2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# Unsupervised Generative Adversarial Network for Plantar Pressure Image-to-Image Translation

Mona Ahmadian[1], Mohammad TH. Beheshti[2], Ahmad Kalhor[3], and Amir Shirian[4]

*Abstract*— **Analyzing human gait from plantar pressure is critical for human health. The majority of works focus on classifying the healthy plantar pattern from unhealthy ones. Different from previous works, we adopt a generative adversarial network to produce healthy plantar pressure image for individual patients. In this work, we do not have pairs of images for training thus we cast the problem as an unsupervised generative adversarial learning task. Our network benefits from multiple components: an encoder-decoder generator, a convolution-based discriminator, a convolution-based evaluation network, and a new term in the loss function to preserve the person's gait style. Our method achieves high performance** (99.8%) **on the CAD WALK databases which have patients with hallux valgus disease.**

## I. INTRODUCTION

Gait is one of the most frequently used forms of human movement during daily activity. The feet together contain more than 50 bones, 60 joints, and 200 muscles, tendons, and ligaments that hold them together and help us move. Any deformation in one of the aforementioned parts can cause disability. Thus, plantar pressure analysis has been used years to diagnose gait movement (or pathology) [1], [2], diabetic foot [3], skin injury [4], and Parkinson disease [5]. Between the feet deformities, hallux valgus is a very common pathological condition in people which causes many disabilities [6]. In this work, we focus on hallux valgus deformity.

The majority of work on plantar pressure analysis follows a two-step approach. First, extracting a set of handcrafted features and then a machine learning system is employed to recognize the deformation class [7], [8]. Although plantar pressure is widely used in the literature, the use of motion capture sensors such as accelerometers has shown to be effective [9]. For recognition, one line of research in gait movement analysis focuses on hand-crafted features [10]. While owning to the latest trend in machine learning autoencoders [11], recurrent neural networks (RNN) [12], and other deep learning methods [13], [14] have also been used for gait classification. Among the different deep learning methods, convolutional neural networks (CNN) have received the most attention owing to the prosperity of CNNs in various tasks.

[1]Mona Ahmadian is with Faculty of Electrical Engineering, Tarbiat Modares University, Tehran, Iran m_ahmadian@modares.ac.ir
[2]Mohammad TH. Beheshti with the Department of Electrical Engineering, Tarbiat Modares University, Tehran, Iran mbehesht@modares.ac.ir
[3]Ahmad Kalhor with the Department of Electrical Engineering, University of Tehran, Tehran, Iran akalhor@ut.ac.ir
[4]Amir Shirian with the Department of Computer Science, University of Warwick, Coventry, United Kingdom Amir.Shirian@warwick.ac.uk

Generative adversarial networks (GAN) have become a hot research topic recently and used in various applications by way of a two-player minimax game[15], [16], [17]. GAN research is divided into two categories. The first is to estimate the density either explicitly or implicitly and generate a new sample from a noise [18]. The other group assumes a joint probability distribution between the source and the target domain and the input image is transformed to the target domain [16], [19], [20], [21]. This image-to-image translation can be studied in a supervised and unsupervised learning setting. In supervised learning, the pair of corresponding images in different domains are available [16], [19]. On the other hand, in the unsupervised learning, there are two sets of source and target domain and there are no paired samples showing how the image in the source domain could be translated to the corresponding target domain [21].

Efforts on gait analysis are primarily concentrated on building a better machine learning system. Although after diagnosis, the clinical treatment also is very important in the pathological process. Due to the lack of healthy plantar pressure for patients, clinical treatment can not efficiently be done. In this work, by designing a generative adversarial network (GAN) based model, we were able to produce healthy plantar pressure images for patient individuals in an unsupervised setting. This problem can be categorized as an image-to-image translation. A key challenge in this problem is to learn a joint distribution of images between the source (patients) and target (healthy people) domain.

This work can be used in the treatment process and to inform the progression of the course for a person with a foot deformity; with the aim of treatment being to improve the person's gait and achieve a healthy walking state. To this end, we propose an encoder-decoder GAN architecture and evaluate the model on the CAD WALK databases [22], [23]. Moreover, due to the lack of ground truth for the produced healthy images, we used a convolutional-based network to evaluate our model. In summary, the main contributions of this paper are:

- To the best of our knowledge, this work is the first work trying to produce healthy plantar pressure patterns for patients as help for medical centers.
- Novel loss function terms to help achieve better image quality.
- Achieving high performance on the CAD WALK database.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

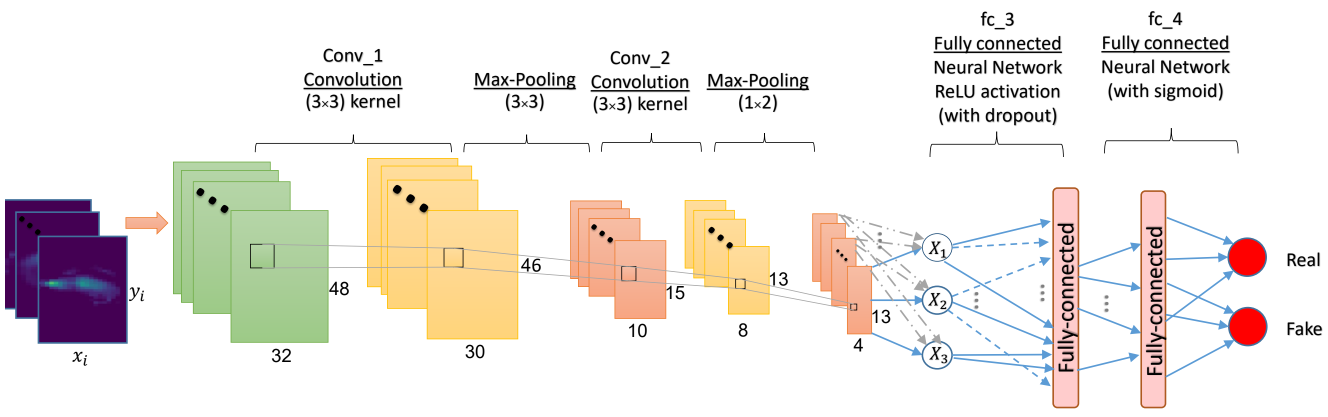


Fig. 1. Our discriminator architecture, consists of convolutional and max pooling layers followed by fully connected layers.

## II. Method

In this section, we describe our generative adversarial-based network to produce healthy plantar pressure images for patients. First, we proposed an encoder-decoder architecture for GAN to capture joint distribution between two domains. Next, we propose a convolutional network for evaluating our method. Below, we describe the components in detail.

### A. Generative Adversarial Network

Following Goodfellow *et. al* [18], we define our generative learning framework. Generative adversarial networks are a clever way to train a model. They do this by framing the problem as a learning problem with two sub-models and consist of two separate neural networks. The general idea behind these two sub-models is that it allows the generator to produce new samples and the discriminator that attempts to classify the samples as real or fake. Both models are trained together, and this continues until the discriminator has been fooled, meaning that the generator model has produced reliable samples. In these neural networks, we have two factors that act opposite and become stronger in competition with each other over time. In fact, learning this framework is like a min-max problem:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{X}\sim p_{data_H}(\mathbf{X})}\Big[\log D(\mathbf{X})\Big] + \mathbb{E}_{\mathbf{Y}\sim p_{data_P}(\mathbf{Y})}\Big[\log(1-D(G(\mathbf{Y})))\Big] \quad (1)$$

where $data_H$ and $data_P$ are healthy and patient data samples. $V$, $D$, $G$, $\mathbf{X}$ and $\mathbf{Y}$ refer to Value function, Discriminator, Generator, a sample image from a healthy set, and a sample image from patient set respectively. Unlike the original GAN paper [18], our generator does not map a fixed prior distribution (noise) to the data distribution $p_{\mathbf{X}}$. In fact, in this work, we are trying to learn the joint distribution between healthy and patient data samples.

Previous studies [24] have shown that each person can be identified by the way they walk. Motivated by this fact, we proposed a new term in the loss function. This term tries to preserve the patient plantar pressure style and transfer it to the generated sample by the generator. Therefore, the GAN loss function becomes:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{X}\sim p_{data_H}(\mathbf{X})}\Big[\log D(\mathbf{X})\Big] + \mathbb{E}_{\mathbf{Y}\sim p_{data_P}(\mathbf{Y})}\Big[\log(1-D(G(\mathbf{Y}))) + \lambda\|\mathbf{Y}-G(\mathbf{Y})\|_2^2\Big] \quad (2)$$

where $\|.\|$ is L2 norm. In what follows, we will explain the structure of our generator and discriminator.

**Generator.** Various generator structure has been used in previous works [18], [19], [25] and it more depends on the input data domain. Following [25], we decided to use autoencoders for generator. Autoencoders are simple neural architectures with strong power for extracting embedding features. The CNN-based generators also are shown to be very effective for having high-quality generated images [26]. Thus, we decided to combine these two networks and adopt asymmetric autoencoder architecture. The architecture detail is shown in Table. I. Unlike [25], To have better quality and quicker learning, we pretrained the encoder-decoder with all our data samples including healthy and patient individuals by the mean square error (MSE) for error estimation function. The generator's weights are initialized and they will update during training.

**Discriminator.** To follow generator architecture, discriminator gets input image of $32\times48$ then two convolutional layers with $3\times3@3$ kernels follow by two $3\times3$ max-pooling layers. At the end, two fully connected layers with 100 and 2 features respectively. Moreover, the dropout was used between the last convolutional layer and the first fully connected layer. The discriminator architecture also is shown in Fig. 1.

The discriminator ingests a healthy image and the image produced by the generator. The discriminator is responsible for detecting whether the image produced by the generator is healthy or not, and network training is continued until

TABLE I
AUTOENCODER ARCHITECTURE. *Conv* AND *TransConv* STAND FOR CONVOLUTION AND TRANSPOSED CONVOLUTIONAL LAYERS RESPECTIVELY. ALL LAYERS HAVE KERNEL SIZE OF $3 \times 3$ AND STRIDE SIZE OF $1$

| | **Layer** | **Output size** | **# of Kernels** |
|---|---|---|---|
| **Encoder** | Image | $32 \times 48$ | - |
| | Conv1 | $30 \times 46$ | 6 |
| | Conv2 | $28 \times 44$ | 16 |
| | Conv3 | $26 \times 42$ | 26 |
| **Decoder** | Embed | $26 \times 42$ | 26 |
| | TransConv1 | $28 \times 44$ | 16 |
| | TransConv2 | $30 \times 46$ | 6 |
| | TransConv3 | $32 \times 48$ | 1 |

the network reaches its equilibrium by minimizing the loss function in Eq. 2.

### B. Evaluation Network

The evaluation of GANs can be considered as an effort to measure the dissimilarity between the real distribution and the generated distribution because there is no corresponding ground truth. To this end, we decided to use a network to learn both data distributions, healthy and unhealthy classes. It means we trained a network on our original data which can discriminate between the healthy and unhealthy data distribution. Our GAN generates a healthy form of unhealthy plantar pressure image. The generated image from GAN is given to the evaluation network to predict its label. Motivated by the success of CNNs, we propose to adopt a CNN for this purpose with the same architecture as the discriminator in our GAN. We also decided to train the strong baselines for the evaluation network which has been used in the previous plantar pressure classification works including support vector machine (SVM) and decision tree.

TABLE II
(A) OUR GAN PERFORMANCE ON THE PATIENT TEST DATA BY USING EVALUATION NETWORK. THE ACCURACY VALUES ARE CLASSIFICATION ACCURACY. (B) PERFORMANCE COMPARISON BETWEEN THE EVALUATION NETWORKS ON CAD WALK DATABASE IN TERMS OF CLASSIFICATION ACCURACY(%).

| **Data** | **Accuracy(%)** |
|---|---|
| Test data | 99.80 |
| Generated data | 99.52 |

(a)

| **Model** | **Accuracy(%)** |
|---|---|
| SVM with linear kernel | 81.28 |
| SVM with RBF kernel | 89.31 |
| Desicion tree | 98.84 |
| CNN | **99.78** |

(b)

## III. EXPERIMENTS AND DISCUSSION

### A. Database

We evaluate our model on the CAD WALK database [22], [23]. This database was collected over two different sessions including healthy people and who have Hallux Valgus. For healthy people, it contains the raw dynamic plantar pressure measurements of 55 healthy individuals which for each of them 24 dynamic plantar pressure measurements are provided for both feet. It means each pixel of the image contains the pressure value corresponding to the pressure sensor. For the patients, it contains the raw dynamic plantar pressure measurements of 50 individuals with Hallux Valgus which for each individual, between 8-15 dynamic plantar pressure measurements are provided for both feet.

### B. Implementation Details

The databases we used, provide only raw plantar pressure images with different sizes and sampling rates. To obtain the same size for images, we supposed a fixed image size which the size is the biggest image size in the databases. The smaller images are simply zero-padded. In the end, All images are resized to $32 \times 48$. The network weights are initialized following the Xavier initialization [27]. We set $\lambda = 0.001$ (see Eq. 2). We also used Adam optimizer with a learning rate of $0.01$ and decay rate of $0.5$ after each $50$ epochs for 20,000 iterations. For all experiments, we perform 5-fold cross-validation and report average accuracy. We used Pytorch for implementing our model and the evaluation network, and an NVIDIA RTX-2080Ti GPU for all experiments.

### C. Results and Analysis

To follow the evaluation framework discussed in Sec. II-B, we designed a convolutional network to evaluate the performance of our GAN. To this end, we trained our proposed method and the baselines with the same evaluation method (5-fold evaluation). To show the superiority of CNN, we compared it with the strong baselines in Table. II (b). Clearly, the proposed CNN outperforms all the baselines. Our model performs better than the SVM-based architecture - a class of classifiers most commonly used in classification problems.

To evaluate the effectiveness of the proposed GAN, we generate healthy plantar pressure images for test data and then check the accuracy of the CNN-based classifier for them in Table. II (a). As it is shown, the generated plantar pressure images have high accuracy and are comparable with real images which support the effectiveness of our method.

Another common method for showing the performance of GANs is to compare the input and output data of the generator. The visualized results are illustrated in Fig. 2. The first row contains the actual patient plantar pressures. The second row is the corresponding healthy generated samples by the generator.

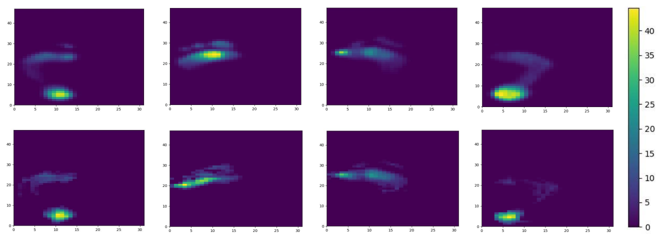

Fig. 2. Qualitative results showing the deformity plantar pressure and corresponding healthy pattern. the first row contains plantar pressure images from the patients and the second row has the corresponding healthy plantar pressure.

## IV. Conclusion

We proposed an unsupervised generative adversarial network to produce healthy plantar pressure image for patients who have hallux valgus disease. To preserve the patients' gait style, we also proposed a new term in the loss function. Moreover, due to problems for performance evaluation of generative models, we proposed a CNN-based classifier and showed that the proposed GAN works with high performance.

In this work, we used the raw plantar pressure images as input with no constraint. This makes our method directly applicable to use for any kind of disease which changes the plantar pressure distribution. One of the applications of this work is to use in physical therapy clinics. Our method can produce a healthy gait plantar pattern for those who have foot disease. This helps physiotherapists to follow the treatment process, even it shows how far the patient is from the healthy and normal pattern.

## References


[1] H. Becker, D. Rosenbaum, L. Claes, and H. Gerngro, "Measurement of plantar pressure distribution during gait for diagnosis of functional lateral ankle instability.," *Clinical biomechanics (Bristol, Avon)*, vol. 12, no. 3, pp. S19–S19, 1997.

[2] L. Wafai, A. Zayegh, J. Woulfe, S. M. Aziz, and R. Begg, "Identification of foot pathologies based on plantar pressure asymmetry," *Sensors*, vol. 15, no. 8, pp. 20392–20408, 2015.

[3] F. Abouaesha, C. H. van Schie, G. D. Grifffths, R. J. Young, and A. J. Boulton, "Plantar tissue thickness is related to peak plantar pressure in the high-risk diabetic foot," *Diabetes care*, vol. 24, no. 7, pp. 1270–1274, 2001.

[4] M. J. Mueller, D. Zou, and D. J. Lott, ""pressure gradient" as an indicator of plantar skin injury," *Diabetes Care*, vol. 28, no. 12, pp. 2908–2912, 2005.

[5] H.-S. Jeon, J. Han, W.-J. Yi, B. Jeon, and K. S. Park, "Classification of parkinson gait and normal gait using spatial-temporal image of plantar pressure," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4672–4675, IEEE, 2008.

[6] N. Wülker and F. Mittag, "The treatment of hallux valgus," *Deutsches Ärzteblatt International*, vol. 109, no. 49, p. 857, 2012.

[7] C. Pradhan, M. Wuehr, F. Akrami, M. Neuhaeusser, S. Huth, T. Brandt, K. Jahn, and R. Schniepp, "Automated classification of neurological disorders of gait using spatio-temporal gait parameters," *Journal of Electromyography and Kinesiology*, vol. 25, no. 2, pp. 413–422, 2015.

[8] N. S. Koktas, N. Yalabik, and G. Yavuzer, "Combining neural networks for gait classification," in *Iberoamerican Congress on Pattern Recognition*, pp. 381–388, Springer, 2006.

[9] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys (CSUR)*, vol. 43, no. 3, pp. 1–43, 2011.

[10] A. K. Buldt, S. Forghany, K. B. Landorf, P. Levinger, G. S. Murley, and H. B. Menz, "Foot posture is associated with plantar pressure during gait: A comparison of normal, planus and cavus feet," *Gait & posture*, vol. 62, pp. 235–240, 2018.

[11] S. Potluri, S. Ravuri, C. Diedrich, and L. Schega, "Deep learning based gait abnormality detection using wearable sensor system," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3613–3619, IEEE, 2019.

[12] X. Li, Y. He, X. Zhang, and Q. Zhao, "Plantar pressure data based gait recognition by using long short-term memory network," in *Chinese Conference on Biometric Recognition*, pp. 128–136, Springer, 2018.

[13] Y. Li, D. Zhang, J. Zhang, L. Xun, Q. Yan, J. Zhang, Q. Gao, and Y. Xia, "A convolutional neural network for gait recognition based on plantar pressure images," in *Chinese Conference on Biometric Recognition*, pp. 466–473, Springer, 2017.

[14] P. Terrier, "Gait recognition via deep learning of the center-of-pressure trajectory," *Applied Sciences*, vol. 10, no. 3, p. 774, 2020.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

[17] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 327–340, 2001.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.

[19] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[21] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in neural information processing systems*, pp. 700–708, 2017.

[22] G. B. Brian, L. K. Noël, H. Toon, and S. Jan, "The cad walk healthy controls dataset," 2018.

[23] G. B. Brian, L. K. Noël, H. Toon, and S. Jan, "The cad walk hallux valgus pre-surgery dataset," 2019.

[24] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 209–226, 2016.

[25] A. Makhzani and B. J. Frey, "Pixelgan autoencoders," in *Advances in Neural Information Processing Systems*, pp. 1975–1985, 2017.

[26] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2016.

[27] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.