# Balanced Contrastive Learning for Imbalanced Gait Recognition

**Abstract**

Reproducing the referenced article is representation learning for unbalanced data. The main content of this article is to apply the BCL method to the WiFi-based gait recognition task. Since the data set of the gait recognition task is unbalanced, the BCL method is an improved method of SCL for long-tail data sets. Results Experimental results show that the results of using BCL are better than SCL, but the accuracy of using traditional cross-entropy loss is higher than that of using BCL loss, which exposes some problems.After the learning from this experiment, the future research direction is to use self-supervised contrastive learning for the gait recognition based human identification (HI) problem to achieve better results through more fine-grained experimental arrangement and planning.

**Keywords:** Contrastive learning, Gait recognition, WiFi sensing, Human identification

## 1 Introduction

Various sensor forms, such as cameras, wearable inertial sensors, and Wi-Fi signals emitted by wireless devices, have been shown to extract a user's gait for person recognition. Among these sensors, images and videos have the risk of privacy leakage, and inertial sensors require users to actively carry mobile devices [15]. In contrast, Wi-Fi signals are more conducive to gait-based person identification because Wi-Fi infrastructure is widely distributed and can work without the user's knowledge. In addition, research points out that WiFi signals in the form of channel state information (CSI) are very promising for a variety of device-free human sensing tasks, such as occupancy detection, activity recognition, fall detection, gesture recognition, human identification, people counting and posture estimation [12]. Unlike coarse-grained received signal strength, CSI for WiFi records finer-grained information on signal propagation between WiFi devices and its reflection relative to the human environment. However, since WiFi signals (2.4 or 5 GHz) lie in the invisible band of the electromagnetic spectrum, WiFi CSI-based human sensing is inherently more privacy-friendly than camera-based surveillance. Therefore, it has attracted great attention from academia and industry.

Gait is the way a person walks, can be considered as a physical and behavioral characteristic that can be used for human body recognition [16]. Gait recognition is a human body biometric identification technology, based on the principles of sports physiology, human movement mechanics and other disciplines. It uses deep machine learning and neural network algorithms to intelligently analyze characteristics such as human stride length, cadence, and foot swing cycle, thereby achieving accurate target identification. Its main advantages are

non-contact, non-intrusive, already aware, difficult to hide and disguise, etc. It has broad application prospects and economic value in related fields such as security monitoring, access control systems, and medical diagnosis.

Most state-of-the-art deep learning models are developed for computer vision tasks (e.g., human activity recognition) and natural language processing (e.g., text translation). These models have demonstrated their ability to handle high-dimensional and multi-modal data. These methods inspire the use of deep learning in WiFi sensing to achieve data pre-processing, network design, and learning objectives. Therefore, deep learning is increasingly being developed for WiFi sensing.

## 2 Related works

### 2.1 Gait Recognition

Most gait recognition methods are based on computer vision, where gait data are represented as images. Algorithms based on visual gait recognition can be roughly divided into two categories, one is appearance-based methods, and the other is model-based methods. This article focuses on how WiFi sensing devices represent people's gait characteristics through reflected signals when people are walking. Due to the multi-path effect of wireless signals, WiFi signals reflected by the human body will produce unique CSI dynamics [14]. Intuitively, through a series of signal processing techniques, gait features extracted from CSI can be applied to identify humans. However, path dependence results in that if a subject walks along an arbitrary path, the gait information derived from the received signal will vary arbitrarily with random changes in the walking path. This results in dramatic changes in features derived from gait information, and gait recognition based on derived features becomes very challenging.

### 2.2 Supervised Contrastive Learning

Contrastive learning is a technique that enhances the performance of visual tasks, by using the principle of contrasting samples with each other, to learn properties that are common between data classes and that distinguish one class from another.SimCLR [3] and MoCo [5], proposed in 2020, are the two representatives of applying self-supervised contrastive learning to deep learning models. Supervised contrastive learning (SCL) [6]leverages label information for fully supervised representation learning, resulting in state-of-the-art image classification performance.

## 3 Method

### 3.1 Balanced Contrastive Learning

BCL solves the problem of class imbalance through ***class-averaging*** and ***class-complement***. Due to the unbalanced category distribution of long-tail data sets, there are too many head categories, resulting in a long-tail distribution of samples in each batch. During the training process, the model continues to increase the distance between other categories and the head category, so the distance between the tail categories is forced closer, resulting in poor feature separability. Since there are more training samples for the head category, the gradient signal of the head category may dominate, making the gradient of the tail category relatively small and

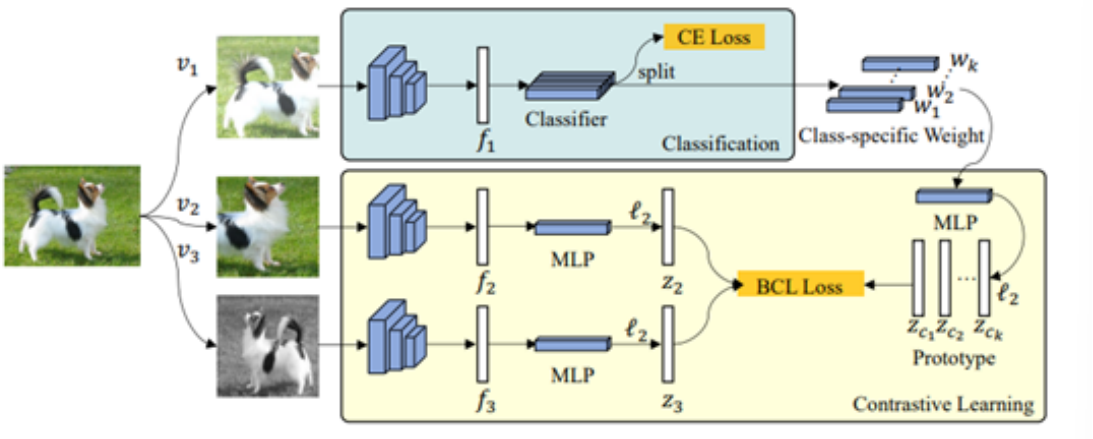


Figure 1. Overview of the proposed framework [18]

likely to be overwhelmed by the gradient of the head category. Therefore, the training loss focuses more on optimizing the head class, so the model parameters of the tail class are difficult to effectively learn. In response to the above problems, ***class-averaging*** averages the instances of each category in the mini-batch to ensure that each category has an approximate contribution to the optimization. The modified loss function is:

$$\mathcal{L} = \frac{-1}{|B_y| - 1} \sum_{p \in\ B_y\{i\}} \log \frac{\exp(z_i \cdot z_p)}{\sum_{j \in y_B} \frac{1}{|B_j|} \sum_{k \in B_j} \exp(z_i \cdot z_k)}$$

With ***class-averaging***, the head class no longer dominates the entire training. However, since each class is not sampled with equal probability, class-complement introduces class center representations or prototypes so that all classes appear in each batch for balanced contrastive learning.

$$\mathcal{L}_{BCL} = \frac{-1}{|B_y|} \sum_{p \in \{B_y \setminus \{i\} \cup \{c_y\}\}} \log \frac{\exp(z_i \cdot z_p)}{\sum_{j \in y} \frac{1}{|B_y|+1} \sum_{k \in B_j \cup \{c_y\}} \exp(z_i \cdot z_k)}$$

Optimizing the loss function is conducive to forming a stable structure between categories, and the corresponding decision boundaries can be better found between categories.

### 3.2 Framework

The overall framework is shown in Figure 1. The framework used in the article contains two main components: the classification branch and the comparative learning branch. Both branches share the same feature extractor. Unlike traditional contrastive learning, which uses a two-stage training strategy, BCL is an end-to-end model. We use the category-specific weights in the classification branch as the prototype after undergoing a nonlinear transformation MLP. The results of the last classification layer of the output logit are usually biased due to data imbalance. The purpose of Logit compensation is to remove the bias caused by data imbalance and learn the correction of the boundaries [2]. The final training loss function is:

$$\mathcal{L} = \lambda \mathcal{L}_{LC} + \mu \mathcal{L}_{BCL}$$

$\lambda$ and $\mu$ are used to control two losses respectively. Furthermore, the contrastive learning branch is only used to allow the backbone to learn the required feature embeddings.

Table 1. Gait-WiFi dataset for training

| user-id | number |
|---|---|
| user1 | 2400 |
| user2 | 2400 |
| user3 | 2400 |
| user4 | 2400 |
| user5 | 1857 |
| user6 | 2400 |
| user7 | 2399 |
| user8 | 1200 |
| user9 | 1200 |

# 4 Implementation details

## 4.1 Comparing with the released source codes

The code of this article is open source, and the code mainly has a complete code display for the two data sets of ImageNet-LT and INaturalist 2018. Due to the limited computing power, the part where the method in this article is applied to the CIFAR-LT dataset is reproduced. Secondly, my idea is to use this method on the gait recognition task to see how effective it is. So this part of the code was modified. Then, the three methods SupCE, SupCL, and BCL were applied to the Gait-WiFi data set to compare which method is more suitable for this data set.

## 4.2 Experimental environment setup

The entire experimental environment uses a server equipped with two NVIDIA GPUs and uses Pytorch 2.1.2 CUDA 11.7 as the main deep learning framework.

The data sets used are CIFAR-10 [7], CIFAR-100, and Gait-WiFi [15]. Among them, CIFAR-10 and CIFAR-100 need to be preprocessed to obtain the long-tail distribution data set under the corresponding imbalance factor.CIFAR-10-LT is the subset of CIFAR-10 consisting of 50,000 images for training and 10,000 images for validation, each color image in the size of 32×32 from 10 categories. CIFAR-100 has the same data size as CIFAR-10 except for 100 classes, both of them are the subset of the Tiny Images dataset. Gait-WiFi dataset contains channel state information (CSI) collected by 11 users using WiFi, which is used for human gait recognition tasks. In the field of WiFi communication, CSI data captures the propagation information of wireless signals in the physical environment after diffraction, reflection and scattering by describing the channel properties of the communication link which can be considered as a ”WiFi picture” of the propagation environment.

Obviously, this training dataset is an unbalanced dataset, in which the number of samples in each category is shown in table 1. The test dataset, illustrated in table 2, is divided into test1 containing the data of user1 and user2, and test2 containing the data of user1, user2, user10, and user11. Since the data of user10 and user11 do not appear in the training dataset, test2 is used to test the generalization ability of the method.

Table 2. Gait-WiFi dataset for testing

| user-id | number |
|---|---|
| user1 | 480 |
| user2 | 480 |
| user10 | 480 |
| user11 | 480 |

Table 3. Top-1 accuracy of ResNet-32 on CIFAR dataset

| Method/Dataset | CIFAR-10-LT | CIFAR-LT-100 |
|---|---|---|
| LDAM-DRW [2] | 77.03 | 42.04 |
| ResLT [4] | 82.40 | 48.21 |
| Hybird-SC [11] | 81.40 | 46.72 |
| **BCL** [18] | **84.32** | **51.93** |
| BCL | 76.640 | 42.290 |

The deep learning model utilized in the experiment is ResNet-32, serving as the backbone. To maintain consistency with [18], identical hyperparameters are employed. Specifically, $\lambda$ is set to 2.0, $\mu$ is 0.6, and the temperature $\tau$ is configured as 0.1. The batch size is defined as 256, with a weight decay of 5e-4 . The training duration spans 200 epochs.

### 4.3 Main contributions

Considering the task of combining imbalanced data sets and WiFi-sensing-based gait recognition has greater practical research significance. Datasets collected in real life are often unbalanced, and sampling bias may exist during the data collection process. For example, recognition through gait information, whether it is the direction and frequency of walking, vary from person to person. Based on the original open source code, this article first reproduces the effect of this method on the CIFAR-LT data set. Subsequently, the author modified the code and applied the method to the gait recognition task, extending the method to the field of human behavior recognition. Finally, by comparing the three methods SupCE, SupCL and BCL on the Gait-WiFi dataset, their applicability in the gait recognition task was systematically evaluated. This research provides new ideas and empirical support for the expansion of methods in the field of gait recognition, and provides a useful reference for further research and application.

## 5 Results and analysis

Table 3 summarizes the results of the experimental reproductions in comparison to some of the results in the article, where the results obtained were not as stunning as the article demonstrated.

According to Table 4, it can be seen that The improved BCL does work better than the SCL. Opposite to what we expected, the Top-1 accuracy obtained by the traditional cross-entropy loss approach is instead higher than that obtained by BCL, both on test1 and test2.

Table 4. Top-1 accuracy of ResNet-32 on Gait-WiFi dataset

| Method/Test | Test-1 | Test-2 |
|---|---|---|
| **SupCE** | **42.08** | **32.92** |
| SCL | 22.19 | 16.41 |
| BCL | 29.896 | 27.604 |

## 6 Conclusion and future work

In this work, applying BCL to the task of gait recognition under WiFi sensing, we find that the results are rather not as good as the traditional cross entropy method. Different loss functions and methods may produce different results for different types of data and tasks. Further studies on the CSI dataset yielded the following inspiration:

### 6.1 Improvements in experimental settings

The first is an understanding of the data set used in the experiment. This article treats CSI data as "images" to process. CSI data can also be viewed as time series [9], Time-Frequency diagrams [8], frequency-domain features [10], etc. Especially with the advancement of deep models, the data patterns they can process are not single, which also expands ideas for gait recognition research. At the same time, the data set has not been preprocessed, and the entire experimental process is still relatively rough. The CSI data collected in the data set reflects not only the gait characteristics but also the interference signals of the surrounding environment. Therefore, when WiFi devices are in different environments, the data collected on the same person may be quite different. Based on the datasets obtained in the exact environment, the recognition model may be trained to overfit, so its generalization ability is poor, and it may need to be retrained in new environments. Some studies such as CrossSense [13] and TransferSense [1] use transfer learning to complete gait recognition tasks in different monitoring environments.

In addition to environmental effects, the characteristics of a person's gait in a WiFi signal still depend on how the person moves relative to the WiFi device. Many studies are based on some data sets with big assumptions. These data collection methods require users to walk in specific directions to obtain corresponding walking trajectories or have fixed WiFi receiver layouts, etc. The research results obtained based on these models have limited practicality, making it difficult to achieve deployment in real life [17]. Therefore, for the experiment to have certain practical significance, the experiment needs to construct a data set with sufficient data volume (number of samples, data dimensions, etc.).

Furthermore, when this experiment was conducted, there was no clear understanding of the category classification effect before and after the experiment. Some visualization methods are necessary for further analysis.

Finally, according to the experimental results in the [12] on some deep neural networks using supervised learning on datasets (UT-HAR, Widar, NTU-Fi HAR, NTU-Fi Human ID), the Widar dataset, which also uses CSI data, shows that among the highest accuracies are ResNet18 (71.70), CNN-5 (70.19). In this paper, we are using ResNet-32 for training, and other options are also worth trying.

### 6.2 Self-supervised contrastive learning in gait-based human identification

As a biometric identification method, gait recognition has a wide range of applications in the fields of identity verification, security surveillance and health monitoring. However, CSI (Channel State Information) based gait identification faces difficulties such as diversity, dynamic environment and individual differences. The self-supervised technique of contrast learning provides a possible solution to these challenges, further exploring the research value of unsupervised contrast-based learning applied to identity recognition:

First, contrast learning provides an unsupervised feature learning mechanism for gait detection, which helps the system to automatically extract and learn gait features through self-similarity learning, as well as by distinguishing the differences between different gaits of the same person or between different individuals under the same gait, thus alleviating the problems of data scarcity and individual differences, and improving the accuracy of identity recognition.

Second, contrast learning helps to achieve environmental adaptation and generalization performance. By learning the self-similarity of gait patterns, the model can be better adapted to different environments and devices, mitigating the effects of environmental differences.

Finally, contrast learning can learn without real identity labels, which helps protect user privacy.

Based on these ideas, it is expected to make innovative contributions in improving recognition performance, reducing the need for labeled data, and enhancing model generalization performance.

## References


[1] Qirong Bu, Xingxia Ming, Jingzhao Hu, Tuo Zhang, Jun Feng, and Jing Zhang. Transfersense: towards environment independent and one-shot wifi sensing. *Personal and Ubiquitous Computing*, 26, 06 2022.

[2] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. 2019.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

[4] Jiequan Cui, Shu Liu, Zhuotao Tian, Zhisheng Zhong, and Jiaya Jia. Reslt: Residual learning for long-tailed recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3695–3706, 2023.

[5] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735, 2020.

[6] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[7] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009.

[8] Lokesh Sharma, Chung-Chieh Chao, Shih-Lin Wu, and Mei-Chen Li. High accuracy wifi-based human activity classification system with time-frequency diagram cnn method for different places. *Sensors (Basel, Switzerland)*, 21, 2021.

[9] Yong Tian, Chen Chen, Qiyue Zhang, Ying Li, Sirou Li, and Xuejun Ding. Multidimensional information recognition algorithm based on csi decomposition. *IEEE Internet of Things Journal*, 10(10):9234–9248, 2023.

[10] Yong Tian, Sirou Li, Chen Chen, Qiyue Zhang, Chuanzhen Zhuang, Xuejun Ding, and Xin Liu. Small csi samples-based activity recognition: A deep learning approach using multidimensional features. *Sec. and Commun. Netw.*, 2021, jan 2021.

[11] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 943–952, 2021.

[12] Jianfei Yang, Xinyan Chen, Dazhuo Wang, Han Zou, Chris Xiaoxuan Lu, Sumei Sun, and Lihua Xie. Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing. *Patterns*, 4(3), 2023.

[13] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. Crosssense: Towards cross-site and large-scale wifi sensing. 2018.

[14] Lei Zhang, Cong Wang, and Daqing Zhang. Wi-pigr: Path independent gait recognition with commodity wi-fi. *IEEE Transactions on Mobile Computing*, 21(9):3414–3427, 2022.

[15] Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. Gaitid: Robust wi-fi based gait recognition. page 730–742, 2020.

[16] Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. Gaitsense: Towards ubiquitous gait-based human identification with wi-fi. *ACM Trans. Sen. Netw.*, 18(1), oct 2021.

[17] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-effort cross-domain gesture recognition with wi-fi. page 313–325, 2019.

[18] J. Zhu, Z. Wang, J. Chen, Y. Chen, and Y. Jiang. Balanced contrastive learning for long-tailed visual recognition. pages 6898–6907, jun 2022.