Interference Motion Removal for Doppler Radar Vital Sign Detection Using Variational Encoder-Decoder Neural Network

Mikolaj Czerkawski, Christos Ilioudis, Carmine Clemente, Craig Michie, Ivan Andonovic, Christos Tachtatzis
Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK

Abstract—The treatment of interfering motion contributions remains one of the key challenges in the domain of radar-based vital sign monitoring. Removal of the interference to extract the vital sign contributions is demanding due to overlapping Doppler bands, the complex structure of the interference motions and significant variations in the power levels of their contributions. A novel approach to the removal of interference through the use of a probabilistic deep learning model is presented. Results show that a convolutional encoder-decoder neural network with a variational objective is capable of learning a meaningful representation space of vital sign Doppler-time distribution facilitating their extraction from a mixture signal. The approach is tested on semi-experimental data containing real vital sign signatures and simulated returns from interfering body motions. It is demonstrated that the application of the proposed network enhances the extraction of the micro-Doppler frequency corresponding to the respiration rate.

Index Terms—Doppler radar, heart rate monitoring, respiration rate monitoring, vital signs, random body movement, variational autoencoder.

I. INTRODUCTION

The application of radar for vital sign detection has been subject to extensive research over the last decade [1], [2] owing to the many advantages inherent within non-contact health monitoring. A radar-based system has the potential for the continuous and unobtrusive tracking of respiration rate and heart rate of one or multiple living targets with no invasion. A solution requiring physical contact is also undesirable in a variety of applications, such as patients with compromised skin or infant monitoring.

A number of core challenges in the context of the system design can be identified; here, the problem of vital sign extraction in the presence of interfering returns from additional body motions (e.g. limbs) is addressed. The extraction of vital sign signals using radar relies on mechanical oscillations of the body surface induced by displacements owing to both respiration and heart beat. These motions can be extracted from the Doppler shifts contained in a received radar signal, particularly difficult in the presence of other non-stationary reflections in the field of view. Further, interference may not necessarily be confined to separate objects as it may also originate from the other parts of the monitored target. The limbs of an ostensibly stationary living target still move in a fairly random fashion, or perform approximately static

activities such as moving their arm and fingers while using a mobile phone; this class of interference motions is generally described as Random Body Movement (RBM).

One of the earliest reports on the blind separation of human vital signs in Doppler radar signals was presented in [3]. A Single Input Multiple Output (SIMO) Continuous-Wave (CW) radar system with one transmitter and two receivers was utilised to separate multiple contributions in the signal with Analytical Constant Modulus Algorithm (ACMA), Real Analytical Constant Modulus Algorithm (RACMA) and Independent Component Analysis (ICA) algorithms explored. Empirical Mode Decomposition (EMD) was applied in [4] to the received single-channel bandpass-filtered signal. In [5] an analytically derived method for respiration signal separation was presented; however, the method is not appropriate for signals with noise. In [6] and [7] a similar blind source separation approach is applied with different clustering methods relying on Non-Negative Matrix Factorization (NMF). In [8] an unsupervised machine learning framework is presented and, similarly to [6], [7] rely on the sparsity of the cardiac signal in time. In [9] a deep neural network is used to process an unwrapped phase signal and remove the interference of body movement. A review of methods for RBM cancellation (as well as Radar Self-Motion (RSM)) is provided in [10], ranging from the use of multiple radar transceivers [11], [12], through a camera-assisted radar [13] to a matched filter approach. [14] However, the majority of the approaches rely on the application of multiple sensors, increasing the complexity and the practical implementation of the systems significantly.

A body of research has been published in the recent years on the application of Variational Autoencoder (VAE) architectures to source separation, predominantly in the audio domain. In [15] a separate VAE model is trained for each expected source within the mixed signal in order to extract each contribution from a magnitude of Short-Time Fourier Transform (STFT). Both [16] and [17] operate on complex STFT outputs instead, addressing the challenge of determined and under-determined separation, respectively. In the latter, a generalised version of the multi-channel VAE approach is introduced. A somewhat similar problem of audio source separation with weak class supervision is addressed in [18] using the architecture from [19], where the first layer of the convolutional network has kernels that extend over an entire (yet single) instantaneous spectrum.

The contribution presented in this work constitutes the first application of a variational encoder-decoder neural network for interference removal from vital signs radar returns. The framework operates solely on single channel complex radar signals and the network operation can be flexibly adjusted by controlling the datasets used for training. This means that the same method can be applied for a wide range of other applications where certain contributions in a radar signal are to be removed. Furthermore, the probabilistic nature of the architecture allows the model to learn a more realistic representations of the signal, where a given input is matched with a distribution of possible satisfactory output samples rather than a single sample.

Section II contains the problem formulation, which is followed by Section III that describes the proposed solution. The results and a demonstration of micro-Doppler based respiration rate extraction are presented in Section IV. Section V contains concluding remarks.

II. PROBLEM FORMULATION

Respiration and heart beat result in mechanical oscillations of the body surface in several locations of the human body. Since both contributions are generally expected to be present on the chest, this is the area where a radar device is directed in the reported experiments. The respiration component is not only of higher amplitude but also present over a larger area, each contribution is modelled as separate scattering point $s_b(t)$, similarly to [20]. Both $s_b(t)$ and $s_h(t)$ are described in a similar manner, with the range traces $R_b(t)$ and $R_h(t)$ corresponding to the motion associated with the respiration and the heart beat, respectively. For a carrier wavelength of λ and effective component magnitudes $|X_b|$ and $|X_h|$ the two components are described as:

$$s_b(t) = |X_b|e^{j4\pi \frac{R_b(t)}{\lambda}} \quad (1)$$

$$s_h(t) = |X_h|e^{j4\pi \frac{R_h(t)}{\lambda}} \quad (2)$$

Random body movement is most often composed of several moving parts. However, since the velocity trace of each reflecting point can be approximated by a mono-component signal, all random body movement $i(t)$ can be treated as a sum of N reflecting components with different strengths $|X_k|$ and varying range traces $R_k(t)$:

$$i(t) = \sum_{k=1}^N |X_k|e^{j4\pi \frac{R_k(t)}{\lambda}} \quad (3)$$

Finally, the total contribution $\rho(t)$ from a human target can be described as a sum of the vital sign contributions and the random body movements:

$$\rho(t) = |X_b|e^{j4\pi \frac{R_b(t)}{\lambda}} + |X_h|e^{j4\pi \frac{R_h(t)}{\lambda}} + \sum_{k=1}^N |X_k|e^{j4\pi \frac{R_k(t)}{\lambda}} \quad (4)$$

The range or velocity traces can be extracted from the radar range-time and velocity-time maps respectively. In the case of

Fig. 1. Sample synthesis pipeline

non-modulated CW radar, only velocity-time maps can be derived. The vital signs contributions and the RBM are expected to have overlapping Doppler bands and significant differences in their power levels. Additionally, the RBM and other sources of interference can have fairly complex structure making them difficult to model using non-stochastic processing.

III. PROPOSED SOLUTION

The proposed solution utilises a variational convolutional neural network (CNN) as an interference removal system in a fashion analogous to VAE. The system is trained and tested in a semi-experimental environment, where the vital sign STFTs are derived from real recordings and the interference motion is synthesized according to a widely applied model [21].

A. Vital Signs Dataset

The real vital signs samples are from an open dataset of “GUARDIAN Vital Sign Data” [22]. The dataset contains a number of reference signals, such as Phonocardiogram (PCG), Electrocardiogram (ECG) and respiration sensor in addition to the recording of radar signals[23]. The radar dataset is not balanced as it contains samples from varying scenarios, sampling rates and sensor placement configurations at different proportions. In order to render the dataset compatible with the needs of the development of a machine learning solution, a balanced subset of the data has been selected, confined to samples with both the PCG sensor and the radar positioned directly in the front of the target. The selection is further restricted to samples contained directly in the corresponding directory and ignores the sub-folders containing more unusual scenarios (such as apnea, post-exercise, speech, or angle variation). A relatively low target sampling rate of 100 Hz is specified and each source signal is decimated to match that rate (this includes a 30-point FIR low-pass filter of degree 16 with Hamming window).

Radar data within the dataset is supplied in the quadrature format, with separate traces for I and Q signals. The power ratio between this signal and the simulated interference motion is controlled by first normalising power of both signals to achieve zero-mean and unit-variance followed by scaling of the interference signal to match a given Signal-to-Interference Ratio (SIR) ratio.

Fig. 2. Architecture diagram of the utilised network

TABLE I
PARAMETERS USED FOR INTERFERENCE SIMULATION

Forward Motion	Disabled
Duration	11
Height	[1.2, 1.8]
Relative Velocity	[0.2, 1.0]
Radar Location	(0,10,0)
λ	$\frac{c}{24.17GHz}$
Range Resolution	0.01

TABLE II
TRAINING PARAMETERS

Batch Size	4
Optimizer	AdamW
Initial Learning Rate	1e-3
Scheduler	ReduceLROnPlateau
Scheduler Patience	32

The temporal quadrature signal needs to be transformed into a two-dimensional time-frequency distribution. Here, STFT is applied and the output distribution can be interpreted as the levels of energy reflected at given Doppler frequencies for a range of time instants. A Blackman window of 200 ms (with 120 ms overlap) is applied with 128 spectral bins to obtain STFT of a whole signal. Then, a segment of 128 instantaneous spectra corresponding to 10.24 seconds is randomly selected. Thus, a complex 128 by 128 STFT representation is obtained with the size appropriate for the network input.

B. Human Walking Simulator

The simulator is based on a common model derived from experimental bio-mechanical data [24] and the scripts therein are based on the MATLAB scripts written by Chen [21]. A more flexible version of the simulator¹ has been implemented in Python allowing for direct calls and integration within the PyTorch framework. The simulator allows the specification of the configuration file selecting the limbs to be used for synthesis, in turn allowing for a more versatile generation of signals, for example, the swinging of the arms only. Additionally, a script for the generation of static datasets containing human walking radar returns is also featured.

The Python framework can be configured to synthesise functions with a desired sampling rate and duration of the signal, as well as velocities and heights (or uniform sampling ranges thereof). The simulator produces a range-time map, and since the vital signs data is derived from a CW radar, all of the range bins are integrated, creating a one-dimensional signal. To increase the speed of training, a dataset of 1024 samples is synthesised beforehand. In this work, the dataset contains samples with the only interference being the left and right foot walking motions with no net forward motion of the human target. Simulation parameters are listed in Table I.

C. Vital Signs Combined with Interference Motion

The body motion radar return is simulated at 100 Hz for 11 seconds to match the length of the vital signs STFT. After both one-dimensional signals are normalized to zero-mean and unit-variance, as shown in Fig. 1, the interference signal is multiplied by a factor yielding a specific value of SIR. The signals with adjusted power are then subject to a STFT with the same parameters. A random 128-unit long temporal segment is then extracted from each STFT and both segments are summed to create a 128 by 128 STFT image.

D. Architecture

The architecture shown in Fig. 2 is designed to operate on complex STFT input with two channels for the real and imaginary component. This way, both channels contain information about the magnitude and phase of the STFT, which has been experimentally found to be beneficial. The encoder consists of 5 convolutional layers with depths of 32, 64, 128, 256, and 512 respectively. The output of the last convolutional layer is flattened and passed through two fully connected variational segments with a width of 128, one for the mean estimation and the other for the variance estimation. The latent code is obtained by sampling from the distribution defined by the output of the two estimation units. The decoder takes in the latent code and reconstructs the output image.

E. Training

The network is trained in a single stage (details in Table II) for 1024 epochs using a sum of two losses - a reconstruction loss and a variational loss. The weight of the latter has been reduced to 10^{-6} from the value suggested in [25] to promote the reconstruction loss during training. The number has been selected experimentally; higher tested values would result in a very low Kullback-Leibler Divergence (KLD) loss but the output of the network would be noisy. Checkpointing was applied to restore the weights from epoch 779 that resulted in the lowest validation loss.

¹<https://doi.org/10.5281/zenodo.4245158>

Fig. 3. (a) Unseen sample containing vital signs mixed with motion interference (b) Corresponding network output. (c) The ground truth reference. The brightness corresponds to the magnitude of STFT.

IV. RESULTS

The feasibility of the presented approach is demonstrated in two contexts. Firstly, it is shown that the designed network is capable of generalisation beyond the training dataset and can produce outputs with low loss for unseen samples. This intermediate domain is convenient because it is not difficult to compute the loss of the network output and hence provide a direct measure of the system performance. This context is not sufficient in the more general scope where the monitoring of vital signs is of interest. To address that, the second part of the results illustrates how the respiration rate estimation can be carried out using the network output.

A. Network Operation

The operation of the network can be illustrated in Fig. 3. The network takes a complex STFT sample in quadrature format containing interference motions (a) and estimates the vital signs contribution to the STFT (b). Note that the images shown in Fig. 3 correspond to the magnitude of the complex samples. Some degree of magnitude blurring is apparent in the network output image, however, as it is demonstrated in Section IV-B, this does not prevent a successful extraction of the respiration rate.

Fig. 4. Reconstruction loss influence comparison

The performance of the network can be further summarised by the results illustrated in Fig. 4. The network is evaluated by computing a metric of interest for a range of SIR values (0 to -9 dB) and for a range of levels of Gaussian noise (standard deviation between 0.0 to 0.45) added to the I and Q channels independently. A direct way of measuring the network performance would be to do this for the mean reconstruction loss. The resulting horizontal gradient observed in Fig. 4 demonstrates that the level of added noise is the primary influence over the loss. The minimal differences along the vertical axis suggest that the network can successfully remove the interference at various levels of SIR while still producing a reconstruction of equivalent quality.

B. Respiration Rate Extraction

The utility of using the network for micro-Doppler oriented applications can be assessed by testing the same respiration rate extraction technique for three STFT representations: 1) a clean vital signs sample 2) sample with added interfering motion contribution 3) the image of 2) processed by the network. The applied metric is the frequency bin error E_b that is defined as the absolute difference between the output of function $f(x_{[t]})$ for a reference signal $r_{[t]}$ and for a signal derived from the network output $d_{[t]}$. The network output is a complex matrix and $d_{[t]}$ is obtained by summing all of the Doppler bins.

$$E_b(d_{[t]}, r_{[t]}) = |f(d_{[t]}) - f(r_{[t]})| \quad (5)$$

For a 1-D signal $x_{[t]}$, the function $f(x_{[t]})$ measures the location of the dominant spectrum peak of a temporal signal:

$$f(x_{[t]}) = |f_{DC} - \text{argmax}(|\mathcal{F}\{x_{[t]}\}|)| \quad (6)$$

where f_{DC} indicates the index of the DC component in the Fourier Transform of the signal $\mathcal{F}\{x_{[t]}\}$.

Fig. 5 illustrates how the mean value of E_b changes for three input types (clean vital signs, mixture with interference, and mixture processed by the network) over the same range of SIR and additive noise level as in the Fig. 4. The scale

Fig. 5. Bin error influence comparison. (a) Clean vital signs sample. (b) Mixture sample. (c) Mixture sample processed by the network. Logarithmic scaling is applied to the displayed error values.

and extent are preserved among the three figures, where the brightness of each segment corresponds to the value of mean E_b computed for the entire validation dataset. Note that the error values have been converted to logarithm scale in order to make the three graphs easier to compare.

The results for the clean STFT could, in fact, be presented in one dimension, since the input remains the same regardless of the SIR level. However, due to the probabilistic nature of the network, the output for each sample may vary a little at each iteration of the sweep. This is manifested by minute changes of bin error along the vertical axis. Still, the most significant influence over the clean STFT extraction error is the level of added Gaussian noise. The error logarithm stays close to -1 before it begins to increase for the additive noise σ values above 1.0.

In contrast, Fig. 5(b) represents the error computed for the same extraction technique with an STFT input containing a mixture of the vital signs and the interfering motion. The dominance of the added interference influence is represented by a significant vertical gradient and the highest levels of error (error logarithm near 4.0). A slight increase due to levels of added noise is observed for the highest levels of tested SIR but this effect becomes negligible as the SIR increases.

Finally, Fig. 5(c) illustrates the error achieved for an input containing the mixed sample processed by the network. The error remains approximately constant for the tested range of SIR values. The consistent increase of the error logarithm for higher levels of noise (from -1 to around 0) demonstrates that the network derives the output from the relevant features in the input. Furthermore, the maximum value of mean E_b in Fig. 5(c) is significantly lower than in Fig. 5(a) demonstrating the noise suppression property of the network.

V. CONCLUSION

The presented contribution demonstrates the first application of a variational encoder-decoder CNN for interference removal from radar returns containing vital signs of human targets. The evaluation has been carried out in a semi-experimental mode, where an open real dataset source has been combined

with simulated human walking motions. The realised SIR sweeps demonstrate that the network reduces the influence of the interference motion on the vital sign extraction while displaying sensitivity to global additive noise. The latter observation suggests that the network does indeed derive the output from the input features instead of trivialising the task to generation of likely samples. The approach presented in this manuscript could be extended to more challenging interference scenarios and consequently, form a basis for a fully experimental framework.

REFERENCES

- [1] M. Alizadeh, G. Shaker, J. C. M. D. Almeida, P. P. Morita, and S. Safavi-Naeini, "Remote monitoring of human vital signs using mm-Wave FMCW Radar," *IEEE Access*, vol. 7, pp. 54 958–54 968, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2912956.
- [2] D. Yang, Z. Zhu, and B. Liang, "Vital Sign Signal Extraction Method Based on Permutation Entropy and EEMD Algorithm for Ultra-Wideband Radar," *IEEE Access*, vol. PP, p. 1, 2019. DOI: 10.1109/ACCESS.2019.2958600.
- [3] N. Petrochilos, M. Rezk, A. Høst-Madsen, V. Lubecke, and O. Boric-Lubecke, "Blind separation of human heartbeats and breathing by the use of a Doppler radar remote sensing," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 1, 2007, ISBN: 1424407281. DOI: 10.1109/ICASSP.2007.366684.
- [4] I. Mostafanezhad, E. Yavari, O. Boric-Lubecke, V. M. Lubecke, and D. P. Mandic, "Cancellation of unwanted doppler radar sensor motion using empirical mode decomposition," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1897–1904, 2013, ISSN: 1530437X. DOI: 10.1109/JSEN.2013.2238376.
- [5] L. J. Dirksmeyer, D. Schmiech, A. Marnach, and A. R. Diewald, "Separation of Two Close Targets in CW-Radar Measurement in the Example of Respiration Monitoring," vol. 1, 2019.

- [6] C. Ye, K. Toyoda, and T. Ohtsuki, "Blind Source Separation on Non-Contact Heartbeat Detection by Non-Negative Matrix Factorization Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 482–494, 2020, ISSN: 15582531. DOI: 10.1109/TBME.2019.2915762.
- [7] C. Ye, G. Gui, and T. Ohtsuki, "Deep clustering with lstm for vital signs separation in contact-free heart rate estimation," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.
- [8] P. Wang, F. Qi, M. Liu, F. Liang, H. Xue, Y. Zhang, H. Lv, and J. Wang, "Noncontact Heart Rate Measurement Based on an Improved Convolutional Sparse Coding Method Using IR-UWB Radar," *IEEE Access*, pp. 1–1, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2950423. [Online]. Available: <https://ieeexplore.ieee.org/document/8887436/>.
- [9] C. Gu, J. Wang, and J. Lien, "Deep Neural Network based Body Movement Cancellation for Doppler Radar Vital Sign Detection," 2019, pp. 1–3, ISBN: 9781728107165. DOI: 10.1109/iee-iws.2019.8803973.
- [10] F. Zhu, K. Wang, and K. Wu, "Doppler Radar Techniques for Vital Signs Detection Featuring Noise Cancellation," *IEEE MTT-S 2019 International Microwave Biomedical Conference, IMBioC 2019 - Proceedings*, pp. 25–28, 2019. DOI: 10.1109/IMBIOC.2019.8777824.
- [11] C. Li and J. Lin, "Random body movement cancellation in doppler radar vital sign detection," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, Dec. 2008, pp. 3143–3152. DOI: 10.1109/TMTT.2008.2007139.
- [12] F. K. Wang, T. S. Horng, K. C. Peng, J. K. Jau, J. Y. Li, and C. C. Chen, "Single-antenna doppler radars using self and mutual injection locking for vital sign detection with random body movement cancellation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 12 PART 2, pp. 3577–3587, 2011, ISSN: 00189480. DOI: 10.1109/TMTT.2011.2171712.
- [13] C. Gu, G. Wang, Y. Li, T. Inoue, and C. Li, "A hybrid radar-camera sensing system with phase compensation for random body movement cancellation in doppler vital sign detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4678–4688, 2013, ISSN: 00189480. DOI: 10.1109/TMTT.2013.2288226.
- [14] Q. Lv, L. Chen, K. An, J. Wang, H. Li, D. Ye, J. Huangfu, C. Li, and L. Ran, "Doppler Vital Signs Detection in the Presence of Large-Scale Random Body Movements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 9, pp. 4261–4270, 2018, ISSN: 00189480. DOI: 10.1109/TMTT.2018.2852625.
- [15] L. Pandey, A. Kumar, and V. Namboodiri, "Monaural audio source separation using variational autoencoders," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2018-Sept, 2018, pp. 3489–3493. DOI: 10.21437/Interspeech.2018-1140.
- [16] H. Kameoka, L. Li, S. Inoue, and S. Makino, *Supervised determined source separation with multichannel variational autoencoder*, 2019. DOI: 10.1162/neco_a_01217. arXiv: arXiv:1808.00892v3.
- [17] S. Seki, H. Kameoka, L. Li, T. Toda, and K. Takeda, "Underdetermined Source Separation Based on Generalized Multichannel Variational Autoencoder," *IEEE Access*, vol. 7, pp. 168 104–168 115, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2954120.
- [18] E. Karamatli, A. T. Cemgil, and S. Kirbiz, "Audio Source Separation Using Variational Autoencoders and Weak Class Supervision," *IEEE Signal Processing Letters*, vol. 26, no. 9, pp. 1349–1353, 2019, ISSN: 1070-9908. DOI: 10.1109/lsp.2019.2929440. arXiv: 1810.13104.
- [19] W. N. Hsu, Y. Zhang, and J. Glass, "Learning latent representations for speech generation and transformation," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2017-Augus, 2017, pp. 1273–1277. DOI: 10.21437/Interspeech.2017-349. arXiv: 1704.04222.
- [20] M. Czerkawski, C. Ilioudis, C. Clemente, C. Michie, I. Andonovic, and C. Tachtatzis, "On Models and Approaches for Human Vital Signs Extraction from Short Range Radar Signals," *14th European Conference on Antennas and Propagation, EuCAP 2020*, 2020. DOI: 10.23919/EuCAP48036.2020.9135189.
- [21] V. Chen, *The Micro-doppler Effect in Radar*, ser. Artech House radar library. Artech House, 2011, ISBN: 9781608070589. [Online]. Available: <https://books.google.co.uk/books?id=eJ7eMHpxt30C>.
- [22] K. Shi and S. Schellenberger, *Guardian vital sign data*, Aug. 2019. DOI: 10.6084/m9.figshare.c.4633958. [Online]. Available: https://figshare.com/collections/GUARDIAN_Vital_Sign_Data/4633958/1.
- [23] K. Shi, S. Schellenberger, C. Will, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelpin, "A dataset of radar-recorded heart sounds and vital signs including synchronised reference sensor signals," *Scientific Data*, vol. 7, no. 1, pp. 1–12, 2020, ISSN: 20524463. DOI: 10.1038/s41597-020-0390-1.
- [24] R. Boulic, N. M. Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification," *The Visual Computer*, vol. 6, no. 6, pp. 344–358, 1990, ISSN: 1432-2315. DOI: 10.1007/BF01901021. [Online]. Available: <https://doi.org/10.1007/BF01901021>.
- [25] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, no. M1, pp. 1–14, 2014. arXiv: 1312.6114.