# Bridging Class Imbalance and Partial Labeling via Spectral-Balanced Energy Propagation for Skeleton-based Action Recognition

Yandan Wang[1,†], Chenqi Guo[1,†], Yinglong Ma[1,*], Jiangyan Chen[1], Yuan Gao[1], Weiming Dong[2]
[1]School of Control and Computer Engineering, North China Electric Power University
[2]State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS),
Institute of Automation, Chinese Academy of Sciences

## Abstract

*Skeleton-based action recognition faces class imbalance and insufficient labeling problems in real-world applications. Existing methods typically address these issues separately, lacking a unified framework that can effectively handle both issues simultaneously while considering their inherent relationships. Our theoretical analysis reveals two fundamental connections between these problems. First, class imbalance systematically shifts the eigenvalue spectrum of normalized affinity matrices, compromising both convergence and accuracy of label propagation. Second, boundary samples are critical for model training under imbalanced conditions but are often mistakenly excluded by conventional reliability metrics, which focus on relative class differences rather than holistic connectivity patterns. Built upon these theoretical findings, we propose SpeLER (**Spe**ctral-balanced **L**abel Propagation with **E**nergy-based Tightened **R**eliability), which introduces a spectral balancing technique that explicitly counteracts spectral shifts by incorporating class distribution information. Meanwhile, a propagation energy-based tightened reliability measure is proposed to better preserve crucial boundary samples by evaluating holistic connectivity patterns. Extensive experiments on six public datasets demonstrate that SpeLER consistently outperforms state-of-the-art methods, validating both our theoretical findings and practical effectiveness. Codes are available at:* https://github.com/ydanwang/SpeLER.

## 1. Introduction

Skeleton-based action recognition (SAR) has gained increasing attention due to its compact yet expressive representation of human actions and rich temporal trajectory information [6, 12, 17, 26, 42]. While recent SAR approaches have achieved remarkable performance, they predominantly rely on balanced, fully-annotated datasets [13, 21, 33, 49, 55]. However, in the real world, SAR often faces class imbalance issues and partial labeling problems due to the inherent variability in action frequencies and the high cost of manual annotation [8, 57]. To enhance the classification performance for minority classes, several data-level and algorithm-level approaches have been proposed in SAR. To generate accurate pseudo-labels for unlabeled samples, graph-based transductive methods like Label Propagation (LP) are used [7, 20, 25, 28, 38, 43], effectively leveraging data's geometric structure for better performance in smaller datasets, which is often the case in SAR compared to image classification. Most of existing studies have addressed class imbalance or incomplete labeling in SAR separately, while few works address both problems simultaneously [18, 51].

Existing studies have enhanced graph-based pseudo-labeling by incorporating class-imbalance awareness during propagation [35, 40, 56], aiming to improve the propagation quality for minority classes. Meanwhile, semi-supervised learning strategies have been explored to focus on generating reliable pseudo-labels for minority classes to address the class imbalance problem [11, 24]. Most existing methods adopt an iterative optimization paradigm [23, 48], where pseudo-labeling algorithms generate annotations for unlabeled data, subsequently used for model training. However, the intricate interplay between these two steps, i.e., pseudo-label generation and imbalanced model training, remains theoretically underexplored.

In this work, we systematically investigate the interaction mechanisms between graph-based pseudo-labeling and model training under class-imbalanced conditions, while identifying two critical factors that mediate their relationship: the spectral properties of normalized affinity matrices and boundary samples. Specifically, we identify that class imbalance fundamentally alters the spectral characteristics of affinity graphs used in label propagation. Since graph-

*Corresponding author: yinglongma@ncepu.edu.cn
†These authors contribute equally to this work. This work was solely supported by State Grid Corporation of China Science and Technology Program under Grant (SGDK0000AIJS2502122).

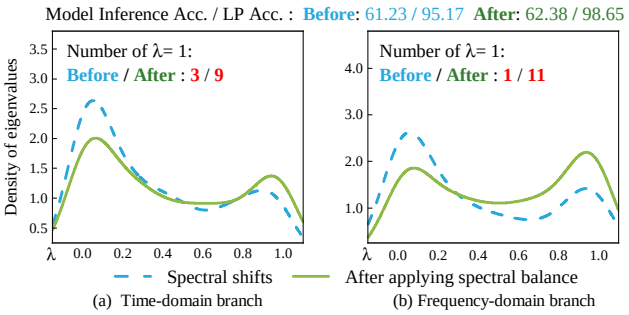


Figure 1. Eigenvalue($\lambda$) density distributions of normalized affinity matrices, model inference and label propagation (LP) accuracy (Acc.) on NTU-120 X-sub dataset (imbalance ratio=60) before and after applying our spectral-balanced approach. Our method improves the partial-label learning and subsequent model training under class imbalance through concentrating towards mid-to-high ranges, rather than being heavily skewed towards low values.

based pseudo-labeling relies on affinity graphs whose properties are typically analyzed using spectral theory, we use spectral perturbation as an indicator to quantify the interference of class imbalance on LP. As shown in Fig. 1, class imbalance distorts the eigenvalue distributions of the normalized affinity matrix constructed from biased features. We theoretically demonstrate that these spectral shifts directly impair the convergence and accuracy of label propagation algorithms, which consequently hinders model training.

Besides, in graph-based propagation, minority samples near decision boundaries often connect with majority samples due to feature similarity, playing a crucial role in propagation performance. However, the scarcity of labeled minority samples results in insufficient coverage and low confidence in pseudo-label generation for minority classes, further exacerbating the class imbalance. Consequently, these critical boundary samples are often misclassified as noise and mistakenly filtered out. Traditional relative connection-based reliability assessment methods like softmax probabilities tend to treat these boundary samples as noise with low confidence, creating a negative feedback loop where fewer pseudo-labeled minority samples exist, leading to more severe imbalance in subsequent training iterations.

To address these challenges, we propose **Spe**ctral-balanced **L**abel Propagation and **E**nergy-Based **R**eliability Assessment (SpeLER) approach that iterates between class-imbalanced model training and transductive label propagation. Motivated by our theoretical analysis, we propose a spectral-balanced propagation method that effectively mitigates class imbalance-induced eigenvalue distortions. We further propose the tightened reliability based on propagation energy, which captures the holistic connectivity patterns in the label propagation state space, thereby enhancing both the quality of pseudo-supervision and subsequent model learning. Our framework incorporates a cross-domain network that preserves minority class features through complementary time and frequency representations (Detailed in Sec. 1 of supplementary material). The proposed architecture, when implemented with a Graph Convolutional Network (GCN) backbone, achieves state-of-the-art performance on SAR tasks. Moreover, when adapted with a CNN-based backbone, it demonstrates superior results on classical time-series classification tasks such as EEG bio-signal analysis, highlighting its exceptional versatility and effectiveness as a general-purpose solution.

To our knowledge, this is the first work to theoretically reveal and address the intricate interplays between model training and partial-label learning under class imbalance through spectral-balanced energy propagation. Our main contributions are summarized as follows:

- We demonstrate how class imbalance distorts eigenvalue distributions, and reveal conventional reliability assessments mistakenly filter out essential boundary samples.
- We propose a spectral-balanced propagation and energy-based tightened reliability to improve label propagation while preserving reliable pseudo-labeled samples.
- Experimental results on six public classification datasets demonstrate our model's superiority and generalizability.

## 2. Related work

**Class-imbalanced learning methods.** Existing SAR approaches can be categorized into CNN, GCN and GCN-Transformer hybrid models [36, 39, 41, 46, 47]. Recently, data-level [52, 54] and algorithm-level learning strategies [19, 29] have been adopted to address class imbalance. However, their performance degrades in label-scarce scenarios, which exacerbates the class imbalance. And also, they typically neglect discriminative time-frequency patterns essential for recognizing minority classes.

**Partial-label learning methods.** Existing partial-label learning approaches can be divided into inductive and transductive methods [9, 48]. Inductive methods generate pseudo-labels based on model confidence or prediction consistency [29, 36, 41, 47], while transductive methods, like Label Propagation as a representative one, have been utilized to generate more reliable pseudo-labels for model training when data is limited [11, 56]. However, these methods cannot effectively preserve informative borderline samples [20, 24], which are critical for decision boundary learning. To address this, we introduce propagation energy to select reliable borderline samples in label-scarce class-imbalanced scenarios, particularly for minority classes.

**Class-imbalanced partial-label learning methods.** Few methods are available for class-imbalanced partial-label learning to solve the two issues simultaneously in SAR applications [5, 18, 51]. Recent advancements in class-imbalanced label propagation have shown potential to gen-

erate pseudo-labels and alleviate the imbalance problem in other computer vision field tasks [16, 20, 40]. However, theoretical understanding of the intricate interactions between partial-label learning and model training under class imbalance remains largely unexplored. We propose a framework that bridges these aspects with theoretical guarantees for mitigating spectral shifts while preserving critical boundary samples that are typically neglected in imbalanced scenarios, enabling robust feature learning across all classes.

## 3. Problem Statement and Theoretical Insights

### 3.1. Problem statement

We focus on skeleton-based action recognition task in the scenarios where class imbalance and partial-label coexist. Suppose the sequential action dataset consists of a labeled subset $\mathcal{D}_L=\{(\boldsymbol{x}_i^L, y_i^L)\}_{i=1}^{N_L}$ and an unlabeled subset $\mathcal{D}_U=\{\boldsymbol{x}_j^U\}_{j=1}^{N_U}$, where $N_L$ and $N_U$ denote the number of labeled and unlabeled instances, respectively. Each sample in $\mathcal{D}_L$ is annotated with $y^L\in\{1,2,\ldots,C\}$. The labeled ratio of dataset is defined as $\gamma=\frac{N_L}{N_L+N_U}$. In this imbalanced dataset, we assume that the number of samples per class $n_i(1\leq i\leq C)$ decreases as the class index $i$ increases, i.e., $n_1>n_2>\ldots>n_C$. The class-imbalance ratio is $\pi=\frac{n_1}{n_C}$.

### 3.2. How class imbalance impacts on partial-label learning

We seek to elucidate the interaction mechanism about how class imbalance compromises the efficacy of partial-label learning algorithms. Taking label propagation as a representative partial-label learning method, we provide theoretical analysis demonstrating how class imbalance fundamentally affects label propagation through its impact on the spectral properties of normalized similarity graphs. Specifically, we demonstrate that class imbalance induces systematic shifts in the eigenvalue distribution of normalized similarity/affinity matrices, which directly impacts both the convergence dynamics and solution accuracy of label propagation, thereby providing insights for developing principled solutions to class-imbalanced partial-label learning.

**Iterative Solution.** Let $\boldsymbol{S}\in\mathbb{R}^{B\times B}$ represent the normalized affinity matrix of multi batches of samples, $\lambda$ denote eigenvalues of $\boldsymbol{S}$, and $\boldsymbol{Y}\in\mathbb{R}^{B\times C}$ denote the distribution of true labels. The label matrix in the $(t+1)$-th iteration is expressed as $\boldsymbol{M}(t+1)=\alpha\boldsymbol{S}\cdot\boldsymbol{M}(t)+(1-\alpha)\boldsymbol{Y}$, where $\alpha\in(0,1)$ is the propagation factor, and $\boldsymbol{M}(0)=\boldsymbol{Y}$. LP updates the label matrix iteratively until convergence. Typically, convergence is defined either by reaching a specified number of iterations or when the difference between successive iterations falls below a predefined threshold.

To analyze the convergence and stability of LP, we denote $\tilde{\boldsymbol{M}}$ as the steady state of iterative learning. We discover that the following theorem holds:

**Theorem 3.1** *Let $\epsilon(t)=\boldsymbol{M}(t)-\tilde{\boldsymbol{M}}$ be the error at the $t$-th iteration of LP. The L2-norm of the iterative error satisfies*

$$\|\epsilon(t)\| \leq B\cdot\alpha^t\cdot|\lambda_r|^t\cdot\sum_{i=r}^{B}|\eta_i|, \tag{1}$$

*where $\sum_{i=r}^{B}|\eta_i|$ is the constant associated with the initial error, and $\lambda_r=\max\{|\lambda_i|:|\lambda_i|<1, i\in\{1,\ldots,B\}\}$.*

Eq. (1) indicates that the disturbance caused by class imbalance (i.e., the initial error) hinders the rapid convergence of the LP. Moreover, in $\boldsymbol{S}$, the absolute value of the largest eigenvalue excluding $\lambda=1$, affects the convergence of LP.

**Closed-form solution.** To enhance the computational efficiency of LP, we employed the closed form solution $\boldsymbol{M}=(\boldsymbol{I}-\alpha\boldsymbol{S})^{-1}(1-\alpha)\boldsymbol{Y}$. To evaluate the accuracy of this solution, let $\Delta\boldsymbol{Y}$ and $\Delta\boldsymbol{M}$ denote the class-imbalance perturbation and the solution error, respectively. We discover that the following theorem holds:

**Theorem 3.2** *Let $\frac{\|\Delta\boldsymbol{Y}\|}{\|\boldsymbol{Y}\|}$ represent the relative perturbation, and $\frac{\|\Delta\boldsymbol{M}\|}{\|\boldsymbol{M}\|}$ denote the relative error induced by imbalance perturbations. Then, $\frac{\|\Delta\boldsymbol{M}\|}{\|\boldsymbol{M}\|}$ satisfies*

$$\frac{\|\Delta\boldsymbol{M}\|}{\|\boldsymbol{M}\|}\leq\frac{1-\alpha\lambda_{\min}}{1-\alpha}\cdot\frac{\|\Delta\boldsymbol{Y}\|}{\|\boldsymbol{Y}\|}, \tag{2}$$

*where $\alpha\in(0,1)$. $\lambda_{\min}$ denotes the smallest eigenvalue of $\boldsymbol{S}$.*

Consequently, the accuracy of the closed-form solution of LP in class-imbalanced scenarios is linked to the eigenvalue distribution of $\boldsymbol{S}$, and the degree of class imbalance. Combined with what we get for the iterative solution, this means that to explicitly understand how class imbalance impacts both the convergence of iterations and the accuracy of solutions, it is essential to analyze its effect on the eigenvalue distribution of $\boldsymbol{S}$.

**Spectral shifts induced by class imbalance.** According to the Bauer-Fike theorem [3], for any eigenvalue $\lambda$ of $\boldsymbol{S}$, we propose that the following holds:

$$\min_{\tilde{\lambda}\in\sigma(\tilde{\boldsymbol{S}})}|\lambda-\tilde{\lambda}|\leq\frac{1-\alpha}{\alpha}\cdot\|\Delta\boldsymbol{Y}\|\cdot\|\boldsymbol{M}^{-1}\|, \tag{3}$$

where $\sigma(\tilde{\boldsymbol{S}})$ is the eigenvalues of $\tilde{\boldsymbol{S}}$ under class-balanced scenarios.

Proofs of Theorems 3.1, 3.2 and Eq. (3) are provided in Secs 2.1∼2.3 of supplementary material. In class-imbalanced scenarios, the disparity between well-represented and under-represented classes creates heterogeneous connectivity patterns across the similarity graph [22]. This structural asymmetry manifests as systematic shifts in the spectral properties, where certain regions of the eigenvalue spectrum become more densely populated while others become more sparse.

In summary, the severity of class imbalance directly correlates with greater spectral shifts, which in turn negatively affects the convergence and accuracy of LP. Based on these theoretical insights, we propose a principled approach that explicitly incorporates class distribution information into the affinity matrix construction, theoretically proven to mitigate the adverse effects of spectral shifts.

### 3.3. Propagation energy-based tightened reliability beyond softmax confidence

Reliability confidence in LP predominantly relies on relative class affinities (e.g., softmax-based measures), which examine pairwise relationships between nodes and individual classes. However, this relative connection-based perspective can lead to suboptimal reliability estimation, particularly for boundary samples that exhibit complex connectivity patterns across multiple classes.

For the $i$-th sample, $M_{ij}$ represents the score for class $j$, $j \in \{1, \ldots, C\}$. When properly propagated, the propagation score $M_{ic}$ for the true class $c$ should be the highest among all class scores. We define the propagation score margin as $\delta_k = M_{ic} - \max_{k \neq c} M_{ik}$.

**Softmax-based confidence.** For a borderline minority class sample $x_i$ that is strongly connected to both its true class and a majority class, we have $\delta_k \to 0$. Then, we can use the Taylor series expansion to approximate the softmax-based confidence for it:

$$\text{Confidence}(x_i) = \max_j \left( \frac{\exp M_{ij}}{\sum_{k=1}^{C} \exp M_{ik}} \right) \approx \frac{1}{2} + \frac{\delta_k}{4}. \quad (4)$$

Due to its dependence on relative class affinities, softmax-based confidence exhibits a strong smoothing effect that disproportionately penalizes borderline samples. This is particularly detrimental for minority class samples near decision boundaries, as their legitimate class connections are overshadowed by the presence of connections to majority classes. Consequently, these crucial minority samples are often prematurely discarded despite having meaningful class-specific relationships, exacerbating the class imbalance issue and limiting the model's ability to effectively leverage them for class-imbalanced learning.

Therefore, we propose that sample reliability in LP should be evaluated through the perspective of holistic topological properties rather than solely through class-specific connections. This insight motivates our introduction of propagation energy as a holistic reliability metric.

**Propagation energy-based tightened reliability.** We define the *propagation energy* of a node as a holistic topological measure that quantifies its overall activation strength within the propagation graph. For sample $\boldsymbol{x}_i$, we introduce class-dependent weight $p_j$ for each predicted class $j$, and its propagation energy is defined as:

$$E_p(\boldsymbol{x}_i) = -\log \sum_{j=1}^{C} \exp \frac{M_{ij}}{p_j}, \quad (5)$$

where $E_p : \mathbb{R}^C \to \mathbb{R}$ maps from the space of class propagation scores to a scalar reliability measure, and $p_j$ denotes the proportion of samples in class $i$ relative to the total number of samples across all classes, i.e., $p_j = \frac{n_j}{\sum_{k=1}^{C} n_k}$.

Lower energy values indicate stronger overall connectivity and higher reliability in the propagation graph, irrespective of whether the sample belongs to the majority or minority class. This formulation is inspired by energy-based models in uncertainty quantification [31], which are specifically adapted to capture holistic connectivity patterns in class-imbalanced label propagation graphs.

For justification, as proved in Sec. 2.4 of supplementary material, $E_p(\boldsymbol{x}_i)$ can be further approximated as:

$$E_p(\boldsymbol{x}_i) \approx -\frac{M_{ic}}{p_c} - \frac{\delta_k}{2p_k} - \log 2, \;\; \delta_k \to 0. \quad (6)$$

This is to say, even as $\delta_k \to 0$, the propagation energy can still accurately reflect the strong connection for $c$.

Consequently, our propagation energy-based approach effectively captures both relative class relationships and holistic graph connectivity patterns. This comprehensive perspective enables the identification of reliable samples based on their overall activation strength in the propagation graph, rather than just their propagation score differences.

## 4. Proposed Method

Inspired by our theoretical analysis, our framework adopts an iterative learning strategy that alternates between model optimization and pseudo-label generation during semi-supervised training. As illustrated in Fig. 2, in model training phase, the network processes all samples to produce predictions through our time-frequency enhanced architecture:

$$y^{pred} = \phi^c \left( \phi^e \left( \boldsymbol{x} \right) \right), \quad (7)$$

where $\phi^e$ and $\phi^c$ represent the feature encoder and classifier of our time-frequency enhanced architecture, respectively.

Subsequently, the reliable pseudo-label matrix is computed using our spectral-balanced energy propagation, leveraging the rich representations extracted by the time-frequency enhanced feature encoder:

$$\boldsymbol{Q}_{Re} = E_p \left( \left( \boldsymbol{I} - \alpha \cdot SpeL \left( \phi^e \left( \boldsymbol{x} \right) \right) \right)^{-1} \boldsymbol{Y}^L \right), \quad (8)$$

where $SpeL(\cdot)$ computes our spectral-balanced normalized affinity matrix to address class imbalance, $E_p(\cdot)$ evaluates the propagation energy of propagated labels, and $\boldsymbol{Y}^L$ contains the ground-truth labels for labeled samples. These reliable pseudo-labels guide the learning process for unlabeled samples during subsequent model training.

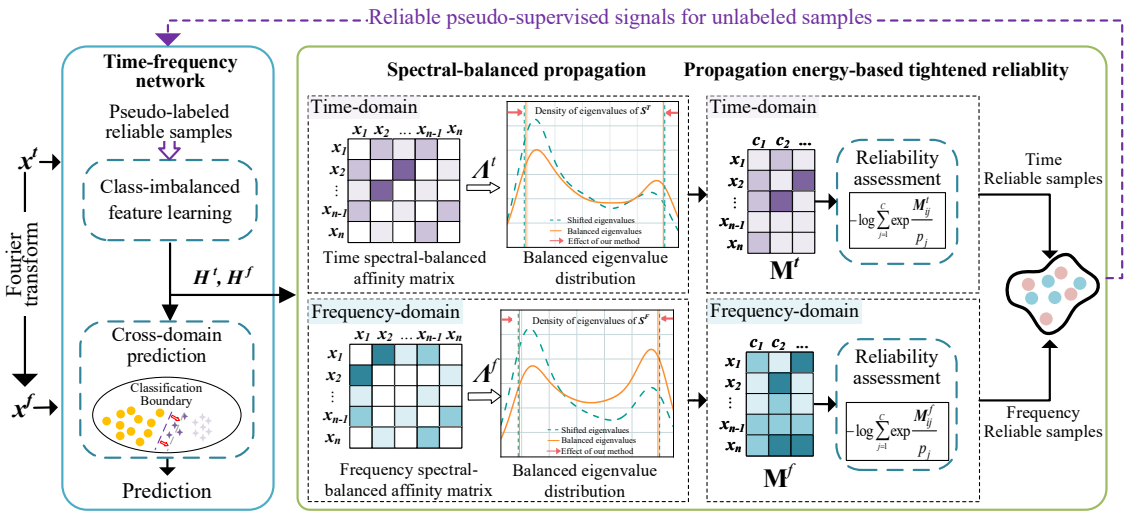


Figure 2. Framework of our model. We employ a spectral-balanced propagation algorithm, along with a propagation energy-based tightened reliability to generate pseudo-labels for unlabeled data and train a network under class imbalance. Our semi-supervised learning process iterates between these two steps. The learned features $\boldsymbol{H}^t$ and $\boldsymbol{H}^f$ are utilized to calculate the spectral-balanced normalized affinity matrix, which exhibits a more balanced eigenvalue ($\lambda$) distribution concentrated in the mid-to-high ranges, rather than the typical skew toward lower values observed in imbalanced datasets. Subsequently, the propagation energy of each pseudo-labeled sample is calculated and the reliable cases are selected for subsequent model training. By learning with reliable boundary samples, the biased feature space and decision boundary can be alleviated.

### 4.1. Time-frequency enhanced network

Our model consists of a time- and a frequency-domain branch, which is detailed in supplementary material. The original input and its frequency representation, transformed via Fourier transform, are fed into their respective encoders and linear classifiers for feature learning and classification.

Our encoders implement an $l$-layer network architecture to extract hierarchical representations from the input, with $l$=5 in our experiments. To enhance the discriminative power of features for minority classes, we introduce strategically designed attention-based feature fusion mechanisms at two critical levels of the network. The shallow features $\boldsymbol{H}^{(1)}$ are propagated to the middle layer through a shallow attention mechanism $Att_s$, while the middle-level representations $\boldsymbol{H}^{(\frac{l}{2}+1)}$ are integrated into the deeper layers via middle attention $Att_m$, creating enriched pathways for information flow. The computational process for each domain encoder are outlined as follows:

$$\boldsymbol{H}^{(\frac{l}{2}+1)} = \phi^{e,(\frac{l}{2}+1)}(Att_s(\boldsymbol{H}^{(\frac{l}{2})}, \boldsymbol{H}^{(1)})), \tag{9}$$

$$\tilde{\boldsymbol{H}}^{(l-1)} = Att_m(\boldsymbol{H}^{(l-1)}, \boldsymbol{H}^{(\frac{l}{2}+1)}), \tag{10}$$

where $Att_s$ and $Att_m$ denote shallow and middle attention-based feature fusion, respectively. $\phi^{e,(\frac{l}{2}+1)}(\cdot)$ refers to the middle layer. Subsequently, the fusion $\tilde{\boldsymbol{H}}^{(l-1)}$ is processed through the final layer in both branches, yielding the final domain representations $\boldsymbol{H}^{t,(l)}$ and $\boldsymbol{H}^{f,(l)}$.

To ensure that both domain branches develop a comprehensive understanding of the input, we introduce a cross-domain fusion operation. This operation focuses on the data patterns most critical for classification, thereby enhancing the deep representations in the corresponding domains.

$$\boldsymbol{H}^t = CrossAtt^t(\boldsymbol{H}^{f,(1)}, \boldsymbol{H}^{t,(l)}), \tag{11}$$

and similarly for the frequency domain $\boldsymbol{H}^f$. $CrossAtt^t$ ($CrossAtt^f$) indicates the cross-domain attention-based feature fusion in time (frequency) domain.

Finally, class probability distributions are computed by passing domain-specific representations through linear projections followed by softmax function. While our method employs transductive learning to generate high-quality pseudo-labels during training, inference follows an inductive paradigm, averaging predictions from both time and frequency branches for unseen samples.

### 4.2. Spectral-balanced energy propagation

Here, we first present a straightforward yet effective spectral-balanced approach to tackle the spectral shift of the normalized affinity matrix $\boldsymbol{S}$ induced by class imbalance, then theoretically justify its efficacy. Following that, we detail the propagation energy-based tightened reliability(ER) to provide reliable pseudo-supervision for network training.

**SpeL and its effects on spectral shifts.** Specifically, for both the time and frequency branches, the representations extracted by the corresponding encoders are fed into the

spectral-balanced label propagation module. Formally, the SpeL similarity between each sample pair is weighted as

$$\hat{w}_{ij}=w_{ij}\cdot\left(p_{c(\boldsymbol{x}_i)}\cdot p_{c(\boldsymbol{x}_j)}\right)^{-\frac{1}{2}}, \tag{12}$$

where $w_{ij}$ represents the Euclidean distance between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The spectral-balanced affinity matrix $\hat{\boldsymbol{W}}$ is then normalized through the degree matrix $\boldsymbol{D}$.

To theoretically justify the effectiveness of SpeL in achieving fine control over the impact of each class on eigenvalues, we define $p_m$ as the proportion of samples in the smallest class, i.e., $\min_{c\in\{1,\dots,C\}} p_c$. For a class $c$ containing sample $\boldsymbol{x}_i$, we further define $\Delta_i$ as the difference in proportion between $p_{c(\boldsymbol{x}_i)}$ and $p_m$, i.e., $p_{c(\boldsymbol{x}_i)}=p_m+\Delta_i$.

Then in theory, for any $p_m>0$, with our SpeL, the derivative of $\lambda(\lambda\neq 1)$ of $\boldsymbol{S}$ with respect to $p_m$ can be expressed as:

$$\frac{\partial\lambda}{\partial p_m}=-\lambda\sum_{i,j}^{n}\frac{(2p_m+\Delta_i+\Delta_j)}{(p_m+\Delta_i)\cdot(p_m+\Delta_j)}, \tag{13}$$

Detailed derivation is provided in Sec. 2.5 of supplementary material. As indicated by the derivative in Eq. (13), it is clear that as $p_m$ decreases, the influence of weight adjustments on the eigenvalue becomes more pronounced. In contrast, higher values of $\Delta_i$, corresponding to the majority classes, are associated with smaller absolute derivatives, meaning their impact on the eigenvalues is reduced. Therefore, weights assigned by SpeL to minority class samples amplify their impact on the eigenvalues while diminishing the influence of majority class samples, thereby alleviating the spectral shifts, resulting in a more balanced eigenvalue density distribution and strengthening the internal connections within the minority class.

**Propagation energy-based tightened reliability.** Building upon our theoretical analysis, we introduce a propagation energy-based approach for precise assessment of sample reliability, especially the reliable boundary samples. First, we derive the pseudo-labels $\boldsymbol{Q}=[q_1,\dots,q_B]^\top$ by applying an argmax operation to the averaged propagation matrices $\boldsymbol{M}^t$ and $\boldsymbol{M}^f$ from the time and frequency domains.

With the propagation score assigned to each sample, the energy distributions $\boldsymbol{\xi}^{\{t,f\}}\in\mathbb{R}^B$ can be derived through Eq. (5). Using the energy thresholds, the label matrix of reliable samples is calculated by:

$$\mathcal{Q}_{Re}=\{q_k\mid k\in\mathcal{R}\}, \tag{14}$$

where $\mathcal{R}$ is the index set of reliable samples satisfying:

$$\mathcal{R}=\{i\mid \xi_i^t\leq\tau_i^t\vee\xi_i^f\leq\tau_i^f, i\in\{1,\dots,B\}\}, \tag{15}$$

To strengthen the reliable boundary samples selection and avoid misclassification, energy thresholds $\boldsymbol{\tau}^{\{t,f\}}$ for each class are adapted dynamically. Specifically, the thresholds are adjusted at each epoch based on the base threshold $\hat{\boldsymbol{\tau}}$ and an adjusting factor derived from $\mathcal{I}\left(\frac{n_i}{\max(n_k)}\right)$, where $k\in\{1,\dots,C\}, \mathcal{I}(v)=\frac{v}{2-v}$. Thus, by adaptively adjusting the energy threshold for sample selection, the SpeLER can better discriminate reliable pseudo-labeled samples.

In summary, our framework establishes a mutually reinforcing cycle between model training and pseudo-label learning under class imbalance. Our spectral-balanced energy propagation generate high-quality pseudo-labels that guide model training, while simultaneously enhancing the network's representational capability. This approach effectively alleviates spectral shifts at the feature level, resulting in a robust framework that addresses the challenges of class-imbalanced semi-supervised learning.

### 4.3. Training objective

To enhance the class-imbalanced feature learning capability of our model, we introduce intra-domain and inter-domain contrastive learning losses. The cross-domain contrastive loss is computed by $-\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp\left(\cos\left(\boldsymbol{z}_i^{d1},\boldsymbol{z}_{i^+}^{d2}\right)/\tau^{d1,d2}\right)}{\sum_{k=1}^{B}\exp\left(\cos\left(\boldsymbol{z}_i^{d1},\boldsymbol{z}_k^{d2}\right)/\tau^{d1,d2}\right)}$. Here, $d1$ and $d2$ denote different domain representations (time or frequency), $\tau^{d1,d2}$ controls the concentration level of the distribution, and $\boldsymbol{z}_{i^+}^{d2}$ corresponds to features from domain $d2$ that belong to the same class as $\boldsymbol{z}_i^{d1}$ from domain $d1$.

During the supervised learning phase, the cross-entropy loss and contrastive loss are calculated for labeled samples in a domain branch. Except for the loss of each branch, the cross-domain contrastive loss $\mathcal{L}_{CL}^{t,f}$, weighted by the balancing factor $\mu$, is incorporated. During the semi-supervised learning phase, the supervisions for prediction consist of known labels for labeled samples and the pseudo-labels for reliable unlabeled samples. We utilize the reliability weight $\omega(E_{\boldsymbol{p}}(\cdot))$ to weight the contribution of pseudo-labeled samples in the loss function, which is defined as:

$$\omega(E_p(\boldsymbol{x}))=\frac{\max(0,-E_p(\boldsymbol{x}))}{\max_{\boldsymbol{x}'\in\mathcal{B}}\max(0,-E_p(\boldsymbol{x}'))}, \tag{16}$$

where $\mathcal{B}$ represents current mini-batch. Eq. (16) ensures that samples with higher reliability (lower energy) contribute more significantly to the training process.

## 5. Experimental Setup and Results

### 5.1. Experimental setup

**Datasets.** We evaluate our approach on three major skeleton-based benchmarks and a wearable sensor-based human action recognition dataset: NTU RGB+D, NTU RGB+D 120 [34], Kinetics-Skeleton [4] and UCI HAR dataset [2]. For NTU RGB+D (NTU-60), we follow standard Cross-Subject (X-sub) and Cross-View (X-view) evaluation protocols. NTU RGB+D 120 (NTU-120) extends

Table 1. Experimental results on NTU-60, NTU-120 and Kinetics-Skeleton with $\gamma$=25%. To evaluate model robustness, we conduct experiments across various class-imbalanced scenarios. The best and second-best results are highlighted in bold and underlined, respectively. Models marked with ‡ are supervised learning models and label propagation algorithm that have been adapted for our research.

| Model | NTU-60 X-sub | | NTU-60 X-view | | NTU-120 X-sub | | NTU-120 X-setup | | Kinetics-Skeleton | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\pi$=60 | $\pi$=30 | $\pi$=60 | $\pi$=30 | $\pi$=60 | $\pi$=30 | $\pi$=60 | $\pi$=30 | $\pi$=60 | $\pi$=30 |
| CrosSCLR | 57.84$_{\pm 0.42}$ | 58.45$_{\pm 0.34}$ | 59.92$_{\pm 0.36}$ | 67.22$_{\pm 0.15}$ | 52.51$_{\pm 0.42}$ | 52.95$_{\pm 0.33}$ | 54.34$_{\pm 0.46}$ | 61.03$_{\pm 0.72}$ | 24.25$_{\pm 0.87}$ | 25.96$_{\pm 1.25}$ |
| FreqMix.‡ | 59.19$_{\pm 0.13}$ | 60.01$_{\pm 0.15}$ | 61.32$_{\pm 0.43}$ | 67.39$_{\pm 0.33}$ | 53.85$_{\pm 0.41}$ | 54.10$_{\pm 0.94}$ | 56.84$_{\pm 0.12}$ | 61.57$_{\pm 1.04}$ | 26.13$_{\pm 1.34}$ | 27.84$_{\pm 1.39}$ |
| BRL‡ | 61.30$_{\pm 0.42}$ | 61.42$_{\pm 0.31}$ | 62.38$_{\pm 0.23}$ | 68.02$_{\pm 0.15}$ | 54.19$_{\pm 0.46}$ | 55.93$_{\pm 0.37}$ | 58.00$_{\pm 0.71}$ | 62.15$_{\pm 0.82}$ | 25.98$_{\pm 2.33}$ | 27.69$_{\pm 1.16}$ |
| AimCLR | 61.98$_{\pm 0.25}$ | 62.40$_{\pm 0.22}$ | 63.91$_{\pm 0.41}$ | 69.19$_{\pm 1.00}$ | 55.96$_{\pm 0.30}$ | 57.03$_{\pm 0.15}$ | 57.02$_{\pm 0.21}$ | 62.91$_{\pm 0.93}$ | 26.25$_{\pm 1.34}$ | 28.33$_{\pm 2.21}$ |
| CTR-GCN‡ | 62.50$_{\pm 0.44}$ | 63.14$_{\pm 0.25}$ | 64.98$_{\pm 0.50}$ | 70.30$_{\pm 0.26}$ | 57.10$_{\pm 0.53}$ | 58.91$_{\pm 0.33}$ | 59.74$_{\pm 0.35}$ | 63.03$_{\pm 0.16}$ | 26.64$_{\pm 1.15}$ | 29.01$_{\pm 1.54}$ |
| MAC | 63.01$_{\pm 0.21}$ | 64.91$_{\pm 0.26}$ | 65.13$_{\pm 0.22}$ | 71.18$_{\pm 0.25}$ | 58.81$_{\pm 0.99}$ | 60.04$_{\pm 0.29}$ | 60.88$_{\pm 0.91}$ | 63.98$_{\pm 0.17}$ | 27.16$_{\pm 1.25}$ | 29.83$_{\pm 1.07}$ |
| Shap-Mix | 63.83$_{\pm 0.53}$ | 65.39$_{\pm 0.21}$ | 66.58$_{\pm 0.97}$ | 71.99$_{\pm 0.42}$ | 60.18$_{\pm 0.78}$ | 61.33$_{\pm 0.31}$ | 62.06$_{\pm 0.16}$ | 64.01$_{\pm 0.22}$ | 27.71$_{\pm 1.46}$ | 29.73$_{\pm 1.81}$ |
| protoLP‡ | 64.10$_{\pm 1.14}$ | 65.22$_{\pm 0.93}$ | 66.98$_{\pm 1.01}$ | 71.83$_{\pm 0.95}$ | 60.59$_{\pm 1.71}$ | 62.36$_{\pm 1.36}$ | 62.44$_{\pm 0.81}$ | 63.97$_{\pm 0.64}$ | 28.34$_{\pm 1.78}$ | 30.13$_{\pm 1.27}$ |
| SCD-Net | 64.32$_{\pm 0.94}$ | 66.01$_{\pm 0.15}$ | 67.13$_{\pm 0.31}$ | 72.95$_{\pm 0.25}$ | 61.05$_{\pm 0.14}$ | 62.18$_{\pm 0.16}$ | 62.85$_{\pm 0.17}$ | 64.39$_{\pm 0.25}$ | 28.29$_{\pm 1.72}$ | 30.75$_{\pm 1.33}$ |
| **SpeLER** | **65.02**$_{\pm 0.37}$ | **67.00**$_{\pm 0.24}$ | **68.75**$_{\pm 0.22}$ | **73.26**$_{\pm 0.23}$ | **62.98**$_{\pm 0.31}$ | **63.87**$_{\pm 0.23}$ | **64.20**$_{\pm 0.12}$ | **65.38**$_{\pm 0.11}$ | **30.01**$_{\pm 1.32}$ | **31.36**$_{\pm 1.32}$ |

the original dataset with 60 additional action classes, utilizing Cross-Subject (X-sub) and Cross-Setup (X-setup) settings for evaluation. We further demonstrate the generalizability of our model on two classical time-series classification (TSC) datasets: SleepEDF and Epilepsy [1, 14].

**Implementation Details.** Building on existing methods for capturing spatio-temporal features, we use CTR-GCN [6] as the backbone for skeleton action recognition, and FCN [44] for classical TSC tasks. Top-1 accuracy is used as the metric to evaluate the classification performance. We compare our method with *i*) Partial-label learning methods, including CrosSCLR [27], MAC [36], AimCLR [41], FreqMixFormer (FreqMix.) [47], TS2Vec [50], CA-TCC [15], SimMTM [10], TS-TFSIAM [30], MTFC [45], TS-TFC [32], SCD-Net [46], TS-CoT [53] and FixMatch [37], and *ii*) class-imbalanced learning methods, including BRL [29], Shap-Mix [52] and protoLP [56]. For fair comparison, the single-domain baselines CTR-GCN, FreqMix., protoLP and FixMatch are adapted to cross-domain structures. Hyperparameter settings and more validations are provided in supplementary materials.

## 5.2. Experimental results analysis

**Overall Performance.** Table 1 presents comprehensive comparisons of our SpeLER approach with state-of-the-art methods for SAR under varying class-imbalanced scenarios. Under the challenging scenario with 25% labeled data, SpeLER demonstrates significant performance advantages. With imbalance ratio $\pi$=60, SpeLER achieves 65.02% and 68.75% accuracy on NTU-60 under x-sub and x-view protocols respectively The performance gains are particularly pronounced on the more complex NTU-120 and Kinetics-Skeleton datasets. These consistent and substantial improvements across diverse benchmarks conclusively demonstrate the efficacy of our cross-domain learning in extracting discriminative representations from limited, skewed data. Moreover, our experiments reveal that

Table 2. Results on NTU-60 and NTU-120 with $\gamma$=50%.

| Model | NTU-60 | | | | NTU-120 | | | |
|---|---|---|---|---|---|---|---|---|
| | X-sub | | X-view | | X-sub | | X-setup | |
| | $\pi$=100 | $\pi$=60 | $\pi$=100 | $\pi$=60 | $\pi$=100 | $\pi$=60 | $\pi$=100 | $\pi$=60 |
| CrosSCLR | 59.46 | 61.19 | 60.69 | 62.21 | 53.39 | 55.13 | 54.98 | 56.40 |
| FreqMix.‡ | 61.10 | 62.93 | 60.06 | 62.83 | 54.91 | 56.29 | 56.90 | 57.59 |
| BRL‡ | 61.78 | 63.51 | 62.31 | 64.22 | 55.39 | 57.12 | 57.34 | 59.25 |
| AimCLR | 62.82 | 64.25 | 63.09 | 65.43 | 55.91 | 58.02 | 58.59 | 60.43 |
| CTR-GCN‡ | 63.52 | 64.77 | 64.35 | 66.44 | 57.32 | 59.49 | 59.44 | 61.49 |
| MAC | 63.64 | 64.24 | 63.27 | 65.03 | 58.93 | 60.65 | 61.71 | 62.68 |
| Shap-Mix | 65.12 | 66.81 | 63.56 | 65.98 | 60.75 | 61.19 | 62.66 | 63.71 |
| protoLP‡ | **66.49** | 67.24 | 67.46 | 69.32 | 62.92 | 63.97 | 63.78 | 64.07 |
| SCD-Net | 66.39 | 67.59 | 69.61 | 71.44 | 64.47 | 63.24 | 63.34 | 64.55 |
| **SpeLER** | 66.45 | **68.67** | **70.06** | **72.75** | **64.71** | **66.02** | **64.10** | **66.75** |

Table 3. Results on wearable sensor-based action recognition and classical TSC datasets with $\gamma$=10%. † denotes the reproduced results, while the remaining results are reported in [53].

| Model | HAR($\pi$=4.00) | SleepEDF($\pi$=6.34) | Epilepsy ($\pi$=1.38) |
|---|---|---|---|
| CA-TCC† | 92.12 | 82.44 | 92.91 |
| TS-TFC† | 93.77 | 82.00 | 97.04 |
| TS-CoT | 94.27 | 81.91 | 97.62 |
| SimMTM† | – | – | 97.65 |
| MTFC† | – | – | 97.39 |
| TS-TFSIAM | 89.00 | 84.00 | 95.50 |
| SpeLER-FCN | **96.56** | **84.31** | **98.80** |

FreqMix., despite its time-frequency modeling adapted for semi-supervised settings, exhibits performance degradation under complex scenarios, indicating that the mere incorporation of frequency domain features is insufficient for addressing imbalance challenges. SpeLER consistently outperforms protoLP, which employs prototype learning as its primary mechanism for addressing class imbalance. These performance advantage demonstrates the effectiveness of our cross-domain learning approach with spectral-balanced energy propagation in addressing the challenges of partial

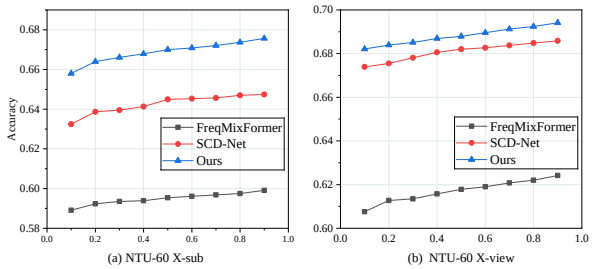


Figure 3. Accuracy of SpeLER, FreqMix. and SCD-Net with different labeled ratios on NTU-60 X-sub and X-view($\pi = 30$).

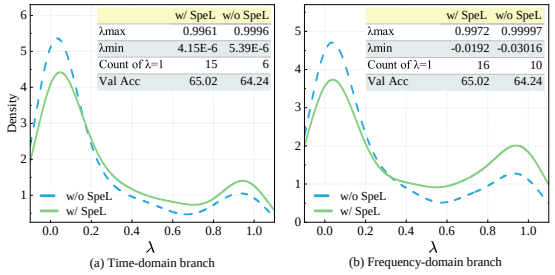


Figure 4. Density estimations of eigenvalues of normalized similarity matrix in each domain branch on NTU-60 ($\pi$=60, $\gamma$=25%).

Table 4. Ablation study of SpeLER components. w/o indicates removal of specific components. Models with $\triangledown$ use the original LP method, while model with $\star$ replace LP with FixMatch.

| Model variants | NTU-60 X-sub | SleepEDF | HAR |
|---|---|---|---|
| | $\pi = 30, \gamma = 25\%$ | $\gamma = 10\%$ | $\gamma = 10\%$ |
| SpeLER | **65.02** | **84.31** | **96.56** |
| *w/o SpeL & ER* $\triangledown$ | 63.30 | 82.41 | 95.10 |
| *w/o SpeL & ER* $\star$ | 64.39 | 82.55 | 94.39 |
| *w/o SpeL* | 64.24 | 83.23 | 95.27 |
| *w/o ER* | 64.36 | 83.37 | 95.33 |
| *w/o Cross-domain Fusion* | 64.19 | 82.26 | 95.06 |

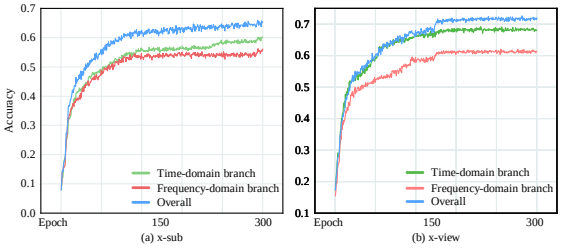


Figure 5. Model accuracy on NTU-60 X-view ($\pi = 60$).

labeling and class imbalance.

**Model Robustness Analysis.** Table 2 shows that SpeLER maintains its performance advantage with 50% labeled data across both NTU-60 / 120 datasets under various protocols and imbalance ratios. Fig. 3 demonstrates SpeLER's robustness across different labeled data percentages compared to FreqMix., and SCD-Net. While the performance gap diminishes as labeled data increases, SpeLER consistently outperforms competing methods in all scenarios, confirming its effectiveness in handling various levels of label scarcity.

**Model Generalizability Analysis.** Table 3 demonstrates SpeLER-FCN's SOTA performance on classical TSC benchmarks with only 10% labeled samples. These improvements substantiate the efficacy of SpeLER in extracting discriminative features from extremely limited labeled data across different domains and application scenarios, highlighting its strong generalization capability.

### 5.3. Ablation study

**Component effectiveness.** Table 4 shows that removing SpeL and ER together causes the largest performance drop, confirming their critical role in addressing class imbalance. Replacing our LP with the standard consistency-based method FixMatch [37] yields inferior results, demonstrating SpeLER's superior handling of imbalanced data. Cross-domain fusion also proves essential, highlighting the importance of integrated time-frequency information.

**Spectral balance.** Fig. 4 demonstrates that our SpeL approach redistributes eigenvalues toward the mid-to-high range, effectively preventing their detrimental concentration at extremely low values. This spectral rebalancing mechanism fundamentally addresses propagation distortion induced by class imbalance, substantially enhancing the robustness and stability of the label propagation process.

**Cross-domain learning.** Fig. 5 shows that integrating time and frequency domains enables our model to capture multi-dimensional structural patterns in skeleton data, leading to more robust feature representations and enhancing classification accuracy, particularly in class-imbalanced scenarios.

## 6. Conclusion

In this work, we theoretically reveal the intricate relationships between partial-label learning and model training under class imbalance. In view of this, we propose SpeLER, which features a spectral-balanced affinity matrix construction that theoretically counteracts eigenvalue distortions through explicit modeling of class distributions, coupled with a propagation energy-based sample reliability assessment mechanism that preserves crucial boundary instances. Comprehensive evaluations demonstrated that SpeLER consistently achieves SOTA performance under varying conditions of class imbalance and label scarcity. Future work will extend our investigation to representation learning in scenarios with abundant labeled data, where spectral properties and boundary sample dynamics may exhibit distinct characteristics.

## References


[1] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001. 7

[2] D Anguita, A Ghio, L Oneto, X Parra, JL Reyes-Ortiz, et al. A public domain dataset for human activity recognition using smartphones. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 437–442, 2013. 6

[3] F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2(1):137–141, 1960. 3

[4] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017. 6

[5] Kaixuan Chen, Lina Yao, Dalin Zhang, Xianzhi Wang, Xiaojun Chang, and Feiping Nie. A semi-supervised recurrent convolutional attention model for human activity recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5):1747–1756, 2020. 2

[6] Yuxin Chen, Ziqi Zhang, Chunfeng Yuan, Bing Li, Ying Deng, and Weiming Hu. Channel-wise topology refinement graph convolution for skeleton-based action recognition. In *IEEE/CVF International Conference on Computer Vision*, pages 13359–13368, 2021. 1, 7

[7] Yao Cheng, Caihua Shan, Yifei Shen, Xiang Li, Siqiang Luo, and Dongsheng Li. Resurrecting label propagation for graphs with heterophily and label noise. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, page 433–444, 2024. 1

[8] Hyung-gun Chi, Myoung Hoon Ha, Seunggeun Chi, Sang Wan Lee, Qixing Huang, and Karthik Ramani. Infogcn: Representation learning for human skeleton-based action recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20186–20196, 2022. 1

[9] Willian Dihanster Gomes de Oliveira and Lilian Berton. A systematic review for class-imbalance in semi-supervised learning. *Artificial Intelligence Review*, 56(2):2349–2382, 2023. 2

[10] Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm: A simple pre-training framework for masked time-series modeling. In *Advances in Neural Information Processing Systems*, 2024. 7

[11] Guodong Du, Jia Zhang, Ning Zhang, Hanrui Wu, Peiliang Wu, and Shaozi Li. Semi-supervised imbalanced multi-label classification with label propagation. *Pattern Recognition*, 150:110358, 2024. 1, 2

[12] Haodong Duan, Yue Zhao, Kai Chen, Dahua Lin, and Bo Dai. Revisiting skeleton-based action recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2969–2978, 2022. 1

[13] Haodong Duan, Mingze Xu, Bing Shuai, Davide Modolo, Zhuowen Tu, Joseph Tighe, and Alessandro Bergamo. Skeletr: Towards skeleton-based action recognition in the wild. In *IEEE/CVF International Conference on Computer Vision*, pages 13634–13644, 2023. 1

[14] Emadeldeen Eldele, Zhenghua Chen, Chengyu Liu, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, page 809–818, 2021. 7

[15] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. Self-supervised contrastive representation learning for semi-supervised time-series classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12), 2023. 7

[16] Zeping Ge, Youlong Yang, and Zhenye Du. Integrated self-supervised label propagation for label imbalanced sets. *Applied Intelligence*, 54(17):8525–8544, 2024. 3

[17] Jingwen Guo, Hong Liu, Shitong Sun, Tianyu Guo, Min Zhang, and Chenyang Si. Fsar: Federated skeleton-based action recognition with adaptive topology structure and knowledge distillation. In *IEEE/CVF International Conference on Computer Vision*, pages 10400–10410, 2023. 1

[18] Rebeen Ali Hamad, Wai Lok Woo, Bo Wei, and Longzhi Yang. Self-supervised learning for activity recognition based on datasets with imbalanced classes. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 1–14, 2025. 1, 2

[19] Hexiang Huang, Xupeng Guo, Wei Peng, and Zhaoqiang Xia. Micro-gesture classification based on ensemble hypergraph-convolution transformer. In *IJCAI Workshop: Challenge on Micro-gesture Analysis for Hidden Emotion Understanding*, 2023. 2

[20] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2, 3

[21] Haoyu Ji, Bowen Chen, Xinglong Xu, Weihong Ren, Zhiyong Wang, and Honghai Liu. Language-assisted skeleton action understanding for skeleton-based temporal action segmentation. In *European Conference on Computer Vision*, page 400–417, 2024. 1

[22] Chiraag Kaushik, Ran Liu, Chi-Heng Lin, Amrit Khera, Matthew Y Jin, Wenrui Ma, Vidya Muthukumar, and Eva L Dyer. Balanced data, imbalanced spectra: Unveiling class disparities with spectral imbalance. In *International Conference on Machine Learning*, pages 23343–23366, 2024. 3

[23] Qiuhong Ke, Mohammed Bennamoun, Hossein Rahmani, Senjian An, Ferdous Sohel, and Farid Boussaid. Learning latent global network for skeleton-based action prediction. *IEEE Transactions on Image Processing*, 29:959–970, 2020. 1

[24] Brody Kutt, Pralay Ramteke, Xavier Mignot, Pamela Toman, Nandini Ramanan, Sujit Rokka Chhetri, Shan Huang, Min Du, and William Hewlett. Contrastive credibility propagation for reliable semi-supervised learning. In *AAAI Conference on Artificial Intelligence*, pages 21294–21303, 2024. 1, 2

[25] Michalis Lazarou, Tania Stathaki, and Yannis Avrithis. Iterative label cleaning for transductive and semi-supervised

few-shot learning. In *IEEE/CVF International Conference on Computer Vision*, pages 8751–8760, 2021. 1

[26] Jungho Lee, Minhyeok Lee, Suhwan Cho, Sungmin Woo, Sungjun Jang, and Sangyoun Lee. Leveraging spatio-temporal dependency for skeleton-based action recognition. In *IEEE/CVF International Conference on Computer Vision*, pages 10255–10264, 2023. 1

[27] Linguo Li, Minsi Wang, Bingbing Ni, Hang Wang, Jiancheng Yang, and Wenjun Zhang. 3d human action representation learning via cross-view consistency pursuit. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4739–4748, 2021. 7

[28] Ping Li, Jiachen Cao, and Xingchao Ye. Prototype contrastive learning for point-supervised temporal action detection. *Expert Systems with Applications*, 213:118965, 2023. 1

[29] Hongda Liu, Yunlong Wang, Min Ren, Junxing Hu, Zhengquan Luo, Guangqi Hou, and Zhenan Sun. Balanced representation learning for long-tailed skeleton-based action recognition. *Machine Intelligence Research*, 22(3):466–483, 2025. 2, 7

[30] Songbai Liu, Hongru Li, Youhe Huang, and Shuang Wen. Ts-tfsiam: Time-series self-supervised learning with time-frequency siamesenet. *Knowledge-Based Systems*, 288: 111472, 2024. 7

[31] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *Advances in Neural Information Processing Systems*, pages 21464–21475, 2020. 4

[32] Zhen Liu, Qianli Ma, Peitian Ma, and Linghao Wang. Temporal-frequency co-training for time series semi-supervised learning. In *AAAI Conference on Artificial Intelligence*, pages 8923–8931, 2023. 7

[33] Haoxuan Qu, Yujun Cai, and Jun Liu. Llms are good action recognizers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18395–18406, 2024. 1

[34] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+d: A large scale dataset for 3d human activity analysis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1010–1019, 2016. 6

[35] Yunhao Shi, Hua Xu, Zisen Qi, Yue Zhang, Dan Wang, and Lei Jiang. Sttmc: A few-shot spatial temporal transductive modulation classifier. *IEEE Transactions on Machine Learning in Communications and Networking*, 2:546–559, 2024. 1

[36] Xiangbo Shu, Binqian Xu, Liyan Zhang, and Jinhui Tang. Multi-granularity anchor-contrastive representation learning for semi-supervised skeleton-based action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7559–7576, 2023. 2, 7

[37] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, pages 596–608, 2020. 7, 8

[38] Vladan Stojnić, Yannis Kalantidis, and Giorgos Tolias. Label propagation for zero-shot classification with vision-language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23209–23218, 2024. 1

[39] Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, and Jun Liu. Human action recognition from various data modalities: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3200–3225, 2023. 2

[40] Long Tian, Jingyi Feng, Xiaoqiang Chai, Wenchao Chen, Liming Wang, Xiyang Liu, and Bo Chen. Prototypes-oriented transductive few-shot learning with conditional transport. In *IEEE/CVF International Conference on Computer Vision*, pages 16317–16326, 2023. 1, 3

[41] Guo Tianyu, Liu Hong, Chen Zhan, Liu Mengyuan, Wang Tao, and Ding Runwei. Contrastive learning from extremely augmented skeleton sequences for self-supervised action recognition. In *AAAI Conference on Artificial Intelligence*, 2022. 2, 7

[42] Xinshun Wang, Zhongbin Fang, Xia Li, Xiangtai Li, Chen Chen, and Mengyuan Liu. Skeleton-in-context: Unified skeleton sequence modeling with in-context learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2436–2446, 2024. 1

[43] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. Instance credibility inference for few-shot learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 1

[44] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *International Joint Conference on Neural Networks*, pages 1578–1585, 2017. 7

[45] Chixuan Wei, Zhihai Wang, Jidong Yuan, Chuanming Li, and Shengbo Chen. Time-frequency based multi-task learning for semi-supervised time series classification. *Information Sciences*, 619:762–780, 2023. 7

[46] Cong Wu, Xiao-Jun Wu, Josef Kittler, Tianyang Xu, Sara Ahmed, Muhammad Awais, and Zhenhua Feng. Scd-net: Spatiotemporal clues disentanglement network for self-supervised skeleton-based action recognition. In *AAAI Conference on Artificial Intelligence*, pages 5949–5957, 2024. 2, 7

[47] Wenhan Wu, Ce Zheng, Zihao Yang, Chen Chen, Srijan Das, and Aidong Lu. Frequency guidance matters: Skeletal action recognition by frequency-aware mixed transformer. In *ACM International Conference on Multimedia*, page 4660–4669, 2024. 2, 7

[48] Yinghao Xu, Fangyun Wei, Xiao Sun, Ceyuan Yang, Yujun Shen, Bo Dai, Bolei Zhou, and Stephen Lin. Cross-model pseudo-labeling for semi-supervised action recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2959–2968, 2022. 1, 2

[49] Tingbing Yan, Wenzheng Zeng, Yang Xiao, Xingyu Tong, Bo Tan, Zhiwen Fang, Zhiguo Cao, and Joey Tianyi Zhou. Crossglg: Llm guides one-shot skeleton-based 3d action recognition in a cross-level manner. In *European Conference on Computer Vision*, pages 113–131, 2025. 1

[50] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: To-

wards universal representation of time series. In *AAAI Conference on Artificial Intelligence*, 2022. 7

[51] Hongcheng Zhang, Xu Zhao, and Dongqi Wang. Semi-supervised learning for multi-label video action detection. In *ACM International Conference on Multimedia*, pages 2124–2134, 2022. 1, 2

[52] Jiahang Zhang, Lilang Lin, and Jiaying Liu. Shap-mix: Shapley value guided mixing for long-tailed skeleton based action recognition. In *International Joint Conference on Artificial Intelligence*, 2024. 2, 7

[53] Weiqi Zhang, Jianfeng Zhang, Jia Li, and Fugee Tsung. A co-training approach for noisy time series learning. In *ACM International Conference on Information and Knowledge Management*, pages 3308–3318, 2023. 7

[54] Tianxiang Zhao, Xiang Zhang, and Suhang Wang. Imbalanced node classification with synthetic over-sampling. *IEEE Transactions on Knowledge and Data Engineering*, 36 (12):1–14, 2024. 2

[55] Yuxuan Zhou, Xudong Yan, Zhi-Qi Cheng, Yan Yan, Qi Dai, and Xian-Sheng Hua. Blockgcn: Redefining topology awareness for skeleton-based action recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 1

[56] Hao Zhu and Piotr Koniusz. Transductive few-shot learning with prototype-based label propagation by iterative graph refinement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23996–24006, 2023. 1, 2, 7

[57] Yisheng Zhu, Hu Han, Zhengtao Yu, and Guangcan Liu. Modeling the relative visual tempo for self-supervised skeleton-based action recognition. In *IEEE/CVF International Conference on Computer Vision*, pages 13913–13922, 2023. 1