Word-level Sign Language Recognition Using Linguistic Adaptation of 77 GHz FMCW Radar Data

M. Mahbubur Rahman[1], Robiulhossain Mdrafi[2], Ali C. Gurbuz[2], Evie Malaia[3], Chris Crawford[4], Darrin Griffin[5], Sevgi Z. Gurbuz[1]

[1]*Dept. of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA*

[2]*Dept. of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA*

[3]*Dept. of Communication Disorders, The University of Alabama, Tuscaloosa, AL, USA*

[4]*Dept of Computer Science, The University of Alabama, Tuscaloosa, AL, USA*

[5]*Dept. of Communication Studies, The University of Alabama, AL, USA*

mrahman17@crimson.ua.edu, szgurbuz@ua.edu

***Abstract*—Over the years, there has been much research in both wearable and video-based American Sign Language (ASL) recognition systems. However, the restrictive and invasive nature of these sensing modalities remains a significant disadvantage in the context of Deaf-centric smart environments or devices that are responsive to ASL. This paper investigates the efficacy of RF sensors for word-level ASL recognition in support of human-computer interfaces designed for deaf or hard-of-hearing individuals. A principal challenge is the training of deep neural networks given the difficulty in acquiring native ASL signing data. In this paper, adversarial domain adaptation is exploited to bridge the physical/kinematic differences between the copysigning of hearing individuals (repetition of sign motion after viewing a video), and native signing of Deaf individuals who are fluent in sign language. Domain adaptation results are compared with those attained by directly synthesizing ASL signs using generative adversarial networks (GANs). Kinematic improvements to the GAN architecture, such as the insertion of micro-Doppler signature envelopes in a secondary branch of the GAN, are utilized to boost performance. Word-level classification accuracy of 91.3% is achieved for 20 ASL words.**

***Index Terms*—ASL, sign language, gesture recognition, RF sensing, radar, micro-Doppler, deep learning**

I. INTRODUCTION

Over the past three decades, most ASL recognition research as focused on the use of wearable and camera-based [1]–[3] sensors, including RGB-D [4], [5] cameras, such as Kinect or Leap Motion [6]. However, camera-based technologies are ineffective in the dark and raise privacy concerns due to constant video monitoring, whereas wearable sensors, such as sensor augmented gloves [7], [8], EMGs [9], and IMUs [10] capture motion that is highly localized to the body part upon which they are worn. Thus, they cannot capture the subtle, rapid changes in the motion trajectories of multiple articulators (hands, head and body) in sign languages [11]; and multiple on-body sensors [12] are required to achieve higher accuracies. This, in turn, limits the signer's freedom in conducting daily activities. The limitations of current technologies have motivated the exploration of non-invasive sensing modalities that can perceive the natural language - sign language - of the Deaf community for the purposes of environment control, remote health, and security.

In this context, RF sensors are uniquely desirable because they are non-contact, can operate in the dark or through-the-wall, protect privacy, and collect fine-grained spatiotemporal data that will aid in ASL understanding: namely, the micro-Doppler signature [13], which is reflective of the time-varying velocity profiles of articulator motion. In recent years, the micro-Doppler signatures has been exploited in a variety of applications including fall detection [14], gait abnormality recognition [15], concussion detection [16], physical therapy and rehabilitation [17], non-contact measurement [18] of heart rate [19], and respiration [20] as well as detection of related conditions, such as sleep apnea [21] or sudden infant death syndrome [22]. The unique capability of RF sensors to capture the rapid progression of dynamic sign sequences in a non-contact fashion has led to its recent proposal for ASL recognition with Wi-Fi [23] and low-cost, short-range radar systems [24].

In prior work [24]–[26], we showed that RF sensor data can capture linguistic features characteristic of signing, such as co-articulation, and a greater amount of information, as measured by fractal complexity of motion, relative to daily activities and gestures . This is reflective of the communicative properties of language [27], [28], and underscores the need to distinguish ASL recognition from gesture recognition. While gesture is typically comprised of simple, unidirectional strokes, sign language discourse is a complex signal, with information density and grammatical structure equivalent to those of spoken languages.

This is especially significant when it is considered that many of the studies in the literature (e.g. [6], [23], [29]) have relied on the copy-signing of hearing participants to benchmark the classification performance of deep neural networks. Our prior work [26] has shown that copy-signing is discriminable from native signing using machine learning. On one hand, this is not surprising because kinematic and rhythmic differences can be observed between copy-signers and native signers - both visually and in the RF data. Indeed, it has been reported that it can take learners of sign language *at least 3 years* to produce signs in a manner that is perceived as fluent by native signers [30]. But, because the training of deep neural

networks (DNNs) relies on massive amounts of data [31] and because it is much easier to recruit hearing participants than deaf participants, ASL recognition studies have relied upon copy-signing, despite its kinematic spatiotemporal differences from native ASL signing.

This paper compares the efficacy of two different ways to address the problem of limited native signing data for training deep models: domain adaptation and data synthesis. More specifically, the paper offers three key contributions:

1) Analysis of optimal RF transmit waveform parameters to ensure accurate kinematic capture of ASL with RF sensors.
2) Transformation of copy-signing data to better resemble native signing data, and thus improve model training. This approach is shown to boost word-level recognition accuracy of 20 signs from 46.2% to 88%.
3) RF micro-Doppler signature synthesis using a multi-branch generative adversarial network (MB-GAN) [32]. This approach is shown to surpass that of domain adaptation, yielding a word-level recognition accuracy of 91.3%.

In Section II, the experimental datasets, data processing and optimal system parameters are presented. Section III describes domain adaptation from imitation to native using cycleGAN [33], while section IV presents the data synthesizing using Wasserstein GAN (WGAN) and Multi-Branch Discriminator GAN (MBGAN) along with their respective classification performance. Finally, in Section V, conclusions and future work are discussed.

II. Experimental Datasets

A. RF ASL Datasets

The RF data of ASL (ASL-R) used in this study were acquired by a TI AWR1642BOOST 77 GHz frequency modulated continuous wave (FMCW) transceiver. Measurements were made of both native signing from deaf or child-of-deaf-adult (CODA) participants fluent in ASL, and copy-signing from hearing participants based on videos of fluent signers. To find the suitable RF transmitter settings that can effectively capture fine grain finger motions in mD representation, data were collected for different bandwidths:

- *20-word Native ASL-R Dataset:* 980 samples (49 per class) from 5 deaf/CODA participants were acquired at bandwidth settings of 750 MHz and 4 GHz.
- *20-word Copy-signing ASL-R Dataset:* 1550 samples from 10 hearing copy-signers were acquired at bandwidth settings of 1.5 GHz and 4 GHz.

Participants were presented with a random ordering of single-word signs, listed in Table 1 to foster independence in each repetition of the signs, and avoid consistent coarticulation.

B. RF Data Processing

The received signal of a radar unit is a complex I/Q time series, from which line-of-sight distance and radial velocity maybe computed. The amplitude and phase of those complex data can be related to the electromagnetic scattering and kinematics of the target being observed. The micro-Doppler signature, or spectrogram, is found from the square modulus ofthe Short-Time Fourier Transform (STFT) of the continuous-time input signal. It reveals the distinct patterns caused by micro-motions, such as hand gestures and human activity. The STFT itself is computed using Hanning windows with 50% overlap to reduce sidelobes in the frequency domain and convert the 1D complex time stream into a 2D mD signature. Reflection from static objects can be removed using moving target indicator (MTI) filters, whereas sensor noise and artifacts were mitigated using a thresholding algorithm.

TABLE I
20-words ASL-R dataset

Breath	Car	Come	Drink
Earthquake	Engineer	Friend	Go
Health	Hello	Help	Hospital
Knife	Lawyer	Mountain	Push
Walk	Well	Write	You

The mD signatures for native signing versus copy-signing signatures are illustrated in Figure 1a. Notice that some kinematic differences can be visually observed between the signatures. For example, the peak Doppler is higher, indicating a faster movement, by the copy-signer, while tandem motions not normally done by native signers are erroneously done by the copy-signer. Indeed, it is not uncommon to see an incorrect number of repetitions, simplified handshapes and motion trajectories, and fluctuating cadence/tempo in the production of signs by copy-signers. Consequently, the feature space of copy-signing and native signing are different enough to allow for their discrimination with machine learning [26].

C. Optimal RF transmitter parameters

Data was collected over 500 coherent processing intervals (CPI) with transmitter bandwidth of 750 MHz, 1.5 GHz and 4 GHz. For each bandwidth setting, two different chirp rate

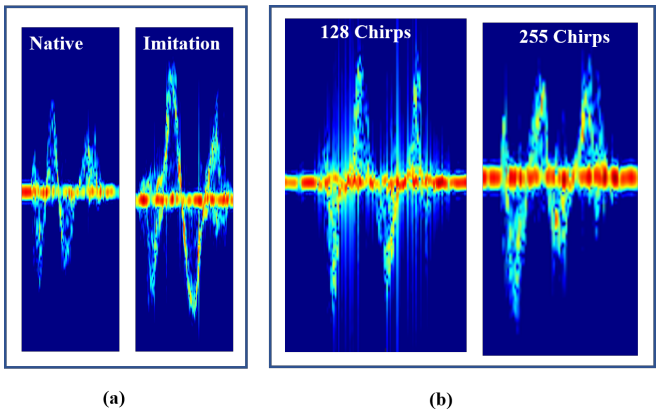


Fig. 1. mD signature for the ASL word BREATH; (a) Native and copy-signing (imitation) spectrograms (b) Spectrograms with 128 and 255 chirps per CPI at 4 GHz transmit bandwidth.

settings were compared: 128 chirps per CPI and 255 chirps per CPI. In each case, the number of ADC samples was 256. It was observed that, with 128 chirps being transmitted per CPI, the signatures appear smeared across Doppler, with some vertical lines, unrelated to kinematics, also appearing, as shown in Figure 1(b). In contrast, when 255 chirps are transmitted per CPI, the mD signatures are more distinct, with no spurious artifacts. This is due to the improved Doppler resolution that occurs with increased dwell time. Therefore, we found that operating the TI 77 GHz FMCW transceiver with a bandwidth of 4 GHz and 255 chirps per CPI is an optimal settings for the acquisition of ASL.

III. Distribution Transformation through CycleGAN

Compilation of large datasets for training state-of-the-art deep neural networks is difficult when human subjects are involved, not only because of the time involved in measuring numerous iterations of each class, but also because it can be difficult to recruit participants, especially if from a minority population, such as the Deaf community. In previous work [26], 20 native ASL signs were classified with an accuracy of 72.5% using minimum-redundancy maximum-relevance (mRMR) selection of 150 handcrafted features extracted from a five node multi-frequency RF sensor network and a random forest classifier. To surpass this performance with just a single sensor, recent advances in deep learning, which have yielded great advances in related fields [31], can be applied. However, deep neural networks (DNNs) rely on large amounts of training data to learn the underlying representations of each class.

One possible approach could be to try to address the data scarcity problem by training the DNN on copy-signing data, while testing on native ASL data. Unfortunately, this approach is not effective because copy-signing does have kinematic characteristics that render them distinguishable from native ASL signing. This is evidenced not only by their discriminability using a support vector machine classifier [26], but also by the poor classification accuracy attained when deep learning is applied. When a convolutional neural network (CNN) is pre-trained on copy-signing data and fine tuned with 80% of the native ASL-R dataset, only 46.15% accuracy was attained when testing on the remaining 20% of native ASL-R data.

Consequently, two approaches are pursued in this paper: 1) using adversarial learning to bridge the gap between copy-signing and native signing and to transform copy-signing data to resemble true native samples, and 2) synthesis of RF signatures for training using a kinematically-enhanced multi-branch GAN (MBGAN). These approaches and signal pre-processing are summarized in Figure 2.

A. Imitation to native domain translation

There are various approaches for image-to-image translation in the literature. We found that the CycleGAN [33] architecture offered better performance when compared to alternatives, such as TravelGAN [34]. CycleGAN translates an image from

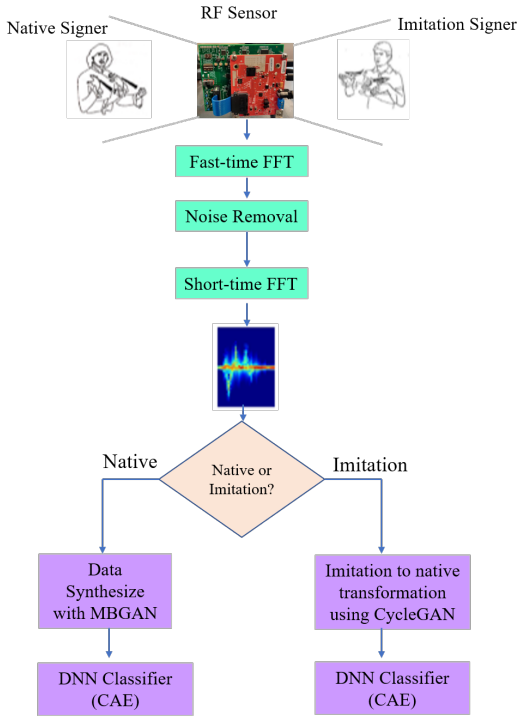


Fig. 2. ASL recognition system Flowchart.

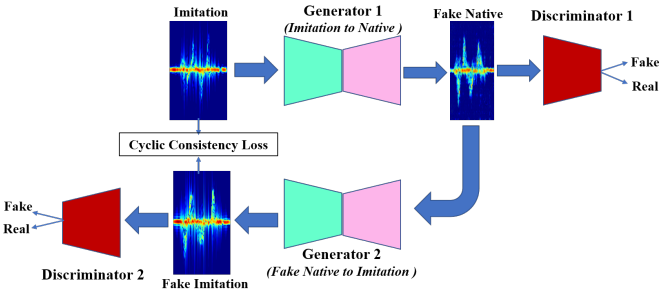
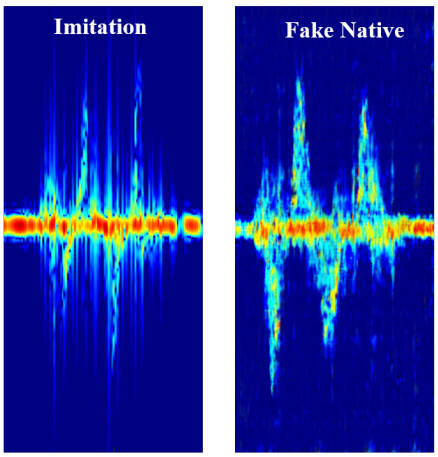
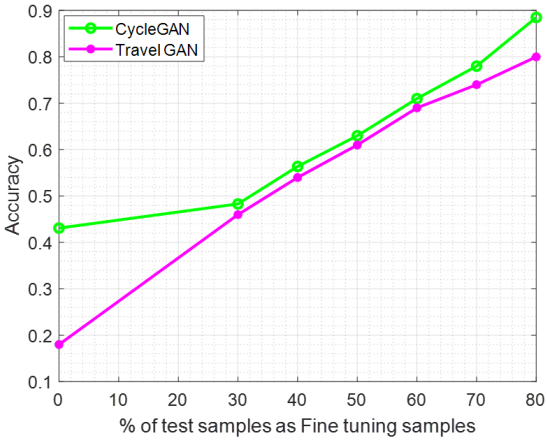


Fig. 3. CycleGAN Architecture for imitation to fake native transformation

a source domain A to a target domain B by forming a series connection between two GANs to form a "cycle": the first GAN tries to synthesize "fake native" from the copy-signing data, while the second GAN works to reconstruct the original sample, synthesizing "fake imitation" samples. Thus, the network tries to minimize the cycle consistency loss, i.e. the difference between the input of the first GAN and the output of second GAN. The functional building blocks of cycleGAN architecture is shown in Figure 3 and an example of copy-signing signature and resulting fake native signature is shown in Figure 4. In this study, 25% of the copy-signing ASL-R data and 40% of the native ASL-R data was reserved for testing, while the remaining was used during training. The resulting fake native signatures were then used to pre-train a three-block convolutional autoencoder (CAE).



(a) Imitation to Fake native sample



(b) Finetuning accuracy for CAE while trained on CycleGAN and Travel GAN transformed data

Fig. 4. (a)CycleGAN transformed native-like sample 'Breath'; (b)Fine-tuning accuracy with CAE trained on CycleGAN transformed native-like samples.

B. Classification Results with Transformed Data

Convolutional Autoencoders (CAEs) have been shown to be effective when small, yet reasonable, amounts of real data are available for training [35]. A deep three-block convolutional autoencoder (CAE) first uses unsupervised pre-training to initialize the network near a good local minima. In each block, a filter concatenation technique [36] is employed in which a filter size of 3×3 and 9×9 were concatenated to take advantage of multilevel feature extraction. After training the CAE model, the decoder was removed and two fully connected layers with 128 neurons followed by a dropout of 0.55 were added after flattening the output of the encoder. At the end, a softmax layer with 20 nodes is employed for classification.

To classify CycleGAN transformed data, the CAE was trained with CycleGAN generated native like samples (output of GAN 1) and tested with real native ASL samples. As shown in Figure 4, when 80% of the native ASL-R data is utilized an accuracy of 88% is achieved. One noticeable thing from the accuracy plot of Figure 4 is, the performance is not that great when the fine tuning samples are less, This indicates that, even after transforming the imitation data to native domain using CycleGAN, there are still significant difference between native samples and fake native samples. The performance only increases when more native ASL samples are being used as fine tuning data.

IV. Data Synthesis with GANs

A. Data Generation

Another approach for dealing with the problem of limited training data is to generate synthetic samples from a small amount of native ASL samples using GANs. In general, the architecture of GANs consists of two competing neural networks i.e., generator and discriminator playing a min-max game. The generator network samples a predefined latent space and upsamples via transposed or deconvolutional layers to produce a synthetic image whereas the discriminator network takes that synthetic images as input and attempts to classify them as being real or fake. As the discriminator gets increasingly better, the gradient vanishes, meaning there is no gradients to update the loss during the training process. Wasserstein GAN (WGAN) [37] treats this vanishing gradient problem by applying a gradient penalty (GP) after every gradient update on the discriminator/critic function and hence the name WGAN-GP comes up. To provide a more stable training process, with proven convergence of the loss function, WGAN-GP uses a new loss function derived from the Wasserstein distance; as the asymmetric Kullback–Leibler (KL) divergence causes buggy results when the intention is just to measure the similarity between two equally important distributions and the Jenson–Shannon divergence fails to provide a meaningful value when two distributions are disjointed [38].

However, with the WGAN generated signatures, it was observed that there is a lack of kinematic fidelity in a significant percentage of signatures generated [32]. Many of these samples have features that are deviant from the typical properties of micro-Doppler, such as high frequency components disconnected from the low-frequency micro-Doppler, negative micro-Doppler corresponding to motion in the reverse direction etc. To maintain critical kinematic features of the data and to better capture gross properties of the micro-Doppler signature in adversarial learning process, MBGAN was proposed in [32] that incorporated the signature's envelope as an additional, second branch into the discriminator of a Wasserstein GAN (WGAN).

The generator model in MBGAN is a DCNN with 8 convolutional layers, each layer being followed by a batch normalization with momentum 0.9 and RELU activation function. On the other hand, the main Branch of discriminator is a 5 layer DCNN where each layer followed by a leaky-RELU activation function and in second branch, 3 1D convolutional layer is used on envelope before it has been concatenated with the flattened output of main discriminator branch. This architecture is shown in Figure 5.

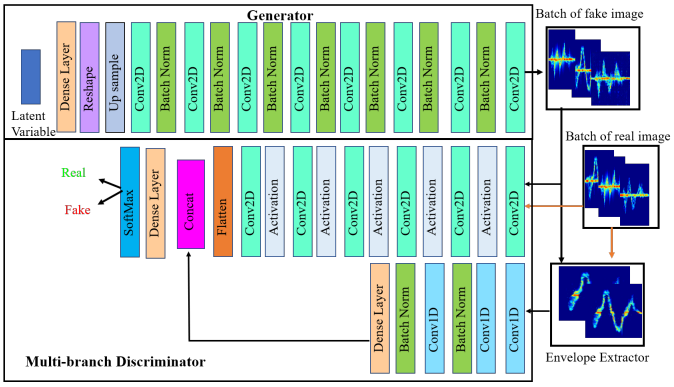


Fig. 5. MBGAN Architecture.

To generate synthetic data, Both WGAN and MBGAN were trained with 75% of native ASL samples and a total of 10000 synthetic samples across 20 classes were generated for both of them .

B. Classification Results with Synthetic Data

The CAE model described in previous section has been trained seperately with the synthetic data generated from WGAN and MBGAN. As the performance is tested on native ASL samples, all the synthetic data have been utilized to train the CAE classifier. When tested with native ASL samples, a classification accuracy of 74.28% and 80.82% were achieved respectively for WGAN and MBGAN across all the 20 classes. By fine tuning with 30% native samples, the performance increases for both WGAN and MBGAN to 86% and 91.30% respectively. The results are tabulated on Table II and the confusion matrix for MBGAN and WGAN are shown in Figure 6. Finally, a flowchart summarizing all the steps from data collection to native ASL words classification is shown in Figure 2.

TABLE II
CLASSIFICATION PERFORMANCE FOR GAN GENERATED DATA

GAN Data	Test Data	Accuracy	Finetuning Accuracy
WGAN	Native samples	74.28%	86%
MBGAN	Native samples	80.82%	91.30%

V. CONCLUSIONS

This paper has presented two approaches for word-level ASL recognition when few native samples are available for model training. Many ASL recognition studies rely on copy-signers due to the ease with which hearing participants can be recruited versus Deaf participants. However, copy-signing and native signing data have significant kinematic differences, which precludes copy-signing data from being used when conducting linguistic studies or to directly train a classifier of native signing data. This paper compared the ability of RF sensors to capture signing kinematics for different transmit waveform parameters, and compared the efficacy of training

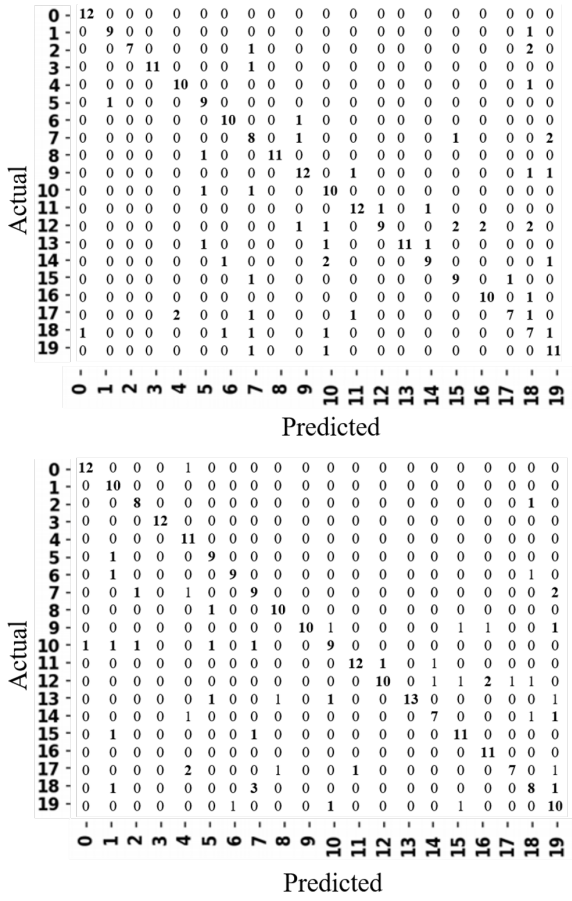


Fig. 6. WGAN (top) and MBGAN (bottom) Confusion matrix

with copy-signing data transformed to resemble native signing data, versus synthesis of RF signatures using adversarial learning. A physics-aware, multi-branch GAN was found to yield to most effective training dataset, for which a three-block CAE yielded a classification accuracy of 91.3% on 20 native ASL RF micro-Doppler signatures. The results demonstrate the potential of ASL recognition with non-invasive, remotely operable RF sensors, especially when limited native signing data is available.

VI. ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation (NSF) CPS Program Award #1932547, #1931861 and NSF NCS Program Award #1734938. Human studies research was conducted under the University of Alabama Institutional Review Board (IRB) Protocol #18-06-1271.

REFERENCES

[1] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, "Using multiple sensors for mobile sign language recognition," in *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, 2003, pp. 45–52.

[2] T. Nguyen and S. Ranganath, "Recognizing continuous grammatical marker facial gestures in sign language video." in *Lecture Notes in Computer Science*, vol. 6495, 2010.

[3] F. Csóka and J. Polec, "Algorithm for real-time finger spelling alphabet recognition in video sequences," in *2017 Signal Processing Symposium (SPSympo)*, 2017, pp. 1–4.

[4] N. Pugeault and R. Bowden, "Spelling it out: Real-time asl fingerspelling recognition," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, pp. 1114–1119.

[5] Cao Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 44–52.

[6] J. C. B. Fang and M. Zhang, "Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation," *SenSys '17: Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, no. 5, pp. 1–13, 2017.

[7] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 526–533, 2015.

[8] B. G. Lee and S. M. Lee, "Smart wearable hand device for sign language interpretation system with sensors fusion," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1224–1232, 2018.

[9] J. Wu, L. Sun, and R. Jafari, "A wearable system for recognizing american sign language in real-time using imu and surface emg sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1281–1290, 2016.

[10] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign language recognition using intrinsic-mode sample entropy on semg and accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2879–2890, 2009.

[11] E. Malaia, J. D. Borneman, and R. B. Wilbur, "Information transfer capacity of articulators in american sign language," *Language and speech*, vol. 61, no. 1, pp. 97–112, 2018.

[12] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[13] V. Chen, *The Micro-Doppler Effect in Radar*. Artech House, 2019.

[14] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[15] A. Seifert, A. M. Zoubir, and M. G. Amin, "Radar classification of human gait abnormality based on sum-of-harmonics analysis," in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 0940–0945.

[16] J. W. Palmer, K. F. Bing, A. C. Sharma, and E. F. Greneker, "Detecting concussion impairment with radar using gait analysis techniques," in *2011 IEEE RadarCon (RADAR)*, 2011, pp. 222–225.

[17] O. Postolache, J. M. D. Pereira, V. Viegas, and P. S. Girão, "Gait rehabilitation assessment based on microwave doppler radars embedded in walkers," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, 2015, pp. 208–213.

[18] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, "A review on recent advances in doppler radar sensors for noncontact healthcare monitoring," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 5, pp. 2046–2060, 2013.

[19] W. Massagram, V. M. Lubecke, A. HØst-Madsen, and O. Boric-Lubecke, "Assessment of heart rate variability and respiratory sinus arrhythmia via doppler radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 10, pp. 2542–2549, 2009.

[20] A. Rahman, V. M. Lubecke, O. Boric–Lubecke, J. H. Prins, and T. Sakamoto, "Doppler radar techniques for accurate respiration characterization and subject identification," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 2, pp. 350–359, 2018.

[21] Y. S. Lee, P. N. Pathirana, C. L. Steinfort, and T. Caelli, "Monitoring and analysis of respiratory patterns using microwave doppler radar," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 2, pp. 1–12, 2014.

[22] E. G. Ziganshin, M. A. Numerov, and S. A. Vygolov, "Uwb baby monitor," in *2010 5th International Confernce on Ultrawideband and Ultrashort Impulse Signals*, 2010, pp. 159–161.

[23] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018.

[24] S. Gurbuz, A. Gurbuz, C. Crawford, and D. Griffin, "Radar-based methods and apparatus for communication and interpretation of sign languages," in *U.S. Patent Application No. US2020/0334452 (Invention Disclosure filed Feb. 2018; Provisional Patent App. filed Apr. 2019.)*, October 2020.

[25] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. Crawford, M. M. Rahman, R. Aksu, E. Kurtoglu, R. Mdrafi, A. Anbuselvam, T. Macks, and E. Ozcelik, "A linguistic perspective on radar micro-doppler analysis of american sign language," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 232–237.

[26] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using rf sensing," *IEEE Sensors Journal*, pp. 1–1, 2020.

[27] E. Malaia, J. D. Borneman, and R. B. Wilbur, "Assessment of information content in visual signal: analysis of optical flow fractal complexity," *Visual Cognition*, vol. 24, no. 3, pp. 246–251, 2016.

[28] J. D. Borneman, E. Malaia, and R. B. Wilbur, "Motion characterization using optical flow and fractal complexity," *Journal of Electronic Imaging*, vol. 27, no. 5, p. 051229, 2018.

[29] L. Pigou, M. Van Herreweghe, and J. Dambre, "Gesture and sign language recognition with temporal residual networks," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 3086–3093.

[30] J. S. Beal and K. Faniel, "Hearing l2 sign language learners: How do they perform on asl phonological fluency?" *Sign Language Studies*, vol. 19, no. 2, pp. 204–224, 2018.

[31] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[32] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Synthesis of micro-doppler signatures for abnormal gait using multi-branch discriminator with embedded kinematics," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 175–179.

[33] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[34] M. Amodio and S. Krishnaswamy, "Travelgan: Image-to-image translation by transformation vector learning," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8975–8984, 2019.

[35] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709–1723, 2018.

[36] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-based human gait recognition using dual-channel deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9767–9778, 2019.

[37] T. Zhang, Z. Li, Q. Zhu, and D. Zhang, "Improved procedures for training primal wasserstein gans," in *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019, pp. 1601–1607.

[38] P. Bhargava, "Better generative modelling through wasserstein gans," in *Intel AI Develop Program Blog*, 2018.