Sensor signal processing

SCLAiR: Supervised Contrastive Learning for User and Device Independent Airwriting Recognition

Ayush Tripathi[1], Arnab Kumar Mondal[2], Lalan Kumar[1,3*], and Prathosh AP[4]

[1]*Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India*
[2]*School of Information Technology, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India*
[3]*Bharti School of Telecommunication, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India*
[4]*Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, Karnataka 560012, India*
[*]*Member, IEEE*

Manuscript received November 18, 2021; revised December 18, 2021; accepted December 26, 2021. Date of publication December 31, 2021; date of current version February 3, 2022.

Abstract—Airwriting recognition is the problem of identifying letters written in free space with finger movement. It is essentially a specialized case of gesture recognition, wherein the vocabulary of gestures corresponds to letters as in a particular language. With the wide adoption of smart wearables in the general population, airwriting recognition using motion sensors from a smart band can be used as a medium of user input for applications in human–computer interaction. There has been limited work in the recognition of in-air trajectories using motion sensors, and the performance of the techniques in the case when the device used to record signals is changed has not been explored hitherto. Motivated by these, a new paradigm for device and user-independent airwriting recognition based on supervised contrastive learning is proposed. A two-stage classification strategy is employed, the first of which involves training an encoder network with supervised contrastive loss. In the subsequent stage, a classification head is trained with the encoder weights kept frozen. The efficacy of the proposed method is demonstrated through experiments on a publicly available dataset and also with a dataset recorded in our lab using a different device. Experiments have been performed in both supervised and unsupervised settings and compared against several state-of-the-art domain adaptation techniques. Data and the code for our implementation will be made available at https://github.com/ayushayt/SCLAiR.

Index Terms—Sensor signal processing, airwriting, domain adaptation, smart band, supervised contrastive learning, wearables.

I. INTRODUCTION

A. Background

Airwriting may be defined as the process of writing letters in free space using unrestricted finger movements [1], [2]. It can be used to provide a user with a fast and touchless input option which can be employed for applications in human–computer interaction [3]. The recognition of writing from motion sensors has garnered attention over the past few years and numerous algorithms have been proposed for the task [4]–[10]. The studies involving the use of motion sensors can be broadly divided into two categories, the first involving the use of dedicated devices such as a wearable glove [11], Wii remote [12]–[14], smartphone [15] and vision-based [16]. This approach, however, involves the user to carry an extra physical device which may be cumbersome for the users. To mitigate this, the second category of approaches aim to recognize writing movements by using wearable devices such as a ring worn on the index finger [17] and smart bands [18].

B. Related Work

There have been several attempts made at recognizing gestures of the palm by using wrist-worn devices [19], [20]. However, most of the work around this area has been focused on the setting, wherein a stable, flat surface is used during the writing process [21]–[23]. The presence of a flat surface for writing leads to stabilization of the hand, thereby decreasing noise and also providing a degree of feedback to the user. Close to the current work, in [24], the authors explored the problem of airwriting recognition using a wrist-worn device. The study involved a single participant and dynamic time warping (DTW) was used as a distance measure for classification, thereby making the study user dependent. In [18], the authors have proposed a convolutional neural network-based user independent framework in addition to a DTW-based user dependent method for recognizing airwritten English uppercase alphabets. While all these methods are shown effective in airwriting recognition, they are limited in terms of the user and device dependent nature of the studies performed hitherto. Such a system has the drawback of not generalizing to large-scale population due to intersubject variability in writing the same character. Furthermore, difference in the acquisition sensor characteristics across different devices may prove to be detrimental for the performance of airwriting recognition systems. Motivated by this, the current study explores a new paradigm for both user and device independent recognition of airwritten alphabets using signals obtained from 3-axis accelerometer and gyroscope sensors.

C. Objectives and Contributions

In this letter, we explore a supervised contrastive learning-based architecture [25] for airwriting recognition. The main idea behind the approach is that the hidden representations of signals for the same alphabet are expected to be similar to each other, while the representation corresponding to signals of different alphabets will be fairly different. Given this intuition, a 2-stage classification approach is

Corresponding author: Ayush Tripathi (e-mail: eez208477@ee.iitd.ac.in).
Associate Editor: Leyre Azpilicueta.
Digital Object Identifier 10.1109/LSENS.2021.3139473

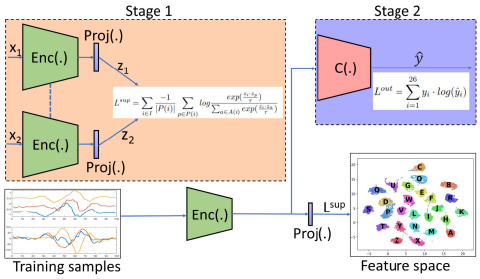


Fig. 1. Block diagram depicting our method. In the first stage an encoder with projection head based on deep network (called Enc(·) and Proj(·), respectively) is trained using supervised contrastive loss. Next, the projection head is discarded and a linear classifier ($C(\cdot)$) is cascaded on top of the feature learner and trained with cross entropy loss. The dashed lines indicate that the encoders have shared weights.

adopted wherein, in the first stage, an encoder (with a projection head) is trained with using the contrastive loss. Subsequently, the projection head is discarded and a classification head is trained, while keeping the encoder weights fixed. We validate the efficacy of the proposed approach by performing leave-one-subject-out (LOSO) experiments on two different datasets: A publicly available dataset [18] (source dataset) and a dataset collected in the lab (target dataset). We also analyze the performance of the approach in two different settings: 1) an unsupervised setting in which the model is trained solely using the source domain data and evaluated on the target domain samples, and 2) a supervised setting in which the encoder is trained using source domain samples and the classifier head is fine-tuned using target domain samples. We demonstrate that the supervised contrastive learning-based approach leads to improved recognition accuracy on the target dataset, compared to other domain adaptation techniques, namely, DANN [26], DRCN [27], and DeepJDOT [28] in both supervised and unsupervised settings.

II. PROPOSED METHOD

The proposed framework consists of two stages, an encoder network trained using contrastive loss and a subsequent classifier head trained using the cross-entropy loss. Fig. 1 depicts the overall network.

A. Problem Description

A multivariate time-series recorded using 3-axis accelerometer and gyroscope of an inertial measurement unit (IMU) placed on the wrist of the dominant hand of a human subject, while writing uppercase English alphabets forms the input for the method. The signal is recorded over three axes (X, Y, and Z), resulting in a total of six time series for each sample. Suppose we have a batch of N sample and label pairs represented as $\{x_k, y_k\}_{k=1,2,..N}$, where $x_k \in \mathcal{X}$ is the input time series corresponding to the alphabet $y_k \in \{A, B, ..Z\}$. The encoder network, $\text{Enc}(.) : \mathcal{X} \rightarrow R^{D_E}$ maps the input x to a D_E dimensional representation vector $r = \text{Enc}(x)$. Subsequently, the vector r is normalized to unit hypersphere in R^{D_E}. There can be various choices of the encoder architecture that can be adopted without any constraints. Subsequently, the projection head maps the encoded representation r to a vector $z = \text{Proj}(r) \in R^{D_P}$. This comprises of the first stage of the training and involves learning the parameters of the encoder along with the projection head, which is done by minimizing the supervised contrastive loss. It is to be noted that at the end of this stage of training, the projection network is discarded, while the encoder weights are retained.

The second stage is the classifier $C(.)$, which is a mapping from the latent space obtained from the encoder (r) to different alphabets. The predicted output is, therefore, given by $\hat{y} = C(r) = \sigma(W^T r)$. Here, W is the classifier weight matrix, and $\sigma(.)$ is the softmax activation function. Since the projection head is discarded after Stage 1 of the training process, at inference time, the overall model has the same number of parameters as a standard classification model trained using cross-entropy loss.

B. Supervised Contrastive Loss

Let $i \in I \equiv \{1, 2, ..N\}$ be the index of an arbitrary sample from a batch of the training dataset, $P(i)$ be the set of indices of samples belonging to the same class as the ith sample in the batch and $A(i) \equiv I \setminus \{i\}$ be the set of indices of all samples from the batch other than the ith sample, and $z_i = \text{Proj}(\text{Enc}(x_i))$ be the output of the projector head for input sample x_i. Supervised contrastive loss [25] is defined as

$$L^{\text{sup}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\frac{z_i \cdot z_p}{\tau})}{\sum_{a \in A(i)} \exp(\frac{z_i \cdot z_a}{\tau})} \tag{1}$$

In the above equation, τ is a scalar temperature parameter, and $x \cdot y$ represents the inner product between vectors x and y. Minimizing the aforementioned loss function encourages the encoder to learn closely aligned representation vectors for samples belonging to the same class by using the inner product as a measure of similarity between samples. The gradient of the loss function with respect to z_i can be computed as

$$\frac{\delta L_i^{\text{sup}}}{\delta z_i} = \frac{1}{\tau} \left\{ \sum_{p \in P(i)} z_p \left(P_{ip} - \frac{1}{|P(i)|} \right) + \sum_{n \in N(i)} z_n P_{in} \right\}. \tag{2}$$

Here, $N(i)$ is the set of indices of samples not belonging to the same class as the ith sample in the batch, and $P_{ix} = \exp(z_i \cdot z_x/\tau)/\sum_{a \in A(i)} \exp(z_i \cdot z_a/\tau)$. It can be shown that easier positive and negative pairs have small contribution toward the gradient compared to hard positive and negative pairs that have a larger contribution, giving rise to intrinsic hard positive/negative mining.

C. Model Architecture

We explore the following different architectures for the encoder block. Ablations on the hyperparameters for different encoder architectures can be seen in Figs. 5(b) and 6 (Supplementary Material).

1) *1DCNN*: The architecture comprises of four convolutional layers interleaved by a pooling layer and followed by a global average pooling layer. For the first couple of convolution layers, the number of filters is set to be 100, while in the later stage at 160, while the filter length for each of the layers is 10.
2) *LSTM and BiLSTM*: Here, vanilla LSTM and BiLSTM with 256 number of units are used.
3) *1DCNN-LSTM/BiLSTM*: Here, the input signal is passed through a couple of 1-D convolutional layers, followed by a pooling layer, and the extracted features are then fed to a 256-unit LSTM/BiLSTM layer. The specifications of the convolutional layers are same as that of the first layers as used in the 1DCNN-based encoder.

The projection head is taken to be a single layer comprising of 128 neurons, activated by the ReLU activation function. The parameters of the encoder and the projection head are learned by minimizing the supervised contrastive loss using the Adam optimizer. Further, the final classifier network is a single fully connected layer comprising 26 neurons and activated by softmax activation function with 50% dropout. The classification network is trained using the usual cross entropy Loss. The effect of varying the projection head dimension and the temperature parameter τ is presented in Fig. 5(a) (cf. Supplementary Material).

III. EXPERIMENTS AND RESULTS

A. Datasets

1) *Source Dataset*: We use the publicly available dataset from [18] that consists of recordings obtained from 55 subjects (28 female, 27 male; 46 right-handed, nine left-handed), while writing English uppercase letters (15 times). A dedicated app was built for the purpose of data collection which the users operated by using their nondominant hand, while the signals were recorded from 3-axis accelerometer and gyroscope of a Microsoft band 2 worn on the dominant hand. The signals were recorded at the maximum sampling rate of 62 Hz supported by the smart band.
2) *Target Dataset*: A dataset consisting of 3-axis accelerometer and gyroscope recordings of 20 subjects (11 male, nine female) was collected in a lab setting with the consent of the participants. A Noraxon Ultium EMG sensor (having an internal IMU) [29] was placed on the wrist of the dominant hand of the participant who was then asked to write the English uppercase letters (repeated 10 times). A user interface built using the Tkinter module in Python was used for providing the participant with visual feedback for the letter to be written and also for automating the annotation process of the recordings. In Fig. 3 (Supplementary Material), we present the sample data collection setup. The signals were recorded at a sampling rate of 200 Hz as supported by the IMU sensor and later downsampled to 62 Hz in order to match to that of the source dataset. Our data will be made publicly available.
   Another target dataset comprising of recordings from 10 subjects was recorded using a different device with sampling rate 400 Hz (later downsampled to 62 Hz) with the same experimental setup as described above.

B. Experimental Details

The recorded signals are obtained from different users writing at different speed and character sizes. Therefore, the following preprocessing steps were employed.

1) The samples are fixed to same length by padding zeros if the length of the sample is less than L (taken to be 155 as in [18]), while discarding the extra samples otherwise.
2) To the fixed length samples, we apply the usual Z-score normalization for each of the six individual signals.

The first set of experiments involves LOSO validation on both the source and target datasets to avoid the user bias. The training data in each fold is split into a training set and a validation set having 80:20 ratio. A minibatch training process with a batch size of 32 is employed and early stopping with a patience of five epochs.

In the next set of experiments, cross entropy loss-based classifier and supervised contrastive loss-based classifier have been compared in both supervised and unsupervised settings. In the unsupervised setting, the model is trained on the source dataset and evaluated on the target dataset. In the supervised case, the classifier head is fine-tuned using the labeled target dataset in a LOSO fashion. The proposed approach has been compared with different domain adaptation techniques, by using the public implementations of the algorithms: DANN, DRCN, and DeepJDOT.

C. Results and Comparison

Table 1 lists the performance of different model architectures by using both cross entropy (CE) and supervised contrastive loss (SCL) for LOSO experiments with the accuracies averaged across all the subjects. It is seen that the SCL approach outperforms CE-based approach for all model architectures. The results obtained by using this approach (mean accuracy of 88.05% also outperforms the best reported accuracy 83.2%) on the given dataset [18]. It is also seen that the 1DCNN-BiLSTM model performs the best among all the model architectures. This may be attributed to the fact that the features extracted by the convolutional layer are beneficial for predicting the written alphabet. In Table 2, the results of LOSO experiments performed on the target dataset are tabulated with similar trends.

Table 1. Mean Recognition Accuracy for Leave-One-Subject-Out Experiment on Source Dataset.

Architecture	CE	SCL
1DCNN	0.8441	0.8620
LSTM	0.8298	0.8446
BiLSTM	0.8460	0.8760
1DCNN-LSTM	0.8480	0.8723
1DCNN-BiLSTM	0.8592	0.8805

Table 2. Mean Recognition Accuracy for Leave-One-Subject-Out Experiment on Target Dataset.

Architecture	CE	SCL
1DCNN	0.8365	0.8619
LSTM	0.7901	0.7928
BiLSTM	0.8015	0.8384
1DCNN-LSTM	0.8307	0.8455
1DCNN-BiLSTM	0.8269	0.8619

Table 3. Mean Recognition Accuracy on Target Dataset in the Unsupervised Setting.

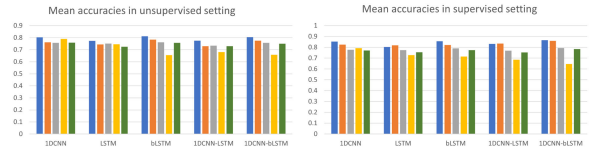
Architecture	DANN	DeepJDOT	DRCN	CE	SCL
1DCNN	0.7569	0.7898	0.7583	0.7617	**0.8030**
LSTM	0.7519	0.7463	0.7248	0.7440	**0.7744**
BiLSTM	0.7619	0.6557	0.7576	0.7840	**0.8128**
1DCNN-LSTM	0.7336	0.6809	0.7296	0.7296	**0.7754**
1DCNN_BiLSTM	0.7575	0.6588	0.7500	0.7757	**0.8053**

Table 4. Mean Recognition Accuracy on Target Dataset in the Supervised Setting.

Architecture	DANN	DeepJDOT	DRCN	CE	SCL
1DCNN	0.7775	0.7921	0.7713	0.8240	**0.8528**
LSTM	0.7746	0.7273	0.7536	**0.8178**	0.8029
BiLSTM	0.7909	0.7151	0.7746	0.8207	**0.8565**
1DCNN-LSTM	0.7685	0.6853	0.7528	**0.8344**	0.8319
1DCNN_BiLSTM	0.7928	0.6463	0.7830	0.8590	**0.8653**

The evaluation of the models on the target domain samples in the unsupervised setting is presented in Table 3. It is seen that in both CE and SCL-based models, there is a decrement in the recognition accuracy when compared to the LOSO experiments in which both training and evaluation was done using samples from the target dataset. However, it can be seen that SCL-based approach outperforms CE-based approach in this scenario as well, thereby motivating the use of the approach for out-of-domain airwriting recognition system in case labeled target dataset is not available. The performance of the approach is also compared against prevalent domain adaptation techniques and it is observed that SCL approach yields superior performance.

Furthermore, in Table 4, we present the results for the set of experiments in which the classifier head is fine-tuned using the labeled target dataset. As expected, the recognition accuracies are greatly improved when compared with the unsupervised setting. Also, it is observed that there is a marginal improvement from the LOSO case. The performance of CE and SCL-based approaches are found to be quite close, while outperforming the baseline domain adaptation techniques. The improvement in the recognition accuracies can be attributed to the fact that fewer parameters are required to be learnt on tuning the classifier head in contrast to learning the parameters for the entire model. Therefore, the encoder parameters learnt on the relatively large source dataset along with the classifier parameters learnt on the small target dataset collectively boost the classification performance. While the trend of improved performance in supervised



(a) (b)

Fig. 2. Mean accuracy on the target dataset using different transfer learning approaches in the (a) unsupervised and (b) supervised setting.

Table 5. Mean Accuracy on Target Dataset 2 Without Fine-Tuning.

Architecture	CE	SCL
1DCNN	0.9307	0.9581
LSTM	0.8923	0.9007
BiLSTM	0.8850	0.9342
1DCNN-LSTM	0.9392	0.9476
1DCNN-BiLSTM	0.9076	0.9576

Table 6. Mean Accuracy on Target Dataset 2 With Fine-Tuning.

Architecture	CE	SCL
1DCNN	0.9273	0.9573
LSTM	0.9084	0.9130
BiLSTM	0.9088	0.9046
1DCNN-LSTM	0.9342	0.9388
1DCNN-BiLSTM	0.9331	0.9538

setting compared to the unsupervised case is in general valid for all the techniques, as illustrated in Fig. 2(a) and (b) the performance of the fine-tuning approach is superior to DANN, DRCN and DeepJDOT. In Tables 7 to 10 (Supplementary Material), we present the top five most confusing letter pairs for the SCL-based approach using the 1DCNN-BiLSTM architecture. In Tables 5 and 6, we present the results of experiments performed by using different device on same subjects without and with fine-tuning. As it can be seen in the tables, SCL-based approach outperforms CE-based approach in a device-independent but user-dependent study.

IV. CONCLUSION

In this letter, we explored supervised contrastive loss-based framework for airwriting recognition using accelerometer and gyroscope signals obtained from a motion sensor worn on the wrist. It is seen that the proposed approach outperforms the state-of the-art accuracy on a publicly available dataset (source dataset), while also boosting the recognition accuracy on on unseen target dataset. It is seen that the supervised contrastive loss-based approach outperforms existing domain adaptation techniques. This implies that our approach can be used for developing an airwriting recognition system, the performance of which is not greatly hampered by variations in the device and user. Future work will focus on exploring explicit methods of bias removal for dataset/user-style specific biases.

ACKNOWLEDGMENT

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

REFERENCES

[1] M. Chen, G. AlRegib, and B.-H. Juang, "Air-writing recognition—Part I: Modeling and recognition of characters, words, and connecting motions," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 403–413, Jun. 2016.

[2] M. Chen, G. AlRegib, and B.-H. Juang,, "Air-writing recognition—Part II: Detection and recognition of writing activity in continuous stream of motion data," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 436–444, Jun. 2016.

[3] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern., Part C.*, vol. 37, no. 3, pp. 311–324, May 2007.

[4] C. Amma, M. Georgi, and T. Schultz, "Airwriting: A wearable handwriting recognition system," *Pers. Ubiquitous Comput.*, vol. 18, no. 1, pp. 191–203, 2014.

[5] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, " uWave: Accelerometer-based personalized gesture recognition and its applications," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2009, pp. 1–9.

[6] D.-W. Kim, J. Lee, H. Lim, J. Seo, and B.-Y. Kang, "Efficient dynamic time warping for 3-D handwriting recognition using gyroscope equipped smartphones," *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5180–5189, 2014.

[7] D. Lu, K. Xu, and D. Huang, "A data driven in-air-handwriting biometric authentication system," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2017, pp. 531–537.

[8] S. Patil, D. Kim, S. Park, and Y. Chai, "Handwriting recognition in free space using WIMU-based hand motion analysis," *J. Sensors*, vol. 2016, 2016, Art. no. 3692876.

[9] M. Alam *et al.*, "Trajectory-based air-writing recognition using deep neural network and depth sensor," *Sensors*, vol. 20, no. 2, 2020, Art. no. 376.

[10] L. Ardüser, P. Bissig, P. Brandes, and R. Wattenhofer, "Recognizing text using motion data from a smartwatch," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, 2016, pp. 1–6.

[11] C. Amma and T. Schultz, "Airwriting: Bringing text entry to wearable computers," *XRDS*, vol. 20, no. 2, pp. 50–55, Dec. 2013.

[12] S. Xu and Y. Xue, "Air-writing characters modelling and recognition on modified CHMM," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2016, pp. 001510–001513.

[13] Y. Li, H. Zheng, H. Zhu, H. Ai, and X. Dong, "Cross-people mobile-phone based airwriting character recognition," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 3027–3033.

[14] S. Xu, Y. Xue, X. Zhang, and L. Jin, "A novel unsupervised domain adaptation method for inertia-trajectory translation of in-air handwriting," *Pattern Recognit.*, vol. 116, 2021, Art. no. 107939.

[15] C. Li, C. Xie, B. Zhang, C. Chen, and J. Han, "Deep Fisher discriminant learning for mobile hand gesture recognition," *Pattern Recognit.*, vol. 77, pp. 276–288, 2018.

[16] P. Roy, S. Ghosh, and U. Pal, "A CNN based framework for unistroke numeral recognition in air-writing," in *Proc. 16th Int. Conf. Front. Handwriting Recognit.*, 2018, pp. 404–409.

[17] L. Jing, Z. Dai, and Y. Zhou, "Wearable handwriting recognition with an inertial sensor on a finger nail," in *Proc. 14th Int. Assoc. Pattern Recognit. Int. Conf. Document Anal. Recognit.*, vol. 1, 2017, pp. 1330–1337.

[18] T. Yanay and E. Shmueli, "Air-writing recognition using smart-bands," *Pervasive Mobile Comput.*, vol. 66, 2020, Art. no. 101183.

[19] A. Levy, B. Nassi, Y. Elovici, and E. Shmueli, "Handwritten signature verification using wrist-worn devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, Sep. 2018, Art. no. 119.

[20] H. Wen, J. Ramos Rojas, and A. K. Dey, "Serendipity: Finger gesture recognition using an off-the-shelf smartwatch," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2016, pp. 3847–3851.

[21] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 21, pp. 545–552, 2008.

[22] Y. Li, K. Yao, and G. Zweig, "Feedback-based handwriting recognition from inertial sensor data for wearable devices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 2269–2273.

[23] X. Lin, Y. Chen, X.-W. Chang, X. Liu, and X. Wang, "SHOW: Smart handwriting on watches," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, 2018, Art. no. 151.

[24] D. Moazen, S. A. Sajjadi, and A. Nahapetian, "AirDraw: Leveraging smart watch motion sensors for mobile human computer interactions," in *Proc. 13th IEEE Annu. Consum. Commun. Netw. Conf.*, pp. 442–446, 2016.

[25] P. Khosla *et al.*, "Supervised contrastive learning," 2020, *arXiv:2004.11362*.

[26] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[27] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Proc. Eur. Conf. Comput. Vision.*, 2016, pp. 597–613.

[28] B. B. Damodaran, B. Kellenberger, R. Flamary, D. Tuia, and N. Courty, "DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 447–463.

[29] Noraxon, Scottsdale, AZ, USA, Ultium EMG. Accessed: Sep. 10, 2021 [Online]. Available: https://www.noraxon.com/our-products/ultium-emg/

[30] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.