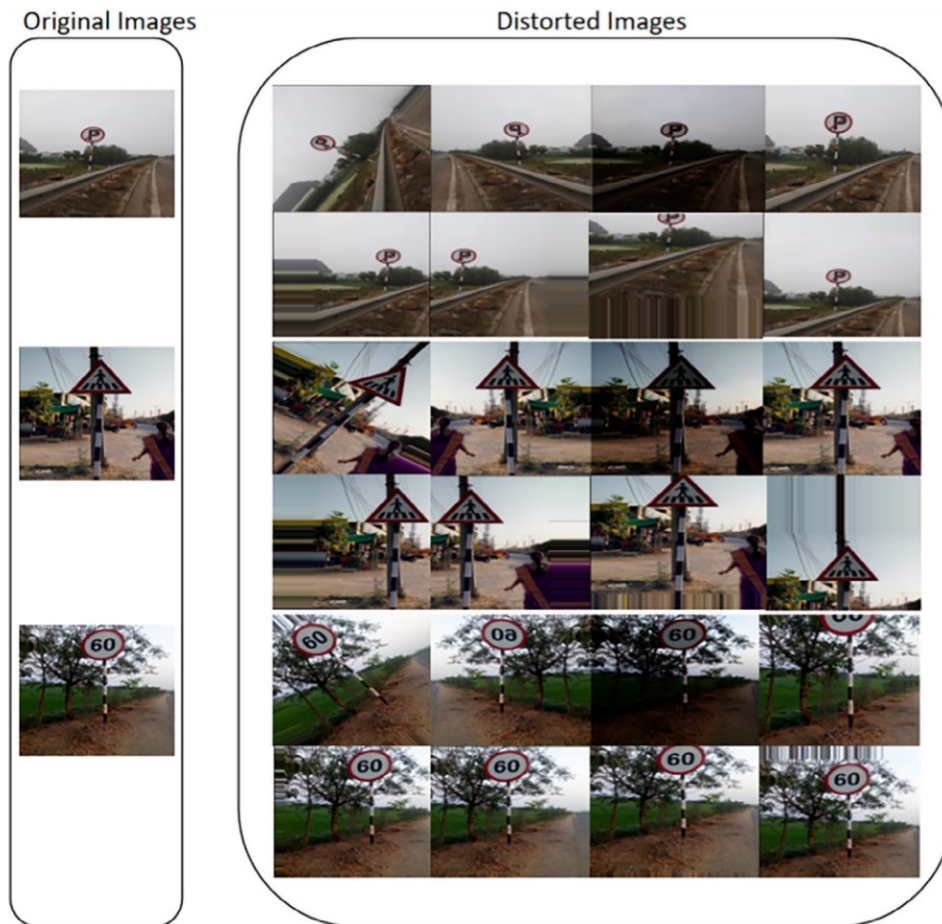


Figure (4): Examples of Data Augmentation Techniques

4. Feature Extraction

Initially, background images are removed from the raw data. The first stage involves preprocessing the traffic dataset to prepare it for the proposed model. The dataset is then augmented using techniques such as enhancement, rotation, translation, and scaling to address class size and complexity. The article presents a detailed architecture outlined below. The model's output classifies traffic signs into various categories, including speed bumps, left turns, right turns, and others. It should be noted that the methodology in Figure 1 does not contribute independently. However, the primary contribution of this paper lies in two classification optimizations.

4.1. Image Filtering

The main objective of this stage is to eliminate false candidates mistakenly identified as potential traffic signs, which may originate from advertisements, vehicles, or other objects with similar colors. To remove single pixels erroneously identified as noise, morphological operations are applied to segmented images. The two basic operations used in this paper are erosion and dilation. Figure 5 illustrates this process for a binary segmented image.

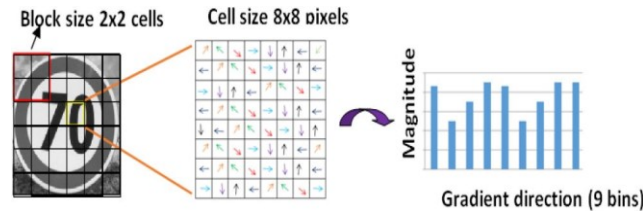


Figure (5): HOG Feature Extraction Process

To improve segmentation results, many identified regions of interest are eliminated based on their size and aspect ratio. Regions with excessively small or large areas are considered false candidates. Large regions are unsuitable for representing traffic signs, while small regions lead to unsuccessful identification. The minimum accepted area and the aspect ratio between the horizontal and vertical sides of the bounding box are defined as follows:

$$\left\{ \begin{array}{l} 16000 \text{ pixels} > \text{area} > 2000 \text{ pixel} \\ 0.4 < \text{ration} < 1.9 \end{array} \right\} \quad (1)$$

4.2. Feature Extraction

One of the techniques used for feature extraction is the Histogram of Oriented Gradients (HOG), proposed by Dalal and Triggs [38] for pedestrian detection. We utilized the HOG

technique (see Figure 3) to extract features from detected regions by analyzing the magnitude and orientation distribution of gradients. To compute HOG features, the traffic sign patch is divided into small regions called cells [24-25]. These cells are grouped into blocks, and block normalization is performed on the final histograms. To calculate the image gradient for each pixel in horizontal and vertical directions, the following equations were used:

$$\begin{aligned} G_x(x, y) &= H(x+1, y) - H(x-1, y) \\ G_y(x, y) &= H(x, y+1) - H(x, y-1) \end{aligned} \quad (2)$$

Thus, the magnitude $G(x, y)$ and orientation $\alpha(x, y)$ of pixel (x, y) were calculated as follows:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (3)$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (4)$$

In this work, detected regions were first normalized to 36×36 pixels. The normalized patches were then divided into 8×8 blocks, resulting in a total of 64 blocks. Each block was further divided into 2×2 cells, each containing 4×4 pixels, yielding 256 normalized histograms. Each histogram had 9 orientations, resulting in a feature vector with 2,304 elements for each image.

5. Sparse representation

Recently, sparse representation has been used widely in applications such as image sampling, compressing, representation, recovering and classification [26] and [27]. The success of sparse representation is due to the fact that by considering some specific basics, most of the natural signals such as image or voice possess the sparse representation. This method was first presented by Olshausen [28].

Assume $X = [x_1 \dots x_n]^T \in R^n$. This signal can represent an image. Sparsity implies that this information can be stated as a linear combination of a few basis. If k represent the number of bases which are important in representing the data, then ($k \ll n$).

Bases used to represent the data, form a Dictionary matrix. Column of this matrix are used to form the whole or part of the data. Each column of the dictionary matrix is called an atom.

If atoms of the dictionary were increased, then it would become an over-complete dictionary. If D represent the dictionary, then the data x in $x = D\alpha$ is stated by linear combination of atoms and α represents the linear combination. Because the dictionary is over-complete, the linear system of equations is underdetermined for α and it will have infinite number of solutions. Since we are looking for sparsest representation, we use an additional condition.

More precisely, we have to add a limitation to the problem to limit the number of non-zero components of sparse representation. The solution of optimization equation (5), yields the sparse representation of signal x .

$$\min_{\alpha} \|\alpha\|_0, \text{ subject to } x = D\alpha \quad (5)$$

In this equation, α is vector that contains the sparse representation coefficients of the signal. Zero norm $\|\alpha\|_0$ counts the number of non-zero components of the matrix. This number indicates the Sparsity of representation. This problem is NP-Hard. One the best methods to approximately solve this problem is OMP method [24]. An approach could be using zero-norm instead of 1-norm to solve this problem, which will change equation (6) as follows:

$$\min_{\alpha} \|x - D\alpha\|_2, \text{ s.t. } \|\alpha\|_1 = L \quad (6)$$

It has been shown that in most applications, the solution to above problem and equation (6) are equal, and that the above problem is an acceptable approximation to sparse representation of signal.

5.1. Classification by sparse representation

Most of the papers presented in sparse representation in machine vision and image processing, focus on image recovery, sparse representation of signals is also used in classification applications that aim to intensify the discrimination against different classes. Thus, the proposed methods can be divided into two groups; recovery aimed methods and discrimination aimed methods. In signal recovery problems, all of the signals are in the same class, so it is easy to define a dictionary. However, in classification problems, it is a more complex task to learn the dictionary. Because the data representation by this dictionary must improve the discrimination against different classes.

5.2. Dictionary learning by Fisher Criterion

One of the primary approaches to create a classification model is by the use of Fisher Discrimination Criterion. Fisher linear classification, transfers the data by linear transformation to a space in which the Fisher Discrimination Criterion is maximum. This criterion is calculated by dividing inter-class scatter by intra-class scatter. Thus data sets belonging to the same class will get closer to each other and farther from other classes. In 2011, a method called “discrimination dictionary learning” was presented based on Fisher Discrimination criterion. In this method, a structure has been considered for dictionary which is based on data labels [29]. We place the problem data in a column in matrix $X = [X_1, X_2, \dots, X_c]$, X_i represents the data belonging to class i . In this method a similar structure has been considered for dictionary atoms. When D represents the dictionary, we can say: $D = [D_1, D_2, \dots, D_c]$, where matrix D_i consists of atoms which are suitable for recovering the data related to class i . Based on the mentioned structured dictionary, an optimization equation is

defined for signal recovery and also separating the maximum according to Fisher discrimination criterion. Equation (7) presents the optimization of the mentioned method.

$$f(D, A) = \min_{D, A} r(D, A, x_i) + \sum_i \|A_i\|_1 + f(A) \quad (7)$$

Where A is sparse representation and f is a function defined based on sparse representation, which optimizes the Fisher discrimination criterion. Also in this equation $\|A_i\|_1$ is 1-norm summation for each matrix column. Function $r(D, A, X)$ applies the signal recovery error on optimization, considering the structured dictionary as shown in equation (8):

$$r(D, A, x_i) = \|x_i - DA_i\|^2 + \|x_i - DA_i\|^2 + \sum_{i \neq j} \|D_j A_i\|^2 \quad (8)$$

When the new dictionary representation of the data is saved in matrix X, equations (9) and (10) show the intra-class and inter-class scatter, respectively. Where m_i represents the class i average and m represents the total average.

$$S_W(X) = \sum_{i=1}^c \sum_{k \in i} (x_k - m_i)(x_k - m_i)^T \quad (9)$$

$$S_B(X) = \sum_{i=1}^c n_i (x_k - m_i)(x_k - m_i)^T \quad (10)$$

Hence the function f is defined as equation (11). The first two term are from Fisher criterion and the third term is to make the problem convex and solve the optimization.

$$f(A) = tr(S_W(A)) - tr(S_B(A)) + \eta \|A\|_F^2 \quad (11)$$

5.3. Local linear embedding LLE

One of the common problems in all sparse representation-based classification methods is ignoring the local neighborhood of the data. More precisely, these methods obtain the sparse representation of data of each class without considering which data sets are in local neighborhood of other data sets. In other words, it is possible that two neighboring data sets belonging to the same class, (X neighbors Y, if and only if Euclidean distance between X and Y is shorter than the Euclidean distance between Y and any other Data) become non-neighbors after being transferred to sparse representation space. This, increases the probability of classifier overfitting. In order to overcome the local neighborhood problem in sparse representation-based classifications, we use the local linear embedding concept. The main idea of the algorithm [30], which is mostly used in dimension reduction, is as follows:

Assume we have N data sets $[X_1, \dots, X_N]$, each of which are D-dimensional. This algorithm transfers the data by a transform matrix (reduces the dimensions of the data) so that the local distance between data sets are maintained. The method conducted by this algorithm to

maintain the distance is that it finds k-nearest neighbor of each data set, then it estimates the level of maintenance of local distance using equation (12).

$$W^* = \arg \min_{w=[w_{ij}]} \sum_{i=1}^N \|X_i - \sum_{x_j \in N_k(x_i)} W_{ij} x_j\|^2, \forall i, \sum_{x_j \in N_k(x_i)} W_{ij} x_j = 1 \quad (12)$$

Where W is a N×N matrix which represents the local neighborhood level. In [14] is used to solve the above optimization problem.

6. The proposed method

6.1. Combining LLE algorithm and sparse representation-based classification methods

Generally, in sparse representation-based classification methods, sparse representation of the data ($\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$), dictionary (D), and classifier (f) are learnt from the data. Learning these three concepts is expressed as an optimization problem. Each of the proposed methods optimize a different objective function which is represented by F (A, D, F). For example, the objective function for discrimination dictionary learning method using Fisher criterion is as equation (7). By applying LLE algorithm, the objective function will change as equation (13).

$$F'(A, D, f) = F(A, D, f) + \gamma \sum_{i=1}^N \alpha_i - \sum_i W_{ij}^* x_j \quad (13)$$

Where γ , is the regularization coefficient which is learnt through cross validation.

By adding the second right term in the above equation, we add a limitation to Sparsity codes that not only they must be able to recover the principal signals properly, but also the sparse codes correspondent to neighboring signals, must be similar. Since the term added to F is a quadratic function with respect to sparse coefficients, if F was convex, then the convex form of the objective function would be maintained, hence optimization of the above equation would be performed easily.

Since the method presented in [29] is one of the best methods in sparse representation-based classification, we presume the value of F (A, D, f) in the above equation to be equal to the proposed objective function in the paper. (Notice that the proposed method can be combined with any function F) So the proposed ultimate objective function (FDLLE) is as equation (15):

$$\min_{\alpha, D} \sum_i (\|X_i - D A_i\|^2 + \|X_i - D \alpha_i\|^2 + \sum_{i \neq j} \|D_j \alpha_i\|^2) + tr(S_W(\alpha)) - tr(S_B(\alpha)) + \lambda_2 \|\alpha\|_F^2 + \gamma \sum_{i=1}^N \alpha_i - \sum_i W_{ij}^* \alpha_j \quad (14)$$

Objective function is presented for optimization of sparse and dictionary coefficients values. The method presented in [29] is used to solve the above equation. We use [31] to obtain the primary dictionary coefficients and [32] to obtain the primary sparse coefficients. The

obtained coefficient values in each stage are put in the objective function and are reduced as long as the algorithm allows us to. Finally, the ultimate coefficients are obtained. Algorithm performance of a test data Y is as follows:

$$\alpha = \arg \min_{\alpha} \|y - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (15)$$

$$label = \arg \min_i \|y - D_i \alpha\|_2^2 + w \|\alpha - m_i\|_2^2 \quad (16)$$

First, sparse coefficients of the test data is obtained by equation (15) and then the test data label is obtained by equation (16).

The number of correct labels is counted and is considered as our proposed system performance.

7. Results

Table 2 compares the accuracy and average processing time between SVM [33], CNN [34], and the proposed methods. The average execution time for the entire system, including detection and classification stages, is 138 milliseconds, approximately 7.2 frames per second.

Table (2): Accuracy and Average Processing Time

Method	Recognition rate %	Total processing time, ms/frame
SVM	95.71	43 ms
CNN	97.63	159 ms
SPLLE	99.2	138 ms

As shown in Table 2, the average processing time of the SVM-based system is better than that of our proposed method. However, our proposed system achieved higher average accuracy compared to the SVM method. Additionally, the proposed system was faster than other CNN-based methods, making it suitable for real-time processing.

We trained and evaluated two models based on sparse representation and sparse representation with Fisher criterion and LLE for traffic sign detection using a large-scale

traffic sign dataset. The performance of the models is shown in Table 3, indicating the accuracy of each model on the validation set.

Table (3): Accuracy of Each Model

Model	Accuracy%
SP	98.7
SPLLE	99.2

The results demonstrate that both models achieved high accuracy on the validation set. The sparse representation model with Fisher criterion and LLE outperformed the sparse representation model in terms of accuracy. The higher accuracy of the SPLLE model can be attributed to its deeper architecture and ability to address the problem of deep degradation.

In an ideal scenario, our system is also capable of detecting and recognizing multiple traffic signs appearing in a single image. The primary reasons for misclassification errors include the small size of traffic signs, partial ambiguity, and color degradation. In some cases, traffic signs were unrecognizable even to humans.

This paper presents a high-accuracy algorithm for detecting Iranian traffic signs using HOG features and sparse classification based on the Fisher criterion and LLE, designed for street-level traffic sign detection. The SPLLE model employs a two-stage processing architecture. The first stage involves a precise dictionary as a feature extractor and a Fisher criterion-based separator for training traffic sign detection. The second stage focuses on recognition. Specifically, HOG features can extract robust and informative multi-scale features, significantly aiding in the localization and identification of traffic sign instances under various conditions, particularly small signs. Sparse representation is a novel method for classifying data, especially images. Its primary advantage is creating a nonlinear model for data representation. By optimizing the L1 norm, an appropriate subspace for each data point is learned. As explained earlier, one limitation of these methods is their disregard for the local structure of data after projecting it into a sparse subspace. The proposed method mitigates this issue to some extent using the LLE algorithm. With limited data, the proposed method significantly outperforms other sparse methods. While this may initially seem contradictory to the method's local focus and goal of preserving local structure, it is important to note that local data is used to train the dictionary, and learning the subspace with consideration of local structure improves classification accuracy. The proposed method achieves a higher detection rate than other methods. As mentioned in the proposed method section, sparse representation-based methods fail to preserve the local structure of data, increasing the likelihood of

overfitting on training data. However, the proposed method incorporates an LLE term to prevent the loss of local data structure.

This translation maintains the technical accuracy and structure of the original document while ensuring fluency and adherence to English academic writing conventions. Let me know if further refinements or specific adjustments are needed.

8. References

- [1] Yu Yongtao, Jiang Tao, Li Yinyin, Guan Haiyan, Li Dilong, Chen Lianghai, Yu Changhui, Gao Li, Gao Shangbing, Li Jonathan.(2022). SignHRNet: Street-level traffic signs recognition with an attentive semi-anchoring guided high-resolution network. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, Vol 192 , pp.142–160.
- [2] Hechri Ahmed, Mtibaa Abdellatif. (2020).Two-stage traffic sign detection and recognition based on SVM and conVolutional neural networks. IET Image Process., 2020, Vol. 14 Iss. 5, pp. 939-946.
- [3] Wei, Y.Z., Gao, M.Y., Xiao, J., Liu, C.X., Tian, Y.H. and He, Y. (2023). Research and Implementation of Traffic Sign Recognition Algorithm Model Based on Machine Learning. Journal of Software Engineering and Applications,2020, Vol.16, pp.193-210.
- [4]Cireşan, D., Meier, U., Masci, J., et al.(2012).Multi-column deep neural network for traffic sign classification', Neural Netw., 2012, 32, pp. 333–338.
- [5] Zhang, D. and Zhou, F.F. (2023). Self-Supervised Image Denoising for Real-World Images with Context-Aware Transformer. IEEE Access , 11, 14340-14349.
- [6] Liu, C., Li, S., Chang, F., Wang, Y., (2019). Machine vision based traffic sign detection methods: Review, analyses and perspectives. IEEE Access 7, 86578–86596.
- [7] Wali, S.B., Abdullah, M.A., Hannan, M.A., Hussain, A., Samad, S.A., Ker, P.J., Mansor, M. B., (2019). Vision-based traffic sign detection and recognition systems: Current trends and challenges. Sens. 19 (9), 2093.
- [8] World Health Organization.(2018). Global Status Report on Road Safety 2018', December 2018.
- [9] Deshmukh, V.R., Patnaik, G.K, Patil, M.E.(2013). Real-time traffic sign recognition system based on colour image segmentation', Int. J. Comput. Appl., 2013, Vol.83, (3), pp. 30–35.
- [10] Hechri, A., Mtibaa, A.: 'Automatic detection and recognition of road sign for driver assistance system'. Proc..16th IEEE Mediterranean Electrotechnical Conf. (MELECON '12), Yasmine Hammamet, Tunisia, March 2012, pp. 888–891.
- [11] Miura, J., Kanda, T., Shirai, Y.: 'An active vision system for real-time traffic sign recognition'. Proc. IEEE intelligent transportation system, Dearborn, MI, USA, USA, 1–3 October 2000, pp. 52–57.
- [12] Shaposhnikov, W.D., Shaposhnikov, D.G.N.L., Golovan, E.-V., et al.(2002). Road sign recognition by single positioning of space-variant sensor. Proc.15th Int.Conf. on Vision Interface, Calgary, Canada, 2002, pp. 213–217
- [13] Vitabile, S., Gentile, A., Sorbello, F.(2002). A neural network based automatic road signs recognizer. Proc. Int. Joint Conf. on Neural Networks, Honolulu, HI, USA, USA, 12–17 May 2002
- [14] Fleyeh, H.(2004). Color detection and segmentation for road and traffic signs. Proc. IEEE Conf. Cybernetics and Intelligent Systems, Singapore.
- [15] Garrido, G., Sotelo, M.A., Martm-Gorostiza, E.(2006). Fast traffic sign detection and recognition under changing lighting conditions. Proc. In Intelligent Transportation Systems Conf., Toronto, Canada, 2006, pp. 811–816

- [16] Yakimov, P., Fursov, V. (2015). Traffic signs detection and tracking using modified hough transform'. Proc. 12th Int. Conf. on Signal Processing and Multimedia Applications, Colmar, France, Vol.20, pp. 22–28
- [17] Barnes, N., Zelinsky, A.(2004).Real-time radial symmetry for speed sign detection'. Proc. IEEE Intelligent Vehicles Symp., Parma, Italy,2004, Vol.1, pp. 566–571.
- [18] Loy, G., Zelinsky, A. (2003). Fast radial symmetry for detecting points of interest',IEEE Trans. Pattern Anal. Mach. Intell., 2003, Vol.25, pp. 959–973
- [19] Loy, G., Barnes, N.(2004). Fast shape-based road sign detection for a driver assistance system'. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), Sendai, Japan, 2004, pp. 70–75
- [20] Paulo, C.F., Correi, P.L. (2007) Automatic detection and classification of traffic signs'. Proc. Eighth Int. Workshop on Image Analysis for Multimedia Interactive Services, Santorini, Greece, June 2007
- [21] Hechri, A., Mtibaa, A.: 'Lanes and road signs recognition for driver assistance system', Int. J. Comput. Sci. Issues, 2011, Vol.8, pp. 402–408.
- [22] Rajesh Kannan Megalingam, Kondareddy Thanigundala, Sreevatsava Reddy Musani, Hemanth Nidamanuru, Lokesh Gadde.(2023). Indian traffic sign detection and recognition using deep learning. International Journal of Transportation Science and Technology, 2023, Vol.12, pp.683–699.
- [23] Chen Hui, Mohammed A.H. Ali, Nukman Yusoff, Abd Razak Bushroa, Turaev Sherzod, Chen YiHan.(2024). Computational methods for automatic traffic signs recognition in autonomous driving on road: A systematic review. Results in Engineering, 2024, Vol.24, pp.1-19.
- [24] Dalal, N., Triggs, B.(2005). Histograms of oriented gradients for human detection. Proc. the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05), San Diego, California, 2005, pp.886– 893.
- [25] Soetedjo, A., Somawirata, I.K.(2014). An efficient algorithm for implementing traffic sign detection on low cost embedded system. Int. J. Innov. Comput., Inf. Control, 2014, Vol.14, pp. 1–14
- [27] Ping Zhang, Bui, Tien D. and Suen, Ching Y.(2007). A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition, 2007. Vol.40 , pp.3415-3429,
- [26] Rajashe kararadhya, S.V, (2008). Neural network based handwritten numeral recognition ofKannada and Telugu scripts. IEEE,2008, PP. 4244-2409.
- [29] Yang, M., Zhang, L., Feng, X., Zhang, D.(2011). Fisher discrimination dictionary learning for sparse representation, ICCV.
- [28] Chu-Kiong Loo, Ali Memariani.(2012). Object Recognition Using Sparse Representation of Overcomplete Dictionary. IEEE, 2012, Vol.76, pp 75-82.
- [30] M. Dehghan and K. Faez, (1997). Farsi Handwritten Character Recognition With Moment Invariants. Proceedings of 13th International Conference on Digital Signal Processing, 1997, Vol.2, pp. 507-510.
- [31] José M. Bioucas-Dias, Mário A. T. Figueiredo,(2007). A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration. IEEE Transactions on Image Processing , 2007, Vol.16 , pp.1-12.
- [32] Bayat Amir Bahador .(2014). Sparse Representation-based Classification of Farsi Handwritten Digits Using Fisher Discrimination Criterion and Local Linear Embedding (LLE). Journal of Pattern Recognition and Intelligent Systems, 2014, Vol. 2, pp.56-65.
- [33] Xie, K., Ge, S., Ye, Q., et al.(2016). Traffic sign recognition based on attributerefinement cascaded convolutional neural networks'. Proc. the Pacific Rim Conf. on Multimedia, Xi'an, China, 2016, pp. 201–210
- [34] Shao, F., Wang, X., Meng, F., et al.(2018). Real-time traffic sign detection and recognition method based on simplified gabor wavelets and CNNs. Sensors, 2018, 18, p. 3192