Spatio-Temporal Domain Adaptation for Gait Based User Recognition from Radar Data

Kalvik Jakkala
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
kjakkala@uncc.edu

Chen Chen
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
chen.chen@uncc.edu

Minwoo Lee
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
minwoo.lee@uncc.edu

Arupjyoti Bhuyan
Idaho National Laboratory
Idaho Falls, ID 83401
arupjyoti.bhuyan@inl.gov

Zhi Sun
Department of Electrical Engineering
University at Buffalo
Buffalo, NY 14260
zhisun@buffalo.edu

Pu Wang
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
pu.wang@uncc.edu

Abstract

Radar-based biometric identification is an emerging user identification platform that exploits radar return signals to capture human biometrics (such as gait, gesture, lip motion, and cardiac motion), which can be used to predict a user's identity. Despite its unique advantages (such as privacy-preserving and resilience to weather/lighting conditions), the generalization performance of this technology is still unknown and greatly hinders its practical deployment. To address this challenge, we collect and investigate a non-synthetic dataset, which revealed the existence of distinct spatial and temporal domain shifts in radar-based gait biometric data. We show that spatio-temporal domain shifts, when not addressed jointly, can significantly degrade identification accuracy. Moreover, we propose a data-efficient yet straightforward domain shift mitigation approach for tuning deep learning models over their entire life cycle. Our approach exploits an unsupervised domain shift detector to measure the malignancy of domain shifts. Such metrics allow us to determine the domains that maximize the net contributions upon adapting to, after which an appropriate domain adaption method is utilized to improve both spatial and temporal generalization. We show that our approach improves data efficiency by reducing the number of domains that necessitate adaptation while maintaining the generalization performance of a blind approach that uses data from all domains.

1 Introduction

As an emerging user identification technology, radar-based biometric identification exploits radar signals scattered and returned by human subjects to capture a variety of behavioral and physiological biometrics such as gait [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], gesture [11, 12], lip motion [13, 14], and cardiac motion [15, 16]. These biometrics characterize the movement patterns unique to each individual.

Radar-based biometric identification systems have several unique advantages. First, they are privacy-preserving as they do not rely on images of human subjects. Next, they can operate in adverse weather and lighting conditions such as heavy fog, smoke, rain, and zero-light conditions. Finally, they can see through some opaque objects depending on its material and the frequency of the radar [17, 18].

A rich body of research has established the viability of this new technology [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. However, the spatial and temporal generalization performance of this technology is still unknown. We found that deep neural networks (DNNs) trained on radar-based gait data suffer from substantial performance degradation with the progression of time and in new locations. Most radar-based biometric systems proposed in the literature, test their performance on a dataset collected alongside the training data and fail to uncover the temporal degradation. Furthermore, they seldom address the spatial generalization issues from introducing new environments. These issues are prevalent in other biometric solutions, such as images as well. Nevertheless, curating a dataset of monumental proportions often overcomes this, which is especially true for face recognition, as they have datasets containing millions of images. A brute force approach is not viable for most of the other biometric identification platforms, such as radar-based ones. Unlike images, most other platforms require specialized devices that are not omnipresent and exclusively designed for user recognition, which further complicates data collection campaigns on a large scale.

The root cause of this generalization issue is dataset shift [19]. Simply put, it occurs when the testing data distribution differs from that of the training data, i.e., the i.i.d. assumption is violated. Depending on the type of difference in the test data distribution, it can be categorized as a covariate, posterior probability, or concept shift. Mitigating dataset shift, also known as dataset drift or domain shift, is known to be one of the hardest problems in machine learning and is prevalent in numerous application domains. This leads us to the ethos of Domain Adaptation, addressed by myriads of methods in literature [20].

After an in-depth analysis of dataset shifts and domain adaptation, we found the current taxonomy of domain shifts is not adequate to explain the full extent of our problem. Apart from the manifestations of domain shifts mentioned above, we seek to distinguish them further. Concretely, we acknowledge the presence of spatial and temporal domain shifts (SDS/TDS). We treat TDS as shifts that arise temporally as a consequence of the inherent dynamics of a domain while, treating SDS as shifts induced from introducing new spatial locations. This distinction allows us to understand and explain the behavior of the shifts that manifest in radar-based gait data.

Before addressing a shift, we need to be aware of its presence. An overwhelming amount of domain adaptation methods only address shifts that behave similar to SDS [20], i.e., they only consider explicitly introduced domains. So, the problem of detecting the presence of a shift, which is a sine qua non for TDS, has not received the attention it deserves. Moreover, among the work available on drift detection [21], a majority of them make assumptions about the behavior of softmax classifiers, which have long been invalidated [22, 23]. We draw attention to a simple yet effective unsupervised domain shift detector based on metric learning [24]. Our shift detector enables a data-efficient domain shift mitigation approach, where the shift malignancy of each domain is first measured, and domains that maximize the net contributions upon adapting are selected. An appropriate domain adaption method is then applied only to the selected domains to improve model generalization. Our prominent contributions and findings are as follows.

- We curated a non-synthetic dataset consisting of radar-based gait biometric data. This dataset, for the first time, allows one to study and improve the spatio-temporal generalization performance of a radar-based biometric identification system.

- We introduced the distinction of TDS and SDS, which enables us to explain the system performance degradation.

- We uncovered a correlation between TDS and SDS, which unveiled significant ramifications for data collection and domain adaptation. In particular, we found SDS can be mitigated to a certain extent by using the TD data and vice-versa, but both SDS and TDS have to be addressed explicitly for consistent generalization performance.

- We revealed that if multiple sources of domain shifts appear, each source of domain shifts yields a different net contribution to generalization upon adapting to that domain.

- We elucidated the effectiveness of metric learning based shift detectors and reaffirm the limitations of softmax thresholding for out of domain detection.

- Finally, we exploit our shift detector to develop a data-efficient domain shift mitigation approach.
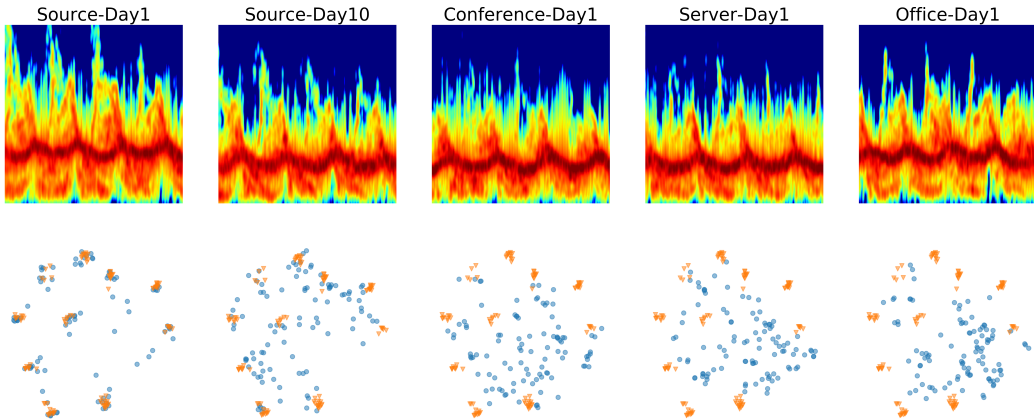
2 Dataset

Figure 1: First row: gait spectrograms of the same person. Second row: t-SNE embeddings generated from each domain of our dataset. Orange triangle clusters represent the training data in source location on day 1, where each cluster represents a distinct human subject. Blue circles represent each domain's testing data, where the data points were sampled from all the classes of a given domain.

In light of the novelty of our problem, we curated a dataset to study it properly. We collected gait data from 10 volunteers between the ages of 18-35. Each subject's data was collected in four different locations. The source location was a research space with cubicles, and the other three areas consisted of a server, conference, and an office room. By maintaining four distinct locations, we introduce SDS. In the source location, data was collected on ten different days for each subject. Five separate days of data was acquired for each of the three other locations, which are used as target domains.

The data collection was limited to 100 data samples in the source location and 50 data samples in each of the target locations on any given day. Unlike cameras, radars collect data actively by broadcasting signals into the environments. Unique movement patterns (e.g., gait) can induce different micro-Doppler frequency shifts in radar return signals, which can be represented as a spectrogram. Thus, a radar spectrogram can serve as the biometric print for user identification [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Apart from biometric traits, radar spectrogram data could also contain spatio-temporally varying noises from signal reflections and interference, which cannot be completely removed. This leads to environment-induced TDS and SDS. To add on, human gait, although unique to each individual, could contain minute variations, potentially as a consequence of a subject's mood, clothing, footwear, or some other similar aspect, which combined with the environment-induced shifts, further aggravates TDS. Changing the number of days in the train and test sets changes the amount of TDS. Similarly, using a subset of the available locations, SDS can be controlled, which makes it possible to study how the two variants of shifts interact with each other.

We use the widely-adopted radar data preprocessing schemes to minimize environment-induced shifts [1, 4, 5][1]. It is worth to note in Fig. 1 that the spectrograms from all locations look very similar. This is because they are spectrograms representing the same action (gait). Most of the environmental data is filtered out by the preprocessing methods, but the minuscule amounts of environmental information that seeps through was enough to induce domain shifts. We show the t-SNE [25] embeddings for each domain in our dataset alongside the training data embeddings. The model used to generate the embeddings was trained on a single day's data from the source location. The embeddings show a domain shift in each new domain. However, this might not be very evident in the absence of the knowledge of the data collection procedure. We address this issue in Section. 4.5.
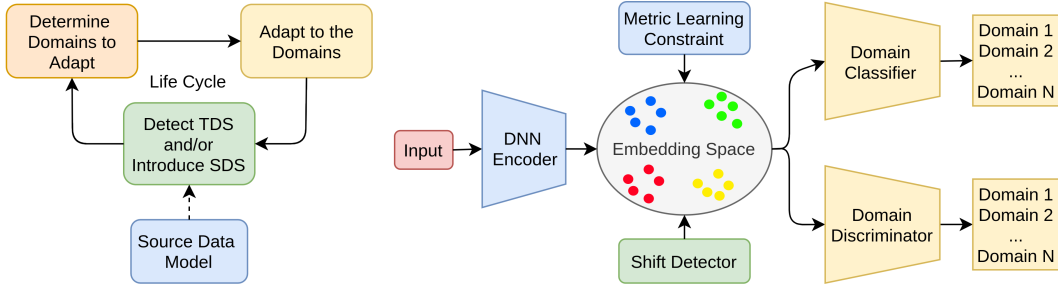
Figure 2: Left: high-level procedure overview. Right: detailed system model

3 Method

Overall Domain Shift Mitigation Procedure. Given the radar data in the source location, we utilize metric learning to detect the presence of any TDS and SDS and measure the shift malignancy. Using these metrics, we determine the domains that mandate domain adaptation and ignore the rest. Then, labeled or unlabeled data from the corresponding domains is collected, and an appropriate domain adaptation method is applied. The entire procedure shown in Fig. 2 is repeated throughout the life cycle of the deployed model. We explain each component of the proposed procedure as follows.

Metric Learning for the Base Model. Given the user recognition task at the core of our dataset, we started with a metric learning-based model [24]. Training networks with metric learning allows one to introduce new classes after training a base model, without having to train the model again, which translates to adding new user identities without the need for retraining. Another benefit of using metric learning is its theoretical underpinning, which is the cluster assumption. The cluster assumption states that the decision margins lie in low density or relatively unoccupied regions of the classification manifold. This assumption has previously been exploited in numerous domain adaptation techniques and is known to promote adaptation under certain constraints. Metric learning reaffirms this assumption by maximizing inter-class distances and minimizing intra-class distances.

Shift and Malignancy Detection. Upon further investigation, we found metric learning-based models are well suited for detecting domain shifts as well. Domain shift detection is the process of identifying shifts in the testing data. Unlike SDS, the presence of TDS is not always known. A few methods have been proposed for domain shift detection, but most either fail to detect shifts in an online fashion [26, 21] or base their predictions on softmax thresholding. However, it is well established that [22, 23] softmax classifiers are not suitable for detecting domain shifts, especially for out of domain data.

To address aforementioned challenge, we rely on thresholding embeddings generated by metric learning models. Softmax classifiers are optimized to make closed-set predictions [27], i.e., pick one of the classes present in the training dataset as the prediction. Such an assumption is too strong and grossly underestimates the likelihood of seeing unknown classes or out of distribution data during testing. Metric learning methods do not make such assumptions or, at the very least, not to the extent softmax based solutions do. Metric learning optimizes a metric on embeddings generated by a base classifier, which potentially allows the classifier to ignore the concept of closed set predictions, thereby overcoming overconfident predictions and calibration issues.

We exploit this premise to detect domain shifts. By maintaining dictionaries of class-wise mean embeddings and thresholding any new data point's distance from known embeddings, we can recognize shifts and their magnitude. Furthermore, such an approach does not require any specific offline training and can be deployed on online data sources. Not to mention, the detector will also be unsupervised, as no labeled data is required to detect a shift. The notion of detecting outliers or novel classes from embeddings is certainly not unheard of. We re-purpose it for detecting the presence and extent of domain shifts. Our experiments show that the approach agrees with labeled shift detectors.

Data-efficient Domain Shift Mitigation. Exploiting the proposed domain shift detector enables us to develop a data-efficient domain shift mitigation approach. Instead of blindly collecting a large

---

[1]Refer to the supplementary material for additional details

amount of labeled or unlabeled data from all spatial and temporal domains, we use the shift detector to measure the shift malignancy of each domain. Next, domains with insignificant shifts are discarded. Based on the cost of collecting labeled/unlabelled TD/SD data in different domains, a heuristic approach is utilized to determine and collect an appropriate type and amount of data. After this, the available domains are adapted to using an appropriate domain adaption strategy. A hybrid domain adaption scheme can also be applied, where supervised domain adaption is used to address certain shifts and unsupervised methods for the others.

For unsupervised domain adaptation, we experimented with three prominent methods. In the first method, we combine metric learning and a domain discriminator based on a gradient reversal layer (GRL) [28]. As anticipated, the approach worked well. The addition of metric learning boosted the model performance by a sizable amount. In the second approach, we replace the GRL with a Generative Adversarial Network [29], which can also significantly improve model generalization performance (which is detailed in section 4.6). Finally, we used a student-teacher based approach [30] but failed to procure any significant improvement. We conjecture that the failure is due to the presence of metric learning. The recent findings of [31] have a strong resemblance to combining student-teacher with metric learning. Since label smoothing behaves similar to metric learning, we speculate the presence of metric learning impedes the need for student-teacher approaches.

4 Experiments

All of our experiments utilized an 18-layer Resnet architecture [32]. The models were trained using an Adam optimizer [33] with an initial learning rate of $1e^{-3}$ and polynomial decay[1]. Additionally, we used Constrictive Annular Loss (CA-Loss) [34] for metric learning. It is a regularised variant of Additive Margin Softmax [35]. Nevertheless, we speculate that any metric learning based approach could be used in place of CA-Loss to draw similar conclusions as we do here.

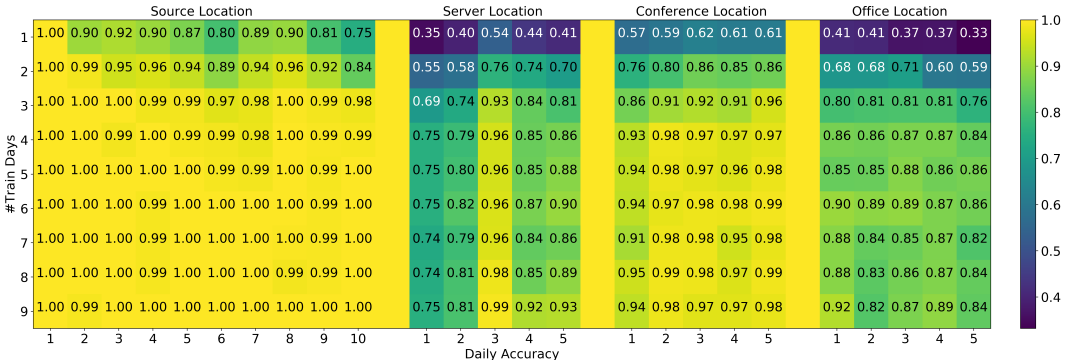## 4.1 Establishing the Presence of TDS and SDS

Figure 3: Results obtained from training only on source location data for varying number of days.

Our first experiment establishes the presence of TDS and SDS in our dataset. We trained a model on data only from the first day of our source location. Test accuracies of data from the remaining nine days of the source location and all target locations are reported in Fig. 3. The temporal degradation in the source and target domains is evident in the *first row* of Fig. 3.

The target domains also suffer from SDS apart from TDS, which is evident from the overall performance degradation in the target domains. This unequivocally establishes the presence of both TDS and SDS in our dataset. We also found a rather unexpected outcome from this experiment. There is a temporal degradation both in the source and target locations, but the deterioration does not strictly correlate with time progression. We conjecture that this behavior is a consequence of the malignancy of shifts. That is, not every shift is equally harmful. We further elaborate on this in Section 4.4.

5

4.2 Mitigating TDS and SDS

Mitigating SDS via TD data. We now show that it is not possible to entirely mitigate SDS by introducing source domain's temporal data alone and vice-versa. Limitations of improving SDS generalization by using TD data are evident in Fig. 3. We trained models from two, all the way up to nine days of labeled source location data, which is followed by an evaluation of daily accuracies. By introducing more temporal data, the source location generalization improves as time progresses. But as one might expect, we start to see diminishing returns in terms of source location accuracy as we increase the number of training days. Moreover, after introducing more source domain's temporal data, the target location performance also has diminishing returns so that each target location's performance is seldom on par with the source.
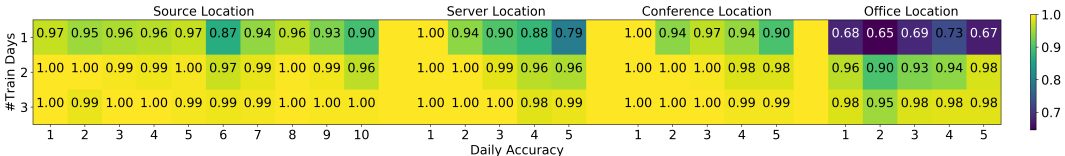

Figure 4: Results obtained from training on source, server, and conference location data for varying number of days. The data from office location is not used for training and only for testing

Mitigating TDS via SD data. We move on to establish the limitation of introducing more target/spatial domain data to alleviate TDS. We trained a model on a single day of labeled data from the source, conference, and server locations. From the *first row* of Fig. 4, it is clear that even when data from target domains is introduced, individual location's temporal shifts are not completely mitigated. However, we do find that the model performance in the office location improves when compared to introducing only temporal data. Since no data from the office was used in training, we interpret it as the true SDS generalization of the model. This implies that the primary benefits of training on data from a particular type of shift is confined to easing similar kinds of shifts.

4.3 Correlation of TDS and SDS

We now draw attention to the correlation between TDS and SDS. It is evident from our prior experiments that TD or SD data individually cannot mitigate both. Nevertheless, from Fig. 3 and Fig. 4, and the findings mentioned above, a correlation between the two is evident. It might be the reason no clear distinction has been established so far, but the implication of this is rather profound. Collecting TD and SD data might incur varying costs. *When the cost disparity is substantial, it is possible to trade one for the other to a certain extent*. This brings us to one of the most exciting findings of our paper. That is, *TDS and SDS have to be addressed jointly to bring substantial generalization improvement*. We trained models with both TD and SD data along with the initial source data. We find in Fig. 4 that when both shifts are addressed together, we attain the best generalization in both TD and SD. By introducing multiple (2 - 3) days of data from every location, we address TDS individually in each location. Introducing data from multiple target locations (server and conference rooms) addresses SDS. Moreover, such jointly trained model generalizes well to the unseen target domain (office).

4.4 Domain Importance

As mentioned in Section 4.1, we notice a disparity in the malignancy of each domain shift. From the plethora of research conducted on adversarial learning [36], it is evident that not every kind of data noise is equally harmful. Networks can overcome small amounts of noise in the data. This is true even when the models have not been explicitly trained with adversarial methods. Similarly, adversarial training with arbitrary data perturbations is not always a precursor for domain generalization.

We find this to be the case for domain adaptation as well. Not all available domains are worth adapting to. In our dataset, the conference location exhibits a domain shift, but it is not equally worth adapting to when compared to server room data. We found that by adapting to the server location alone, we can generalize to all other locations shown in Fig. 5. The server room can be considered more adversarial than the conference room. When we train a model on the source and server data, the information discrepancy amongst the two domains is substantial compared to the source and conference room
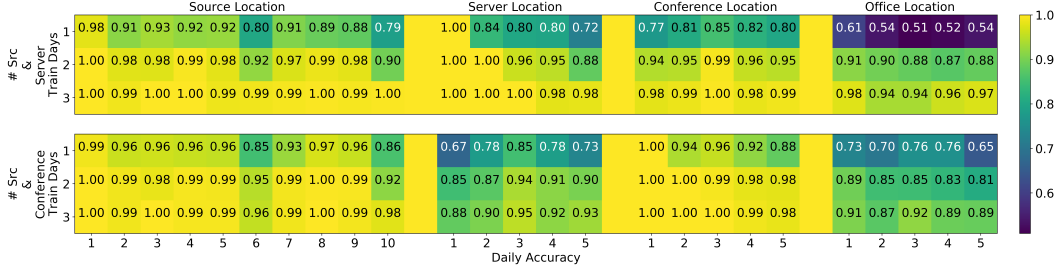
Figure 5: (Top) Results obtained from training on the source and server data for a varying number of days. (Bottom) Results obtained from training on the source and conference data for a varying number of days. Data from office location is for testing only

data. This results in spatial generalization with only half of the spatial target domain data. Admittedly, the model trained with source and server data is not as good as the model trained with source and conference data in every aspect. However, the overall generalization is better. Depending on the cost of data collection, it might be well worth the marginal loss in accuracy. This raises the question of how one can estimate the potential gains in generalization upon adapting to a domain, and we address this issue in Section 4.5.
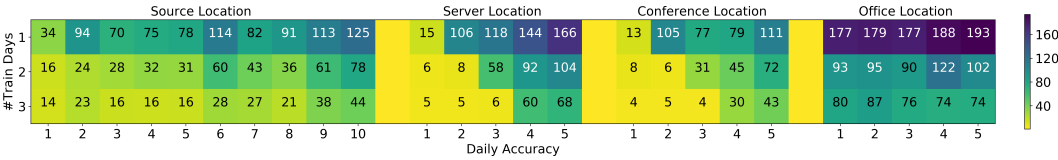
4.5 Shift Detection

Figure 6: Mean distance from known class centers obtained from the source, server, and conference data for a varying number of days.

As mentioned in Section 3, before addressing domain shifts, we must be aware of the existence of a shift in their data. So far, we used an offline approach for shift detection, where the accuracy of the model trained using ground truth labels is used to uncover the presence of domain shifts. We now propose to sidestep this issue. By thresholding the distance to known training data embeddings, we acquire results similar to our labeled, accuracy based observations. Not only does this need no ground truth labels, but it can also be deployed online without any specific training for the problem. We note that a separate holdout dataset will be required to tune the thresholds[2]. An added advantage of such an approach is its capability of detecting the malignancy of a shift. The accuracy-based approach does not necessarily convey the amount of shift introduced [22, 23]. We amend this issue by interpreting the magnitudes of the distance from class centers as an indicator of its malignancy. In Fig. 6, we present the results of Fig. 4, but with metric learning based shift metrics[3]. We find that the results strongly correlate with our hypothesis along with insights[3] into the malignancy of each domain.

4.6 Data-efficient Domain Shift Mitigation

All the experiments we reported so far were exclusively trained with labeled data. Nevertheless, that need not be the case. We utilized a domain discriminator trained as a GAN[2] to adapt to TDS and SDS. We trained a network with three days of labeled source location data and three days of unlabeled server location data, which makes TDS adaptation in the source domain supervised. However, TDS and SDS adaptation in the target domains is addressed in an unsupervised manner. Table 1 shows the results of our approach (Hybrid) and results from training a model without a discriminator on three days of labeled source, conference, and server data. We reduced the amount of labeled data

---

[2]Refer to the supplementary material for additional details

[3]Refer to the supplementary material for metric learning based results of all our experiments along with t-SNE embeddings, and elaboration of additional findings from the magnitude of shifts.

Table 1: Model Accuracies (%). (Labeled) The model trained on three days of labeled source, server, and conference data. (Hybrid) The model trained on three days of labeled source data and three days of unlabeled server data.

	Source-Temporal	Server-Test	Conference-Test	Office-Test
Labeled	99.59	98.85	99.37	97.51
Hybrid(Ours)	99.61	98.70	99.70	97.47

by utilizing unsupervised domain adaptation techniques. Furthermore, utilizing the lessons learned from our findings so far, the number of domains to adapt was reduced, i.e., we only use the server location data with the highest domain shifts, while ignoring data from conference and office locations. Furthermore, we conjecture that by incorporating few-shot learning methods, the source location labeled data could also be reduced. A single day of source location data currently consists of only 1000 labeled data samples, which hinders the performance as it is not sufficient to train a robust model. We addressed this issue by using labeled data from three days of the source location. However, using other techniques such as initialization with pre-trained model weights might also work well, thereby reducing the source location's labeled data.

5 Discussion

With all our experiments and findings being detailed, we now move on to discussing their implications.

Detecting Shifts. The first and foremost question is to detect shifts. By incorporating metric learning, one can detect shifts with thresholding. Merely being aware of the proper tools to detect a shift will be a great take away from our work. Since softmax based methods are invalid, our method is a straightforward alternative. However, a thorough investigation into how accurate these metrics are on a wider variety of datasets, with longer time horizons, is needed.

Handling Shifts. Upon detecting a shift, the question of how to handle it naturally arises, which entails figuring out how much and which data should be acquired. As mentioned in Section 4.4, we suggest one should take account of the malignancy and the type of shift in such decisions. Being aware of the malignancy of every shift allows one to determine which domains need to be adapted to, thereby significantly reducing any resources involved in the adaptation process. Shift malignancy could also be used to determine the amount and frequency of adaptation that is required. Finally, knowing the type of shift– SDS, TDS, or both–allows us to trade one for the other. The underpinning theory of this approach is the distinction of SDS and TDS, which requires us to address them properly. Our approach, as shown in Fig. 2, lays out a straightforward solution for realizing spatio-temporal generalization in a data-efficient manner.

Limitations. One limitation of our work is in addressing TDS explicitly on a daily basis. Currently, in our hybrid approach, we address TDS by aggregating data from multiple days in each location and treating it as an SDS. TDS in the source location is explicitly addressed because we use supervised classification. However, in the target domains, TDS generalization is a byproduct of explicit SDS adaptation. This can be improved upon by using a discriminator to classify the day label, which will address TDS explicitly in the target domains without any class labels. But, given the limited amount of data we collected in each domain, we did not address TDS with such an approach. It would certainly be interesting to study the impact of a daily data discriminator on a dataset with larger quantities of data and perhaps on a longer time horizon as well.

6 Related Works

[19] is one of the only known works which surveyed domain shifts and formally suggested a unifying taxonomy. Indeed, it does mention continuous shifts similar to the TDS observed in our data. However, they failed to include such shifts in their final taxonomy and did study the correlation shifts could have with one other.

[26] introduced one of the most prominent works on shift detection and correction. But, their work is limited to label shifts and relies on softmax predictions. We find their work similar to ours in their use

of a black box classifier. In fact, in [26]'s **Lemma 1**, if we replace the use of a confusion matrix with a pairwise similarity matrix for embeddings, their approach and bounds hold for our metric learning based black box predictor while circumventing the limitations of softmax classifiers.

7 Conclusion

Our work has established the need to differentiate spatial and temporal shifts. Contrary to some beliefs, we show that these drifts can be particularly harmful, which is especially true for radar-based gait recognition, in which the presence of the issue was not well known. The dataset we introduced has unveiled the previously unbeknown relation between TDS and SDS. We further uncovered the impact of adapting to one domain could have on other domains and introduced the promising yet straightforward avenue of methods that use metric learning to detect and estimate a shift's potential impact. Finally, we show our proposed life-cycle to tune and decide which domains to adapt to substantially optimized our adaptation efforts. Our overall approach to handling shifts establishes a promising layout for improving the generalization performance of radar-based biometric identification systems in a data-efficient manner.

Broader Impact

Even though our primary objective was to address radar-based gait detection, we believe it has significant ramifications for one of the major issues faced in deep learning deployment. Concretely, we suggest Implicit and Explicit Domain Shifts (IDS/EDS). IDSs are the implicit domain shifts that arise inherently due to some known or unknown dynamics of a domain, and EDSs are the explicit domain shifts, introduced only for a given task or application. Under this generalized definition, the TDS we observed in our dataset can be considered to be a form of IDS that happens to correlate with time.

Similarly, SDS is one type of EDS introduced in our dataset. Such a generalized nomenclature implies the possibility of introducing multiple EDSs depending on the type of labels present. Furthermore, various other IDSs could be uncovered and addressed depending on the dynamics of a domain. This distinction could redefine the current pipeline adopted by most machine learning practitioners for deploying models. It would allow for smarter data collection procedures and better utilization of resources. A large scale study to validate the generalization of our suggested taxonomy is required, and it also raises the need to reevaluate the current taxonomy of domain shifts.

Furthermore, we never addressed the issue of model retraining. Currently, we assumed that all the data is always available. However, it is unrealistic to maintain the entire dataset for each retraining cycle as the dataset will keep increasing in size by design in our current workflow. Therefore, it is essential to consider incorporating continual or federated learning based approaches to enable training only on new data while avoiding catastrophic forgetting.

One cause for concern of our work is the development of radar-based user recognition. Unlike most biometric solutions, radar can not only recognize people but can also be used to determine one's actions. Moreover, depending on the type of radar used, they can see through walls to a varying extent, which we believe could be extremely detrimental if used by an adversary. Proper oversight and regulations might be in order when deploying such devices. However, the device also has numerous advantages. Unlike image-based methods, biases such as race, ethnicity, and gender are not present. To the best of our knowledge, the only issue was data, which we addressed in our work here.

Acknowledgments and Disclosure of Funding

References

[1] Wei Wang, Alex X. Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 363 – 373. ACM, 2016.

[2] Akarsh Pokkunuru, Kalvik Jakkala, Arupjyoti Bhuyan, Pu Wang, and Zhi Sun. Neuralwave: Gait-based user identification through commodity wifi and deep learning. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 758–765. IEEE, 2018.

[3] Kalvik Jakkala, Arupjyoti Bhuya, Zhi Sun, Pu Wang, and Zhuo Cheng. Deep csi learning for gait biometric sensing and recognition. *arXiv preprint arXiv:1902.02300*, 2019.

[4] Prabhu Janakaraj, Kalvik Jakkala, Arupjyoti Bhuyan, Zhi Sun, Pu Wang, and Minwoo Lee. Star: Simultaneous tracking and recognition through millimeter waves and deep learning. In *2019 12th IFIP Wireless and Mobile Networking Conference (WMNC)*, pages 211–218. IEEE, 2019.

[5] Jacopo Pegoraro, Francesca Meneghello, and Michele Rossi. Multi-person continuous tracking and identification from mm-wave micro-doppler signatures. *arXiv preprint arXiv:2003.03571*, 2020.

[6] Xuejun Huang, Jinshan Ding, Dongxing Liang, and Liwu Wen. Multi-person recognition using separated micro-doppler signatures. *IEEE Sensors Journal*, 2020.

[7] Yue Lang, Qing Wang, Yang Yang, Chunping Hou, Yuan He, and Jinchen Xu. Person identification with limited training data using radar micro-doppler signatures. *Microwave and Optical Technology Letters*, 2019.

[8] Sherif Abdulatif, Fady Aziz, Karim Armanious, Bernhard Kleiner, Bin Yang, and Urs Schneider. A study of human body characteristics effect on micro-doppler-based person identification using deep learning. *arXiv preprint arXiv:1811.07173*, 2018.

[9] Xingshuai Qiao, Tao Shan, and Ran Tao. Human identification based on radar micro-doppler signatures separation. *Electronics Letters*, 56(4):195–196, 2020.

[10] Yue Lang, Qing Wang, Yang Yang, Chunping Hou, Haiping Liu, and Yuan He. Joint motion classification and person identification via multitask learning for smart homes. *IEEE Internet of Things Journal*, 6(6):9596–9605, 2019.

[11] Muhammad Shahzad and Shaohu Zhang. Augmenting user identification with wifi based gesture recognition. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 2018.

[12] Kenshi Saho, Keisuke Inuzuka, and Keitaro Shioiri. Person identification based on micro-doppler signatures of sit-to-stand and stand-to-sit movements using a convolutional neural network. *IEEE Sensors Letters*, 4(3), 2020.

[13] Jiayao Tan, Xiaoliang Wang, Cam-Tu Nguyen, and Yu Shi. Silentkey: A new authentication framework through ultrasonic-based lip reading. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1), 2018.

[14] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Linghe Kong, and Minglu Li. Lip reading-based user authentication through acoustic sensing on smartphones. *IEEE/ACM Transactions on Networking*, 27(1):447–460, 2019.

[15] Lin Feng, Chen Song, Yan Zhuang, Wenyao Xu, Changzhi Li, and Kui Ren. Cardiac scan: A non-contact and continuous heart-based user authentication system. In *ACM Mobicom*, pages 315 – 328. ACM, 2017.

[16] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. Cardiac scan: A non-contact and continuous heart-based user authentication system. In *Int'l Conf. Mobile Systems, Applications, and Services (MobiSys)*, page 278 – 291, 2017.

[17] Qinyi Xu, Yan Chen, BeiBei Wang, and K. J. Ray Liu. Radio biometrics: Human recognition through a wall. *IEEE Transactions on Information Forensics and Security*, 12(5):1141–1155, 2017.

[18] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Fredo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics*, 34(6), 2015.

[19] Jose G Moreno-Torres, Troy Raeder, RocíO Alaiz-RodríGuez, Nitesh V Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern recognition*, 45(1):521–530, 2012.

[20] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

[21] Stephan Rabanser, Stephan Günnemann, and Zachary Lipton. Failing loudly: an empirical study of methods for detecting dataset shift. In *Advances in Neural Information Processing Systems*, pages 1394–1406, 2019.

[22] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[23] Jasper Snoek, Yaniv Ovadia, Emily Fertig, Balaji Lakshminarayanan, Sebastian Nowozin, D Sculley, Joshua Dillon, Jie Ren, and Zachary Nado. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, pages 13969–13980, 2019.

[24] Mahmut Kaya and Hasan Şakir Bilge. Deep metric learning: a survey. *Symmetry*, 11(9):1066, 2019.

[25] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[26] Zachary C Lipton, Yu-Xiang Wang, and Alex Smola. Detecting and correcting for label shift with black box predictors. *arXiv preprint arXiv:1802.03916*, 2018.

[27] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[28] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[29] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[30] Zhong Meng, Jinyu Li, Yifan Gong, and Biing-Hwang Juang. Adversarial teacher-student learning for unsupervised domain adaptation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5949–5953. IEEE, 2018.

[31] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. When does label smoothing help? In *Advances in Neural Information Processing Systems*, pages 4696–4705, 2019.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[33] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[34] Jun-Bo Liu, Ya-Ping Huang, Qi Zou, and Sheng-Chun Wang. Learning representative features via constrictive annular loss for image classification. *Applied Intelligence*, 49(8):3082–3092, 2019.

[35] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7):926–930, 2018.

[36] Anirban Chakraborty, Manaar Alam, Vishal Dey, Anupam Chattopadhyay, and Debdeep Mukhopadhyay. Adversarial attacks and defences: A survey. *arXiv preprint arXiv:1810.00069*, 2018.

Supplementary Material of Spatio-Temporal Domain Adaptation for Gait Based User Recognition from Radar Data

Anonymous Author(s)
Affiliation
Address
`email`

Contents

1 Dataset

Our dataset was collected using a TI IWR1642 mmWave sensor [1]. The radar chirp configuration used in [2] was utilized to collect data from 4 different locations. The primary location, which was used as the source domain, was a research lab with cubicles. Each of the ten subjects was asked to walk back and forth in the room with a radar pointed at them. The subject's path was pre-determined and was always perpendicular to the radar signal's propagation direction. Moreover, each day of a user's data contains 100 data samples. 4-seconds of walking data was considered as an individual data sample.

Apart from the source domain, three other locations were used to collect data. These locations were used as the target domains. The locations consisted of a server, conference, and office room. Of the three locations, the office location was reserved for the test set, and none of the experiments utilized the office data for training. The radar device was deployed at all three target locations, and the walking path setup is similar to the source domain. Furthermore, unlike the source domain, the users generated only 50 data samples on each day at each target location. In addition, the data from all four locations collected each day was not necessarily from the same day. The source location's day one for a user could be today, while the data from the same user at each target location could be collected on three different days. Similarly, all ten users were not required to generate data on the same day for a given location. This allowed us to simulate a real-world setup, such as an office with new employs generating data, upon being hired or given access to a restricted space.

2 Model

All our experiments used an 18-layer Resnet as defined in [3]. The last fully-connected layer was replaced with a Multi-Layered Perceptron (MLP) with 128 hidden units and a final layer with hidden units equal to the number of classes (10). The 128 dimensional fully connected layer's output was used as an embedding layer. Moreover, a Constrictive Regularizer [4] was utilized as the activity regularizer. Moreover, the last fully connected layer also used a Constrictive Regularizer for the kernel weights instead of the activations. Additionally, given the definition of Additive Margin Loss [5], we do not have any bias variables in the last layer. Finally, the logits were activated following the definition in [5].

In all our supervised experiments, data from each available domain was shuffled together and sampled in batches. This was done irrespective of the domain each sample belonged to. For our unsupervised experiments, data from each available domain was sampled separately, which allowed us to generate embeddings for data from each domain. We generate domain predictions with an MLP with 128 hidden units followed by another layer with hidden units equal to the number of domains being used for training. The domain classifier MLP was either prepended by a gradient reversal layer (GRL) [6] or treated as a discriminator in a Generative Adversarial Network (GAN) [7].

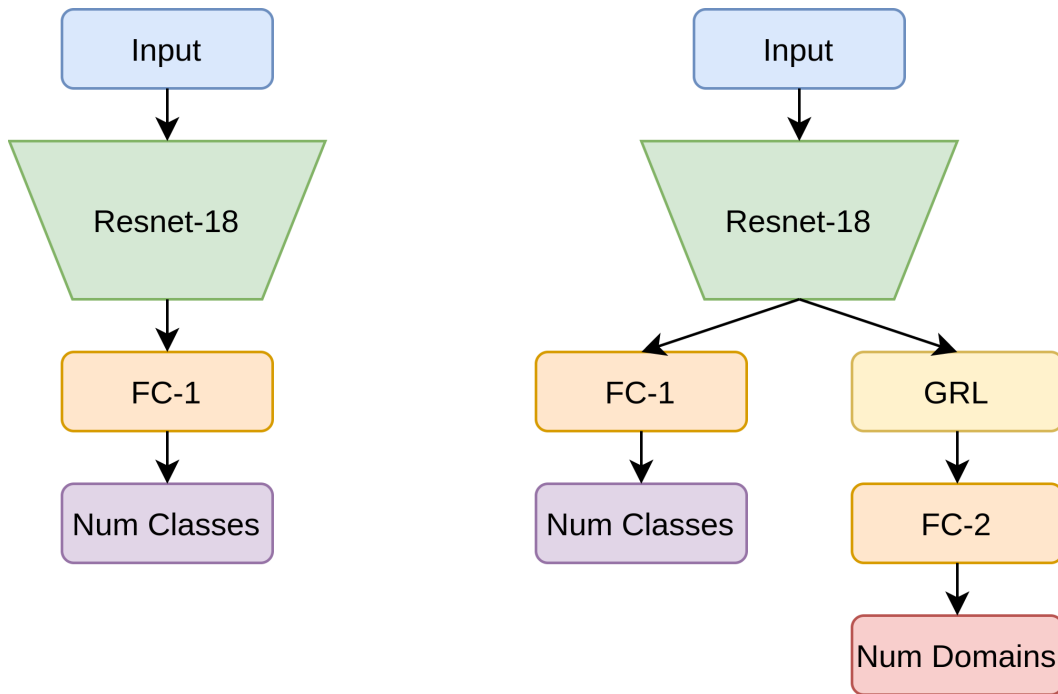Figure 1: (Left) Supervised model architecture. (Right) Unsupervised model architecture.

3 Experiment Embeddings

We present the t-SNE embeddings for all our experiments in the section.

Figure 2: Embeddings from the model trained using 1 to 6 days of source-only data. The number in brackets indicates the day indices. For example, [1 - 6] means the data from day 1 to day 6. Orange triangle clusters represent the training data (Source Location's Day 1), where each cluster represents a distinct human subject. Blue circles represent each domain's testing data, where the data points were sampled from all the classes of a given domain.
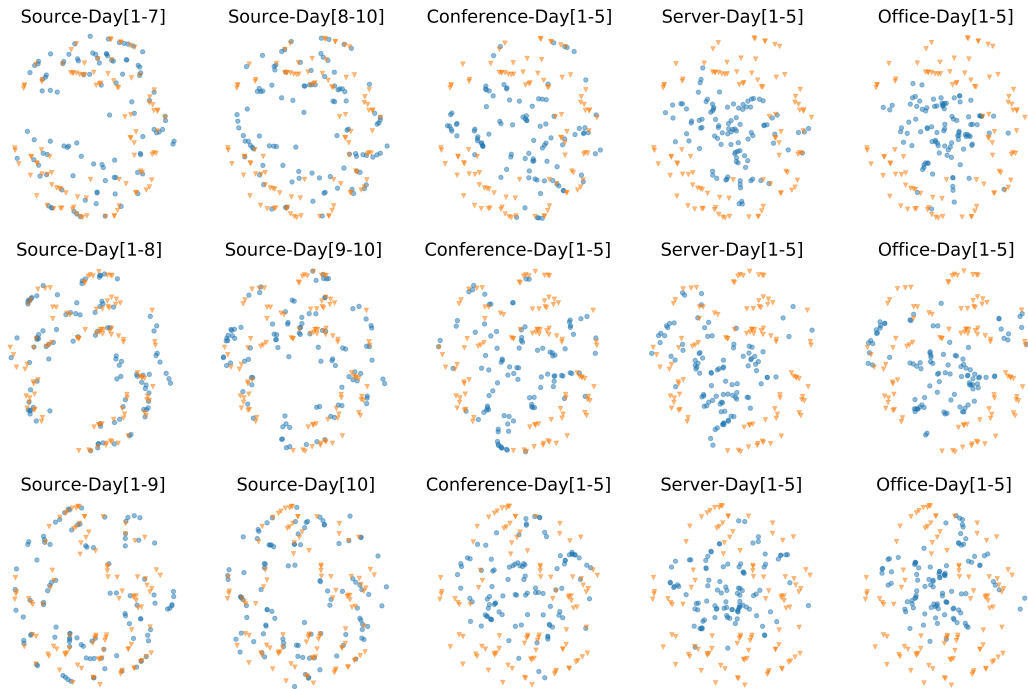
Figure 3: Embeddings from the model trained using 7 to 9 days of source-only data
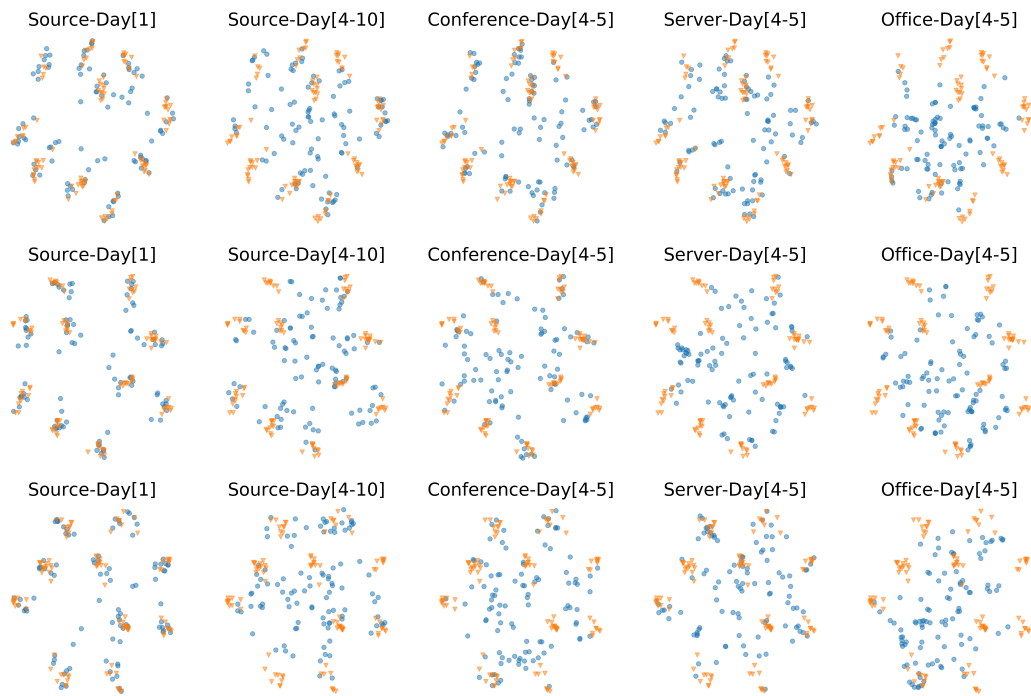

Figure 4: Embeddings from model trained on source and target domain data. (Top) model trained one day of source, server and conference data, (Middle) model trained via one day of source and conference data, (Bottom) model trained via one day of source and server data
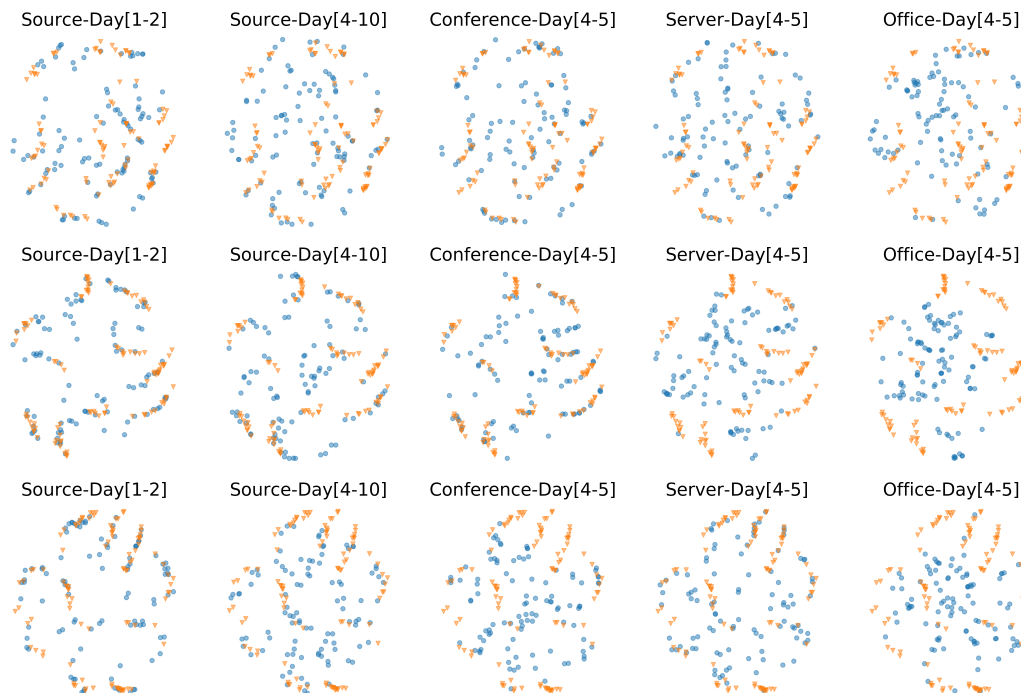
Source-Day[1-2] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Source-Day[1-2] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Source-Day[1-2] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Figure 5: Embeddings from model trained on source and target domain data. (Top) model trained via two days of source, server and conference data, (Middle) model trained via two days of source and conference data, (Bottom) model trained via two days of source and server data
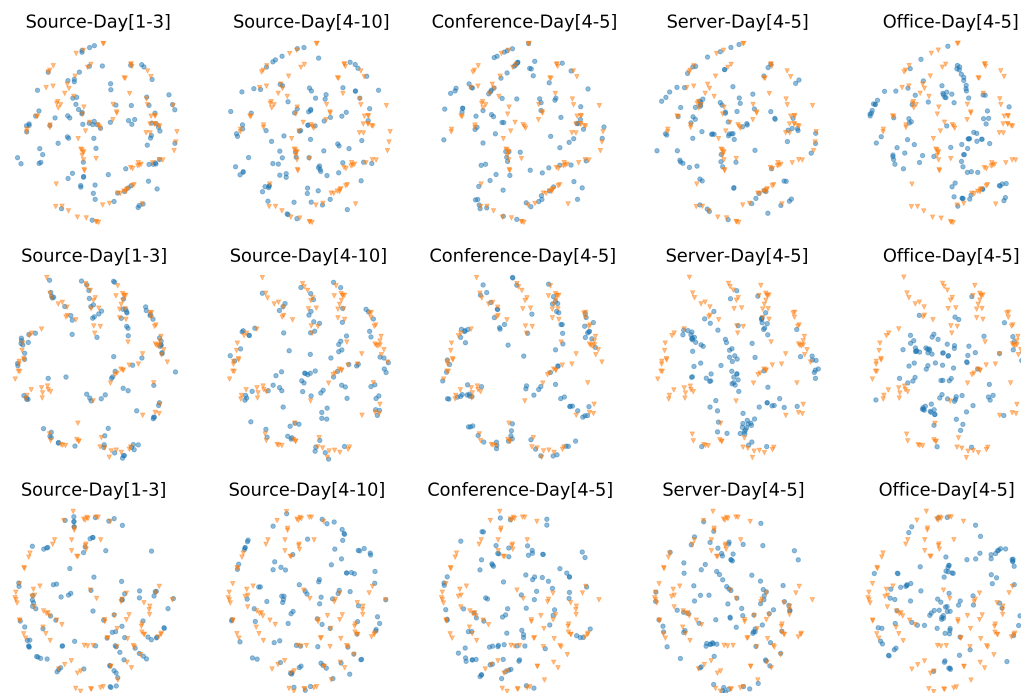
Source-Day[1-3] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Source-Day[1-3] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Source-Day[1-3] Source-Day[4-10] Conference-Day[4-5] Server-Day[4-5] Office-Day[4-5]

Figure 6: Embeddings from model trained on source and target domain data. (Top) model trained via three days of source, server and conference data, (Middle) model trained via three days of source and conference data, (Bottom) model trained via three days of source and server data

Figure 7: Embeddings from model trained on three days of labeled source and three days of unlabeled server domain data. The unsupervised data was trained using a GAN based approach.

4 Experiment Shift Metrics

Here we present the accuracies obtained from all our experiments alongside the corresponding metric learning based distance metrics. We find that not only does the distance metrics agree with the accuracy based findings, it also makes it more evident. Indeed we see an apparent temporal degradation in Fig. 8 in the source domain. Moreover, introducing more training data from the source domain alleviates the SDS to a certain extent, but it also biases the model more towards the source domain in some cases.
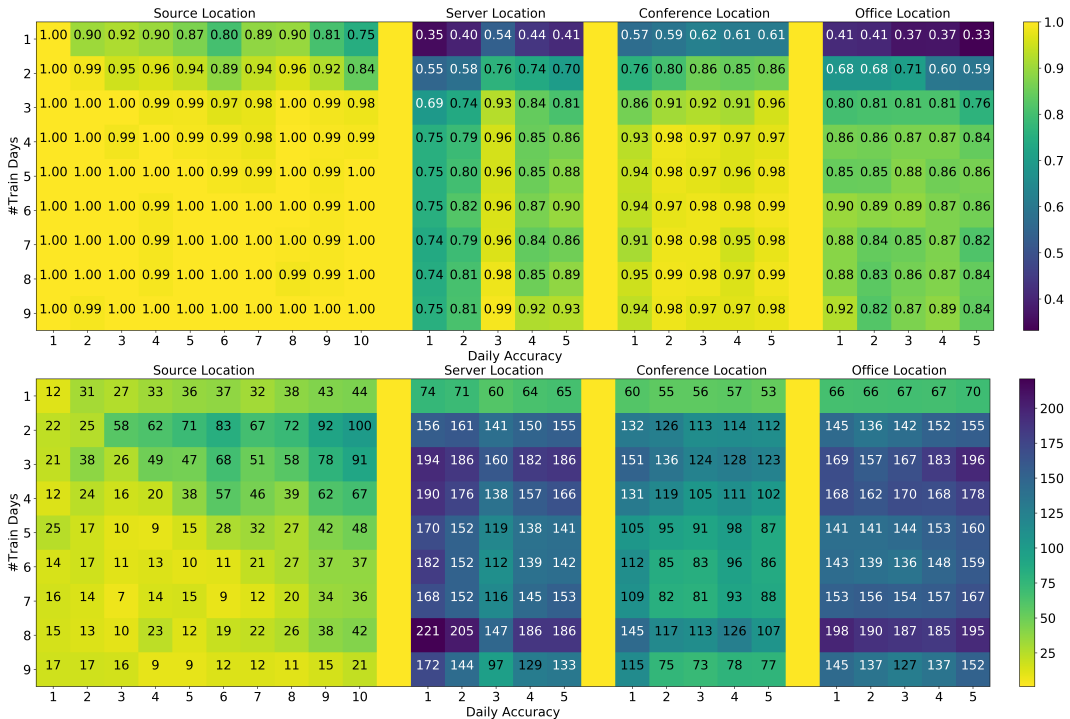

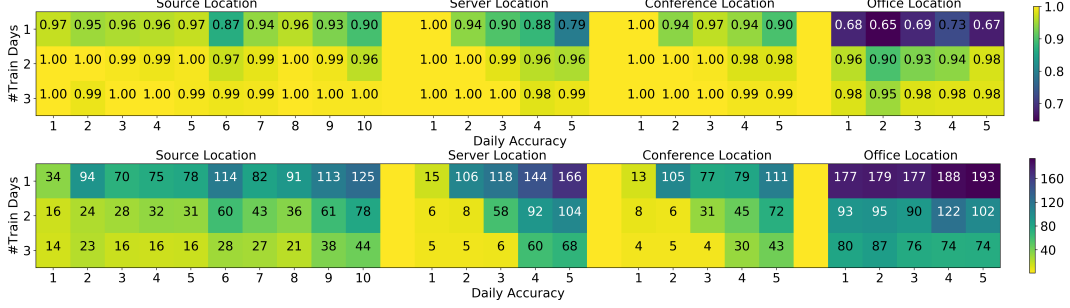Figure 8: Results obtained from training only on source location data for varying number of days.

6

Figure 9: Results obtained from training on source, server, and conference location data for varying number of days. The data from office location is not used for training and only for testing
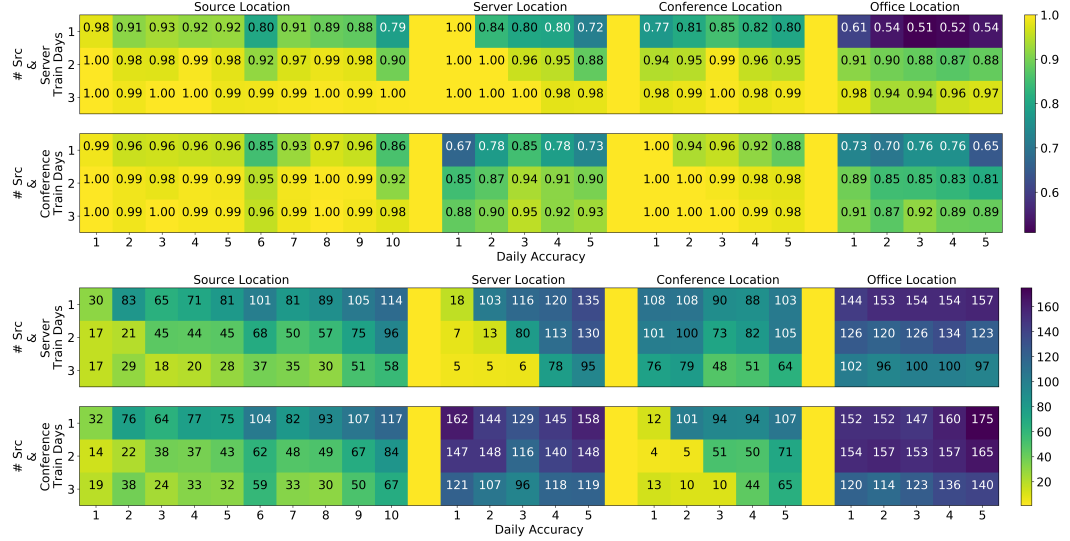

Figure 10: (Top) Results obtained from training on the source and server data for a varying number of days. (Bottom) Results obtained from training on the source and conference data for a varying number of days. Data from office location is for testing only

5 Metric Learning Based Shift and Malignancy Detection Algorithm

We utilize Algorithm 1 to compute thresholds used for detecting outliers. In our experiments, L2 distance was used along with a margin m of 10. Upon computing class thresholds $Threshold$ and trained model M, any new data samples that were observed, were subject to class-wise thresholding $DistanceMetric$ between the known class centers C^{Train} and the generated embedding for the new data point.

Algorithm 1 Algorithm to Compute Class-Wise Thresholds

Input: Training data X_{Train}, validation data X_{Val}, number of classes n, $DistanceMetric$ (L2 or cosine distance), hyperparameter safety margin m.

Output: Class-wise threshold $\{Threshold_j | j = 1, 2, ..., n\}$, M, C^{Train}

1: Train model M on X_{Train} with metric learning
2: Generate class-wise embeddings $\{C_j^{Train} | j = 1, 2, ..., n\}$ from X_{Train} and M
3: Generate class-wise embeddings $\{C_j^{Val} | j = 1, 2, ..., n\}$ from X_{Val} and M
4: Compute $\{Threshold_j | j = 1, 2, ..., n\} = DistanceMetric\,(C_j^{Val} - C_j^{Train} | j = 1, 2, ..., n) +$
m

7

6 GRL, GAN and Student-Teacher Methods for Unsupervised Domain Adaptation

We addressed unsupervised domain adaptation using a GAN based approach. However, we also tried a few other methods. The most notable of which were, based on GRL and Student-Teacher(ST) based methods. We report the accuracies obtained from each method in Table 1. These models were trained using three days of labeled source location data and three days of unlabeled server location data. Apart from the three methods mentioned above, we also tried introducing several other methods. Specifically, in combination with the model with GRL and AMCA loss, we tried each of the methods in [8, 9, 10, 11, 12, 13] including mixup/cutmix regularization, adversarial training, pseudo-labels, multi-adversarial and cyclic adversarial domain adaptations. However, none of these methods amounted to any significant results. We are not sure if these methods failed as a consequence of being used along with metric learning or because the data are in the format of spectrograms. We leave investigations into the matter for our future work.

Table 1: Model Accuracies (%). (GRL) The model trained with GRL based UDA. (GAN) The model trained with GAN based UDA. (ST) The model trained with student-teacher based UDA

	Source-Temporal	Server-Test	Conference-Test	Office-Test
GRL	99.37	97.80	99.40	94.96
GAN	99.61	98.70	99.70	97.47
ST	97.77	86.99	95.50	85.07

References

[1] Iwr1642 single-chip 76-ghz to 81-ghz mmwave sensor integrating dsp and mcu evaluation module iwr1642boost this product has been released to the market and is available for purchase. for some products, newer alternatives may be available.

[2] Prabhu Janakaraj, Kalvik Jakkala, Arupjyoti Bhuyan, Zhi Sun, Pu Wang, and Minwoo Lee. Star: Simultaneous tracking and recognition through millimeter waves and deep learning. In *2019 12th IFIP Wireless and Mobile Networking Conference (WMNC)*, pages 211–218. IEEE, 2019.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] Jun-Bo Liu, Ya-Ping Huang, Qi Zou, and Sheng-Chun Wang. Learning representative features via constrictive annular loss for image classification. *Applied Intelligence*, 49(8):3082–3092, 2019.

[5] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7):926–930, 2018.

[6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[9] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6023–6032, 2019.

[10] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[11] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 2, 2013.

[12] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, and Richard Socher. Augmented cyclic adversarial learning for low resource domain adaptation. *arXiv preprint arXiv:1807.00374*, 2018.