2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

HUMAN-INDEPENDENT ACTIVITY RECOGNITION OF CONSTRUCTION WORKERS

Seongeun Park[1], Hoonyong Lee[2], Changbum Ryan Ahn[1], Moonseo Park[1]
[1]Seoul National University, Seoul, Republic of Korea
[2]University of Michigan, Ann Arbor, USA

Abstract

With recent advancements in sensor and data analysis technology, multiple research on worker activity recognition through wearable sensors have been conducted to solve worker safety and productivity problem at construction sites. However, most rely on pre-trained models which require re-training of each worker to take into account differences between workers. To alleviate this limitation, we propose a human-independent model that can adapt to differences in workers. Our model uses variational-denoising autoencoder with soft parameter sharing to extract common features in different construction activities, achieving 78.64% accuracy which is higher than existing benchmark models.

Introduction

The challenges of low productivity and high incidence of accidents in the construction industry are often addressed through close monitoring of workers (Joshua and Varghese, 2011; Kim and Cho, 2020). Worker activity monitoring has recently become more efficient with advancements in sensor and data analysis technology which enabled the development of wearable Inertial Measurement Units (IMU) sensors that are small and cheap, while boasting high accuracy. IMU sensors collect data on workers' physical movements through acceleration and angular velocity, generating distinct signals (Gil-Martín et al., 2020). Machine learning algorithms are then used to analyze patterns in the signals to classify different construction activities (Gong et al., 2022). Various research have demonstrated the effectiveness and reliability of automated activity recognition of construction workers with IMU data (Kim and Cho, 2020; Lee et al., 2020; Bangaru et al., 2021).

However, one limitation of learning-based models is that their accuracy rapidly decreases when data distribution changes, meaning when learning and prediction objects are different (Cook et al., 2013). This is problematic for monitoring worker activities, as the behavior of construction workers varies depending on their age, proficiency, habits, and body type. In order to address this issue, a large amount of labeled data to cover a wide range of data distribution is needed. However, challenges remain as worker activities also vary greatly across different construction sites.

In the field of Human Activity Recognition (HAR), a model named ODIN (Hussein and Hajj, 2022) was designed to overcome the limitations of supervised learning in classifying daily life activities. It uses a denoising autoencoder structure, which enhances domain-specific and shared characteristics of the source and target domains in multi-variate time series data. It has shown improved performances compared to existing models in cross-user domain adaptation for daily life activity classification. However, its effectiveness for worker activity recognition (WAR) is still unknown as the performance of HAR models varies greatly depending on the activity type, differences in movement specificity, and individual variability between construction and daily life activities (Hussein and Hajj, 2022). A model that simply removes personal characteristics and extracts characteristics of common behaviors from a HAR model may not work well for WAR.

This highlights the need for a human-independent WAR model suitable for construction activities. In this paper, we propose a model that applies Variational AutoEncoder (VAE) to ODIN (Hussein and Hajj, 2022), which generalizes workers' individual characteristics to extract commonalities in construction activity for more flexible cross-domain knowledge transfers. We also utilize VAE as a generative model that generates new samples from learned functions. As depicted in Figure 1, our work addresses the problem of worker domain adaptation in WAR by enabling activity recognition in unseen data. Our approach is a first-of-its-kind effort to tackle the challenge in the construction field. By achieving WAR without the need for large amounts of labeling which can be expensive, this study could contribute to lowering the cost of automation at construction sites.

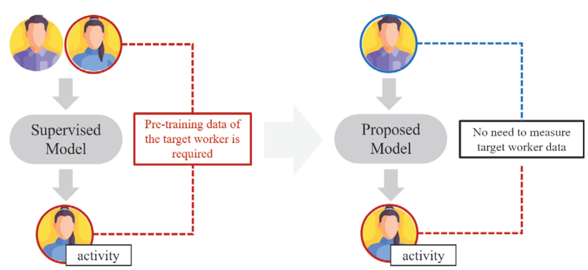

Figure 1: Research Objective

Background

For WAR, IMU sensors are attached to the bodies of subjects (i.e., construction workers) to collect sensor signals containing information about their movements (Kim and Cho, 2020). These sensors provide a continuous stream of data, which are processed and analyzed to detect patterns. Machine learning algorithms are used to classify the data into different activities (Gong et al., 2022). Kim and Cho (2020) proposed a motion recognition model for construction workers using a Long Short-Term Memory (LSTM) network based on the evaluation of the validity of the number and location of motion sensors. Lee et al. (2020) proposed DeTECLoad, which simultaneously predicts load and posture based on the CNN-LSTM algorithm using IMU sensors. DeTECLoad first converts IMU data into image data using Gramian Angular Field, then uses a hybrid Convolutional Neural Network-Long Short-Term Memory to classify load-carrying modes from the image data. Bangaru et al. (2021) proposed a model that recognizes non-engineering activity by fusing electromyographic (EMG) and IMU sensors based on artificial neural network (ANN). These are all supervised learning algorithms that achieve strong performances (Bangaru et al., 2021; Lee et al., 2020; Sherafat et al., 2020; Kim and Cho, 2020; Ahn et al., 2019; Chen et al., 2017; Yang et al., 2017; Jebelli et al., 2016).

However, a major limitation of such learning-based algorithms is that they do not account for potential shifts in data distribution (Sugiyama and Kawanabe, 2012). Shifts in data distribution are common in HAR because activity patterns vary across different people (Cook et al., 2013). Using supervised machine learning approaches to recognize human activities from on-body wearable accelerometers thus requires a large amount of labeled data (Trabelsi et al., 2013), which can be challenging and expensive (Cvetkovic et al., 2011). To overcome this limitation, domain adaptation methods (Pan and Yang, 2010) have been widely used in HAR (Chakma et al., 2021; Li et al., 2021; Sanabria et al., 2021a,b; Chang et al., 2020; Wilson et al., 2020). For example, ODIN (Hussein and Hajj, 2022) uses a soft parameter sharing (SPS) architecture that includes representation learning and the relationship between the parameters of source and target models as non-linear. The adaptation loss function is modeled as the squared Maximum Mean Discrepancy (MMD). The effectiveness of the model was tested using Position Activity Recognition (Sztyler and Stuckenschmidt, 2016) and Heterogeneity Activity Recognition (Stisen et al., 2015) datasets, which measure daily life activity using inertial sensors on smartwatches and smartphones, and it outperformed existing state-of-the-art approaches in cross-user domain adaptation.

Performance of HAR models heavily depends on the type of activity class each model is designed to recognize. Different activity classes have different characteristics, and the model must accurately identify these unique attributes to effectively classify each activity. As a result, a model that performs well for one activity class may not perform as well for another. Therefore, choosing appropriate model for each activity class is key to achieving good performance.

Classifying worker activities at construction sites has different challenges from classifying daily life activities. Construction worker activities have smaller scale movements and greater variability compared to daily life activities (Tian et al., 2022). The scale of movement is smaller because construction work often involves specific tools, while daily life activity involves a wider range of movements. The variability of construction worker activity is high due to differences in age, skill level, habit, and body type. This makes it more difficult to accurately detect and classify construction worker activities, pointing to the need for a new approach.

Methodology

Variational AutoEncoder

The idea behind encoder-decoder-based structures is to compress input data into a compact latent space while preserving information in the data. However, the encoder's choice of data representation may not be very interpretable and may also lead to overfitting where single input examples are mapped to specific numbers without any high-level meaning (Stephan et al., 2021). To promote continuity in the latent space and reduce overfitting, Variational AutoEncoder (VAE) (Kingma and Welling, 2013) can be used instead (Stephan et al., 2021). VAE adds a probabilistic element to AutoEncoder (AE) by modeling the latent vector as a probability distribution (Kingma and Welling, 2013). It maps the input data to the distribution's mean and standard deviation, and a sample is drawn to obtain the latent vector. This allows VAE to generate samples similar to the input data (Kingma et al., 2019). VAE offers several benefits over AE as it is able to handle different domains by modeling and adjusting the distribution of the latent vectors, and can also generate new samples that would reduce overfitting and improve generalization performance (Cai et al., 2019). VAE also provides a measure of uncertainty for its representations, which is useful in determining the extent of domain shift in domain adaptation.

In this paper, we propose a model that extends the structure of ODIN (Hussein and Hajj, 2022) with VAE. ODIN is a domain adaptation model that uses Denoising Autoencoder for source domain learning and Soft Parameter Sharing between source and target domains, as shown in Figure 2. The domain of our study is construction workers for WAR. Various research on daily-life HAR has shown that ODIN outperforms existing models. Our proposed soft parameter sharing denoising autoencoder model using VAE enables effective representation learning for WAR. Unlike AEs which lack the ability to reconstruct learned latent vectors, VAEs have probability distributions of latent vectors from the Encoder. This allows for generation of decoding results similar to the latent vectors, even if the learned latent vectors are distributed close together. Using VAEs can prevent overfitting and extract generalized features.

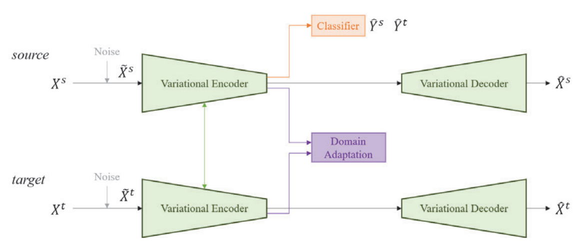

Figure 2: Proposed Architecture

Data

To verify our model, we use VTT-ConIoT dataset (Mäkelä et al., 2021), a publicly available dataset for evaluating activities of workers through inertial sensors in construction operations (Mekruksavanich et al., 2022; Mekruksavanich and Jitpattanakul, 2023). It classifies 16 different activities routinely carried out at a construction site into 6 classes (see Table 1).

Table 1: Activity class of VTT ConIoT (Mäkelä et al., 2021)

Class	Included Activities
Painting	Roll-Painting, Spraying-Paint, Level-Paint
Cleaning	Vaccum-cleaning, Picking-objects
Climbing	Climbing-stairs, Jumping-down, Stairs-Up-Down
Handsup	Laying-back, HandsUp-High, HandsUp-low
FloorWork	Crouch-floor, Kneel-floor
WalkingDisplacements	Walk-straight, Walk-winding, Pushing-cart

The dataset consists of IMU data collected from 12 participants. 3 IMU sensors were attached to the upper part of non-dominant arm, the back of the non-dominant shoulder, and the hip of each participant, as shown in Figure 3.

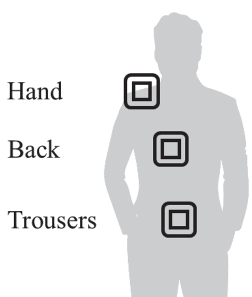

Figure 3: Sensor Place

Each data point is 120 or 180 seconds long, and contains acceleration and angular velocity features for three axes, with a sampling rate of 128 Hz.

Data Preprocessing

From the acceleration and gyroscope measurements collected from the 3 sensor locations, we extracted 18 features which have 3 axes each. Outliers were identified and removed from the first 4 seconds and the last 1 second of the collected data. To further refine the data, we removed high-frequency data exceeding a movement rate of 10 times per second using a low-pass Butterworth filter. The filtered data was then normalized to prevent overfitting to specific features. In addition, we segmented the data to reflect the temporal dependency of the time-series data with moving window of 2 seconds and 1-second interval for data sampling.

Model Training

In addition to our model, we trained the following two models to compare their performance with that of our model: 1) CNN-LSTM, which has shown strong performance in WAR as seen in Lee et al. (2020); 2) ODIN (Hussein and Hajj, 2022), a domain adaptation model designed to classify daily life activities. We selected these two models because they outperform the state-of-the-art in their respective domains (Lee et al., 2020; Hussein and Hajj, 2022). For training, subjects 1, 3, 4, 8, 9, 10, 11, 12, and 13 from the VTT-ConIoT dataset were used as the source domain, and subject 7 as the target domain. All models were trained and validated using the same dataset.

Results

Across five trials conducted under the same conditions, our model outperform ODIN by an average of 13.11% in the case of domain shift (see Table 2). In case of no domain change, our model shows weaker performance (a 4.27% decrease in accuracy) compared to ODIN. This suggests that our model learned more general and subject-independent features than ODIN, highlighting the superiority of VAE in domain adaptation over AE.

Table 3 shows our model outperforms CNN-LSTM (Lee et al., 2020) by 41.39% in environments where domain is shifted. CNN-LSTM achieved high accuracy (99.45%) in the same domain scenarios, but its accuracy dropped to 37.25% when the domain shifted.

To visualize the learned representations of the three models, we use t-distributed stochastic neighbor embedding (t-SNE) projection (Van der Maaten and Hinton, 2008) Each activity class is represented by a unique number, and the two domains are color-coded, as shown in Figures 4, 5, 6, and 7. When the domains are the same (i.e., when the learning data contain the test data), the representation is well-organized for the same classes (Figure 4). However, when domain shift occurs, the representation becomes less organized (Figure 5).

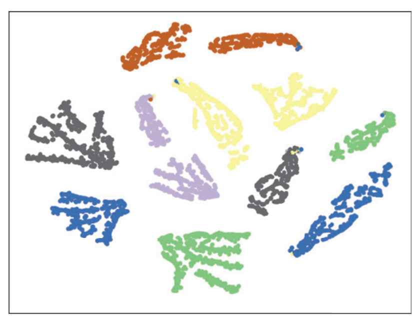

Figure 4: T-SNE plot for CNN-LSTM -same domain

ODIN, on the other hand, demonstrated some improvements in terms of class clustering compared to Figure 5, as seen in Figure 6. Although the clusters are closely grouped together in the same color, blue classes are still not well separated from other classes and are situated close to green and gray classes. The accuracy of ODIN was lower for WAR than HAR, indicating that daily life and construction activity classes have different characteristics.

Our model show improved representation compared to

Table 2: Model Accuracy

Model	Domain	Trial 1	Trial 2	Trial	Trial 4	Trial 5	Average
ODIN	Same	98.06	99.03	98.06	99.51	98.06	98.54
	Shift	61.65	64.08	65.53	71.84	64.56	65.53
Proposed Model	Same	93.69	96.60	92.72	95.63	92.72	94.27
	Shift	71.84	82.04	74.76	84.47	80.10	78.64

Table 3: Comparing Model Accuracy

Model	Accuracy (Same Domain)	Accuracy (Domain Shift)
CNN-LSTM (Lee et al., 2020)	99.44%	37.25%
ODIN (Hussein and Hajj, 2022)	98.54%	65.53%
Proposed Model	94.27%	78.64%

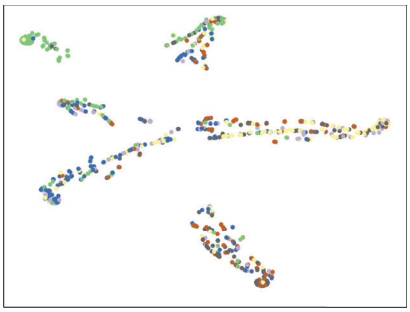

Figure 5: T-SNE plot for CNN-LSTM -domain shift

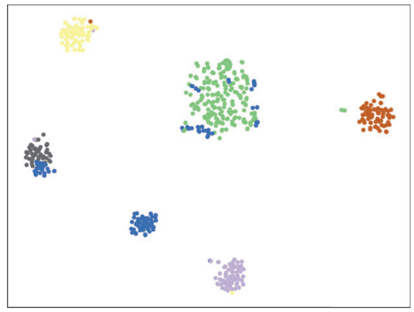

Figure 6: T-SNE plot for ODIN -domain shift

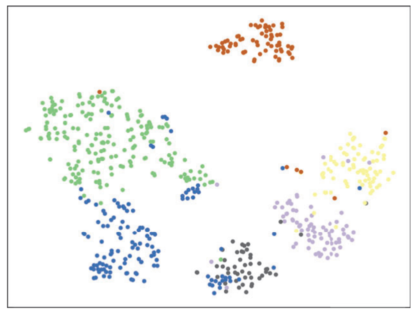

Figure 7: T-SNE plot for our model -domain shift

ODIN (see Figure 7) especially for the blue class, evidenced by its ability to better generalize and its learned decision boundary for the task at hand.

Conclusions

This paper contributes to the field of wearable sensor-based activity recognition of construction workers. Our model shows an improved accuracy of 78.64% over two benchmark models, demonstrating its ability to effectively classify various construction behaviors independent of workers' personal characteristics. The potential impact of this work is significant as it can lead to increased safety and efficiency at construction sites. One limitation, however, is that attaching 3 sensors may have influenced worker behaviors. Future studies could thus explore the performance of our model with fewer sensors and investigate optimal sensor mounting locations for WAR in situations where domain shifts occur. Overall, our model demonstrates the potential for VAE in domain adaptation for WAR, and our findings have important practical implications for the construction industry.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1A2C200299912).

References

Ahn, C. R., Lee, S., Sun, C., Jebelli, H., Yang, K., and Choi, B. (2019). Wearable sensing technology applications in construction safety and health. Journal of Construction Engineering and Management, 145(11):03119007.

Bangaru, S. S., Wang, C., Busam, S. A., and Aghazadeh, F. (2021). Ann-based automated scaffold builder activity recognition through wearable emg and imu sensors. Automation in Construction, 126:103653.

Cai, R., Li, Z., Wei, P., Qiao, J., Zhang, K., and Hao, Z. (2019). Learning disentangled semantic representation for domain adaptation. In IJCAI: proceedings of the conference, volume 2019, page 2060. NIH Public Access.

Chakma, A., Faridee, A. Z. M., Khan, M. A. A. H., and Roy, N. (2021). Activity recognition in wearables using adversarial multi-source domain adaptation. Smart Health, 19:100174.

Chang, Y., Mathur, A., Isopoussu, A., Song, J., and Kawsar, F. (2020). A systematic study of unsupervised domain adaptation for robust human-activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(1):1–30.

Chen, J., Qiu, J., and Ahn, C. (2017). Construction worker's awkward posture recognition through supervised motion tensor decomposition. Automation in Construction, 77:67–81.

Cook, D., Feuz, K. D., and Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. Knowledge and information systems, 36:537–556.

Cvetkovic, B., Kaluza, B., Luštrek, M., and Gams, M. (2011). Semi-supervised learning for adaptation of human activity recognition classifier to the user. In Proc. of Workshop on Space, Time and Ambient Intelligence, IJCAI, pages 24–29. Citeseer.

Gil-Martín, M., San-Segundo, R., Fernandez-Martinez, F., and Ferreiros-López, J. (2020). Improving physical activity recognition using a new deep learning architecture and post-processing techniques. Engineering Applications of Artificial Intelligence, 92:103679.

Gong, Y., Yang, K., Seo, J., and Lee, J. G. (2022). Wearable acceleration-based action recognition for long-term and continuous activity analysis in construction site. Journal of Building Engineering, 52:104448.

Hussein, A. and Hajj, H. (2022). Domain adaptation with representation learning and nonlinear relation for time series. ACM Transactions on Internet of Things, 3(2):1–26.

Jebelli, H., Ahn, C. R., and Stentz, T. L. (2016). Fall risk analysis of construction workers using inertial measurement units: Validating the usefulness of the postural stability metrics in construction. Safety science, 84:161–170.

Joshua, L. and Varghese, K. (2011). Accelerometer-based activity recognition in construction. Journal of computing in civil engineering, 25(5):370–379.

Kim, K. and Cho, Y. K. (2020). Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. Automation in Construction, 113:103126.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Kingma, D. P., Welling, M., et al. (2019). An introduction to variational autoencoders. Foundations and Trends® in Machine Learning, 12(4):307–392.

Lee, H., Yang, K., Kim, N., and Ahn, C. R. (2020). Detecting excessive load-carrying tasks using a deep learning network with a gramian angular field. Automation in Construction, 120:103390.

Li, Z., Cai, R., Ng, H. W., Winslett, M., Fu, T. Z., Xu, B., Yang, X., and Zhang, Z. (2021). Causal mechanism transfer network for time series domain adaptation in mechanical systems. ACM Transactions on Intelligent Systems and Technology (TIST), 12(2):1–21.

Mäkelä, S.-M., Lämsä, A., Keränen, J. S., Liikka, J., Ronkainen, J., Peltola, J., Häikiö, J., Järvinen, S., López, M. B., et al. (2021). Introducing vtt-coniot: A realistic dataset for activity recognition of construction workers using imu devices. Sustainability, 14(1):1–20.

Mekruksavanich, S., Jantawong, P., Hnoohom, N., and Jitpattanakul, A. (2022). Wearable-based activity recognition of construction workers using lstm neural networks. In 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pages 1–4. IEEE.

Mekruksavanich, S. and Jitpattanakul, A. (2023). Automatic recognition of construction worker activities using deep learning approaches and wearable inertial sensors. Intelligent Automation & Soft Computing, 36(2).

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.

Sanabria, A. R., Zambonelli, F., Dobson, S., and Ye, J. (2021a). Contrasgan: Unsupervised domain adaptation in human activity recognition via adversarial and contrastive learning. Pervasive and Mobile Computing, 78:101477.

Sanabria, A. R., Zambonelli, F., and Ye, J. (2021b). Unsupervised domain adaptation in activity recognition: A gan-based approach. IEEE Access, 9:19421–19438.

Sherafat, B., Ahn, C. R., Akhavian, R., Behzadan, A. H., Golparvar-Fard, M., Kim, H., Lee, Y.-C., Rashidi, A., and Azar, E. R. (2020). Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. Journal of Construction Engineering and Management, 146(6):03120002.

Stephan, M., Stadelmayer, T., Santra, A., Fischer, G., Weigel, R., and Lurz, F. (2021). Radar image reconstruction from raw adc data using parametric variational autoencoder with domain adaptation. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 9529–9536. IEEE.

Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., Sonne, T., and Jensen, M. M. (2015). Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In Proceedings of the 13th ACM conference on embedded networked sensor systems, pages 127–140.

Sugiyama, M. and Kawanabe, M. (2012). Machine learning in non-stationary environments: Introduction to covariate shift adaptation. MIT press.

Sztyler, T. and Stuckenschmidt, H. (2016). On-body localization of wearable devices: An investigation of position-aware activity recognition. In 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 1–9. IEEE.

Tian, Y., Li, H., Cui, H., and Chen, J. (2022). Construction motion data library: an integrated motion dataset for onsite activity recognition. Scientific Data, 9(1):726.

Trabelsi, D., Mohammed, S., Chamroukhi, F., Oukhellou, L., and Amirat, Y. (2013). An unsupervised approach for automatic activity recognition based on hidden markov model regression. IEEE Transactions on automation science and engineering, 10(3):829–835.

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. Journal of machine learning research, 9(11).

Wilson, G., Doppa, J. R., and Cook, D. J. (2020). Multisource deep domain adaptation with weak supervision for time-series sensor data. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pages 1768–1778.

Yang, K., Ahn, C. R., Vuran, M. C., and Kim, H. (2017). Collective sensing of workers' gait patterns to identify fall hazards in construction. Automation in construction, 82:166–178.