Applied Soft Computing Journal 180 (2025) 113266

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Multi-granularity ensemble sample selection and label correction for classification with noisy labels[☆]

Kecan Cai [a,b,1], Hongyun Zhang [a,b,*,1], Witold Pedrycz [c,d,e], Duoqian Miao [a,b], Chaofan Chen [a,b]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China
[b] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, PR China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada
[d] Systems Research Institute, Polish Academy of Sciences, 00-901 Warsaw, Poland
[e] Istinye University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Sariyer/Istanbul, Turkiye

ARTICLE INFO

Keywords:
Multi-granularity
Noisy label
Image classification
Sample selection
Label correction

ABSTRACT

Sample selection is crucial in classification tasks with noisy labels, yet most existing sample selection methods rely on a single criterion. These approaches often face challenges, including low purity of selected clean samples, and underfitting due to an insufficient number of selected clean training samples. To address these challenges, this paper proposes GNet-SSLC, a novel multi-granularity network framework that integrates multiple criteria ensemble sample selection (SS) and multiple views label correction (LC). In the SS phase, this paper proposes a metric learning-based dual k-Nearest Neighbor (k-NN) sample selection method. This method first uses corrected soft labels from the initial k-NN round to guide the selection of clean samples in the subsequent k-NN round. To further enhance selection accuracy, we combine this dual k-NN approach with a small loss sample selection technique through a voting mechanism. This multiple criteria ensemble method addresses the issues of low purity and instability inherent in single criterion approaches. In the LC phase, this paper designs a multiple views label correction framework that generates high-quality pseudo-labels for selected noisy samples. A key innovation of the framework is the design of a regularized contrastive learning loss, which optimizes the semi-supervised learning process by leveraging multiple views of training samples. The additional inclusion of training samples with high-quality pseudo-labels can effectively mitigate underfitting caused by a limited number of clean training samples. Experimental results on both synthetic and real-world noisy datasets indicate that GNet-SSLC enhances the purity and stability of the selected clean samples, and significantly improves classification performance. The enhancement is particularly notable with high noise rate dataset, such as CIFAR-100 dataset with 80% noise rate, achieving a 19.3% increase in classification accuracy compared to the baseline method.

1. Introduction

Image classification represents a fundamental yet challenging task in computer vision, attracting considerable attention in recent years [1]. This task relies heavily on the availability of vast amounts of high-quality, labeled data. However, the process of obtaining such well-annotated data sets is not only resource-intensive but also time-consuming, posing significant challenges [2]. To circumvent these issues, cheaper alternatives like crowd-sourced annotations or web-crawled data are often utilized. Nevertheless, these methods frequently lead to compromised data quality due to label noise, which stems from inaccuracies in annotations or unreliable labeling processes [3].

In response to these challenges, there has been a surge in research efforts aimed at developing robust learning methodologies to handle

☆ This work was supported by the National Natural Science Fund of China (62076182, 62376198, 62163016, 62006172), in part by National Key Research and Development Program of China (2022YFB3104700) and in part by Jiangxi Provincial Natural Science Fund (No. 20212ACB202001).

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China.
E-mail addresses: caikecan@tongji.edu.cn (K. Cai), zhanghongyun@tongji.edu.cn (H. Zhang), wpedrycz@ualberta.ca (W. Pedrycz), dqmiao@tongji.edu.cn (D. Miao), imchenchaofan@163.com (C. Chen).
1 Joint first author.

https://doi.org/10.1016/j.asoc.2025.113266
Received 20 January 2024; Received in revised form 9 May 2025; Accepted 10 May 2025
Available online 14 June 2025
1568-4946/

datasets with noisy labels. These strategies encompass a variety of approaches, such as model regularization techniques [4], the refinement of loss functions [5], and sample selection methods [6,7]. Notably, sample selection has become a particularly prominent method for addressing the challenges posed by noisy labels. This technique fundamentally involves separating clean samples (with correct labels) from a noisy sample dataset (with incorrect labels), and subsequently focusing the optimization process of the neural network on these clean samples to minimize the detrimental effects of the noisy data. Studies have identified a memory effect in neural networks, wherein clean samples tend to have smaller losses compared to noisy samples during the early stages of training, a phenomenon known as the small loss strategy [7,8]. Leveraging this observation, several studies have adopted this strategy to select samples with lower loss values, presuming them to be clean [9,10]. However, this reliance on loss values as the sole criterion for sample selection has its drawbacks. It can often lead to inaccurate sample categorization, and this method's effectiveness tends to diminish in the later stages of network training. Moreover, a significant limitation of many current sample selection algorithms is their tendency to completely disregard all samples deemed noisy, especially problematic in scenarios characterized by high noise levels. This often results in a scarcity of clean samples, insufficient for effective model training.

To further overcome these limitations, some researchers have explored the incorporation of metric learning techniques, which focus on sample selection from a feature standpoint [11,12]. However, this approach sometimes neglects the valuable insights that can be gleaned from loss values. Alternatively, other methods have integrated semi-supervised learning frameworks, which utilize both clean and noisy samples [13,14]. Yet, these techniques often do not adequately address the reliability of pseudo-labels derived from noisy data. Consequently, the challenges of accurately identifying and optimally utilizing clean and noisy samples persist as critical research areas in the development of more robust sample selection methodologies.

This paper presents a novel network for sample selection and label correction. The proposed method adopts a multi-granularity ensemble strategy to address classification tasks with noisy labels. It enhances performance by integrating multiple selection criteria and incorporating diverse data views. The approach consists of two stages: (i) multiple criteria ensemble sample selection and (ii) multiple views semi-supervised label correction. The ensemble sample selection method integrates a small loss-based and a novel metric learning-based method through a voting mechanism. This integration aims to address the low purity and instability in clean sample selection, which often arises from reliance on a single criterion. Building upon the results of sample selection and multiple views of training samples generated by various data augmentations, a novel semi-supervised noise label correction framework is designed. This framework generates superior pseudo-labels for selected noisy samples, thus mitigating underfitting due to limited clean training samples.

More specifically, the paper first proposes a metric learning-based dual k-Nearest Neighbor (k-NN) sample selection method. This method utilizes the corrected labels from the first round of k-NN as soft labels, aiding the second round of k-NN for sample selection and alleviating the impact of raw noisy labels. Then, we propose a multiple criteria ensemble sample selection approach, combining the proposed metric learning-based method with a small loss-based method. Both methods undergo multiple data augmentations, with the final clean sample selection executed through voting. Finally, the semi-supervised noise label correction framework is presented. This framework employs a regularized contrastive learning loss from the contrastive learning method CTRR [11] to optimize the semi-supervised loss from the semi-supervised learning method FixMatch [14]. This integration leads to more effective model iteration optimization, and the generation of high-quality pseudo-labels. Samples with pseudo-labels enable the model to exploit valuable information from noisy training samples and prevent model underfitting, particularly in high noise scenarios.

In summary, this paper focuses on enhancing image classification performance with noisy labels through improved sample selection and label correction techniques. The significant contributions of our work are as follows:

- A novel metric learning-based dual k-NN sample selection approach: Differ from existing single k-NN methods, our approach first utilizes corrected soft labels from the initial k-NN round to guide the selection of clean samples in the subsequent k-NN round. This dual-phase strategy reduces the impact of raw noisy labels, effectively addressing the issue of low purity in clean samples selected by current methods.
- A novel multiple criteria ensemble sample selection method: By integrating various criteria, including both the traditional small loss-based technique and our proposed metric learning-based dual k-NN approach, this method offers a new perspective compared to conventional single-criterion techniques. The use of multiple criteria significantly improves the stability of the selection process.
- A multiple views label correction method: We have designed a novel regularized contrastive learning loss based on multiple views, which optimizes the existing semi-supervised learning loss by leveraging diverse views of training samples. By incorporating samples with high-quality corrected pseudo-labels into the training process, this method mitigates the overfitting issue caused by the limited number of selected clean samples.

Overall, the paper presents a comprehensive investigation into sample selection-based noisy labeled image classification. It introduces innovative methods that address key challenges in sample selection and the effective utilization of noisy samples. The remaining sections are organized as follows: Section 2 provides an overview of related works. Section 3 focuses on explaining the proposed model. Section 4 presents experimental validation and analysis of the results to demonstrate the effectiveness of the proposed methods. Finally, Section 5 concludes the paper.

2. Related works

In this section, we present a comprehensive overview of existing literature pertinent to our research domain, with a specific emphasis on studies that align closely with the scope and objectives of our work.

2.1. Multi-granularity

Multi-granularity, a fundamental concept in granular computing [15], involves analyzing, processing, or representing data at multiple levels of granularity or detail. This approach acknowledges that different aspects of a problem may require solutions at different levels [16]. Consequently, this paper adopts multi-granularity as the theoretical foundation and proposes the multiple criteria ensemble sample selection and multiple views label correction method to deal with noisy labels at different levels.

Fig. 1 shows the visualization of the multi-granularity design. Through upward and downward granulation operations, we can obtain data at different granularities. Downward refinement yields granular data containing finer features, while upward coarsening produces granular data with coarser features. These various granular data can be perceived and processed at multiple layers of abstraction and detail, allowing for more versatile approaches to problem-solving. Multi-granularity recognizes that data is diverse. It means that different parts of a complex problem can be better solved at different levels of detail.

The concept of multi-granularity has gained significant attention in recent years [17–20], demonstrating its effectiveness in tackling

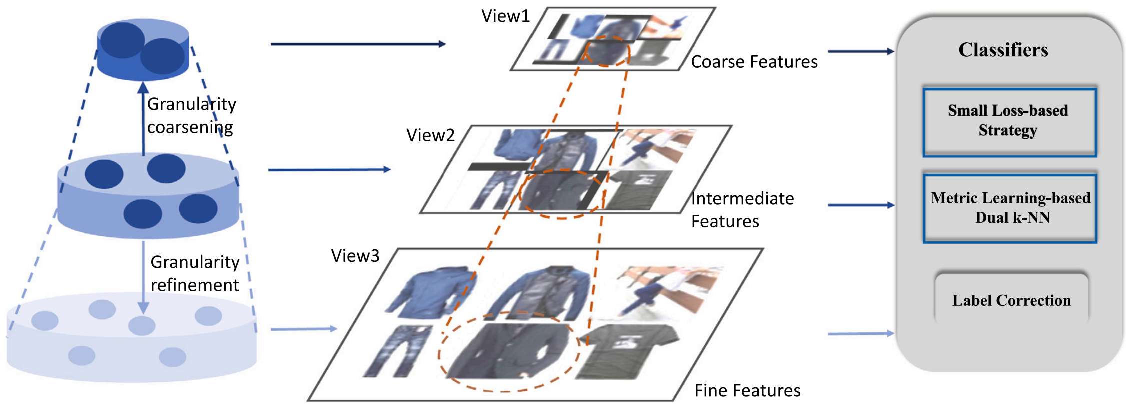

Fig. 1. Visualization of the multi-granularity design, illustrating upward and downward granulation operations to generate multiple views at various levels.

complex challenges in computer vision. Khoder et al. [17] proposes three variants of a novel ensemble learning approach. This method emphasizes the use of multiple extracted feature subsets, obtained from multiple learned linear embeddings, to enhance data representation. Instead of using a single classifier, these multiple feature subsets are concatenated into a single vector for more effective classification. MG-CAP [18] presents a novel multi-granularity canonical appearance pooling method. It addresses visual-semantic discrepancies by employing a granular framework to learn multi-grained features and enhance feature discriminative power. MGDNet [19] introduces a multi-granularity decoupling network, addressing the challenge of imbalanced remote sensing scene training samples. It incorporates methods such as multi-granularity complementary feature representation and a diversity component feature loss function. MGLP [20] presents a multi-granularity-based semi-supervised label propagation method. Developed from a multi-granularity perspective, it enhances label accuracy for unlabeled data by leveraging diverse neighborhood sizes and multi-level neighborhood information granules in graph-based semi-supervised learning.

This research provides a comprehensive examination of existing methodologies in computer vision, focusing specifically on multi-granularity techniques. A significant gap in the literature is observed: most methodologies are not tailored for the crucial task of classifying images with noisy labels. These noisy labels, often due to factors like human error or misinterpretation of image content, pose substantial challenges for accurate classification.

This study aims to fill this gap by developing approaches specifically designed to address the challenges of image classification with noisy labels. Grounded in multi-granularity, we propose a multiple criteria ensemble sample selection and multiple views label correction method. By leveraging the principles of multi-granularity, our research seeks to enhance the robustness and accuracy of network models, particularly in the face of uncertainty and imprecision introduced by noisy labels.

2.2. Small loss sample selection

Research has shown that neural networks exhibit a memory effect [3,7], where clean samples often have smaller losses compared to noisy samples during the early stages of training. Based on a threshold τ, samples with loss values below this threshold are considered potential clean samples. Corresponding algorithm is presented in Algorithm 1.

Algorithm 1 Small loss-based sample selection algorithm.

Input: train dataset $\mathcal{D} = \{(x_i, y_i), i \in (1, \cdots, N)\}$, sample losses= $\{loss_i, i \in (1, \cdots, N)\}$, threshold τ.
Output: clean sample set $\mathcal{D}_c$, noisy sample set $\mathcal{D}_n$.
1: $\mathcal{D}_c = \phi,\ \mathcal{D}_n = \phi$
2: **for** $i = 1$ *to* N **do**
3: **if** $loss_i \le \tau$ **then**
4: $\mathcal{D}_c = \mathcal{D}_c \bigcup \{(x_i, y_i)\}$
5: **else**
6: $\mathcal{D}_n = \mathcal{D}_n \bigcup \{(x_i, y_i)\}$
7: **end if**
8: **end for**
9: **return** $\mathcal{D}_c, \mathcal{D}_n$.

Various methods have been proposed based on the small loss strategy to select samples with lower losses as clean [8–10]. MentorNet [8] reduces overfitting in deep neural networks by employing a secondary network, the teacher network, to guide the loss function of the main network, StudentNet. This approach dynamically adjusts the training curriculum based on data-driven insights, prioritizing samples with accurate labels and effectively managing loss. Co-teaching [9] introduces a deep learning paradigm for handling noisy labels by training two neural networks simultaneously. Each network selects data with potentially clean labels and shares these selections with the other. They mutually determine the most reliable data in each mini-batch for training, updating themselves based on the data chosen by their peer to reduce the impact of noisy labels. O2UNet [10] presents a method for detecting noisy labels in deep neural networks without human annotations. It alternates a network's hyperparameters between overfitting and underfitting states, tracking the losses of each sample during iterations. Samples with higher normalized average losses are more likely to have noisy labels.

Using loss values as the sole criterion for sample selection in neural network training has limitations. Initially, high loss values can identify challenging samples, beneficial for early training stages. However, as training progresses, especially in later stages, loss values become less indicative of sample complexity and representativeness. This reduced effectiveness can lead to suboptimal or erroneous sample selection, potentially causing overfitting to high-loss samples. Thus, it is crucial to integrate alternative approaches to sample selection. Incorporating diversity criteria, or techniques assessing the overall representativeness of samples can provide a more balanced and effective training strategy. This approach not only addresses the limitations of relying solely on loss values but also enhances the model's performance and generalizability in computer vision tasks.

2.3. Metric learning sample selection

Metric learning and its variants have been developed for sample selection to address noisy labels. MOIT [21] combines metric learning with supervised contrastive learning loss to enhance model robustness against noisy labels by directly learning image representations. Building on MOIT, [22] proposes the Sel-CL method, selecting highly confident sample pairs from noisy data to improve supervised contrastive learning robustness. Ref. [23] introduces a method that leverages label correlations without requiring anchor points or precise fitting of noisy class posteriors. It identifies the transition matrix by exploiting label correlation mismatches, solving a bilinear decomposition problem. From a constrained image feature perspective, CTRR [11] introduces a regularization method within the contrastive learning framework, using noisy image features as a regularization term to limit noisy label impact during training. Jo-SRC [12] employs a contrastive learning method that estimates clean sample likelihood, using joint loss with consistency regularization. JoCoR [24] proposes a robust learning paradigm to address noisy labels, emphasizing sample selection and metric learning. It reduces diversity between two networks during training by jointly updating parameters using small-loss examples, guided by a Co-Regularization joint loss. PNP [25] trains two networks to predict labels, identifying noisy samples with tailored loss functions for updates. DMI [26] applies information-theoretic loss functions to train neural networks robust against label noise, focusing on sample selection and metric learning without auxiliary information or noise pattern estimation. LIMIT [27] introduces noise in gradient updates, employing an auxiliary network to predict classifier gradients, reducing label-noise memorization and enhancing generalization through sample selection and metric learning, independent of labels. APL (NCE+RCE) [28] combines two robust losses, creating a framework for robust loss functions in deep neural networks with noisy labels, focusing on sample selection and metric learning to address underfitting and enhance training accuracy. SLN [29] presents a noisy label learning approach, emphasizing sample selection and metric learning. It demonstrates that SLN in SGD improves generalization by influencing the output landscape's sharpness and output probability confidence. INCV [30] presents a method linking test accuracy to dataset noise ratio, partitioning noisy data and removing high-loss samples during training, and applying cross-validation for sample selection and Co-teaching for metric learning to train DNNs robustly against noisy labels.

Inspired by these studies, it is clear that sample features can effectively represent the original samples, independent of their labels. Therefore, this paper proposes a novel metric learning-based sample selection method from a multiple criteria perspective, incorporating metric learning into small loss-based sample selection to identify clean samples. Differ from existing methods, this paper initially combines multiple criteria, including small loss-based and metric learning-based strategies, offering a fresh perspective compared to traditional single-criterion techniques.

2.4. Semi-supervised sample selection

The integration of semi-supervised learning techniques has been pivotal in harnessing the potential of unlabeled data for feature representation learning. Recently, considerable attention has been directed toward improving semi-supervised learning methods through the integration of advanced sample selection strategies, aiming to better exploit noisy samples.

TCL [13] enhances deep learning with limited clean labels by using trusted data to generate reliable soft labels and jointly training the classifier and label aggregator. DMix [14] integrates semi-supervised learning by dividing data into clean (labeled) and noisy (unlabeled) sets using a loss distribution mixture model. It trains two diverged networks on both sets, enhancing semi-supervised learning with label co-refinement and co-guessing strategies. RoCL [31] employs a two-stage strategy: supervised training on selected clean samples, followed by label correction for noisy samples through semi-supervised learning. SELF [32] progressively filters out samples with noisy labels from the training data using semi-supervised learning techniques, correcting the labels of correctable samples and involving them in network training along with samples with low losses. ELR [33] exploits early learning, incorporating semi-supervised methods and specialized regularization to prevent false label memorization in noisy annotation scenarios. SSS-Net [34] integrates loss-similarity-based clustering and shadowed-sets theory for adaptive clean sample selection, introducing a high-quality pseudo-label re-selection method via co-training two networks.

The use of semi-supervised methods in sample selection, particularly in noisy training samples, has shown superior performance over supervised approaches. This paper enhances the proposed ensemble learning sample selection method by integrating label correction to generate pseudo-labels for noisy samples. However, the effectiveness of these methods heavily depends on the accuracy of sample selection. To mitigate the impact of incorrect selections, this study further incorporates contrastive learning to correct noisy sample labels based on multiple views of the input data. Incorporating corrected noisy samples into the training process helps improve the model's performance in noisy labeled image classification.

3. The proposed approach

In this section, we present a detailed introduction to the proposed Multi-Granularity Ensemble Sample Selection and Label Correction method (GNet-SSLC). As shown in Fig. 2, GNet-SSLC primarily consists of two stages: multiple criteria ensemble sample selection (SS) (the yellow section in Fig. 2), and multiple views semi-supervised label correction (LC) (the blue section in Fig. 2). The following subsections provide an in-depth explanation of these two key stages.

3.1. Multiple criteria ensemble sample selection

3.1.1. Overview

Existing sample selection methods, such as small loss-based [7] and metric learning-based [11] sample selection, can filter out clean samples from noisy datasets, thus avoiding the impact caused by noisy samples. However, the stability tends to decrease in the later stages of training, leading to incorrect selections.

To address these challenges, this study presents a multiple criteria ensemble sample selection method. As depicted in Fig. 2 Sample Selection (SS), SS ensembles two criteria through a voting mechanism: the small loss-based criterion and the proposed metric learning-based dual k-NN criterion. More specifically, the procedural details of the small loss-based sample selection criterion are described in Algorithm 1. Following this, a comprehensive explanation of the proposed metric learning-based dual k-NN sample selection approach is provided.

3.1.2. Metric learning-based dual k-NN sample selection

Ref. [35] analyzes the label distribution of each sample's k-nearest neighbors based on feature similarity. It then uses this information to estimate the sample's label and determine whether it is clean. However, the presence of label noise in the original samples can lead to incorrect label estimations based on these original sample labels. Consequently, this paper proposes a metric learning-based dual k-NN sample selection method, considering the use of soft labels $\hat{y}$ obtained from the first round of k-NN. These soft labels replace the original hard labels y in estimating the label distribution of samples, aiming to achieve a cleaner set of samples with higher purity.

The framework of the metric learning-based approach is illustrated in Fig. 3. Specifically, the feature set $\mathcal{G}(x) = \left\{\mathcal{G}(x_i), i \in (1,\ldots,N)\right\}$ of training sample $(x, y) = \left\{\left(x_i, y_i\right), i \in (1,\ldots,N)\right\}$ is extracted using a neural network. For each feature $\mathcal{G}(x_i)$, the algorithm conducts a first round of k-NN to predict the label probability distribution $p\left(t|x_i\right)$ and the soft label $\hat{y}_i$ of x_i, based on Eqs. (1) and (2). Eq. (1) estimates the label probability distribution of sample x_i by calculating the number of samples corresponding to each class within the local nearest neighbor of sample x_i,

$$p\left(t|x_i\right) = \frac{1}{\left|\mathcal{N}_k\left(\mathcal{G}(x_i)\right)\right|} \sum_{j=1}^{N} \mathbb{1}\left\{y_j = t, x_j \in \mathcal{N}_k\left(\mathcal{G}(x_i)\right)\right\},\ t = 1,\ldots,C \tag{1}$$

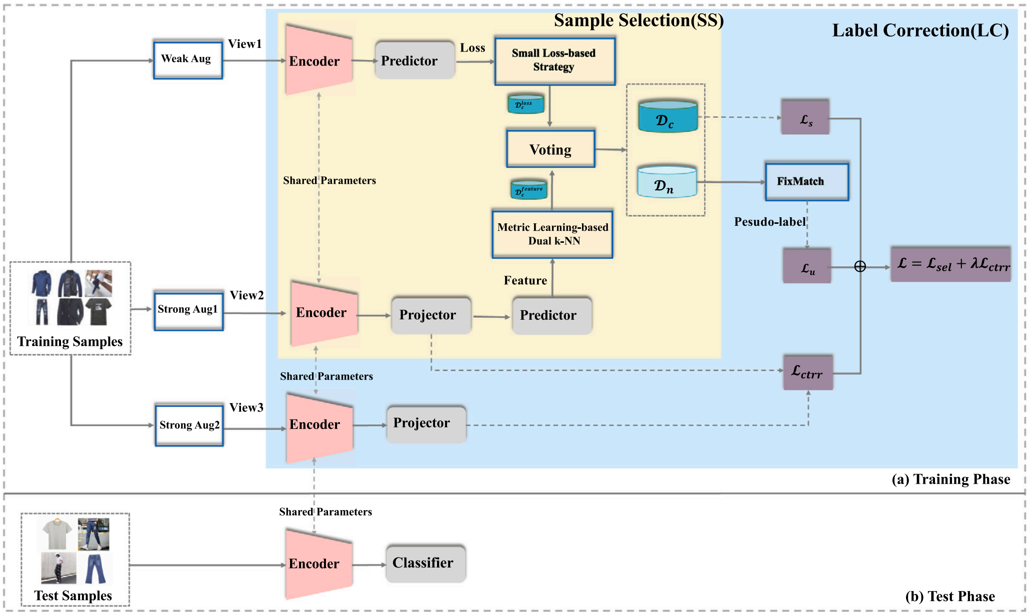

Fig. 2. Overview of the multi-granularity ensemble sample selection and label correction (GNet-SSLC) method, consisting of two key stages: the multiple criteria ensemble sample selection (SS), and the multiple views semi-supervised label correction (LC).

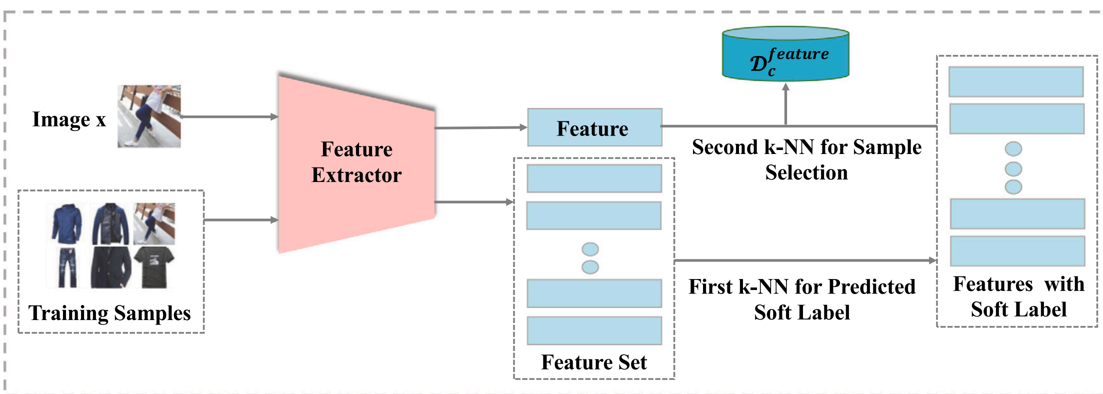

Fig. 3. Flow chart of the metric learning-based dual k-NN sample selection method with two rounds of k-NN prediction.

where C denotes the number of classes and label $y_j \in (1, \ldots, C)$, N represents the total number of samples. $\mathcal{N}_k\left(\mathcal{G}(x_i)\right)$ represents the k nearest neighbor samples of the feature $\mathcal{G}(x_i)$, and $\mathbb{1}\{condition\}$ is an indicator function outputting 1 when the condition holds true, and 0 otherwise. Then, based on the estimated label probability distribution $p\left(t|x_i\right)$ $(t = 1, \ldots, C)$ obtained from the first round of k-NN, soft label $\hat{y}_i$ of sample x_i can be generated as follows.

$$\hat{y}_i = \arg\max p\left(t|x_i\right),\ t = 1, \ldots, C \tag{2}$$

This means when the predicted label probability distribution $p\left(t|x_i\right)$ $(t = 1, \ldots, C)$ reaches its maximum value, the value of the corresponding parameter t becomes the predicted label (class), denoted as $\hat{y}_i$. Subsequently, a second k-NN search is then conducted based on the soft labels $\hat{y} = \left\{\hat{y}_i, i \in (1, \ldots, N)\right\}$ to re-estimate the label probability distribution of samples. This process helps mitigate the influence of estimation based on original noisy labels. The computation is carried out as follows.

$$\hat{p}\left(t|x_i\right) = \frac{1}{\left|\mathcal{N}_k\left(\mathcal{G}(x_i)\right)\right|} \sum_{j=1}^{N} \mathbb{1}\left\{\hat{y}_j = t, x_j \in \mathcal{N}_k\left(\mathcal{G}(x_i)\right)\right\},\ t = 1, \ldots, C \tag{3}$$

For each sample x_i, the distribution difference between the predicted label distribution $\hat{p}\left(t|x_i\right)$ and the original label distribution $q\left(t|x_i\right)$ is measured using cross-entropy loss, as shown in Eq. (4). The distribution $q\left(t|x_i\right)$ is generated based on the original label y_i in the form of one-hot encoding. The smaller this distribution difference is, the more likely it is that sample x_i is a clean sample.

$$disagree_i = H\left(q\left(t|x_i\right), \hat{p}\left(t|x_i\right)\right) = -\sum_{t=1}^{C} q\left(t|x_i\right)\ \log\ \hat{p}\left(t|x_i\right) \tag{4}$$

The specific process of the proposed metric learning-based dual k-NN sample selection algorithm is illustrated in Algorithm 2. The dual k-NN sample selection, based on metric learning, undergoes two rounds of k-NN prediction. In the first round, soft labels are generated as described in lines 2–7 of Algorithm 2, corresponding to the "First k-NN" stage illustrated in Fig. 3. In the second round, the label probability distribution is estimated based on these soft labels for clean sample selection, as described in lines 8–12 of Algorithm 2, corresponding to the "Second k-NN" stage in Fig. 3. Based on the obtained label probability distribution and Algorithm 1, the approach selects potential clean samples, detailed in Algorithm 2, lines 13–17.

Algorithm 2 Metric learning-based dual k-NN sample selection algorithm

Input: train dataset $\mathcal{D} = \{(x_i, y_i), i \in (1, \cdots, N)\}$, features $\mathcal{G} = \{\mathcal{G}(x_i), i \in (1, \cdots, N)\}$, original sample distribution $q = \{q_i, i \in (1, \cdots, N)\}$, threshold τ.
Output: clean sample set $\mathcal{D}_c$, noisy sample set $\mathcal{D}_n$.
$\mathcal{D}_c = \phi$, $\mathcal{D}_n = \phi$
for $i = 1$ *to* N **do**
  for $t = 1$ *to* C **do**
    $p(t|x_i) = \frac{1}{|\mathcal{N}_k(\mathcal{G}(x_i))|} \sum_{j=1}^{N} \mathbb{1}\{y_j = t, x_j \in \mathcal{N}_k(\mathcal{G}(x_i))\}$
  end for
  $\hat{y}_i = \arg\ max p(t|x_i)$ // Generate soft labels
end for
for $i = 1$ *to* N **do**
  for $t = 1$ *to* C **do**
    $\hat{p}(t|x_i) = \frac{1}{|\mathcal{N}_k(\mathcal{G}(x_i))|} \sum_{j=1}^{N} \mathbb{1}\{\hat{y}_j = t, x_j \in \mathcal{N}_k(\mathcal{G}(x_i))\}$ // Generate distribution from soft labels
  end for
end for
for $i = 1$ *to* N **do**
  $disagree_i = -\sum_{t=1}^{C} q(t|x_i)\ log\ \hat{p}(t|x_i)$
  $\mathcal{D}_c = \mathcal{D}_c \bigcup \{(x_i, y_i) | disagree_i \leq \tau\}$ // Select potential clean samples
  $\mathcal{D}_n = \mathcal{D}_c \bigcup \{(x_i, y_i) | disagree_i > \tau\}$
end for
return $\mathcal{D}_c, \mathcal{D}_n$.

3.2. Multiple views label correction

Clean samples are selected based on the proposed multiple criteria ensemble sample selection method introduced above. However, with a high noise rate, insufficient clean training samples prevent adequate model training. To exploit noise samples, the semi-supervised learning approach FixMatch [14] generates pseudo-labels for noise samples to avoid model underfitting. Nevertheless, these pseudo-labels heavily depend on the sample selection results.

This study further enhances the SS using a multiple views semi-supervised label correction approach, as depicted in Fig. 2 Label Correction (LC). LC incorporates a contrastive regularization function $\mathcal{L}_{ctrr}$ [11], to generate pseudo-labels for noise samples and mitigate the potential impact of erroneous sample selection. By learning features from multiple views, the model achieves a more stable correction of sample labels. More specifically, the weak data augmentation branch and the first strong data augmentation branch provide feature and loss values to the ensemble sample selection module for select clean samples. The LC considers the clean samples as labeled samples and uses them to compute the classification cross-entropy loss ℓ_s, while the remaining noisy samples are treated as unlabeled samples and pseudo-labels are generated for them. Given a batch B of clean samples $\mathcal{X} = \{(x_b, y_b), b \in (1, \ldots, B)\}$, the standard supervised cross-entropy loss ℓ_s is computed as follows,

$$\ell_s = \frac{1}{B} \sum_{b=1}^{B} H\left(p_b, p\left(y \middle| x_b\right)\right),\ \ y = 1, \ldots, C \tag{5}$$

where p_b is the original one-hot label distribution of clean samples and $p\left(y \middle| x_b\right)$ is the predicted label distribution for x_b over label y. C denotes the number of classes. When given a batch of noisy samples, denoted as $\mathcal{U} = \{u_b, b \in (1, \ldots, B)\}$, where noisy samples are considered as unlabeled, Fixmatch applies strong augmentation $\mathcal{A}(\cdot)$ and weak augmentation $\alpha(\cdot)$ to these unlabeled samples. The process then computes the predicted label probability distribution for the samples subjected to weak augmentation.

$$q_b = p\left(y \middle| \alpha(u_b)\right),\ \ y = 1, \ldots, C \tag{6}$$

Then, based on the estimated label probability distribution q_b, pseudo-label $\hat{q}_b$ of unlabeled sample u_b can be generated as follows.

$$\hat{q}_b = arg\ max(q_b) = arg\ max\left(p\left(y \middle| \alpha(u_b)\right)\right),\ \ y = 1, \ldots, C \tag{7}$$

When the predicted label probability distribution $p\left(y \middle| \alpha(u_b)\right)$ $(y = 1, \ldots, C)$ achieves its maximum value, the corresponding parameter y value represents the predicted label (pseudo-label), denoted by $\hat{q}_b$. Following this, the one-hot label distortion of pseudo-labels and the label probability distribution of noisy samples in the two strong data augmentation (Strong Aug1 and Strong Aug2) branches form the basis for applying consistency regularization. This process results in the corresponding cross-entropy loss, denoted as ℓ_u, as shown in the formula.

$$\ell_u = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{max(q_b) \geq \tau_1\} H\left(\hat{q}_b, p\left(y \middle| \mathcal{A}(u_b)\right)\right) \tag{8}$$

Only noise samples with class probabilities greater than the given threshold τ_1 will be assigned pseudo-labels. The semi-supervised classification loss $\mathcal{L}_{sel}$ is obtained by weighting both clean samples losses and noisy sample losses according to Ref. [14], where λ_u represents the weight to the loss of noise samples.

$$\mathcal{L}_{sel} = \ell_s + \lambda_u \ell_u \tag{9}$$

Compared to supervised learning methods that update network parameters solely based on clean samples, semi-supervised learning approaches combine the losses from noisy samples to update network parameters, resulting in more robust models.

However, semi-supervised methods heavily rely on the outcomes of sample selection. To mitigate the impact of erroneous sample selection, this study introduces a regularization contrastive loss, $\mathcal{L}_{ctrr}$, based on multiple views. Two strong data augmentation branches generate features through contrastive learning and compute the corresponding contrastive regularization loss $\mathcal{L}_{ctrr}$,

$$\mathcal{L}_{ctrr}(x_i, x_j) = \left(log\left(1 - \langle \tilde{q}_i, \tilde{z}_j \rangle\right) + log\left(1 - \langle \tilde{q}_j, \tilde{z}_i \rangle\right)\right) \mathbb{1}\{p_i^T p_j \geq \tau_2\} \tag{10}$$

where x_i, x_j are feature views of image x obtained through strong data augmentations Aug1 and Aug2, as well as feature transformations encoder and projector. Other variables in this formula are derivable from x_i and x_j, with detailed computations referenced in [11] to prevent repetition. The final loss for model optimization is the weighted sum of the contrastive regularization loss $\mathcal{L}_{ctrr}$ and the semi-supervised classification loss $\mathcal{L}_{sel}$.

$$\mathcal{L} = \mathcal{L}_{sel} + \lambda \mathcal{L}_{ctrr} \tag{11}$$

We follow the training setting described in Refs. [11,14] to optimize the entire network. After training, the classification performance of the model is validated on the test set based on the predictions obtained from the encoder and classifier. To enhance understanding, the pseudocode for the label correction method is provided in Algorithm 3. Initially, multiple views are generated, as detailed in lines 1–3 of Algorithm 3. Subsequently, based on the first two distinct data-augmented views, clean and noisy samples are identified using the ensemble sample selection algorithm, as illustrated in line 4 of Algorithm 3 and depicted in the Sample Selection (SS) module with the yellow background in Fig. 2. The procedure for the multiple criteria ensemble sample selection algorithm is comprehensively discussed in Section 3.1. Following this, the losses corresponding to the clean and noisy samples, derived from the first two views, are calculated, as indicated in lines 6–9 of Algorithm 3. Finally, using the subsequent two strongly augmented views, the contrastive learning loss is computed according to the loss function in Eq. (10). The final loss function used

Algorithm 3 Multiple views semi-supervised label correction algorithm

Input: train dataset $\mathcal{D} = (\mathcal{X}, \mathcal{Y}) = \left\{ \left(x_i, y_i \right), i \in (1, \cdots, N) \right\}$, sample losses, hyperparameters τ, λ, λ_u.
Output: loss for model optimization.
1: $\mathcal{D}_i = Weak\_Aug(\mathcal{D})$
2: $\mathcal{D}_j = Strong\_Aug1(\mathcal{D})$
3: $\mathcal{D}_k = Strong\_Aug2(\mathcal{D})$
4: $\mathcal{D}_c, \mathcal{D}_n = Ensemble(small\_loss(\mathcal{D}_i), dual\_kNN(\mathcal{D}_j))$ // Ensemble via voting mechanism
5: **for** $\mathcal{X} = \left\{ \left(x_b, p_b \right), b \in (1, \cdots, B) \right\}$ in $\mathcal{D}_c, \mathcal{U} = \left\{ u_b, b \in (1, \cdots, B) \right\}$ in $\mathcal{D}_n$ **do**
6: $\ell_s = \frac{1}{B} \sum_{b=1}^{B} H \left(p_b, p \left(y | x_b \right) \right)$
7: $q_b = p \left(y \middle| \alpha \left(u_b \right) \right)$
8: $\ell_u = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ max \left(q_b \right) \geq \tau \right\} H \left(arg\ max(q_b), p \left(y \middle| \mathcal{A} \left(u_b \right) \right) \right)$
9: $\mathcal{L}_{sel} = \ell_s + \lambda_u \ell_u$ // Compute two views semi-supervised loss
10: **end for**
11: **for** $\mathcal{D}_j = \left\{ \left(x_i, p_i \right), i \in (1, \cdots, B) \right\}, \mathcal{D}_k = \left\{ \left(x_j, p_j \right), j \in (1, \cdots, B) \right\}$ **do**
12: $\mathcal{L}_{ctrr} = \sum_{i=1, j=1, i \neq j}^{B} \mathcal{L}_{ctrr}(x_i, x_j)$ // Compute two views contrastive learning loss
13: **end for**
14: $\mathcal{L} = \lambda \ell_{ctrr} + \ell_s + \lambda_u \ell_u$ // Compute multiple views label correction loss
15: backward

to optimize the entire network is presented in line 14 of Algorithm 3, corresponding to the Label Correction (LC) module with the blue background in Fig. 2.

4. Experiments

In this section, we conduct experiments on three popular datasets for noisy label classification task to demonstrate the effectiveness of the proposed multi-granularity ensemble sample selection and label correction approach.

4.1. Experimental setup

4.1.1. Datasets

We validate the performance of the proposed method on synthetic noise datasets CIFAR-10 and CIFAR-100 [36], as well as real-world noise datasets CNWL [37].

- *CIFAR-10* [36] contains 60,000 color images of size 32 × 32 with 10 categories. The training set consists of 50,000 images, and testing set contains 10,000 images.
- *CIFAR-100* [36] includes 60,000 color images of size 32 × 32 with 100 classes. 50,000 images are used for training, and 10,000 images are used for testing.
- *Controlled Noisy Web Labels (CNWL)* [37] consists of images obtained from the web with a certain proportion of real-world noisy labels. We use the Mini-ImageNet dataset as the base dataset. Mini-ImageNet is a subset of the ImageNet dataset and consists of images of size 84 × 84 with 100 classes. 60,000 images for training and 5000 images from the validation set for testing.

Following the protocol described in Refs. [14,21], we artificially add symmetric noise at 20%, 40%, and 80% levels and asymmetric noise at 40% level to CIFAR-10 and CIFAR-100 datasets, respectively. And followed the settings described in Ref. [37] to add 20%, 40%, and 80% real-world web label noise to CNML. Symmetric noise is generated by randomly changing the labels of training data to incorrect class labels based on a noise level percentage. Asymmetric noise, on the other hand, involves replacing labels only with similar classes. Real-world noise refers to label noise that originates from the real world, commonly found in datasets collected through networks.

4.1.2. Implementation details

The baseline for this study is the CTRR network, as presented in Ref. [11]. The semi-supervised approach implemented is FixMatch [14]. All hyperparameters in this experiment are aligned with those of the baseline method, necessitating no additional hyperparameters tuning. Specifically, the training phase employs the SGD optimizer with a momentum of 0.9 and a weight decay of 10^{-4}. The model is trained with a batch size of 256 over 550 epochs. The initial learning rate is set at 0.02, and a cosine learning rate decay strategy is applied throughout the training process. Strong data augmentation techniques, including Gaussian blur, color distortion, random flipping, and random cropping, are utilized to enhance representation learning through contrastive learning. For the computation of classification cross-entropy loss, weaker data augmentation methods, consisting of random flipping and random cropping, are employed. Additionally, consistent with the settings in FixMatch [14], the threshold τ for the semi-supervised module is fixed to 0.95, the parameter λ is set according to the noise ratio, and λ_u is assigned to 1.0. The computer equipment is as follows: a. Memory: 64 GB, b. CPU: 11th Gen Intel(R) Core(TM) i9-11900K @ 3.50 GHz, c. GPU: NVIDIA GeForce RTX 3090*4, d. Operating System: Ubuntu 20.04.

4.2. Experiment results and analysis

4.2.1. Performance on synthetic noisy label datasets

In addition to the comparison with the baseline study [11], this paper also conducts comparisons with other state-of-the-art methods reported in the last five years. Synthetic noisy label datasets CIFAR-10 and CIFAR-100 are generated with symmetric noise levels of 20%, 40%, and 80%, as well as asymmetric noise at a level of 40%.

As shown in Table 1, GNet-SSLC significantly outperforms other methods at all noise levels on CIFAR-10 under the same backbone, particularly at the higher noise rate of 80%. For instance, GNet-SSLC achieves a classification accuracy of 90.79%, which is 4.6% higher than the baseline method. Also, it achieves competitive results even though the comparison methods employ deeper networks, such as ResNet-26 and ResNet-32, as their backbone. This improvement is primarily due to the dual k-NN sample selection method and the effective ensemble sample selection process, which together mitigate the impact of noisy labels and enhance the model's generalization capabilities.

The results in Table 2 highlight that GNet-SSLC demonstrates remarkable robustness in more complex datasets like CIFAR-100. Under 80% noise, GNet-SSLC achieved a 19.3% improvement in classification accuracy compared to the baseline method, indicating its superior ability to handle high noise levels. This substantial improvement can be attributed to the robust label correction mechanism. Since CIFAR-100 encompasses a broader array of categories, intensifying the challenge

Table 1
Comparison of test accuracy on CIFAR-10 dataset under different noise ratios with various methods.

Method	Backbone	Sym			Asym
		20%	40%	80%	40%
TCC-Net [38]	9-layer CNN	91.94	–	73.79	82.71
DMI [26]	ResNet-18	88.33	83.24	43.67	83.99
SLN [29]	ResNet-18	88.77	87.03	63.99	81.02
LIMIT [27]	ResNet-18	89.63	85.39	58.71	83.56
Co-teaching [9]	ResNet-18	92.05	87.73	44.16	77.78
Co-learning [39]	ResNet-18	92.21	–	61.20	81.42
APL [28]	ResNet-18	92.51	89.34	70.52	84.06
SPRL [2]	ResNet-18	92.68	–	57.50	–
CTRR (baseline) [11]	ResNet-18	92.56	90.98	86.16	85.51
INCV [30]	ResNet-32	89.71	84.78	52.27	86.04
FINE [40]	ResNet-34	91.0	–	69.4	89.5
MentorNet DD [8]	ResNet-101	91.23	88.64	46.31	–
O2U-net [10]	ResNet-101	92.57	90.33	37.76	–
DMix [14]	ResNet-18	95.12	94.11	35.36	90.04
Sel-CL+ [22]	–	**95.5**	–	89.2	**93.4**
SELF [32]	ResNet-26	–	93.7	69.91	89.07
RoCL [31]	ResNet-34	–	**94.55**	85.76	92.31
GNet-SSLC(ours)	ResNet-18	93.89	93.46	**90.79**	91.92

Table 2
Comparison of test accuracy on CIFAR-100 dataset under different noise ratios with various methods.

Method	Backbone	Sym			Asym
		20%	40%	80%	40%
PNP [25]	7-layer CNN	64.25	–	31.32	60.25
TCC-Net [38]	9-layer CNN	68.93	–	18.82	54.63
SLN [29]	ResNet-18	55.35	51.39	11.96	–
JoCoR [24]	ResNet-18	53.01	–	15.49	32.70
Jo-SRC [12]	ResNet-18	58.15	–	23.80	38.52
DMI [26]	ResNet-18	58.82	53.22	20.3	46.2
LIMIT [27]	ResNet-18	58.02	49.71	20.01	–
Co-teaching [9]	ResNet-18	65.71	57.64	15.28	–
Co-learning [39]	ResNet-18	66.58	–	35.45	47.62
APL [28]	ResNet-18	68.09	63.46	20.0	52.8
CECL [41]	ResNet-18	69.20	–	36.37	**65.49**
SPRL [2]	ResNet-18	70.93	–	28.53	–
CTRR (baseline) [11]	ResNet-18	71.03	66.6	43.69	52.85
FINE [40]	ResNet-34	70.3	–	25.6	61.7
MentorNet DD [8]	ResNet-101	72.64	67.51	30.12	–
O2U-net [10]	ResNet-101	**74.12**	69.21	39.39	–
DMix [14]	ResNet-18	71.39	70.83	49.52	50.99
Sel-CL+ [22]	–	**76.5**	–	59.6	**74.2**
SELF [32]	ResNet-26	–	71.98	42.09	53.83
RoCL [31]	ResNet-34	–	**74.64**	53.89	73.28
GNet-SSLC(ours)	ResNet-18	72.19	71.68	**62.94**	69.32

in classifying subclasses within identical superclasses. This complexity necessitates a larger sample volume. GNet-SSLC, through the correction of labels for noisy samples and incorporating them into the model optimization process, substantially augments the training samples, thereby strengthening the network's training capacity.

4.2.2. Performance on the real-world dataset

The real-world applicability of GNet-SSLC is validated on the real-world label noise dataset CNWL [37]. Table 3 displays the experimental results, with comparative method results primarily sourced from literature [21]. All methods implemented a basic network architecture of ResNet-18 on this dataset.

The results indicate that the proposed approach generally surpasses comparative methods. Notably, it attains a maximum improvement of approximately 9% and a minimum improvement of around 2% in classification accuracy. Among the methods compared, both DMix [14] and MOIT [21] engaged small loss-based sample selection and semi-supervised learning algorithms. Additionally, MOIT [21] incorporated a contrastive learning framework. Nevertheless, the proposed method exceeded the performance of both. This finding highlights GNet-SSLC's

Table 3
Comparison of test accuracy on CNML dataset under different real-world noise ratios with various methods.

Method	Backbone	Noise rate		
		20%	40%	80%
DMix [14]	ResNet-18	50.3	50.94	35.42
FaMUS [42]	ResNet-18	51.42	48.06	35.50
ELR [33]	ResNet-18	58.1	50.62	41.68
InstanceGM [43]	ResNet-18	58.38	52.24	39.62
MOIT [21]	ResNet-18	63.14	60.78	45.88
MOIT+ [21]	ResNet-18	63.1	61.16	46.78
GNet-SSLC(ours)	ResNet-18	**67.03**	**63.1**	**55.72**

capability not only in handling artificially synthesized label noise but also in demonstrating robust ability against real-world label noise (see Fig. 4).

4.3. Statistical analysis of classification performance

To evaluate the overall performance of the proposed GNet-SSLC framework compared to the baseline method, we calculated the P-values for the classification accuracies obtained on the CIFAR-10 and CIFAR-100 datasets. The goal was to determine whether the observed differences in performance between the two methods were statistically significant. Given the small sample size and the potential non-normal distribution of the data, we employed the Wilcoxon signed-rank test, a non-parametric method suitable for paired sample comparisons.

The accuracy scores for the proposed method and the baseline method were as follows: Baseline Method: [0.9256, 0.9098, 0.8616, 0.8551, 0.7103, 0.666, 0.4369, 0.5285]. GNet-SSLC: [0.9389, 0.9346, 0.9079, 0.9192, 0.7219, 0.7168, 0.6294, 0.6932]. The calculated P-value for the Wilcoxon signed-rank test was 0.0078. This P-value is below the conventional threshold of 0.05, indicating that the performance differences between the proposed method and the baseline are statistically significant. This confirms that the observed improvements in classification accuracy achieved by the GNet-SSLC framework are unlikely to have occurred by chance, thereby validating the effectiveness of our proposed method.

4.4. Ablation study

In this section, ablation studies discuss the impact of different components on model performance. The proposed method comprises two primary components: Sample Selection (SS) and Label Correction (LC). Sample Selection encompasses two subcomponents: one emphasizes small loss-based (Loss) sample selection, while the other concentrates on metric learning-based (Feature) sample selection.

4.4.1. Performance contributions of different components of SS

To validate the effectiveness of various subcomponents of SS, experiments were conducted on the CIFAR-10 and CIFAR-100 datasets. These experiments compared the purity of selected clean samples and the stability of the clean sample selection process using different configurations: (1) Backbone + Loss; (2) Backbone + Feature; (3) Backbone + Loss + Feature.

The purity variation of selected clean samples during the iterative training process on the CIFAR-10 and CIFAR-100 datasets, with 20%, 40%, and 80% symmetric noise and 40% asymmetric noise, is depicted in Figs. 5 and 6. It is observed that the loss-based sample selection method leads to a higher noise rate in selected clean samples on the CIFAR-10 and CIFAR-100 datasets across different noise ratios. Furthermore, the fluctuation range of the noise rate for each iteration is significant, indicating instability. In certain scenarios, such as CIFAR-100 with 80% symmetric noise, the loss-based method demonstrates a slight improvement in purity compared to other methods, yet its

Fig. 4. Visualization of label correction results (right) based on the proposed multi-granularity ensemble sample selection and label correction method on the CIFAR-10 dataset with a 40% symmetric noise ratio (left). Red borders indicate incorrectly labeled images.

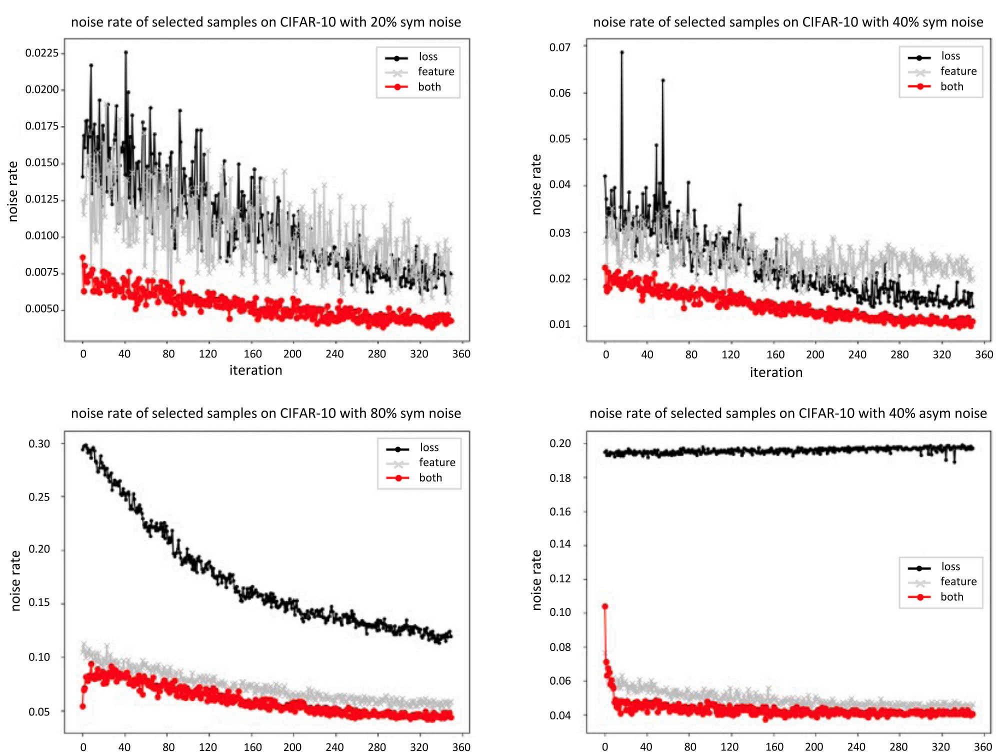

Fig. 5. Changes in purity and stability of selected samples on the CIFAR-10 dataset. The terms "loss", "feature", and "both" correspond respectively to the small loss-based sample selection method, the metric learning-based dual k-NN method, and the ensemble of these two sample selection methods.

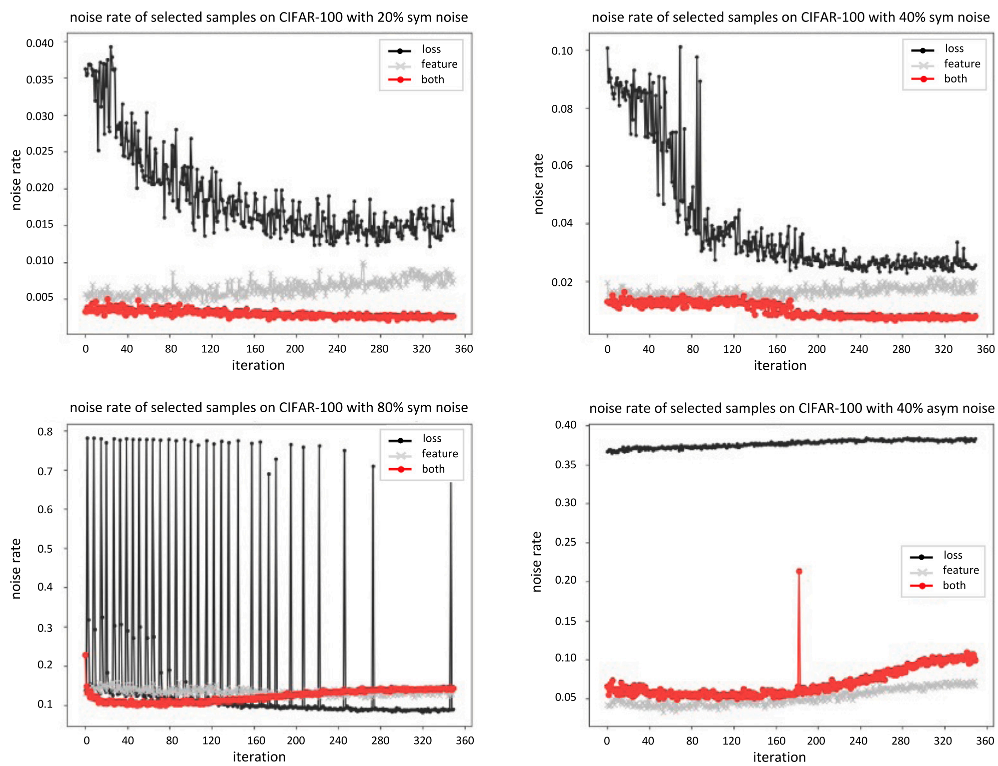

Fig. 6. Changes in purity and stability of selected samples on the CIFAR-100 dataset. The terms "loss", "feature", and "both" correspond respectively to the small loss-based sample selection method, the metric learning-based dual k-NN method, and the ensemble of these two sample selection methods.

Table 4
The performance of different components of the proposed method on CIFAR-10 dataset under different noise ratio.

Method	Sym			Asym
	20%	40%	80%	40%
Baseline	92.56	90.98	86.16	85.51
Baseline + SS	93.83	92.40	89.09	91.50
Baseline + SS + LC	93.89	93.46	90.79	91.92

Table 5
The performance of different components of the proposed method on CIFAR-100 dataset under different noise ratio.

Method	Sym			Asym
	20%	40%	80%	40%
Baseline	71.03	66.6	43.69	52.85
Baseline + SS	72.17	71.32	48.96	58.96
Baseline + SS + LC	72.19	71.68	62.94	69.32

Table 6
Comparison of computational efficiency (average training time) on CIFAR-10 dataset.

	Method	Training time/Epoch
Supervised	**Baseline**	∼35 s
	Baseline+SS(ours)	∼46 s
Semi-supervised	DMix	∼63 s
	Baseline+SS+LC(ours)	∼88 s
	Meta-Learning	∼103 s

stability remains markedly poor. The feature-based sample selection method, in contrast, exhibits superior performance in terms of both sample purity and stability, thereby affirming its effectiveness. Moreover, by integrating the feature-based and loss-based criteria in an ensemble approach, there is an enhancement in detected sample purity and a notable increase in stability. This underlines the superiority of the multiple criteria ensemble method over single criterion methods in terms of stability and sample purity.

4.4.2. Performance contributions of SS and LC

In order to evaluate the effectiveness of the SS and LC components, comprehensive comparative experiments were carried out using the widely recognized CIFAR-10 and CIFAR-100 datasets with different noise ratio. The primary objective of these experiments was to assess and compare the accuracy metrics across three distinct configurations: firstly, the baseline model; secondly, the baseline model augmented with the SS component; and thirdly, the baseline model further enhanced by the integration of both SS and LC components.

The detailed results of these experiments are presented in Tables 4 and 5. These tables provide a clear and quantitative demonstration of the significant improvements in the performance metrics of the baseline model when augmented with the SS and LC modules. A noteworthy observation from the experimental data is the pronounced contribution of the LC module, especially under scenarios characterized by high noise rates in the data. This pronounced improvement can be primarily attributed to the LC module's capability in generating accurate pseudo-labels for samples with label noise. Since LC effectively augments the training dataset with cleaner, more reliable data, as extensively detailed in the comparative experimental section of our study.

4.5. Training time analysis

Table 6 shows a detailed comparative analysis of the computational efficiency exhibited by our approach against other prevalent methodologies. It is noteworthy that our method, due to its integration of techniques such as Sample Selection (SS) and Label Correction (LC), incurs a higher computational overhead. This increase in computational resource consumption is an expected trade-off for the enhanced capabilities. In the specific area of semi-supervised learning sample selection methods, our approach demonstrates slower performance compared to DMix [14] but is faster than Meta-Learning [38] in terms of training speed. Despite the increased computational demand, the enhanced accuracy of the model justifies the additional cost, making it a valuable contribution to computer vision research.

This detailed examination brings to light an essential aspect of our research: the balance between computational efficiency and model performance. Despite the increased computational demand, the marked improvement in accuracy achieved by our model stands as a testament to its efficacy. The trade-off between the computational cost and the accuracy gains is a common theme in the evolution of computer vision technologies. In this context, the enhanced accuracy and the robustness of our model not only justify the additional computational resources required but also underscore its significant contribution to the advancement of research in the computer vision domain. This balance positions our approach as a valuable and innovative addition to the field, offering an understanding of the interplay between computational efficiency and model performance.

4.6. Results and discussion

The experimental results demonstrate that GNet-SSLC significantly enhances the purity and stability of selected clean samples, leading to substantial improvements in classification accuracy across both synthetic and real-world datasets. The method's ability to maintain high performance under high noise conditions is particularly noteworthy.

However, while GNet-SSLC has shown promising results, several limitations can be explored to further enhance its effectiveness and applicability. Firstly, investigating the computational complexity of the dual k-NN and multiple views label correction processes could lead to more efficient implementations. Future work could focus on exploring optimization techniques to reduce these computational demands. Techniques such as model pruning, quantization, or distillation could be explored to reduce the model's complexity without significantly compromising accuracy. Additionally, the experiments mainly utilized ResNet-18 as the backbone network. Future research could explore integrating GNet-SSLC with more advanced architectures like Vision Transformers or EfficientNets, which might further enhance its performance, especially in complex classification tasks.

5. Conclusion

In this study, we proposed an innovative multi-granularity ensemble methodology for classifying image datasets with noisy labels. By integrating multiple criteria and multiple views, our approach effectively overcomes the limitations typically associated with single-criterion methodologies. A key strength of our method lies in its enhanced ability to accurately and stably select clean samples, achieved through the combination of our novel metric learning-based dual k-NN sample selection and a refined small loss-based technique. The inclusion of semi-supervised learning paradigms and contrastive learning-derived regularization techniques across multiple data representations further mitigates the issue of underfitting, particularly in scenarios with limited training data. This makes our method particularly valuable in machine learning contexts where large, well-annotated datasets are not readily available. Empirical validation across various datasets demonstrates that our approach outperforms existing state-of-the-art methods in noisy label classification.

Moreover, it is important to acknowledge the limitations of our methodology, particularly regarding its computational demands, which may present challenges in scaling to extremely large datasets or in real-time processing applications. Despite these limitations, our proposed methodology represents a significant advancement in noisy label classification, offering an effective solution that contributes to the ongoing research in this field. Future work should focus on exploring the computational complexity of the dual k-NN and multiple views label correction processes, potentially leading to more efficient implementations. Additionally, we plan to investigate the integration of GNet-SSLC with advanced architectures like Vision Transformers, which could potentially enhance performance in more complex classification tasks.

CRediT authorship contribution statement

Kecan Cai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hongyun Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Witold Pedrycz:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis, Conceptualization. **Duoqian Miao:** Writing – review & editing, Supervision, Funding acquisition. **Chaofan Chen:** Validation, Software, Methodology, Investigation.

Declaration of competing interest

The authors declare that there is no conflict of interest, such as personal or professional relationships, affiliations, knowledge or beliefs, that may affect the subject matter or materials discussed in the manuscript.

Data availability

Data will be made available on request.

References

[1] Pengcheng Jiang, Yu Xue, Ferrante Neri, Convolutional neural network pruning based on multi-objective feature map selection for image classification, Appl. Soft Comput. 139 (2023) 110229.

[2] Xiaoshuang Shi, Zhenhua Guo, Kang Li, Yun Liang, Xiaofeng Zhu, Self-paced resistance learning against overfitting on noisy labels, Pattern Recognit. 134 (2023) 109080.

[3] Görkem Algan, Ilkay Ulusoy, Image classification with deep learning in the presence of noisy labels: A survey, Knowl.-Based Syst. 215 (2021) 106771.

[4] Hansong Zhang, Shikun Li, Dan Zeng, Chenggang Yan, Shiming Ge, Coupled confusion correction: Learning from crowds with sparse annotations, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, 2024, pp. 16732–16740, 15.

[5] Longrong Yang, Hongliang Li, Fanman Meng, Qingbo Wu, King Ngi Ngan, Task-specific loss for robust instance segmentation with noisy class labels, IEEE Trans. Circuits Syst. Video Technol. 33 (1) (2023) 213–227.

[6] Jiexi Yan, Lei Luo, Cheng Deng, Heng Huang, Adaptive hierarchical similarity metric learning with noisy labels, IEEE Trans. Image Process. 32 (2023) 1245–1256.

[7] Devansh Arpit, Stanis Jastrzundefinedbski, Nicolas Ballas, et al., A closer look at memorization in deep networks, in: Proceedings of the 34th International Conference on Machine Learning - Volume 70, 2017, pp. 233–242.

[8] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li Jia Li, Li Fei Fei, Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels, in: International Conference on Machine Learning, PMLR, 2018, pp. 2304–2313.

[9] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, Masashi Sugiyama, Co-teaching: Robust training of deep neural networks with extremely noisy labels, Adv. Neural Inf. Process. Syst. 31 (2018).

[10] Jinchi Huang, Lie Qu, Rongfei Jia, Binqiang Zhao, O2u-net: A simple noisy label detection approach for deep neural networks, in: IEEE/CVF International Conference on Computer Vision, 2019, pp. 3326–3334.

[11] Li Yi, Sheng Liu, Qi She, A. Ian McLeod, Boyu Wang, On learning contrastive representations for learning with noisy labels, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16682–16691.

[12] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, Zhenmin Tang, Jo-src: A contrastive approach for combating noisy labels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5192–5201.

[13] Shikun Li, Tongliang Liu, Jiyong Tan, Dan Zeng, Shiming Ge, Trustable co-label learning from multiple noisy annotators, IEEE Trans. Multimed. 25 (2021) 1045–1057.

[14] Junnan Li, Richard Socher, Steven C.H. Hoi, Dividemix: Learning with noisy labels as semi-supervised learning, in: 8th International Conference on Learning Representations, ICLR, 2020.

[15] Andrzej Bargiela, Witold Pedrycz, Granular computing, in: HANDBOOK on COMPUTER LEARNING and INTELLIGENCE: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation, World Scientific, 2022, pp. 97–132.

[16] Marco S. Reis, Multiscale and multi-granularity process analytics: A review, Processes 7 (2) (2019) 61.

[17] A. Khoder, F. Dornaika, Ensemble learning via feature selection and multiple transformed subsets: Application to image classification, Appl. Soft Comput. 113 (2021) 108006.

[18] Shidong Wang, Yu Guan, Ling Shao, Multi-granularity canonical appearance pooling for remote sensing scene classification, IEEE Trans. Image Process. 29 (2020) 5396–5407.

[19] Wang Miao, Jie Geng, Wen Jiang, Multigranularity decoupling network with pseudolabel selection for remote sensing image scene classification, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–13.

[20] Shengdan Hu, Duoqian Miao, Witold Pedrycz, Multi granularity based label propagation with active learning for semi-supervised classification, Expert Syst. Appl. 192 (2022) 116276.

[21] Diego Ortego, Eric Arazo, Paul Albert, Noel E O'Connor, Kevin McGuinness, Multi-objective interpolation training for robustness to label noise, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6606–6615.

[22] Shikun Li, Xiaobo Xia, Shiming Ge, Tongliang Liu, Selective-supervised contrastive learning with noisy labels, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 316–325.

[23] Shikun Li, Xiaobo Xia, Hansong Zhang, Yibing Zhan, Shiming Ge, Tongliang Liu, Estimating noise transition matrix with label correlations for noisy multi-label learning, Adv. Neural Inf. Process. Syst. 35 (2022) 24184–24198.

[24] Hongxin Wei, Lei Feng, Xiangyu Chen, Bo An, Combating noisy labels by agreement: A joint training method with co-regularization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13726–13735.

[25] Zeren Sun, Fumin Shen, Dan Huang, Qiong Wang, Xiangbo Shu, Yazhou Yao, Jinhui Tang, Pnp: Robust learning from noisy labels by probabilistic noise prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5311–5320.

[26] Yilun Xu, Peng Cao, Yuqing Kong, Yizhou Wang, L_DMI: A novel information-theoretic loss function for training deep nets robust to label noise, in: Neural Information Processing Systems, 2019.

[27] Hrayr Harutyunyan, Kyle Reing, Greg Ver Steeg, Aram Galstyan, Improving generalization by controlling label-noise information in neural network weights, in: International Conference on Machine Learning, PMLR, 2020, pp. 4071–4081.

[28] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, James Bailey, Normalized loss functions for deep learning with noisy labels, in: International Conference on Machine Learning, PMLR, 2020, pp. 6543–6553.

[29] Pengfei Chen, Guangyong Chen, Junjie Ye, Pheng-Ann Heng, et al., Noise against noise: stochastic label noise helps combat inherent label noise, in: International Conference on Learning Representations, 2020.

[30] Pengfei Chen, Ben Ben Liao, Guangyong Chen, Shengyu Zhang, Understanding and utilizing deep neural networks trained with noisy labels, in: International Conference on Machine Learning, PMLR, 2019, pp. 1062–1070.

[31] Tianyi Zhou, Shengjie Wang, Jeff Bilmes, Robust curriculum learning: From clean label detection to noisy label self-correction, in: International Conference on Learning Representations, 2020.

[32] Duc Tam Nguyen, Chaithanya Kumar Mummadi, Thi Phuong Nhung Ngo, Thi Hoai Phuong Nguyen, Laura Beggel, Thomas Brox, SELF: Learning to filter noisy labels with self-ensembling, in: 8th International Conference on Learning Representations, ICLR 2020, 2020.

[33] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, Carlos Fernandez-Granda, Early-learning regularization prevents memorization of noisy labels, Adv. Neural Inf. Process. Syst. 33 (2020) 20331–20342.

[34] Kecan Cai, Hongyun Zhang, Witold Pedrycz, Duoqian Miao, SSS-net: A shadowed-sets-based semi-supervised sample selection network for classification on noise labeled images, Knowl.-Based Syst. 276 (2023) 110732.

[35] Dara Bahri, Heinrich Jiang, Maya Gupta, Deep k-nn for noisy labels, in: International Conference on Machine Learning, PMLR, 2020, pp. 540–550.

[36] A Krizhevsky, V Nair, G. Hintong, CIFAR-10 and CIFAR-100 datasets[EB/OL]. https://www.cs.toronto.edu/∼kriz/cifar.html.

[37] Lu Jiang, Di Huang, Mason Liu, Weilong Yang, Beyond synthetic noise: Deep learning on controlled noisy labels, in: International Conference on Machine Learning, PMLR, 2020, pp. 4804–4815.

[38] Qiangqiang Xia, Feifei Lee, Qiu Chen, TCC-net: A two-stage training method with contradictory loss and co-teaching based on meta-learning for learning with noisy labels, Inform. Sci. 639 (2023) 119008.

[39] Cheng Tan, Jun Xia, Lirong Wu, Stan Z. Li, Co-learning: Learning from noisy labels with self-supervision, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 1405–1413.

[40] Taehyeon Kim, Jongwoo Ko, JinHwan Choi, Se-Young Yun, et al., Fine samples for learning with noisy labels, Adv. Neural Inf. Process. Syst. 34 (2021) 24137–24149.

[41] Wenhai Wan, Xinrui Wang, Mingkun Xie, Shengjun Huang, Songcan Chen, Shaoyuan Li, Unlocking the power of open set: A new perspective for open-set noisy label learning, 2023, arXiv preprint arXiv:2305.04203.

[42] Youjiang Xu, Linchao Zhu, Lu Jiang, Yi Yang, Faster meta update strategy for noise-robust deep learning, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 144–153.

[43] Arpit Garg, Cuong Nguyen, Rafael Felix, Thanh-Toan Do, Gustavo Carneiro, Instance-dependent noisy label learning via graphical modelling, in: IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 2288–2298.