Reviews of 2487 - *"Adversarial Multi-view Networks for Activity Recognition"*

Reviewer 1 (1AE)

1AE's recommendation to Editors after major revisions

Acceptable with minor (or no) changes

1AE's meta-review after major revisions

Overall, the reviewers felt that the revised submission addressed most of their concerns. I encourage the authors to go through the individual reviews for detailed feedback. The two main points are:
1. The authors were asked to use F1 measures apart from accuracy to report performance. They are advised to use mean and weighted F1 scores because of significant class imbalance
2. The limitation of not working reasonably well for non-repetitive tasks needs to be discussed in the paper.

I also want to point out that while the reviews seem positive, there was a long discussion amongst the reviewers on whether to accept or reject the current submission. The main concern was around the presentation of performance numbers. In the end, the reviewers agreed that we had given the authors enough feedback, and they should be able to add the correct performance numbers to the paper. Once these are added, the external reviewers do not need to take another look at the paper. The 1AE will take a close look at the final set of changes to make sure that the paper adequately addresses the raised concerns.

1AE's recommendation to Editors

Major revisions (enumerated in a subsequent field) are required for this to be publishable

Impact

Significant impact

1AE's meta-review

The submission deals with the important problem of activity recognition. The authors propose the use of an adversarial network and evaluate the performance on existing HAR datasets.

Overall, the reviewers were quite split in their assessment of the work. However, after few rounds of discussions, the reviewers agreed that there is lot to like in the work and it will be a good contribution to the IMWUT community. That said, it needs some major improvements and I have listed the needed revisions at the end of this review.

The reviewers found that the submission is well-written and presents the system in adequate detail. From the writing perspective, one thing missing though is a human-centric discussion. The current manuscript is more ML-heavy. I would encourage the authors to not tone down the ML component because that is going to be useful for the expert readers but add a more human-centric discussion of the implication of this work. The improvements made by the adversarial network must have enabled some new capabilities. How do those capabilities might help the end user? What are the future area of improvements? What kind of capabilities would a user not get until these improvements are made? What is a good application area (with HAR) for the current system? It would be great if the authors are able to make some of these recommendations.

Given the paper does not make a dataset contribution, it needs to make sure that the other contributions are strengthened a bit. An IMWUT paper does not necessarily need to contribute a dataset but the reviewers felt that showing improvement in performance on existing datasets is not enough. The reviewers argue that the authors should try to do a deployment (the authors can decide the right size of this deployment) to see what kind of performance does the new algorithm provide on this new data.

Other than these concerns, there were also some concerns around the evaluation metrics, and R4 made some citation suggestions. The reviewers also felt that that the paper needs to make it clear why adversarial training is necessary here. A quantitative proof of the utility of adversarial network would help.

Overall, this is very strong work and if the authors are able to address some of the concerns that the reviewers have, it should be an easy accept in the next round. The authors can go through individual reviews for detailed comments.

1AE's meta-review: Major / minor revisions

1. Add human-centric implication/discussion section
2. Clarify the key contribution
3. Add a separate, new study to evaluate performance beyond the existing datasets
4. Make it clear that the model does not handle non-repetitive activities well right now

5. Show that adversarial network is needed

The length of the paper is fine. If, in order to make the recommended changes, the authors end up adding 2-3 pages, that should be fine.

Reviewer 2 (2AE)

Reviewer's recommendation to 1AE after major revisions

Acceptable with minor (or no) changes

Comments after major revisions

The authors have significantly improved their manuscript and they have addressed my main concerns through the changes in section 5 and the conclusion section

Contribution to IMWUT

This paper proposes a discriminative adversarial multi-view network for sensor-based human activity recognition. The authors present experimental results of evaluating this model on three benchmark datasets (Mhealth, pamap2 and ucidsads)

Impact

Medium impact

Review

This paper proposes an adversarial multi-view network to recognize activities from sensor data.

The authors empirically evaluate their proposed approach on 3 benchmark datasets.

Unfortunately, the paper right now reads like a standard ML paper. The only contribution if the adversarial multi-view network model. It does not contribute with any new dataset. It does not have any implications for the design of mobile/ubiquitous systems. It also lacks a human-centered discussion.

I recommend the authors to either submit it to an ML conference or to add a more substantial human-centric focus to the paper. Alternatively, they could add a new experimental section where they deploy their models in-the-wild and evaluate their performance from a human-centric perspective

Recommendation(s) to 1AE

Reject

Major / minor revisions (recommendation to 1AE)

(blank)

Confidence

Very confident - I am knowledgeable in the area

Reviewer 3 (Reviewer)

Reviewer's recommendation to 1AE after major revisions

Acceptable with minor (or no) changes

Comments after major revisions

I think the authors have addressed most of my concerns. Yet there are still some key issues, which can be misleading (see below). The workload is no-trivial (should be another major revision), and I hope the authors can clarify the following clearly to avoid any misleading information.

1. I suggested the authors to reported appropriate metrics, (e.g., mean f1, can be together with weighted F1), yet the authors only used confusion matrix as a complementary metric (at FIg. 6) to answer this question. However, the authors still used the 'inappropriate metric' i.e., the weighted F1 and weighted accuracy (i.e., F1/accuracy without considering the data imbalance) as default. The consequence is obvious. In Opportunity dataset (Sec. 5.10.1), the null class takes more than 75%, and the authors still used overall accuracy (which is biased towards the majority null class. If random guess to Null class, the accuracy would give >75% accuracy for a 18-class classification problem) to evaluate an algorithm good or not, and the corresponding conclusion could be very misleading. I suggest the authors report mean-F1 together with weighted-F1, after discussion and used the appropriate metric to get the conclusion (on Fig. 8).
In a different word, this is not a ML paper but HAR paper so more insight (based on appropriate metric may be more important than pure accuracy!)

2. In Sec. 5.10.1, Bi-LSTM was used. Bi-LSTM doesn't take the current cell/hidden state into the next window, so it doesn't take the full advantage of LSTM in memorising the long-term information for non-repetitive activity recognition. The authors should tone down the conclusion they get, or re-investigate this problem (based on appropriate metric, or appropriate LSTM network structure), or having open discussion and list future investigation plans (if they are deciding not to go down this rabbit hole).

3. Sec. 5.8, the authors gave the confusion matrix, which can give class-wise results.

They used class 2 and class 3 which is not intuitive, and the readers need to go back to fig.4 (to understand what activities are they). I suggested the authors gave the activity ID in the fig. 6 description.

Contribution to IMWUT

This is a very well-written paper, and the authors proposed a multi-view learning approach to address subject-independent human activity recognition (task). Several feature extractors were proposed and aggregate together, and adversarial Siamese network were used to reduce the effect of human's personal noise. This is highly relevant to IMWUT

Impact

Medium impact

Review

Comments:
Methodology
Based on different CNN-based network structures, the authors proposed multiple feature extractors that can learn features from different views. An adversarial Siamese network was proposed to reduce the personalised signal to further improve the performance. Three public datasets were tested based on a designed subject-independent test framework, and related results were reported.

Presentation:
The paper is very well-written. Literature and motivation were clearly given, followed by the proposed methods, the experiments including the ablation study were well presented.

Experiments/Discussion.
The authors compared the proposed method with several state-of-the-art. Yet, these three dataset includes long repetitive activities, and CNN is very good at. It was pointed out by [16] that LSTM is better at recognising non-repetitive complex behaviours (e.g., Opportunity dataset), and it is important the authors should have a discussion the limitation of this study, if they are not planning to study on the popular (and harder) Opportunity datasets.
Another issue is the evaluation metric, the authors used the weighted acc/f1. However, this metric ignores the minority class, which should be clarified as HAR may always suffer from data imbalanced problem (where weighted acc/f1 is not appropriate.).

Details:
1. The authors used first name in the literature, instead of last name. E.g., 2.2.1

Liangying et al. should be Peng et al., Nils et al. should be Hammerla et al. Yu et al. should be Guan and Ploetz.
2. Also report the class distribution of the 3 datasets, and clarify the limitation of the weighted acc/f1 when facing imbalanced dataset (e.g., Opportunity dataset).
3. Siamese network was also used to extract features that is robust to various nuisance factors. How does the proposed Siamese Adversarial framework compared with the baseline Siamese? (i.e., is adversarial training necessary?)

Recommendation(s) to 1AE

Major revisions (enumerated in a subsequent field) are required for this to be publishable

Major / minor revisions (recommendation to 1AE)

See comments.

Confidence

Very confident - I am knowledgeable in the area

Reviewer 4 (Reviewer)

Reviewer's recommendation to 1AE after major revisions

Acceptable with minor (or no) changes

Comments after major revisions

The authors have improved their submission and have addressed many of my concerns.

However, there are certain aspects of the paper that need further clarification. While I understand the motivation behind the evaluation on the Opportunity dataset to showcase the system efficacy on the sporadic/non repetitive motions, I find the evaluation of the same lacking. Given the dataset has only 3 participants and class set imbalance, I suggest that the authors remove this study from their paper and rather make it focus on non-repetitive activities.

The findings on the Opportunity dataset can be a part of preliminary explorations with the limitations section reflecting the same.

Contribution to IMWUT

The papers presents a novel neural network architecture, a Discriminative Adversarial MUlti-view Network (DAMUN), for human activity recognition using multimodal sensory streams. The paper presents a neat technical contribution to IMWUT and presents a better evaluation protocol for future studies on activity recognition using wearables.

Impact

Medium impact

Review

The current submission presents a strong contribution to the IMWUT community. The submission is well-written, well-structured, and easy to read. There are various things I like about it: it has a rigorous evaluation, a well motivated machine learning approach and an ablation study showing the merits of each part. While I recommend the paper for acceptance with revisions, there are a few things I would like to see addressed in the final version:

1. Related Work: Optical sensor based related work is missing from the paper. Many of these work in the wild and hence better situate the needs presented by the authors to create a model that is user independent. A few noteworthy findings from the computer vision community are:

* Baradel, Fabien, et al. "Glimpse clouds: Human activity recognition from unstructured feature points." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
* http://activity-net.org/
* Ke, Shian-Ru, et al. "A review on video-based human activity recognition." computers 2.2 (2013): 88-131.

Furthermore the UbiComp community has also utilized cameras for human activity recognition and also sub-activity classification such as exercise recognition, classroom activity monitoring, etc.

* Radu, Valentin, and Maximilian Henne. "Vision2sensor: Knowledge transfer across sensing modalities for human activity recognition." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3.3 (2019): 84.
* Khurana, Rushil, et al. "GymCam: Detecting, Recognizing and Tracking Simultaneous Exercises in Unconstrained Scenes." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2.4 (2018): 185.
* Ahuja, Karan, et al. "EduSense: Practical classroom sensing at Scale." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3.3 (2019): 71.

2. Given the small number of subjects in each dataset (10 subjects at max), there are

only 9 subjects present at anytime for training as the authors are using leave one out cross validation. DL methods could be prone to overfitting with such a limited number of subjects. The paper would benefit with a discussion on the same. Also, given the overlap in some sensory streams across datasets, is cross database testing possible?

3. It will be interesting to see the benefit of each wearable sensory stream and its contribution to the accuracy. Results broken down by each sensor can also help inform future data collection setups for activity recognition for both selection as well as placement of sensors.

4. The current paper reads like a ML paper and lacks a human centric discussion. While I do not expect the authors to deploy their system in the wild and evaluate it, a discussion on the same would be valuable for future researchers. Some pointers along these lines could be the system setup, latency of the system, limitations in real-world settings, etc.

5. Minor fixes:
- Hadamard fusion on page 2 needs citation.
- “useful e multi-mosdre twos” - typo page 3
- “subjectnumber” - page 10
- Fig 4 - page 14 - make the fonts bigger

Recommendation(s) to 1AE

Acceptable with minor (or no) changes

Major / minor revisions (recommendation to 1AE)

(blank)

Confidence

Very confident - I am knowledgeable in the area

Return to submission and reviews