



CASL: Capturing Activity Semantics through Location Information for enhanced activity recognition

Zhang, X., Cui, S., Zhu, T., Chen, L., Zhou, F., & Ning, H. (2023). CASL: Capturing Activity Semantics through Location Information for enhanced activity recognition. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1-9. <https://doi.org/10.1109/TCBB.2023.3238064>

[Link to publication record in Ulster University Research Portal](#)

Published in:

IEEE/ACM Transactions on Computational Biology and Bioinformatics

Publication Status:

Published (in print/issue): 01/01/2023

DOI:

[10.1109/TCBB.2023.3238064](https://doi.org/10.1109/TCBB.2023.3238064)

Document Version

Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

CASL: Capturing Activity Semantics through Location Information for enhanced activity recognition

Xiao Zhang, Shan Cui, Tao Zhu, Liming Chen, *Member, IEEE*, Fang Zhou, and Huansheng Ning, *Member, IEEE*

Abstract—Using portable tools to monitor and identify daily activities has increasingly become a focus of digital healthcare, especially for elderly care. One of the difficulties in this area is the excessive reliance on labeled activity data for corresponding recognition modeling. Labeled activity data is expensive to collect. To address this challenge, we propose an effective and robust semi-supervised active learning method, called CASL, which combines the mainstream semi-supervised learning method with a mechanism of expert collaboration. CASL takes a user’s trajectory as the only input. In addition, CASL uses expert collaboration to judge the valuable samples of a model to further enhance its performance. CASL relies on very few semantic activities, outperforms all baseline activity recognition methods, and is close to the performance of supervised learning methods. On the adlnormal dataset with 200 semantic activities data, CASL achieved an accuracy of 89.07%, supervised learning has 91.77%. Our ablation study validated the components in our CASL using a query strategy and a data fusion approach.

Index Terms—Healthcare, Deep learning, Semantic annotation, Location information, Semi-supervised active learning.



1 INTRODUCTION

SENSOR-BASED recognition of Activities of Daily Living (SbrADL) is a hot topic in digital healthcare, and its research can provide valuable advice for better healthcare and lifestyle. Most SbrADL rely heavily on supervised learning, which requires many labeled sensor data. Usually, we achieve semantic annotation (recognize and label the sensor data) by observing or monitoring the daily life of the participants. In any case, these annotation methods are manually labor-intensive, time-consuming [1]. The lack of labeled sensor data due to cost is known as annotation scarcity [2].

Many scholars put forward methods to solve the problem of annotation scarcity, which can be divided into knowledge-driven methods and data-driven methods. Knowledge-driven methods are to manually build the ontology model and obtain the activity semantics through the rules obtained by the ontology model. Saguna et al. [3] combine ontology modeling, Spatio-temporal modeling with inference to identify crossover and concurrent activities. Ye et al. [4] calculated semantic similarity among activities, objects, and sensor events using the hierarchical relationship of activities and used semantic similarity to segment sensor event sequences of concurrent activities to obtain partial concurrent activity semantics. Ning et al. [5] provided a ontology to achieve sensor data semantization according to

publicly agreed standards. This new ontology can improve its reuse across different models. Even so, knowledge-based methods require a lot of domain knowledge and artificial modeling costs. When the experiment or application environment changes, all the action rules of knowledge engineering need to be changed accordingly. We focus on the data-based approach because of the lack of portability caused by the above problems.

To address the annotation scarcity, we develop a novel model referred to as CASL, which takes the location information contained in the SbrADL datasets as the only input. Although the way activities are performed may change, the location information (the sensor’s location or the hidden location contained in the multimodal dataset [6]) in each experiment environment is relatively fixed and easy to obtain. When we keep the sensitivity of all the sensors consistent, the activity location information can well match the activity category [7]. Besides, location information can protect user privacy better than other features in the SbrADL datasets [7], [8]. The privacy problem is not the focus of our work and will not be discussed in detail.

The main challenge to SbrADL is that the amount of data in public datasets is limited [9]. It is difficult to avoid overfitting in training deep models through SbrADL datasets. Therefore, we combine the mainstream semi-supervised approach to overcome overfitting, including data augmentation, Consistent Regularization [10] into CASL to solve this problem. The semi-supervised part was proven effective by past work [11]. Then we combined the semi-supervised and active deep learning to form the complete CASL. Moreover, because of the limited amount of data, the heuristic method of active learning is prone to sampling anomalies. The active learning method is ineffective after combining semi-

• X. Zhang, S. cui, F. Zhou and H. Ning are with the School of Computer and Communication Engineering, University of Science and Technology Beijing, 100083, Beijing, China.
E-mail: zhoufang@ies.ustb.edu.cn

• T. Zhu is with the computer school, University of South China, 421001, Hunan, China.

• L. Chen is School of Computer Science and Informatics at University of Ulster, BT37, Newtownabbey, United Kingdom.

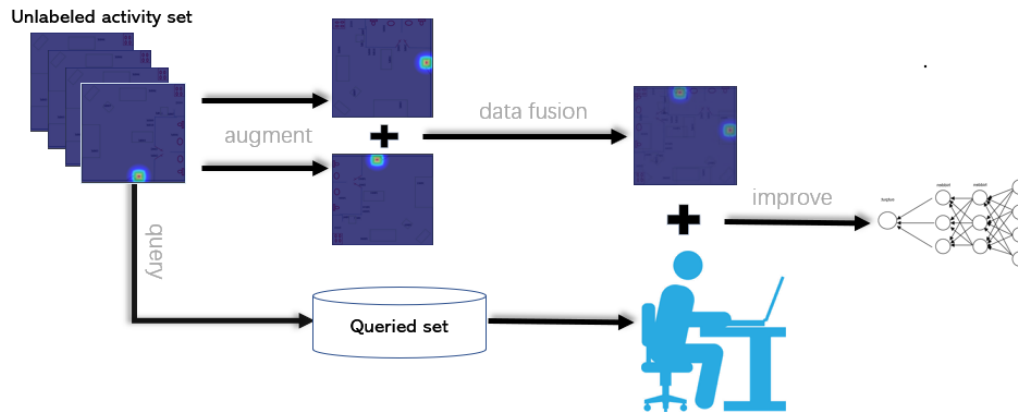


Fig. 1 The heat map of unlabeled sensor data was enhanced several times. The enhanced image is pixelated to augment the data further. Finally, the model is improved through augmented data and expert collaboration.

supervised learning in batch settings. We proposed a novel sampling method called Sampling-based on Distance and Density Tradeoffs (S2DT) to overcome the above problem. This sampling method is based on the different importance of the distance and the sample density in different datasets.

In short, our work is to combine semi-supervised learning, active learning heuristic algorithm and deep model into CASL, with location information as its only input. The CASL is robust and has a reasonable recognition rate even in the case of a small dataset. The main contributions are as follows:

- We propose a new semi-supervised active learning algorithm, which takes location information as input and solve annotation scarcity.
- We develop a novel sampling method to solve the problem of easy sampling of outliers from limited labeled sensor data.
- We demonstrate by ablation experiments that our proposed sampling method applied to CASL achieved state-of-the-art results.

2 RELATED WORK

To set the stage for CASL, we first introduce existing data-driven methods for semantic annotation. Semi-supervised and active learning methods rely on limited labeled data and have shown a growing trend [12].

Semi-supervised methods require less labeled data and a large amount of unlabeled data. Wang et al. [13] proposed a new generative adversarial network framework, called SensoryGAN, which efficiently generates sensor data. However, this approach works poorly for outliers. In order to solve the problem, Zhang et al. [14] proposed semi-supervised GAN. The semi-supervised GAN differs from conventional GANs in that semi-supervised GANs perform $n+1$ classification, including n activity classification and a fake data classification. We borrowed data augmentation from semi-supervised learning and applied it to CASL. Chen et al. [15] proposed a semi-supervised deep model for multi-mode wearable sensor data activity identification. This work addresses the challenges of person-to-person variability and similarity between classes and the problems of finite marker

data and class imbalance. The development of this work provides an idea for us to combine semi-supervised learning with the deep model.

Unlike semi-supervised learning, active learning requires experts or annotators to annotate activities' semantics manually. Active learning aims to select more valuable activities to be judged by people as much as possible and deliver these data to the classifier for judgment. Zhao et al. [16] proposed a new principled active learning instance selection method with stronger robustness to noise in activity semantic annotation. Walter et al. [17] supplement existing recognition systems by using crowd inputs for on-demand, real-time activity recognition to provide robust, deployable activity recognition. Although the sample selected by the query strategy of active learning is crucial for training, the discarded sample is also valuable due to many samples. Therefore, deep active learning approaches are deployed in activity recognition to use information from the missing samples [18], [19]. Hossain et al. [18] embed active learning in the training phase of deep learning to collect activity labels by querying the most information-rich and costliest unlabeled sample points and using low uncertainty instances. The combination of active learning and deep model brings new inspiration to the research of activity recognition and our work. Hossain et al. [19] propose an active learning combinatorial deep model based on joint loss function optimization to update its network parameters. Their work inspires us to choose the combination of active learning and the deep learning model.

The semi-supervised method makes the learner independent of external interaction and automatically uses the potential information of non-semantic sensor data to improve the model performance [20]. However, the model can only give semantic information to the activity according to the most likely situation for those very fuzzy activities [21]. The core idea of active learning is to find the most valuable training samples through some heuristic strategies so that the model can achieve or even exceed the expected effect by labeling as few samples as possible through expert judgment [22]. Combining semi-supervised learning with the active learning method has received relatively little attention but is quite natural [21].

As discussed above, the combination of active learning,

Algorithm 1 Capture Activity Semantics through Location Information

Input: The batch of Labeled activity sequence A_l , unlabeled A_u , $n = \text{batch_size}$ and E is all kinds of activities.

Output: Labeled SbrADL dataset A

```

1: Initialize  $L, A'_l, A'_u$ 
2:  $\tilde{A}_l \leftarrow \text{AutoAugment}(A_l)$ 
3: repeat
4:    $\tilde{A}_u^k \leftarrow \text{AutoAugment}(A_u)$ 
5:   repeat
6:     if  $k = 1$  :  $at_i^* \leftarrow \arg \max_{at \in E} (P_\theta(at|ac_i))$ 
7:      $\tilde{at}_i^k \leftarrow \arg \max_{at \in E} (P_\theta(at|\tilde{ac}_i^k))$ 
8:   until  $i > n$ 
9:    $l := \text{Query}(\tilde{A}_u^k)$  and  $l \notin L$ 
10:   $L \leftarrow L \cup l, A'_u \leftarrow A'_u \cup \tilde{A}_u^k$ 
11: until  $k > 2$ 
12:  $A'_l \leftarrow \text{orde\_disruption}(A_l \cup \tilde{A}_l)$ 
13:  $A'_u \leftarrow \text{orde\_disruption}(A'_u)$ 
14:  $A \leftarrow \text{MixUp}(A'_l, A'_u)$ 
  
```

semi-supervised learning, and the deep model as a novel model can overcome annotation scarcity very well. The novel model uses location information as the input. See section 3 for a detailed description of our model.

3 THE ARCHITECTURE OF CASL

This section introduce CASL, which solves the problem of annotation scarcity. The dataset used by CASL is divided into semantic activity set A_l and non-semantic activity set A_u . CASL learns from A_l , consistency between all augmented sets of A_u , and expert advice from active learning modules. Therefore, the final loss function is shown in Formula (1), consisting of supervised loss, unsupervised loss, and active learning limitation loss.

$$L = L_l + \lambda_u L_u + \lambda_n L_n \quad (1)$$

where L_l is the supervised loss, and L_u is the unsupervised loss that we set up to represent consistency between augmented data of A_u . L_n represents the loss of the active learning module. $\lambda_u \in (0, 1)$ and $\lambda_n \in (0, 1)$ are both hyperparameters. λ_u represents the weight of unsupervised loss. λ_n represents the estimate of expert annotation quality and the weight of active learning part loss.

Algorithm 1 depicts the complete CASL process. We use ac for the specific activity and at for the one-hot encoding of the type of related activity. at^* and at^k represent the predictions of the activity type in A_u and the k^{th} augmented dataset of A_u . E represents all categories of activities under the current SbrADL dataset, such as $E = \{\text{Eat}, \text{Cook}, \text{Sleep}, \text{Callphone}\}$, and $at \in E$. Next, we will explain the components of our CASL and the specific form of each loss item.

3.1 Augmentation for heat map data

It is a challenge to avoid overfitting in training because the amount of SbrADL datasets is generally limited. It is

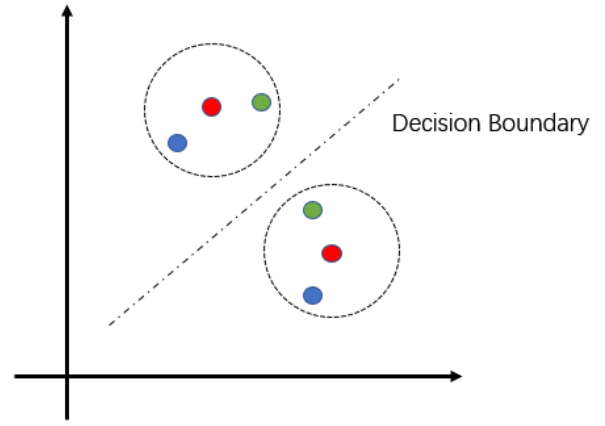


Fig. 2 Classification boundary of Consistent Regularization. The red dot represents the activity (ac) and its label (at), $at \in E$. Blue and green are two augmented versions of red. The dotted circle centered on the red sample is the expected tolerance range. The activities within this range should be considered the same category as the central data points.

necessary to augment the data to increase the robustness of the model. To this end, we generate the heat map of A_l and A_u based on the location information of activities. In particular, the augmentation of A_u is based on Consistent Regularization that requires the model to all augmented activities of the same non-semantic activity is as exact as possible, as shown in Figure 2. To achieve the above objectives, we need to force the classifier to make low entropy predictions for the semantic tag of activity. Therefore, we add an unsupervised loss term L_u into the loss function, as shown in Formula (2).

$$L_u = \frac{1}{|E||A_u|} \sum_{i \in |A_u|} |at_i^1 - at_i^2| \quad (2)$$

In calculating at_i^k , we use the $\text{Softmax}(\text{logits}(ac_i)/\tau)$ to replace the original probability formula $P(at_i|ac_i)$, including τ represent temperature. $\text{logits}(ac_i)$ is the output of the FC layer of the classifier model. Then we obtain the probability p of the category of ac_i by the Softmax function. The purpose of this calculation is to sharpen the distribution of prediction by adjusting τ . The lower temperature usually represents a sharper probability distribution, and entropy is low [23].

3.2 Query strategy

A_l cannot represent the complete information of the entire dataset. We need to select more representative samples with richer information from A_u . Therefore, we take the distance and density of the data distribution as the standard for querying. The queried set is denoted as L (algorithm 1, line7).

Distance: Considering that the initial activity set with semantic already contains rich information, the queried samples need to avoid overlapping with the labeled set $A_l \cup L$. The queried samples representing a less frequent activity should be far enough away from the labeled data in the data space. We use the Mahalanobis distance $d(\cdot)$ to

calculate the distance between two activities. For $ac_i \in A_u$ and $ac_i \notin L$, the minimum distance between it and all samples in the $A_l \cup L$ collection that currently contains labeled information is expressed as follows:

$$Dist(ac_i) = \min_{ac_j \in A_l \cup L} d(ac_i, ac_j) \quad (3)$$

Density: The distance cannot be used as the sole sampling index because we may query meaningless outliers. Therefore, we also need to limit our sampling method by the representativeness of the sample distribution of A_u , which can be expressed in terms of the sample distribution density. We measure the sample distribution density near the candidate sample by the following formula:

$$Density(ac_i) = \frac{1 - d(ac_i, ac_j)}{\sum_{ac_i \in \mathbb{G}_{A_u} L, ac_j \in A_l \cup L} d(ac_i, ac_j)} \quad (4)$$

From what has been discussed above, both distance and density are considered to obtain a more valuable sample. The query strategy is shown as Formula (5). α is a hyper-parameter used to balance the importance of distance and density, $\alpha \in (0, 1)$.

$$\begin{aligned} \arg \max (1 - \alpha) Dist(ac_i) + \alpha Density(ac_i) \\ s.t. ac^* \in \mathbb{G}_{A_u} L \end{aligned} \quad (5)$$

We record the heat map data as ac and expert opinion as at^* which are combined into the element term l_i in L , $i \in (0, n)$. The loss term of the active learning module is calculated as L_n using the average absolute error function, and the calculation method is as follows:

$$L_n = \sum_{ac_i \in A_l \cup L} |Active(ac_i) - f(ac_i)| \quad (6)$$

where $Active(ac_i)$ is the expert's opinion on the activity semantics. The reason for selecting L1 loss is that as the iterations go on, the results of the query strategy may be outliers of data anomalies, while the average absolute value error has a relatively limited penalty on outliers.

3.3 MixUp

We applied MixUp [24] on CASL. MixUp is independent of data and serves as a way of mixing data to augment the data further. It taught CASL a simple linear interpolation function that significantly reduced the complexity of unlabeled data spaces. MixUp obtains the ac' and at' of new virtual samples by mixing pixels of the heat map on behalf of different activities, and we represent at' with the one-hot encoding. The interpolation calculation of pixel fusion and label independent thermal coding is as follows:

$$ac' = \lambda ac_1 + (1 - \lambda) ac_2 \quad (7)$$

$$at' = \lambda at_1 + (1 - \lambda) at_2 \quad (8)$$

As with Mixup, the combination between two activities and labels, λ is sampled randomly from $Beta(\alpha, \alpha)$. In all the experiments described in section 4, α is set to 1 as a hyperparameter. The fusion of active ac fields we used is shown in Figure 3. The virtual samples formed by the new tuples $\langle ac, ac' \rangle$ form two new activity sets (algorithm 1, Line 12 to 15).

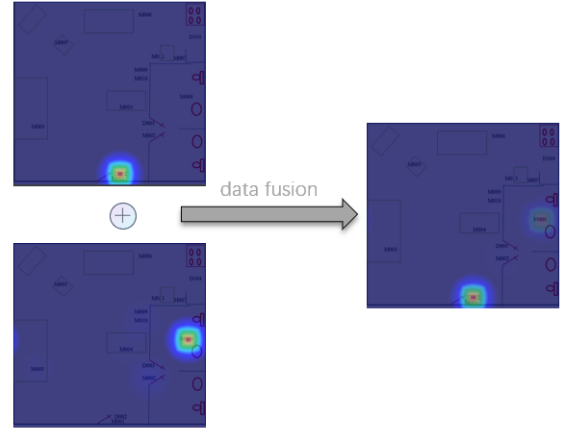


Fig. 3 Pixel fusion of the thermal map of the movement in and out of the toilet ($\lambda = 0.3$), that is, the ac of the new pixel worth the new activity can be obtained by calculating the corresponding pixel according to formula (7).

In addition, this research adopted different methods in the mixing process that mixed different semantic activities instead of randomly sampling. Because there is no profit in mixing samples of the same kind, please see Section 4.3 for details.

3.4 Training Signal Annealing

Even lightweight image recognition models are prone to over-fitting because the amount of SbrADL data is limited. To solve this problem, we introduced a training method: Training Signal Annealing (TSA) [23]. In a time of t , we set up a threshold η_t , $\frac{1}{|E|} \leq \eta_t \leq 1$ slowly release information of A_l during training. If the calculated $p_\theta(at^*|ac)$ of a label is greater than η_t , we remove this activity from the loss calculation. We represent the whole activity set batch with B , the supervision loss with L_l is shown as follows.

$$L_l = \frac{1}{Z} \sum_{ac_i, at_i^* \in B} [-I(p_\theta(at_i^*|ac_i) < \eta_t) \log(p_\theta(at_i^*|ac_i))] \quad (9)$$

where $Z = I(p_\theta(at^*|ac) < \eta_t)$ and $I(\cdot)$ is the indicator function. Suppose T is the total number of training steps, and t is the current number. For the case of the limited amount of data and easy overfitting, we change η_t in an exponential form, as shown in Formula ??.

$$\eta_t = \exp\left(\left(\frac{t}{T} - 1\right) \times 5\right) \times \left(1 - \frac{1}{K}\right) + \frac{1}{K} \quad (10)$$

4 EXPERIMENT

In this section, we begin with a brief description of our experimental setup and then report the results of the evaluation by comparing it with four baseline methods and MixMatch.

4.1 Implementation details

In this section, we introduce the two SbrADL datasets and describe the feature processing strategy and the selection of the image recognition model.

4.1.1 Dataset Description

We validated our method with two public data sets from the SbrADL data set published by the CASAS group. These dataset deployed several environmental sensors in the lab room to monitor movement and sensing water, bed, and door.

HH130 [9]. People in the experimental setting in the first dataset were asked to perform the following SbrADL: *Hygiene(ac1)*, *Leave(ac2)*, *Enter(ac3)*, *Relax(ac4)*, and *Phone(ac5)*. In addition, the *Eat* and *Sleep – out* only accounted for 0.0947% and 0.0330% of these datasets, which were not included in this experiment. The dataset contains 8047 pieces of data.

adlnormal [25]. The data in this dataset represents participants performing five SbrADL activities in the apartment. The five tasks are: *Eat(ac1)*, *Wash – hand(ac2)*, *Wash – dishes(ac3)*, *Call – phone(ac4)* and *Cook(ac5)*. The dataset contains 6425 pieces of data.

4.1.2 Processing of model input

When processing the data, we manually modified the sensitivity of the sensors to keep the sensitivity of all the sensors consistent. We then list the sensors and the number of times they produce messages in the current window. We visualized this list based on the sensor position on the experimental platform to get a heat map. The heat map can uniquely identify an activity. For the resulting heat map, the brightness and color of the pixels can represent the frequency at which the sensor starts, as shown in Figure 3. After this processing, each heat map generated can represent a activity data. We use the modified version of VGG16 [26] to recognize the heat map to predict the activity semantic label.

4.2 Baseline

We compare our algorithms with the following baselines from [12].

Random Forest(RF). RF is used to proceed with activity recognition is to represent the rules in the form of a tree to reason the activity label. In this experiment, RF contains 500 decision trees.

Naive Bayes(NB). This method is characterized by the combination of prior and posterior probability, which avoids the subjective bias of using prior probability only and the over-fitting phenomenon of using sample information alone. In this experiment, we use Gaussian mixture model (DPGMM) [27] to determine the number of Gaussian clusters for each training dataset adaptively.

Co-training [28]. This method is used to assign confidence to the semantics to which the activity belongs. Then select and delete the first few samples from the unlabeled dataset predicted to be positive or negative with high confidence. Each classifier is trained on a different randomly sampled subset of the training dataset. After each classifier is preliminarily trained, it can be used to calculate the confidence of semantic labels of unlabeled activity data. According to the calculated confidence, we query the samples with the highest confidence, and assign the voting results of many classifiers to the detected samples as semantic label.

We put these samples into the labeled samples and use them to train other classifiers.

GLSVM [29]. This method is based on the graph-based label propagation based on the SVM classifier. It uses all activities to construct a graph and the propagation of predicted label on the graph to infer the semantics of activities. The graph is created from the information contained in all the activity data.

4.3 Performance Analysis

In this part, we verified the effectiveness and robustness of our model through comparative experiments, as shown below.

The recognition rate for the overall dataset. The semi-supervised part of CASL is inspired by MixMatch. Therefore, we compared CASL with MixMatch at four feedback times per round (the number of samples obtained from the semantically tag-free activity) and four other methods of semantic acquisition as the baseline methods: Random Forest (RF), Naive Bayes (NB), Co-training [28], GLSVM [29]). To avoid the bias from experts, we adopted the method of 3-fold cross-validation, and Figures 4 and 5 show the result. For the SbrADL dataset, CASL has an obvious advantage over the four baseline methods in activity recognition rate, and the activity recognition rate of CASL is always better than that of MixMatch.

The recognition rate for a single activity. For the ADL recognition scenario, we can not show that the accuracy of one kind of activity identification is exceptionally high, and the accuracy of another kind is relatively low in the specific application. For example, when we applied the final semantic annotation model to the data set of the system containing the warning function of the abnormal state of the elderly at home, our poor recognition rate of the abnormal and dangerous condition of the old Fall-in-bathroom was not allowed. Therefore, we do not just care about the average accuracy of the model across the entire dataset. We also care about how well our model performs in terms of the individual accuracy of each class. Table 2 shows the results of the comparison experiment between CASL, MixMatch, and supervised learning under HH130 with 300 labeled active data. After observing the experimental results, we can draw the following conclusions:

1) For *Relax(ac4)*, other models performed poorly. The activity took place over a long time, and the experimental platform of the HH130 dataset was collecting information of other irrelevant infrared environmental sensors that were sometimes triggered by the activity room. Figures 6 and 7 show fuzzy samples of Relax. Even though CASL has a lower recognition rate, the results show that expert knowledge can also improve model performance on such activities with fuzzy labels.

2) However, it reflects some problems. The recognition rate of *Enter(ac3)* is only 0.12% higher than MixMatch after the expert. In addition, because the data quality of activities like Enter is high, but the improvement rate is still low, Enter is rarely mixed with other activities due to its high quality after outputting the actual category of each sample to be fed back.

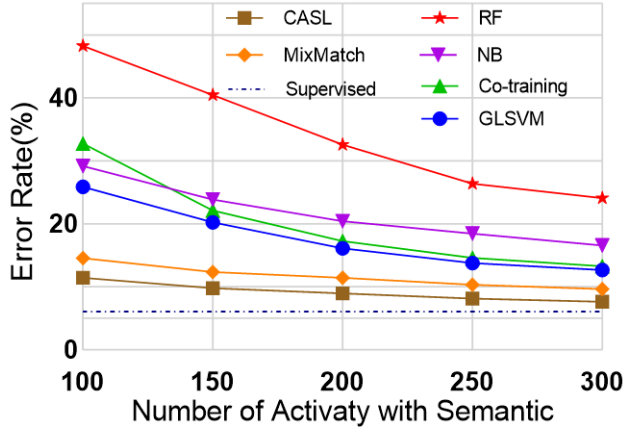


Fig. 4 Error rate line of CASL to MixMatch and baseline methods on HH130 for a varying number of labels.

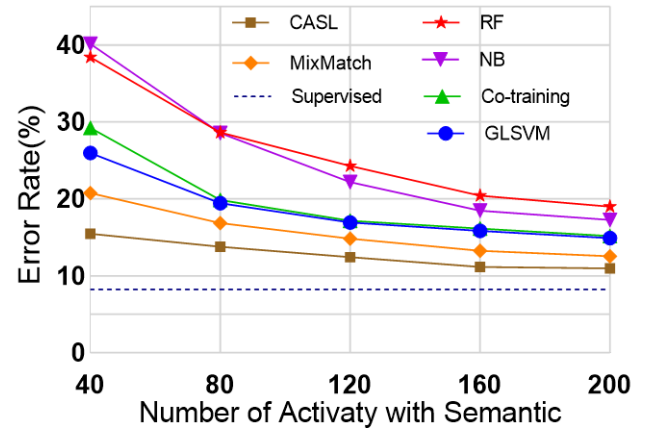


Fig. 5 Error rate line of CASL to MixMatch and baseline methods on adlnormal for a varying number of labels.

TABLE 1 HH130 recognition rate (with 300 labeled activity data)

Activity	Method		
	Supervised	MixMatch	CASL(with 4 feedbacks)
ac1	94.95%	90.15%	92.03%
ac2	95.89%	92.95%	93.78%
ac3	97.77%	95.01%	95.13%
ac4	84.42%	80.91%	84.57%
ac5	96.43%	91.81%	94.22%
Average	94.23%	90.37%	91.95%

4.4 Ablation Study

In order to verify the effect of CASL, we use query strategy, data fusion strategy and feature selection to perform ablation experiments in this section.

Query strategy: In this part, we compare the recognition rate of CASL in the HH130 dataset with five active learning query strategies (Random Sampling (RS), Margin Sampling (MS) [30], Sequence Vote Entropy (SVE) [31], Density Weighted Uncertainty Sampling (DWUS) [32], [33]). We set 300 initial activity data with semantic annotation. The results are shown in Table 2. Regardless of the number of feedback received each time, sampling based on distance and density trade-offs always performs best.

Data fusion strategy: We set up experiments for verifying the effect of MixUp on CASL. The Table 3 shows that CASL, which does not use MixUp, has a recognition error rate of 32.89% and 26.35%, respectively, under different feedback times (2 feedbacks and 4 feedbacks). The recognition error rate reached 16.43% when fusion strategy without restriction, respectively. The average accuracy was 13.73% when the fusion strategy was restricted. Compared with Cutout [34] and CutMix [35], the recognition rate of MixUp is superior to Cutout in the map after indoor activity recognition processing. After we processed the dataset into a heat map, the background of all the pictures was the same layout diagram of the experimental platform. In this case, the other two strategies are less effective than MixUP [24]. From the results, the fusion strategy we chose is effective.

Feature selection: We chose location information as the only input to CASL because it is more general than other information in SbrADL datasets. However, the premise of

this selection is to ensure that its effect is not inferior to other features. We select three feature extraction methods in [36] (LSTM [37], RKHS [38] and DDNN [36]) to compare with heat map methods of CASL, and the results are shown in Figures 6 and 7. Results show that CASL has better performance than other feature extraction methods on adlnormal dataset. However, the heat map method is less accurate than the DDNN on the HH130 dataset. Although the results show that CASL is effective, we need to explore and combine other common information in the SbrADL datasets to improve CASL's performance.

5 CONCLUSION

We develop a novel semantic annotation method, called CASL, to overcome the problem of expensive ADL dataset semantic annotation. This method combines active learning and mainstream semi-supervised methods. Moreover, we conceived a new sampling method to overcome the problem of sampling outliers on limited SbrADL datasets. We extensively studied the method's labor overhead, recognition rate and validity of each component. The results of the experiment found that this method is effective. We hope to apply this method to larger datasets in future work by exploring the new S2DT method for large datasets. In addition, the experiments done by CASL are all based on SbrADL datasets, and we are interested in applying this to other types of ADL datasets. Based on the results of the ablation experiment, we need to explore statistical features, temporal features and spatial correlation features to find the feature that can be combined with or replace location information as input of CASL.

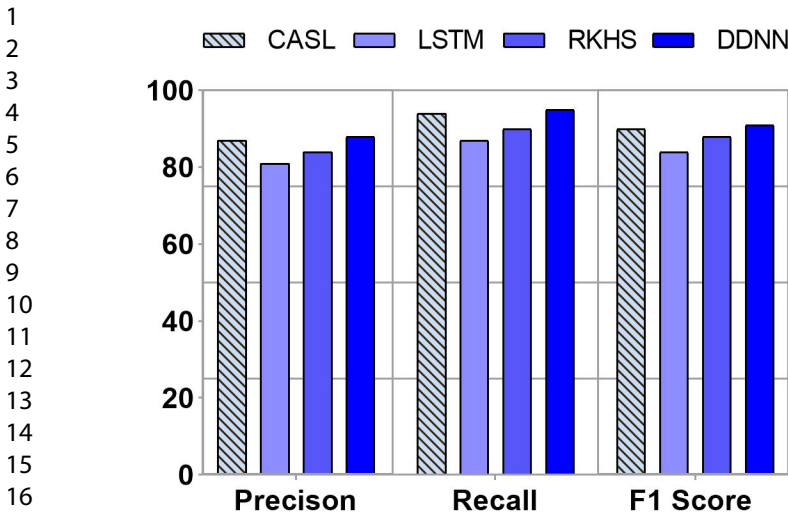


Fig. 6 Comparison of accuracy of different feature extraction methods with CASL on HH130.

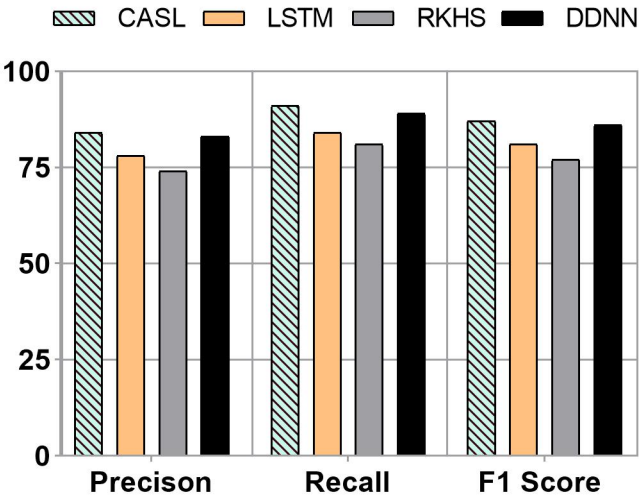


Fig. 7 Comparison of accuracy of different feature extraction methods with CASL on adlnormal.

TABLE 2 Results of ablation experiment for query strategy. All values are error rates on HH130 and CASL with varying feedbacks.

Method	S2DT	RS	MS	DWUS	SVE
2 feedbacks	85.23%	83.53%	84.12%	84.42%	85.43%
4 feedbacks	87.84%	84.52%	85.24%	87.65%	86.26%
8 feedbacks	86.84%	86.12%	86.26%	91.12%	88.63%

TABLE 3 Results of ablation experiment for data fusion strategy. All values are error rates on HH130 (with 200 labeled activity data) and CASL with varying feedbacks.

CASL	2 feedbacks	4 feedbacks
without MixUp	32.89%	26.35%
MixUp without limit	16.43%	13.73%
MixUp on different labels	13.16%	10.25%
with Cutout	17.89%	14.15%
with CutMix	16.85%	13.75%

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (61872038).

REFERENCES

[1] J. Wang, T. Zhu, J. Gan, H. Ning, and Y. Wan, "Sensor data augmentation with resampling for contrastive learning in human activity recognition," *arXiv preprint arXiv:2109.02054*, 2021.

[2] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

[3] A. Zaslavsky, D. Chakraborty et al., "Recognizing concurrent and interleaved activities in social interactions," in *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*. Sydney, NSW, Australia: IEEE, Dec 2011, pp. 230–237.

[4] J. Ye, G. Stevenson, and S. Dobson, "Kcar: A knowledge-driven approach for concurrent activity recognition," *Pervasive and Mobile Computing*, vol. 19, pp. 47–70, 2015.

[5] H. Ning, F. Shi, T. Zhu, Q. Li, and L. Chen, "A novel ontology consistent with acknowledged standards in smart homes," *Computer Networks*, vol. 148, pp. 101–107, 2019.

[6] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, "Cnn-based sensor fusion techniques for multi-modal human activity recognition," in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, New York, USA, Sep 2017, pp. 158–165.

[7] D. J. Cook and N. C. Krishnan, *Activity learning: discovering, recognizing, and predicting human behavior from sensor data*. John Wiley & Sons, 2015.

[8] I. Psychoula, D. Singh, L. Chen, F. Chen, A. Holzinger, and H. Ning, "Users' privacy concerns in iot based applications," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*. IEEE, 2018, pp. 1887–1894.

[9] D. J. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE intelligent systems*, vol. 2010, no. 99, p. 1, 2010.

[10] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *arXiv preprint arXiv:1610.02242*, 2016.

[11] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *arXiv preprint arXiv:1905.02249*, 2019.

[12] L. Yao, F. Nie, Q. Z. Sheng, T. Gu, X. Li, and S. Wang, "Learning from less for better: semi-supervised activity recognition via shared structure discovery," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, USA, Sep 2016, pp. 13–24.

[13] J. Wang, Y. Chen, Y. Gu, Y. Xiao, and H. Pan, "Sensorygans: An effective generative adversarial framework for sensor-based human activity recognition," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[14] X. Zhang, L. Yao, and F. Yuan, "Adversarial variational embedding

- for robust semi-supervised learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 139–147.
- [15] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 5, pp. 1747–1756, 2019.
- [16] L. Zhao, G. Sukthankar, and R. Sukthankar, "Robust active learning using crowdsourced annotations for activity recognition," in *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
- [17] W. S. Lasecki, Y. C. Song, H. Kautz, and J. P. Bigham, "Real-time crowd labeling for deployable activity recognition," in *Proceedings of the 2013 conference on Computer supported cooperative work*, 2013, pp. 1203–1212.
- [18] H. S. Hossain, M. A. Al Haiz Khan, and N. Roy, "Deactive: scaling activity recognition with active deep learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–23, 2018.
- [19] H. S. Hossain and N. Roy, "Active deep learning for activity recognition with context aware annotator selection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1862–1870.
- [20] D. Guan, W. Yuan, Y.-K. Lee, A. Gavrilov, and S. Lee, "Activity recognition based on semi-supervised learning," in *13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007)*. IEEE, 2007, pp. 469–475.
- [21] A. Calma, T. Reitmaier, and B. Sick, "Semi-supervised active learning for support vector machines: A novel approach that exploits structure information in data," *Information Sciences*, vol. 456, pp. 13–33, 2018.
- [22] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, X. Chen, and X. Wang, "A survey of deep active learning," *arXiv preprint arXiv:2009.00236*, 2020.
- [23] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong, and Q. V. Le, "Unsupervised data augmentation for consistency training," *arXiv preprint arXiv:1904.12848*, 2019.
- [24] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.
- [25] D. J. Cook and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment," *Methods of information in medicine*, vol. 48, no. 05, pp. 480–485, 2009.
- [26] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2017.
- [27] D. M. Blei and M. I. Jordan, "Variational inference for dirichlet process mixtures," *Bayesian analysis*, vol. 1, no. 1, pp. 121–143, 2006.
- [28] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*, Madison, Wisconsin, USA, Jul 1998, pp. 92–100.
- [29] M. Stikic, D. Larlus, and B. Schiele, "Multi-graph based semi-supervised learning for activity recognition," in *2009 international symposium on wearable computers*. IEEE, 2009, pp. 85–92.
- [30] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, 2009.
- [31] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Honolulu, Hawaii, USA, Oct 2008, pp. 1070–1079.
- [32] H. T. Nguyen and A. Smeulders, "Active learning using pre-clustering," in *Proceedings of the twenty-first international conference on Machine learning*, Banff, Alberta, Canada, Jul 2004, p. 79.
- [33] P. Donmez, J. G. Carbonell, and P. N. Bennett, "Dual strategy active learning," in *European Conference on Machine Learning*. Warsaw, Poland: Springer, Sep 2007, pp. 116–127.
- [34] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.
- [35] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul/Korea, Oct 2019, pp. 6023–6032.
- [36] H. Qian, S. J. Pan, B. Da, and C. Miao, "A novel distribution-embedded neural network for sensor-based activity recognition," in *IJCAI*, 2019, pp. 5614–5620.
- [37] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–28, 2017.
- [38] H. Qian, S. Pan, and C. Miao, "Sensor-based activity recognition via learning from distributions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.



Xiao Zhang received BE degree from Ludong University in 2015. Then, he is studying, with a master's degree, in School of Computer and Communication Engineering, University of Science and Technology Beijing. His research interests include Evolutionary Computation and Internet of Things.



Shan Cui received her B.S. degree from the School of Mathematics and Information Science, Weifang University in 2015 and her M.S. degree from the School of Mathematics and Statistics, Shandong Normal University, China. she is currently pursuing the Ph.D. degree from School of Computer and Communication Engineering, University of Science and Technology Beijing, China. Her current research interests include artificial intelligence algorithm interactions.



Tao Zhu received the Ph.D. degree from University of Science and Technology of China in 2015, and BE degree from Central South University in 2009. Then, he worked as a post-Ph.D. and a lecturer in School of Computer and Communication Engineering, University of Science and Technology Beijing. Currently, he is with University of South China. His research interests include Evolutionary Computation and Internet of Things.



Liming Chen is a professor in the School of Computer Science and Informatics at University of Ulster, Newtownabbey, United Kingdom. He received his B.Eng and M.Eng from Beijing Institute of Technology (BIT), Beijing, China, and his Ph.D in Artificial Intelligence from De Montfort University, UK. His research interests include data analysis, ubiquitous computing, and human-computer interaction. Liming is a Fellow of IET, a Senior Member of IEEE, a Member of the IEEE Computational Intelligence Society (IEEE CIS), a Member of the IEEE CIS Smart World Technical Committee (SWTC), and the Founding Chair of the IEEE CIS SWTC Task Force on User-centred Smart Systems (TFUCSS). He has served as an expert assessor, panel member and evaluator for UK EPSRC (Engineering and Physical Sciences Research Council, member of the Peer Review College), ESRC (Economic and Social Science Research Council), European Commission Horizon 2020 Research Program, Danish Agency for Science and Higher Education, Denmark, Canada Foundation for Innovation (CFI), Canada, Chilean National Science and Technology Commission (CONICYT), Chile, and NWO (The Netherlands Organisation for Scientific Research), Netherlands.



Fang Zhou received the B.Sc, M.Sc and Ph.D degree in computer science from the University of Science and Technology Beijing, China, in 1995, 2002 and 2012. From 2015 to 2016, she was a Visiting Researcher with the Department of Computer and Information Sciences, Temple University, USA. She is currently an Associate Professor with the Department of Computer Science and Technology, University of Science and Technology Beijing. Her research interests include machine learning, information retrieval and

information safety.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Huansheng Ning received his B.S. degree from Anhui University in 1996 and his Ph.D. degree from Beihang University in 2001. Now, he is a professor and vice dean of the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His current research focuses on the Internet of Things and general cyberspace. He is the founder and chair of the Cyberspace and Cybermatics International Science and Technology Cooperation Base. He has presided many research projects including Natural Science Foundation of China, National High Technology Research and Development Program of China (863 Project). He has published more than 150 journal/conference papers, and authored 5 books. He serves as an associate editor of IEEE Systems Journal (2013-Now), IEEE Internet of Things Journal (2014-2018), and as steering committee member of IEEE Internet of Things Journal (2016-Now).