On the Utility of Virtual On-body Acceleration Data for Fine-grained Human Activity Recognition

Zikang Leng
zleng7@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Yash Jain
yashjain@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Hyeokhyen Kwon
hyeokhyen.kwon@emory.edu
Emory University
Atlanta, Georgia, USA

Thomas Plöetz
thomas.ploetz@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Previous work has demonstrated that virtual accelerometry data, extracted from videos using cross-modality transfer approaches like IMUTube, is beneficial for training complex and effective human activity recognition (HAR) models. Systems like IMUTube were originally designed to cover activities that are based on substantial body (part) movements. Yet, life is complex, and a range of activities of daily living is based on only rather subtle movements, which bears the question to what extent systems like IMUTube are of value also for fine-grained HAR? In this work we first introduce a measure to quantitatively assess the subtlety of human movements that are underlying activities of interest–the motion subtlety index (MSI)–which captures local pixel movements and pose changes in the vicinity of target virtual sensor locations, and correlate it to the eventual activity recognition accuracy. We explore for which activities with underlying subtle movements a cross-modality transfer approach works, and for which not. As such, the work presented in this paper allows us to map out the landscape for IMUTube-like system applications in practical scenarios.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → **Computer vision**.

KEYWORDS

Human activity recognition; Virtual IMU Data; Eating

ACM Reference Format:

Zikang Leng, Yash Jain, Hyeokhyen Kwon, and Thomas Plöetz. 2023. On the Utility of Virtual On-body Acceleration Data for Fine-grained Human Activity Recognition. In *Proceedings of the 2023 International Symposium on Wearable Computers (ISWC '23), October 08–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3594738.3611364

1 INTRODUCTION

The effectiveness of supervised learning methods for deriving human activity recognition (HAR) systems for wearables depends heavily on the availability of curated, i.e., annotated datasets [6]. One major issue with current machine learning solutions in the field is the paucity of labeled datasets. Annotating sensor data in HAR is expensive, often privacy invasive, and often prone to errors or has other practical limitations [7, 12, 16].

Recently, systems like IMUTube [18] have been introduced to tackle the paucity of labeled datasets in developing human activity recognition systems by generating virtual inertial measurement unit (IMU) data from videos, which were used to train the model. With the previous success of systems like IMUTube for recognizing human activities with underlying coarse motions, the next step now is to explore to what extent such approaches generalize to activities with more subtle body movements.

We define a novel metric–the *Motion Subtlety Index* (MSI)–that measures the subtlety of motion of human activities performed in a video sequence by using optical flow and pose estimation methods. With this new measure, we are now in the position to systematically assess methods like IMUTube with regards to their effectiveness on fine-grained activities. We study its effectiveness for virtual wrist sensors focusing on activities with more subtle motions than the coarse motions studied previously, including washing hands, playing instruments, or driving, which are essential activities when it comes to assessing an individual's quality of life [22].

Our experimental evaluation on a range of activities of daily living shows that the MSI extracted from human activity videos is highly correlated to the eventual recognition accuracy of HAR systems that were derived using virtual IMU data extracted from videos (Pearson $r = -0.85$, $p \leq 0.001$). As such, the MSI is an excellent proxy that can be used for the *a-priori* prediction of the potential effectiveness of cross-modal transfer approaches, and we can gain a deeper understanding on when cross-modality transfer approaches like IMUTube are likely to succeed – and when they are likely to fail. The contributions of this paper are two-fold:

(1) We propose a new metric, MSI, to show that the subtlety of movements involved in activities is measurable through videos.
(2) Through the newly introduced quantification of the subtlety of human movements and its correlation to the eventual effectiveness of HAR systems that were derived based on virtual IMU data, we can draw conclusions about application cases for systems like IMUTube.

ISWC '23, October 08–12, 2023, Cancun, Quintana Roo, Mexico

ACM ISBN 979-8-4007-0199-3/23/10.
https://doi.org/10.1145/3594738.3611364

2 BACKGROUND

Virtual IMU Data from Videos: IMUTube was recently introduced for automatically converting large-scale video datasets into virtual IMU data that can be used for training sensor-based HAR systems [18]. Given a video, the IMUTube system automatically selects relevant activity clips, where 3D motion information is estimated that is converted to virtual IMU data for model training. IMUTube was mainly designed, deployed, and validated for human activities with underlying coarse motions, such as locomotion or gym exercises [17, 19]. Using the virtual IMU data collected from large-scale videos, previous studies were able to train a very complex deep learning model with parameters three orders of magnitude larger than state-of-the-art deep learning models [25]. Similar to IMUTube, Rey *et al.* [27] used deep neural networks to estimate IMU data from 3d poses. Xia *et al.* [31] introduced a spring-joint model to improve the quality of the virtual IMU data. Lastly, Santhalingam *et al.* [28] generated virtual IMU data from existing American Sign Language video datasets and applied it to sign language recognition.

Sensor-Based Recognition of Daily Activities: Recent work proposed wearable-based human activity recognition for fine-grained daily activities to capture contexts for how users are situated [20, 22, 23]. For example, some work has focused on designing eating detection systems using wearables, as daily eating behaviors were found to be strongly associated with risks in physical and mental health [2–4, 14, 24, 29, 30]. Many daily activities, including eating, involve very subtle motions. While previous methods for recognizing daily activities were shown to be effective, they are still limited by the amount of lab data available for training. This work proposes a method for predicting the effectiveness of IMUTube, focusing on daily activities involving subtle movements, for understanding the complexity of small data problems.

3 QUANTIFYING MOTION SUBTLETY

For a systematic exploration regarding which practical scenarios can be covered by methods like IMUTube, i.e., to assess when virtual IMU data are of practical value for deriving HAR systems, we introduce an approach for predicting the utility of IMUTube for specific activities that are based on–possibly subtle–body (part) motions. We define subtle motions as those movements that involve one or two body parts moving in a very limited range of distances, which we quantify with our proposed metric. The range of hand and arm movements in hand washing or eating activities, for example, is much smaller than those involved in sports or gym exercises. For the quantification of the subtlety of movements, we introduce the *Motion Subtlety Index* (MSI).

Motion Subtlety in Videos: Figure 1 illustrates how the MSI is calculated for an exemplary video segment that captures a sequence of a writing activity, for example, with the wrist as the target location for the virtual IMU sensor. For each frame in a video segment with T frames, we first compute the optical flow [32] and estimate 2D poses [9]. The estimated optical flow at each pixel and time is normalized according to frame size to take account for different resolution of videos available, $(u_i^t, v_i^t) \rightarrow (u_i^t/H, v_i^t/W)$, where (u_i^t, v_i^t) are vertical and horizontal optical flow at pixel i and time t and (H, W) are height and width of frame size of the video. Next, at each frame, we calculate the average magnitude of the normalized optical flow at the local patch, $K \times K$, in the neighborhood of the wrist keypoint location, which is automatically detected by our 2D pose estimation procedure [10]. To take account of the varying resolution of video frames, the patch size is 2% of the larger dimension of the frame, $K = 0.02 \times max(H, W)$, to include only the pixels in

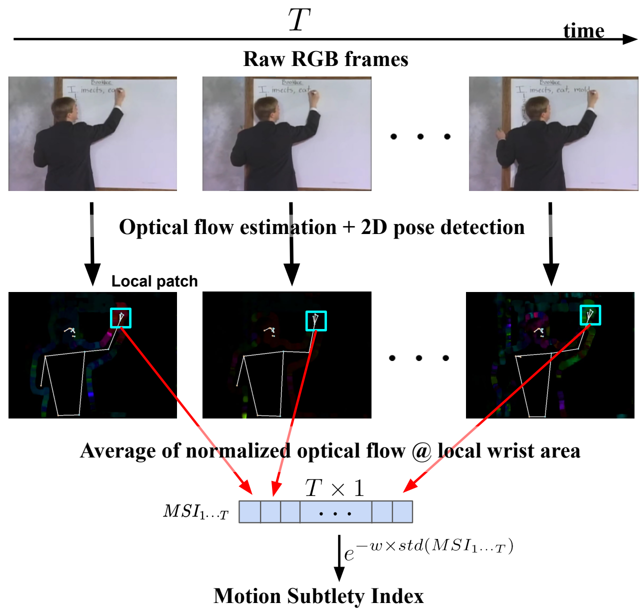

Figure 1: Overview of how Motion Subtlety Index (MSI) is calculated from video frames. For a given video segment, we perform optical flow estimation and 2d pose detection on the video. The estimated optical flow around the sensor location, wrist, is then averaged and normalized for each frame. The exponential of negative standard deviation of the previous computed value for all frames is the MSI of the video. Video shown is taken from Kinetics-400 [13]

the vicinity of wrist keypoints. The MSI is calculated as follows:

$$MSI_t = \frac{1}{N} \sum_{-\frac{K}{2} \geq i,j \geq \frac{K}{2}} \sqrt{(u_{n+i}^t)^2 + (v_{n+j}^t)^2}, \tag{1}$$

where u_n^t and v_n^t are vertical and horizontal components of the normalized optical flow measurements from the keypoint location at time t, and $N = K \times K$. The MSI for the analysis window is then computed as the exponential of negative standard deviation of $MSI_{1\ldots T} = [MSI_1, MSI_2, \cdots, MSI_T]$:

$$MSI = e^{-w \times std(MSI_{1\ldots T})} \tag{2}$$

where $w = 100$ to consider the minimal difference in MSI near zero.

Overall, MSI captures the motion information recorded around the on-body sensor location in the given video sequence. A smaller MSI means more significant motions are involved with ongoing activities, whereas a larger MSI indicates more subtle movements. MSI value calculation takes far less time and resources than virtual IMU data generation. Using an NVIDIA Titan Xp GPU, calculating the MSI value for a one-minute-long video takes approximately two minutes, while generating the virtual IMU data for the same video takes approximately twenty minutes. Therefore, MSI values are a quick way to determine if virtual IMU data is helpful, which saves the time and resources spent on generating virtual IMU data.

Motion Subtlety in Real IMU Data: We now also introduce the real Motion Subtlety Index (rMSI) to quantify the subtlety of

Table 1: Dataset Statistics for eating and daily activity classification experiments.

Task	Dataset	Duration
Daily Activity	Real IMU (HAD-AW)	454 min
	Virtual IMU (Curated)	201 min
	Virtual IMU (In-the-wild)	71 min
Eating	Real IMU (Lab-20)	284 min
	Real IMU (Wild-7)	128 min
	Virtual IMU (Curated)	17 min
	Virtual IMU (In-the-wild)	31 min

motions using real IMU data, which will be compared to MSI calculated from videos. Given a sequence of accelerometry signal in the analysis frame, $X^{T\times3}$, where T is the duration of real IMU data, we first remove the approximated gravity component, $\hat{X}_t = X_t - X_g$, which is the average of accelerometry signal, $X_g = \frac{1}{T}\sum_{0\leq t\leq T} X_t$, following Mizell [21]. Then, the magnitude of acceleration is computed, $\bar{X}_t = \sqrt{\hat{x}_t^2 + \hat{y}_t^2 + \hat{z}_t^2}$, where $\hat{X}_t = [\hat{x}, \hat{y}, \hat{z}]$. Lastly, rMSI is defined as the standard deviation of the magnitude of accelerometry signal without gravity, $rMSI = e^{-std([\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_T])}$. Overall, higher rMSI value indicates more subtle movements are involved for performing the on-going activity.

4 EXPERIMENTS

4.1 Dataset

For studying subtle motions in daily activities, we use two benchmark datasets for wrist sensors: daily activity classification [22] and eating activity classification [29]. We generated virtual IMU data from two video datasets. One is a well-curated video dataset, Kinetics-400 [13], containing videos from specific activity classes trimmed down to clips of 10 seconds. We also collected an in-the-wild dataset containing videos downloaded from YouTube without trimming. Table 1 gives an overview of the dataset used in our experiments.

For daily activities, we used the HAD-AW dataset [22] as our benchmark, which consists of 31 daily activities collected using an Apple Watch. We chose 17 activity classes that were available in our curated video dataset, Kinetics-400 [13], including *playing violin, playing piano, playing guitar, driving automatic, driving manual, reading, writing, eating a sandwich, cutting components, washing dishes, washing hands, showering, sweeping, wiping, drawing, flipping, bed-making*.

For eating, we used the dataset available from Thomaz *et al.* [29] that contains wrist sensor data from both in-lab and in-the-wild settings while eating. The *Lab-20* dataset was collected from 21 participants in the lab for both eating and non-eating activities. The eating moments involve, eating with fork and knife, hand, and spoon. For the in-the-wild scenario, we used the *Wild-7* dataset [29], which was collected from seven participants. From the Kinetics-400 dataset, we collected 417 video clips for the eating class, which were labeled as one of 10 eating-related classes (eating: burger, cake, carrots, chips, doughnuts, hotdog, ice cream, nachos, spaghetti, watermelon) and used for generating the curated virtual IMU data.

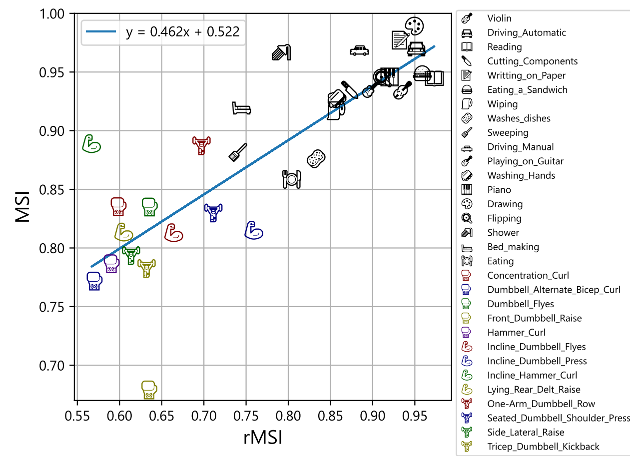

Figure 2: rMSI values plotted against MSI values for 17 daily activities and eating (Black) with 13 gym activities (Colored). Majority of gym exercises, where IMUTube was demonstrated to be very effective previously, show significantly smaller MSI and rMSI compared to those of dailty activities. Pearson Correlation analysis shows strong positive correlation between rMSI and MSI values ($r = 0.83$, $p \leq 0.001$)

4.2 HAR with Wrist Sensors

We followed the approach used in the original IMUTube experiments [18]. Virtual IMU data was extracted for wrist sensors from videos, which are calibrated with (a small amount of) real IMU data used for training. Real IMU data is subsampled to 25 Hz to match the frame rate of videos. As a baseline, we use the model that was trained only using real IMU data.

For daily activities, evaluation is done using a Random Forest classifier and DeepConvLSTM [**?**]. All real and virtual IMU data were segmented with a window length of 3 seconds and step size of 1.5 seconds (identical to previous work [22]). For the random forest classifier, ECDF features [11] were extracted from each window for training a Random Forest classifier. We applied 5-fold stratified cross-validation. We could not test leave-one-user-out cross-validation, as only a few users performed an entire set of activities in the official dataset. We report the average macro F1 score and its standard deviation from all folds for three runs.

For eating, we used the Lab-20 dataset as training set and Wild-7 as testing set following [29]. Accordingly, virtual IMU data was calibrated with the Lab-20 dataset. We segmented the data with a window size of 6 seconds and 50% overlap. From each window, extracted mean, variance, skewness, kurtosis and root mean square features for training a Random Forest classifier. DeepConvLSTM was trained with raw analysis frames. Classfier predictions were fed into a DBSCAN clustering to aggregate eating moments. We report the average binary F1 scores and its normal approximation interval from the three runs on Wild-7 dataset.

5 RESULTS

We first demonstrate the validity of MSI compared with rMSI, the motion subtlety that is actually quantified from the real IMU dataset.

Table 2: Classification results (mean and binary F1 scores) using a Random Forest classifier and DeepConvLSTM.

Task	Virtual	RF	ConvLSTM
Daily Activity	Real-Only	**0.477** ± 0.001	**0.552** ± 0.004
	Curated	0.462 ± 0.001	0.515 ± 0.005
	In-the-wild	0.472 ± 0.001	0.540 ± 0.002
Eating	Real-Only	0.715 ± 0.062	0.763 ± 0.043
	Curated	**0.800** ± 0.027	**0.821** ± 0.019
	In-the-wild	0.775 ± 0.035	0.794 ± 0.023

Then, we show the overall classification performance for classifying activities with subtle motions.

Figure 2 shows a strong positive correlation between rMSI and MSI values for the activities (Pearson $r = 0.60$, $p = 0.01$), indicating that MSI effectively quantifies the subtlety of movements in videos. We calculated the rMSI and MSI for 17 daily activities, where the real IMU data came from HAD-AW [22], and the videos came from Kinetics-400 [13]. We additionally analyzed rMSI and MSI for 13 gym exercises from MyoGym [15], in which IMUTube very effectively improved classification performances in previous work [19]. The majority of gym exercises involving large movements showed significantly smaller both MSI and rMSI compared to daily activities, demonstrating that approximately $MSI = 0.9$ is a potential indicator for considering the target activities involving subtle movements. In the following sections, we further study the relationship between MSI and classification performance for 17 daily activities with subtle motions ($MSI \geq 0.9$). In this analysis, the mode of MSI and rMSI from all frames in each activity class is used to represent the overall subtlety of movements for each activity.

5.1 Classification Analysis

Shown in Table 2, we observed mixed results between daily activity classification and eating detection. For daily activity classification, using additional virtual IMU data resulted in worse classification performance when compared to the baseline where only real IMU data was used. We consider that the model relied heavily on real IMU data to capture very detailed movement patterns for the classification tasks as virtual IMU data could not capture the characteristics of those subtle motions from the video data and acted as noise.

For eating detection, much to our surprise, the addition of virtual IMU data was very effective. Compared to the baseline, using curated and in-the-wild video datasets improved the model performance significantly. We consider that improvements came from the virtual IMU data containing a wide range of fine-grained eating motions from videos using varying utensils or food types that involve large arm motions (MSI = 0.86). This contrasts with eating activity in the HAD-AW dataset used for daily activity classification, which only included "eating a sandwich". Upon examining videos of people eating a sandwich, we found that most of the body movements involve the head. The wrist and arm movements are limited compared to, e.g., eating with a fork and knife.

6 DISCUSSION

We now analyze *when*, i.e., for which activities IMUTube is beneficial, and when not. Having a *cut-off* threshold MSI value would practically help practitioners to decide whether or not to put resources and time into generating virtual IMU data. Thus, we conducted an experiment, using virtual IMU data only for those classes for which the MSI value was determined to be below a certain cut-off value. Recall that the MSI values are calculated on the videos *prior* to generating virtual IMU data using IMUTube.

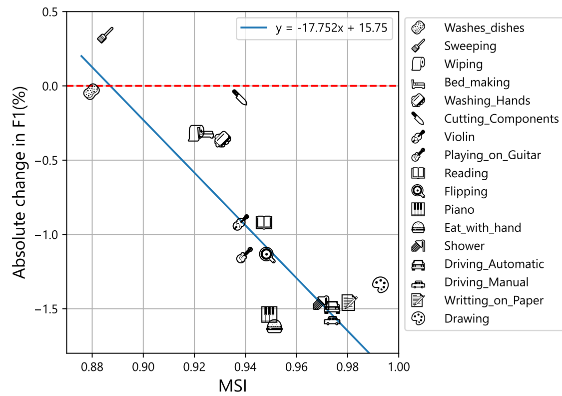

Figure 3: For all data points, we use all 17 activities and compare the difference in F1 score between the model (Random Forest) that is trained only with real IMU data to the one that is based on both real and virtual IMU data. As the MSI threshold (x-axis) increases, we added virtual IMU data only for activity classes that have MSI values lower than that threshold. Pearson Correlation analysis shows strong negative correlation between MSI and changes in F1 score ($r = -0.85$, $p \leq 0.001$).

Figure 3 shows a strong negative correlation (Pearson, $r = -0.85$, $p \leq 0.001$) between the MSI cut-off values and changes in F1 score introduced by using virtual IMU data for the activity classes having MSI below the MSI cut-off value. The zero-crossing of the linear line fit for all data points was $MSI = 0.89$, which marks an approximate decision boundary between activities that benefit from additional virtual IMU data or not. Since eating had a MSI of 0.86, less than the zero-crossing value, the addition of the virtual IMU data greatly benefited eating detection. This supports that MSI can provide a reference to gauge the benefit of using virtual IMU data when classifying activities with subtle motions.

Although MSI was demonstrated to be useful for the 17 daliy activities in this study, those are far short of covering the diverse activities in our daily living. Thus, in our future work, we will include more fine-grained activities [20] to futher evaluate the utility of MSI and IMUTube-like systems. On average, we used 13 minutes of videos per activity to calculate MSI. In the future, we will conduct experiments to determine the minimum video footage required for performing the MSI analysis. Additionally, we will study if our finding holds when applied to virtual IMU data generated with other cross-modality transfer systems based on videos [27, 28, 31]. Also, the proposed MSI, although useful and intuitive, is based on heuristic approach and applies the same hyperparameters across all videos not considering class-specific differences in motion subtlety. To further take into account for variabilities of videos (such as occlusions and camera viewpoints) and activity classes, we will study detecting motion subtleties using data-driven approaches in our future work, such as explainable deep models for videos [1, 5].

7 SUMMARY AND CONCLUSION

In this paper, we showed that the subtlety of motions in activities is quantifiable in video data – through our newly introduced Motion Subtlety Index (MSI), which correlates with the eventual downstream activity recognition accuracy on IMU data. We were able to systematically assess how the addition of virtual IMU data benefits general HAR systems and showed how the *a-priori* calculation of MSI values *on videos* can be used to effectively guide the application of systems like IMUTube, as calculating MSI is 10× faster than running IMUTube. Overall, this study demonstrates that the proposed MSI provides quantifiable approach to pinpoint when IMUTube will fail when it comes to subtle activities. Activity classes beyond $0.9 \leq MSI$ seems to have very subtle and complex motions in hand and wrist movements that are very difficult for IMUTube to capture due to the current limitations of state-of-the-art human motion tracking techniques [8].

REFERENCES

[1] Vineeth N. Balasubramanian. 2022. Toward Explainable Deep Learning. *Commun. ACM* 65, 11 (oct 2022), 68–69. https://doi.org/10.1145/3550491
[2] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. Fitbyte: Automatic diet monitoring in unconstrained situations using multimodal sensing on eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[3] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–20.
[4] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 2015. A wearable system for detecting eating activities with proximity sensors in the outer ear. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 91–92.
[5] Nadia Burkart and Marco F. Huber. 2021. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Int. Res.* (2021).
[6] Wenqiang Chen, Shupei Lin, Elizabeth Thompson, and John Stankovic. 2021. SenseCollect: We Need Efficient Ways to Collect On-body Sensor-based Human Activity Data! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–27.
[7] L. Cilliers. 2020. Wearable devices in healthcare: Privacy and information security issues. *Health information management journal* 49, 2-3 (2020), 150–156.
[8] Y. Desmarais, D. Mottet, P> Slangen, and P. Montesinos. 2021. A review of 3D human pose estimation algorithms for markerless motion capture. *Computer Vision and Image Understanding* 212 (2021), 103275.
[9] S. Dubey and M. Dixit. 2022. A comprehensive survey on human pose estimation approaches. *Multimedia Systems* (2022), 1–29.
[10] H. Fang, S. Xie, Y. Tai, and C. Lu. 2017. RMPE: Regional Multi-person Pose Estimation. In *ICCV*.
[11] N. Y. Hammerla, R. Kirkham, P. Andras, and T. Ploetz. 2013. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In *Proceedings of the 2013 international symposium on wearable computers*. 65–68.
[12] D. Jiang and G. Shi. 2021. Research on data security and privacy protection of wearable equipment in healthcare. *Journal of Healthcare Engineering* 2021 (2021).
[13] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. 2017. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950* (2017).
[14] H. Konttinen, T. Van Strien, S. Männistö, P. Jousilahti, and A. Haukkala. 2019. Depression, emotional eating and long-term weight changes: a population-based prospective study. *International Journal of Behavioral Nutrition and Physical Activity* 16, 1 (2019), 1–11.
[15] Heli Koskimäki, Pekka Siirtola, and Juha Röning. 2017. Myogym: introducing an open gym data set for activity recognition collected using myo armband. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 537–546.
[16] Hyeokhyen Kwon, Gregory D Abowd, and Thomas Plötz. 2019. Handling annotation uncertainty in human activity recognition. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 109–117.
[17] Hyeokhyen Kwon, Gregory D Abowd, and Thomas Plötz. 2021. Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables. *Sensors* 21, 24 (2021), 8337.
[18] Hyeokhyen Kwon, Catherine Tong, Harish Haresamudram, Yan Gao, Gregory D Abowd, Nicholas D Lane, and Thomas Ploetz. 2020. Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–29.
[19] Hyeokhyen Kwon, Bingyao Wang, Gregory D Abowd, and Thomas Plötz. 2021. Approaching the Real-World: Supporting Activity Recognition Training with Virtual IMU Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–32.
[20] Gierad Laput and Chris Harrison. 2019. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[21] David Mizell. 2003. Using gravity to estimate accelerometer orientation. In *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.* Citeseer, 252–252.
[22] Sara Mohammed, Reda Elbasiony, and Walid Gomaa. 2018. An LSTM-based Descriptor for Human Activities Recognition using IMU Sensors. 504–511. https://doi.org/10.5220/0006902405040511
[23] A. Moschetti, L. Fiorini, D. Esposito, P. Dario, and F. Cavallo. 2017. Daily activity recognition with inertial ring and bracelet: An unsupervised approach. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3250–3255.
[24] E. Navarro-Tapia, L. Almeida-Toledano, G. Sebastiani, M. Serra-Delgado, Ó. García-Algar, and V. Andreu-Fernández. 2021. Effects of microbiota imbalance in anxiety and eating disorders: probiotics as novel therapeutic approaches. *International journal of molecular sciences* 22, 5 (2021), 2351.
[25] F. J. Ordóñez and D. Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.
[26] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* (2016).
[27] Vitor Fortes Rey, Peter Hevesi, Onorina Kovalenko, and Paul Lukowicz. 2019. Let There Be IMU Data: Generating Training Data for Wearable, Motion Sensor Based Activity Recognition from Monocular RGB Videos. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers.*
[28] Panneer Selvam Santhalingam, Parth Pathak, Huzefa Rangwala, and Jana Kosecka. 2023. Synthetic Smartwatch IMU Data Generation from In-the-Wild ASL Videos. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2023).
[29] Edison Thomaz, Irfan Essa, and Gregory D Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1029–1040.
[30] T. Van Strien. 2018. Causes of emotional eating and matched treatment of obesity. *Current diabetes reports* 18, 6 (2018), 1–8.
[31] Chengshuo Xia and Yuta Sugiura. 2022. Virtual IMU Data Augmentation by Spring-Joint Model for Motion Exercises Recognition without Using Real Data. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers.*
[32] M. Zhai, X. Xiang, N. Lv, and X. Kong. 2021. Optical flow and scene flow estimation: A survey. *Pattern Recognition* 114 (2021), 107861.