Study on Temperature (τ) Variation for SimCLR based Activity Recognition

Pranjal Kumar*

Abstract Human Activity Recognition (HAR) is a process to automatically detect human activities based on stream data generated from various sensors, including inertial sensors, physiological sensors, location sensors, cameras, time, and many others. Unsupervised contrastive learning has been excellent, while the contrastive loss mechanism is less studied. In this paper, we provide a temperature (τ) variance study affecting the loss of SimCLR model and ultimately full HAR evaluation results. We focus on understanding the implications of unsupervised contrastive loss in context of HAR data. In this work, also regulation of the temperature(τ) coefficient is incorporated for improving the HAR feature qualities and overall performance for downstream tasks in healthcare setting. Performance boost of 1.3% is observed in experimentation.

Keywords Contrastive learning · activity recognition · healthcare

1 Introduction

The purpose of human activity recognition (HAR), which consists of observations and analysis of human behaviour and its environment, is to determine the current behaviour and goals of the human body. HAR research has gained attention by its advantages in smart surveillance

Pranjal Kumar
NIT Hamirpur, H.P, India-177005
Tel.: +918637511985
Fax: +91-1972-223834
E-mail: pranjal@nith.ac.in

systems, healthcare systems, connections between virtual reality, smart homes, aberrant behaviour detection and other areas and its capacity to support and connect with unique disciplines. One of the most widely-discussed research areas is HAR[1] among academics from both academia and industry whose aim is the progress of all-round computing and human computer interaction. The advances in deep learning have made the field a key component of the most smart systems. In the majority of computer vision tasks such as image classification, object detection, image separation and activity recognition, and natural languages processing (NLP). Due to the intensive work required by manually notifying millions of data samples, supervised strategy to learn features from labelled data has nearly been saturated. Though a plethora of information is available, researchers have been urged to find alternative ways of making use of it by lack of annotations.

Unsupervised learning makes it possible for us to learn feature representations without the supervision of human beings.Contrastive learning has reached a state of the arts in a variety of tasks, which was recently proposed as an unsupervised study[2–7]. The main difference from other techniques is that the data transformation and contrastive loss strategy used. In short, most contrastive learning methods construct first a series of augmented data to build positive and negative pairs on an instance level. Similiarity between positive pairs could then be maximised by different contrast losses, such as Triplet[8], NCE[9] and NT-Xent[3], while negative pairs could be minimised. Uniformity helps contrastive learning to learn distinguishable characteristics, but over-pursuit of uniformity makes the contrastive loss unable to tolerate closely correlated samples which break down the underlying structure and damage downstream feature attributes[10, 11]. Re-

cently, SimCLR was incorporated for healthcare and HAR in particular for the first time[12]. Motivated from the work done by Tang et al., in this paper, we lay out a module that would deemed to be beneficial for HAR systems and other healthcare-related applications. Main contribution of this work is summarized below:

– We provide a study for understanding the behaviour of contrastive learning (emphasising on temperature coefficient, τ) in sensor data context for human activity recognition.
– We optimise the SimCLR module by regulating the temperature coefficient in order to enhance the quality of features for downstream tasks.
– Improved performance for overall model[12].

2 Related Work

Many studies have examined the identification of human activities from diverse points of view. These include: by specialised approach [13]; by algorithm type [14], by sensor type[1, 15, 16]; by fuse type [17] or by device type [18], although other analyses have been carried out more generally by the HAR categories[19, 20]. HAR accomplished five primary tasks, namely the recognition of the fundamental activities[21], the recognition of everyday activities[22], uncommon events[23], biometric subjects[24], and energy expenditure predictions[25]. Different sensors such as video cameras, ambient temperature sensors, relative humidity, light, pressure and wearable sensors are used. The major forms of wearable sensors are generally integrated smartphone sensors or sensors incorporated into wearable devices. Dong and Biawas [26] introduced a wearable sensor network designed to monitor human activity. In a similar study, Curone et al. have used wearable triaxial accelerometers to monitor activity[27].

Progress in deep learning has made the field a central part of the smart systems. The ability to learn rich patterns from today's vast amount of data makes the use of deep neural networks (DNNs) an important approach in HAR. The amount of annotated training data available is very reliant on traditional supervised learning approaches. Self-supervised learning methods have recently integrated both generative[28] and contrastive[3] approaches that have been able to use unlabelled data to understand the underlying representations. In recent studies[29, 7, 30–34] on unsupervised feature representation for images, concept known as contrastive learning was incorporated[5]. Contrastive learning (CL) is a discriminatory approach that aims to group similar samples closer and far away. The results after application of contrastive learning are astounding: for example, Sim-

CLR[3] reduce the gap between unsupervised and supervised pre-training representations in linear classification performance.

3 Contrastive Learning

The purpose of contrastive methodology is to comprehend a function that maps the input data features to the features on a hypersphere dimension. Wang et al. depicted that the contrastive loss for a given unlabelled training sample set $X = \{x_1, \ldots, x_N\}$ is given as follows[11]:

$$L(x_i) = -\log \frac{\exp(s_{i,i}/\tau)}{\sum_{k=i} \exp(s_{i,k}/\tau) + \exp(s_{i,i}/\tau)} \quad (1)$$

Where $s_{i,j} = f(x_i)^T g(x_j) \cdot f(\cdot)$ is an extractor that maps pixel space corresponding images onto space in a hypersphere. $g(\cdot)$ could serve the same purpose as of f [3]. τ is a temperature hyper-parameter that helps in distinguishing positive and negative samples. The contrastive loss attempts to attract positive key samples and separate the negative key samples. This goal can also be achieved with a simpler contrastive loss function as shown below[11]:

$$L_{simple}(x_i) = -s_{i,i} + \lambda \sum_{i=j} s_{i,j} \quad (2)$$

The goal of contrastive learning is to learn augmented data alignment and discriminatory embedding. The contrastive loss does not restrict negative sample distribution. The temperature contributes to the control of penalty strength on hard negative samples. Specifically, small contrastive losses tend to penalise much more the most severe negative samples in the form of a more separate local structure in a sample and a more uniform embedding distribution[11].

4 Methodology

SimCLR[3] architecture consists of these primary modules.

– A data incrementation module that randomly transforms a given example of data leading to two correlated views on the same example.
– A network base neural encoder that extracts vectors from enhanced data examples.
– A neural network projection head maps the space where the contrast loss is applied.
– A loss function set to a contrastive prediction task.

Table 1 Comparison with baseline models

Model	Supervised(only)	Self-Supervised	SimCLR(optimised)
Weighted F1	0.922	0.923	**0.955**

$\tau=0.07$

$\tau=0.1$

$\tau=0.2$

$\tau=1$
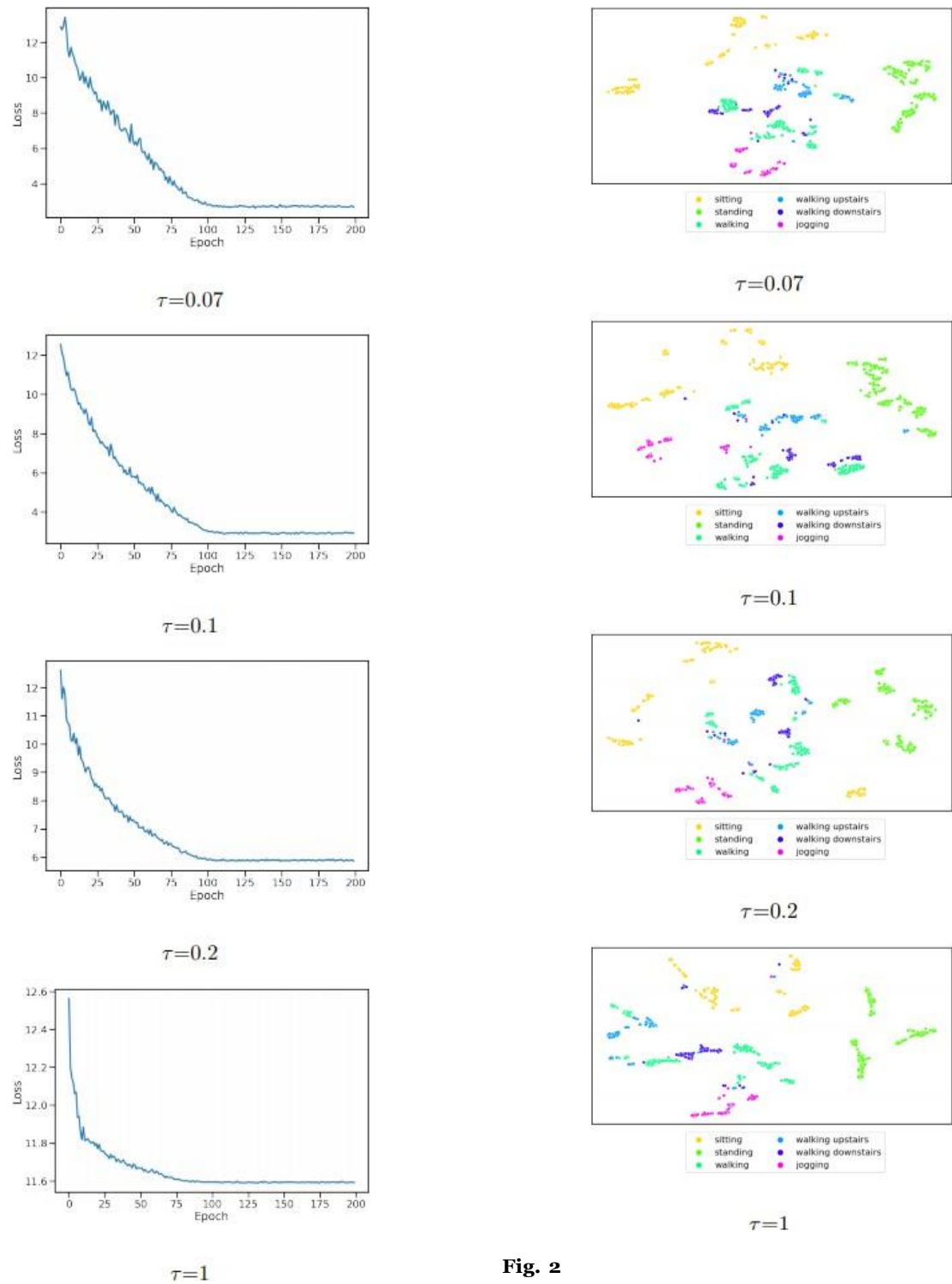
Fig. 1

Loss Variation

$\tau=0.07$

$\tau=0.1$

$\tau=0.2$

$\tau=1$

Fig. 2

T-SNE[35] Plot

Table 2 Quantitative comparison on the MotionSense Validation Dataset.

Temperature(τ)	F1 Macro	F1 Micro	F1 Weighted
0.07	0.858	0.870	0.874
0.1	0.900	0.921	0.922
0.2	**0.935**	**0.954**	**0.955**
0.3	**0.931**	**0.951**	**0.951**
0.7	0.894	0.914	0.916
1	0.879	0.905	0.907

5 Results & Discussion

5.1 Dataset

MotionSense[36] was used in our assessment as a publicly available dataset. This dataset comprises data from 24 individuals who carried an iPhone 6s in the front pocket of their pants and perform 6 different activities: walking downstairs, upstairs, walking, jogging, sitting and standing. In this study 6630 windows, each 400 timestamping and 50 percent overlap, were used for data from a 50% tri-axial accelerometer.

5.2 Experimental Setup

Linear and fine tune evaluation was administered on NVIDIA TESLA V100 SXM2. During pre-entraining for 200 epochs and batch size 512 the SGD optimizer with a cosine decay of learning rate is used. TPN[37] is incorporated as a base encoder for the HAR systems to suit the needs of a comparatively lightweight neural network architecture. The projection head was utilised as a three-layer, fully connected MLP with a loss function of NT-Xent. The base decoder is composed of three temporal (1D) layers, each with 24, 16, 8 and 32, 64 and 96 kernel sizes. During preparation the projection head is composed of 3 fully connected layers with 256, 128 and 50 units, and the grading head is composed of two fully connected layers of 1024 and 6 units in the fine-tunated evaluation. A 0.1 drop-off rate is used to activate the ReLU function. At the end there is an additional global maximum pooling layer. The model is trained at the SGD optimizer for linear assessment for 50 epochs and a learning rate of 0,03. The model is perfectly tuned with Adam optimizer and a study rate of 0.001 for 50 epochs for a finely tuned assessment.

5.3 Quantitative Results

In this section we conduct extensive experimentation on the temperature coefficient, in order to understand the modeling relationship of the proposed network using activity prediction precision as the assessment metric. The effect of the temperature is assessed. In the first place, we try to determine whether the temperature precisely checks the severity of the penalties in severe negative samples. Numerical results are tabulated in Table 2.

- When the temperature is 0.2 or 0.3, the model achieve the best results. Small or large temperature model achieve inadequate performance.
- The current model shows a 1.3% increase in performance than the previous[12].

5.4 Qualitative Results

If the loss value is extremely minimal, the contrastive loss function will inflict substantial penalties on closest neighbours. Semantically similar instances of data will very likely be distributed with the anchor point. Considering the depictions in T-SNE plots in Figure 2, we follow that embedding with $\tau = 0.07$ is distributed better and evenly, although the embedding with $\tau = 1$ is more reasonable and locally clustered and globally separated.

- With the τ decreasing, there is a larger gap from positive samples to other misleading negatives, i.e. more distinguishable positive and negative samples.
- Indeed, as shown in Figure 2, small temperatures tend to increase the impact of the hard negative samples.
- Results demonstrate that the positive samples are more aligned with the increased temperature and that the model tends to develop more invariant features with regard to the different transformations applied to sensor data.

6 Comparative Study with Baseline Models

In this section, we compare our best model with the most advanced methods. A linear and finally defined evaluation was conducted using the MotionSense dataset to evaluate the impact of using different temperature(τ) variances for SimCLR pre-training. Results are shown in Table 1. F1 scores are taken directly from work already carried out by Tang et al. for supervised and self-supervised models.

7 Conclusion

In this work, we have studied one of the most important tasks in digital health applications i.e., HAR and

the SimCLR contrastive learning framework from visual representation learning. We have examined the effect of temperature(τ) changes on contrastive loss in connection with sensor data to improve the feature quality and performance for downstream tasks.

References

1. Salwa Slim, Ayman Atia, M Elfattah, and Mostafa-Sami M Mostafa. Survey on human activity recognition based on acceleration data. *Intl. J. Adv. Comput. Sci. Appl*, 10:84–98, 2019.
2. Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.
3. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.
4. Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.
5. Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006.
6. Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *2021 AAAI Conference on Artificial Intelligence (AAAI)*, 2021.
7. Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
8. Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
9. Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 297–304. JMLR Workshop and Conference Proceedings, 2010.
10. Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning, 2021.
11. Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss, 2021.
12. Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, and Cecilia Mascolo. Exploring contrastive learning in human activity recognition for healthcare, 2021.
13. Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.
14. Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1254, 2018.
15. Umran Alrazzak and Bassem Alhalabi. A survey on human activity recognition using accelerometer sensor. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 152–159. IEEE, 2019.
16. Xinyu Li, Yuan He, and Xiaojun Jing. A survey of deep learning-based human activity recognition in radar. *Remote Sensing*, 11(9):1068, 2019.
17. Antonio A Aguileta, Ramon F Brena, Oscar Mayora, Erik Molino-Minero-Re, and Luis A Trejo. Multi-sensor fusion for activity recognition—a survey. *Sensors*, 19(17):3808, 2019.
18. Wesllen Sousa Lima, Eduardo Souto, Khalil El-Khatib, Roozbeh Jalali, and Joao Gama. Human activity recognition using inertial sensors in a smartphone: An overview. *Sensors*, 19(14):3213, 2019.
19. Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. Human activity recognition: A survey. *Procedia Computer Science*, 155:698–703, 2019.
20. Zawar Hussain, Quan Z. Sheng, and Wei Emma Zhang. A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications*, 167:102738, Oct 2020.
21. Bardia Yousefi and Chu Kiong Loo. Biologically-inspired computational neural mechanism for human action/activity recognition: A review. *Electronics*, 8(10):1169, 2019.
22. Sakorn Mekruksavanich and Anuchit Jitpattanakul. Exercise activity recognition with surface electromyography sensor using machine learning approach. In *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, pages 75–78. IEEE, 2020.
23. Rajesh Kumar Tripathi, Anand Singh Jalal, and Subhash Chand Agrawal. Suspicious human activity recognition: a review. *Artificial Intelligence Review*, 50(2):283–339, 2018.
24. Robertas Damaševičius, Rytis Maskeliūnas, Algimantas Venčkauskas, and Marcin Woźniak. Smartphone user identity verification using gait characteristics. *Symmetry*, 8(10):100, 2016.
25. Tifenn Rault, Abdelmadjid Bouabdallah, Yacine Challal, and Frédéric Marin. A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications. *Pervasive and Mobile Computing*, 37:23–44, 2017.
26. Bo Dong and Subir Biswas. Wearable networked sensing for human mobility and activity analytics: A systems study. In *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)*, pages 1–6. IEEE, 2012.
27. Davide Curone, Gian Mario Bertolotti, Andrea Cristiani, Emanuele Lindo Secco, and Giovanni Magenes. A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity. *IEEE transactions on information technology in biomedicine*, 14(4):1098–1105, 2010.
28. Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. arxiv 2014. *arXiv preprint arXiv:1406.2661*, 2020.

29. Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018.

30. R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670*, 2018.

31. Mang Ye, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. Unsupervised embedding learning via invariant and spreading instance feature. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6210–6219, 2019.

32. Philip Bachman, R Devon Hjelm, and William Buchwalter. Learning representations by maximizing mutual information across views. *arXiv preprint arXiv:1906.00910*, 2019.

33. Olivier Henaff. Data-efficient image recognition with contrastive predictive coding. In *International Conference on Machine Learning*, pages 4182–4192. PMLR, 2020.

34. Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.

35. Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

36. Mohammad Malekzadeh, Richard G Clegg, Andrea Cavallaro, and Hamed Haddadi. Protecting sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, pages 1–6, 2018.

37. Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2):1–30, 2019.