ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447570

CONTACTLESS RADAR HEART RATE VARIABILITY MONITORING VIA DEEP SPATIO-TEMPORAL MODELING

Haoyu Wang, Jinbo Chen, Dongheng Zhang, Zhi Lu, Changwei Wu, Yang Hu, Qibin Sun, Yan Chen*

University of Science and Technology of China, Hefei, China

ABSTRACT

Radar sensing has been a promising solution for contactless monitoring of Heart Rate Variability (HRV), an essential indicator of the cardiovascular and autonomic nervous systems. However, existing works neglect heartbeat-driven body surface motions spreading across the entire body with spatial variations, which limits their accuracy in identifying fine-grid consecutive heartbeat timings and overall HRV performance. In this paper, we propose to exploit the entire body reflections and model the inherent spatial-temporal relationship between these reflections and heartbeats by deep neural network for contactless HRV monitoring. Specifically, a hybrid convolution-transformer-based network is designed to convert the complex multi-dimensional spatial-temporal modeling problem into an efficient sequence modeling process. Experimental results demonstrate its superiority over the baseline method, achieving the median IBI estimation error of 12ms (w.r.t. 98.47% accuracy), RMSDD error of 7.3ms, SDRR error of 2.9ms, pNN50 error of 5.5%.

***Index Terms*—** HRV, Radar Sensing, Deep Learning

1. INTRODUCTION

Heart rate variability (HRV), the variation of the periods between Inter-Beat Intervals (IBI), is identified as an essential indicator of the overall health status of an individual [1, 2]. HRV analysis provides insights into both cardiac health and the Autonomic Nervous System (ANS) [2, 3]. Hence, the development of more convenient, comfortable, and effective HRV monitoring technology has always received significant attention in both academia and industry. Among these technologies, radar sensing has emerged as a promising solution [4, 5, 6, 7, 8, 9, 10, 11] due to the contactless, privacy-preserving, and comfortable user experience.

In these works, various efforts are focusing on utilizing signal processing algorithms to model the relation between RF signals and heartbeat motion [6, 7, 8]. Wang *et al.* [7] propose to optimize the decomposition of the phase of the channel information modulated by the chest movement, thus estimating the heartbeat signal and further evaluating the HRV metrics. Also, with the rapid growth of deep learning models, the number of related works has increased significantly on this topic [4, 5, 9]. Zhang *et al.* [9] propose to decompose nonlinear signal mixing and recover fine-grained heartbeat waveform by leveraging variational encoder-decoder network design.

Typically, radar-based HRV sensing works by separating the reflected signal from a particular place on the body surface, recovering the heartbeat waveform modulated in the signal, and finally calculating HRV metrics from the waveform. However, as highlighted in [12], the inertial cardiac mechanical activity drives the entire torso surface to move in the same heartbeat rhythms but with spatially variant morphology, rather than being a one-dimensional motion at a single body surface place. This indicates that previous methods may be ambiguous in identifying fine-grid consecutive heartbeat timings by considering one-dimensional movement and overlooking the relationship between others on the torso surface. Therefore, for accurate HRV monitoring, we need to jointly consider reflected signals from the entire body surface.

However, addressing the above issue in radar sensing is complex due to two main challenges. Firstly, unlike vision-based photoplethysmography [13] which easily identifies skin regions, radar sensing struggles to accurately separate reflections from the entire torso surface. Secondly, modeling the relationship between the spatially variant body surface motions and the heartbeat timing is challenging, particularly given the notable noise and interference in the radio signal domain.

In this paper, we propose to exploit the entire body reflection signals and model the inherent spatial-temporal relationship between these reflections and heartbeats by deep learning network for contactless HRV monitoring. Specifically, we extract coarse-grid voxel signals around the target body from raw signals and construct the spatial-temporal cardiac signals representation which includes the entire body surface heartbeat movements with spatial redundancy. Then, a hybrid convolution-transformer-based network is designed to convert the complex multi-dimensional spatial-temporal modeling problem into an efficient sequence modeling process. The HRV metrics are finally calculated based on the heartbeat waveform generated by the network. To substantiate the effectiveness of our proposal, we collect a dataset consisting

Corresponding author: Zhi Lu. This work was supported by the National Natural Science Foundation of China (62302471), the National Key R&D Programmes (2022YFC2503405, 2022YFC0869800), and the Postdoctoral Fellowship Program of CPSF (GZC20232565).

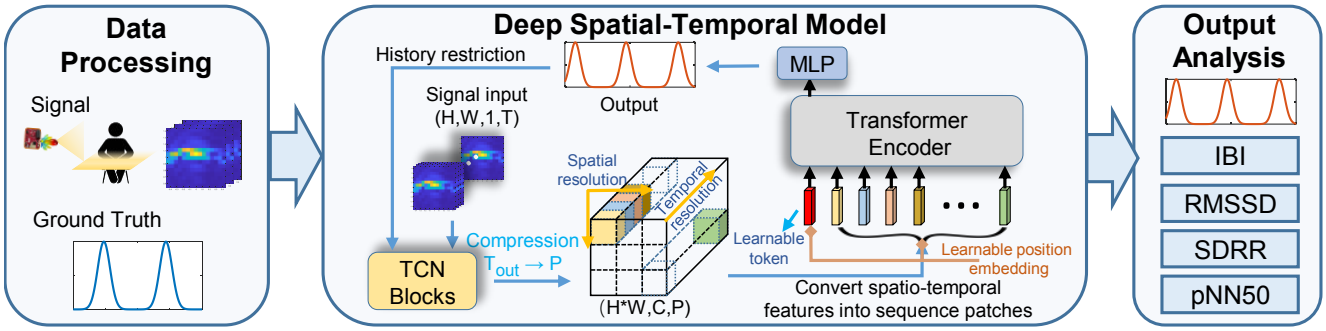


Fig. 1: System overview. First, we construct cardiac spatio-temporal signal representation from the raw data. Then, our deep spatial-temporal model transforms the complex multi-dimensional problem into a sequence modeling process. Finally, we use the generated heartbeat waveform to compute HRV metrics.

of data from 8 different participants. Comparative analyses against state-of-the-art methods consistently demonstrate the superior performance achieved by our approach.

2. SYSTEM DESIGN

2.1. Cardiac Spatio-Temporal Signals Constructing

First, we need to extract cardiac spatial-temporal signals reflected from the entire body surface. The fundamental concept behind using radar sensors for detecting cardiac micro-motions involves extracting the phase variation of received signals reflected from the target. The phase model is mathematically defined as follows:

$$\phi(t) = 2\pi \frac{d(t)}{\lambda}, \tag{1}$$

where λ represents the wavelength of the RF signals, $d(t)$ signifies the distance between the radar sensor and the reflection objects, and t stands for sensing time.

However, it is complex to precisely identify and extract these signals from the received RF signals. Instead, we first extract signals from the target with spatial redundancy to ensure the entire body reflections are included. Then, we emphasize the heartbeat features and suppress other interferences in these signals by temporal filtering. This streamlined processing approach helps us construct a spatial-temporal cardiac signals representation, ensuring that all reflections related to the heartbeat are included with spatial-temporal details embedded.

Specifically, we opt for a coarse-grained spatial range to ensure the inclusion of the entire body reflections using the beamforming and localization techniques in [14]. The beamforming calculation is expressed as:

$$S_{(x,y,t)} = \sum_{n=1}^{N} \sum_{t_f=1}^{T} y_{n,t_f} e^{j2\pi(t_f-1)T_s k \frac{r(x,y,n)}{c}} e^{j2\pi \frac{r(x,y,n)}{\lambda}}, \tag{2}$$

where N is the number of virtual channels, y_{n,t_f} is the received signal at t_f (ADC sampling points within the chirp), T_s is the ADC sampling period, k is the frequency slope, λ is the wavelength, $r(x,y,n)$ is the distance from voxel (x,y) to the antenna pair in channel n, and t is the frame time.

Then, we utilize a second-order differentiator [8] to eliminate noise signals while retaining the heartbeat signals for each voxel. The differentiator is defined as:

$$s_0'' = \frac{(s_{-3}+s_3)+2(s_{-2}+s_2)-(s_{-1}+s_1)-4s_0}{16h^2}, \tag{3}$$

where s_0'' refers to the second derivative at a particular time sample, s_i is the value of the time series i samples away, h is the frame periodicity between consecutive samples.

Finally, we can construct the cardiac spatio-temporal signals in the form of $X_{in} \in R^{H\times W\times L}$. H and W denote the dimensions of the 2D spatial size, L denote temporal length.

2.2. Deep Spatial-Temporal Model

Modeling cardiac spatial-temporal signals with deep neural networks is challenging. This challenge arises from the explosive network input size due to high-frequency radar sampling and the need for long-duration signals with redundant spatial sizes in the modeling process. To address this, we designed a hybrid convolutional-transformer-based neural network architecture as shown in Fig. 1.

Specifically, we first employ a conditional Temporal Convolutional Network (TCN) [15, 16] to conduct feature extraction and compression in the temporal dimension. Within the TCN, Gated Activation Units are used to incorporate the influence of historical features:

$$z = \tanh(W_{f,k} * x + V_{f,k} * f(h)) \odot \sigma(W_{g,k} * x + V_{g,k} * f(h)), \tag{4}$$

where $*$ denotes convolution, $\odot$ represents element-wise multiplication, $\sigma(\cdot)$ is the sigmoid function, k is the layer index, f and g correspond to the filter and gate, W denotes a learnable convolution filter, and $V_{f,k}$ and $V_{g,k}$ are learnable linear projection matrices. $f(h)$ maps the historical sequence to the input feature with the same dimensions.

Then, we can model the entire process as follows:

$$p(X_f|X_{in}) = \prod_{t=1}^{T} p(X_f|y_1, ..., y_{n-1}, X_{in}), \tag{5}$$

where the feature X_f conditions on the sequence ground truth at previous timesteps and the current input X_{in}.

To clearly utilize the physical interpretations of spatial-temporal information, inspired by Vision Transformers [17, 18], we convert multidimensional data into sequence-like patches. Then, we employ a transformer model to capture complex relationships between these patches. Finally, we project these representations to obtain fine-grained heartbeat information. Specifically, given the TCN output as $X_f \in R^{H\times W\times C\times T_{out}}$, where C is the number of output channels, and T_{out} represents the feature length with a global temporal view, we split the T_{out} dimension into P patches using sum-pooling. We add a learnable token to capture the global context, and the spatio-temporal patches are mixed. This process yields mixed spatio-temporal patches $X_{st} \in R^{H*W*P+1\times C}$. After adding position embeddings, we obtain global spatio-temporal features as:

$$X_w = Attention(X_{st}, X_{st}, X_{st}) \in R^C. \quad (6)$$

Following this, a simple two-layer MLP is utilized as:

$$X_{out} = \sigma(X_w \cdot W_1^T + b_1) \cdot W_2^T + b_2, \quad (7)$$

where, W_1 and W_2 are learnable matrices, and σ represents the activation function. X_{out} represents the final output.

Our task involves precise heartbeat timing rather than physiology waveform reconstruction. Inspired by [19], we use the heartbeat waveform as ground truth which is generated by interpolating the time points corresponding to the heartbeat events (R peaks in electrocardiogram, ECG) with Gaussian distribution, as shown in Fig. 2. For the final output, a simple peak-finding process is used to locate heartbeat timing.

3. EXPERIMENTS AND RESULTS

3.1. Implementation

Our non-contact HRV monitoring system employs the TI AWR1843 millimeter-wave radar, equipped with 2 transmitters and 4 receivers. Operating at 100 Hz, it features a 65MHz/us signal slope and a 3.32 GHz bandwidth. The ground truth ECG data is acquired using the TI ADS1292 evaluation board. Our system is implemented in PyTorch, optimized with Adam and L1 loss, employing a learning rate of 0.001 and a mini-batch size of 32. Key hyperparameters include L = 640, C = 128, T_{out} = 128, X_{out} dimension = 100, $H = W$ = 9 (a 0.5×0.5m region), and the TCN with 9 stacks utilizing a dilation factor of 2. Within the Transformer block, we employ an attention dimension of 32 and 4 attention heads.

Experimental setting: To evaluate our system, we conduct experiments with 8 participants. The radar data and ECG signals are collected simultaneously. During experiments, participants are asked to sit and breathe naturally. We conduct measurements when subjects sit at different distances from radar, specifically at 1m, 2m, 3m, and 4m. At each distance, we collect three segments of data, totaling nine minutes. In total, we collect data for nearly 5 hours. Finally,

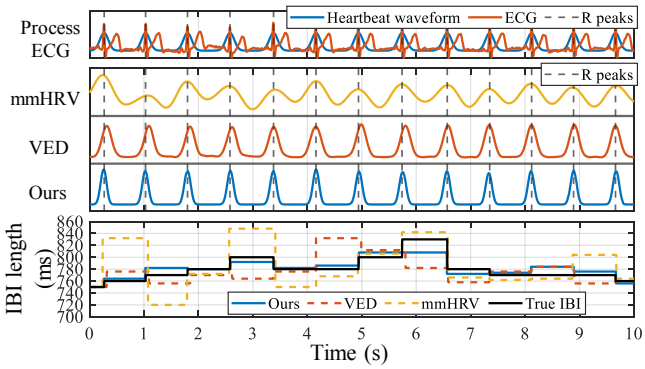


Fig. 2: Results and ground truth visualization. The first row displays the true ECG signal alongside its corresponding heartbeat waveform ground truth. Subsequent rows show results from different methods. The final row presents the IBI results along the sampling dimension.

75% of the data (6 subjects) are used to train the model, then the model is tested upon the remaining 2 subjects.

Baseline: We select mmHRV [7] and VED [9] as baseline methods for comparison. mmHRV is the SOTA radar-based HRV analysis achieved by signal processing design. VED demonstrates promising performance in recovering fine-grid heartbeat waveforms by using a novel deep generative model.

Metrics: We use the following metrics:

- *Absolute IBI Error*: The absolute difference between predicted IBI and its corresponding ground truth.
- *Relative Error*: The ratio between the absolute IBI error and the ground truth value.
- *RMSSD*: Root mean square of successive differences.
- *SDRR*: Standard deviation of all the IBIs.
- *pNN50*: A measure of the percentage of consecutive IBIs differing by more than 50 ms.

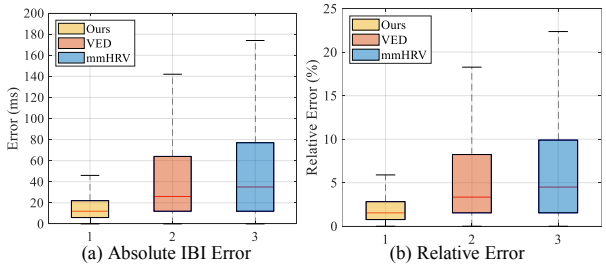
To ensure a fair performance comparison and prevent ambiguity in IBI error calculations caused by differing heartbeat counts in the generated and ground truth, we calculate IBI along time sampling dimension rather than heartbeat sequences, as shown in Fig. 2. Specifically, at each time point, the IBI is determined by the period between the two adjacent waveform peaks. For period-based HRV metrics of RMSDD, SDRR, pNN50, evaluation is achieved through a 60-second sliding window with a 5-second step.

3.2. Performance Evaluation

Fig. 2 demonstrates the output of various methods. Meanwhile, Fig. 3 illustrates the overall IBI estimation accuracy of the three different methods at all distances. mmHRV, VED and our system achieve median absolute IBI errors of 35ms, 26ms, 12ms, and median relative errors of 4.5%, 3.3%, 1.5%. Our method achieves the best overall performance. This is because deep learning algorithms tend to extract more heartbeat features compared to traditional signal processing meth-

Metrics			IBI Error(ms) ↓			RMSSD Error(ms) ↓			SDRR Error(ms) ↓			pNN50 Error(%) ↓		
Methods			mmHRV	VED	Ours	mmHRV	VED	Ours	mmHRV	VED	Ours	mmHRV	VED	Ours
Distance (cm)	100	Mean	19.24	25.45	**12.41**	22.94	19.48	**5.56**	10.39	12.02	**1.57**	19.94	11.00	**5.07**
		Median	12.00	16.00	**10.00**	14.95	16.85	**5.70**	4.75	7.20	**1.30**	13.36	10.14	**5.37**
	200	Mean	52.05	50.51	**18.12**	39.56	45.76	**13.39**	16.44	38.55	**5.38**	35.27	20.08	**6.20**
		Median	38.00	26.00	**12.00**	41.20	37.60	**8.65**	15.85	37.00	**3.25**	35.39	20.38	**4.69**
	300	Mean	61.36	61.74	**18.14**	49.45	50.40	**9.96**	25.97	36.29	**4.40**	39.93	23.05	**7.37**
		Median	51.00	40.00	**12.00**	49.90	51.59	**9.40**	25.95	35.65	**3.30**	39.20	23.29	**7.35**
	400	Mean	68.66	68.22	**24.44**	36.27	49.68	**17.89**	22.65	32.18	**12.40**	29.86	21.87	**12.60**
		Median	59.00	44.00	**16.00**	33.55	56.30	**13.10**	21.65	29.35	**5.50**	28.63	20.59	**8.69**
	All	Mean	50.21	50.93	**18.24**	36.99	40.86	**11.66**	18.83	29.45	**5.80**	31.26	18.79	**7.78**
		Median	35.00	26.00	**12.00**	37.25	37.15	**7.30**	18.15	21.25	**2.90**	31.28	15.89	**5.53**

Table 1: HRV monitoring performance at different distances under different methods.



(a) Absolute IBI Error (b) Relative Error

Fig. 3: Overall performance of IBI.

ods, and the entire body reflections can be processed by deep spatio-temporal models to extract more precise heartbeat timing information.

Table 1 presents the mean and median HRV-related metrics for all methods across various distances. Several noteworthy observations can be discerned. Firstly, as the distance increases, both the median and mean IBI errors for all three methods tend to rise. This can be attributed to the decrease in signal-to-noise ratio with increasing distance. Notably, mmHRV and VED appear more sensitive to distance, possibly because limited body reflection performs poorly when signal quality decreases. Furthermore, our approach outperforms others in HRV metrics, including RMSSD, SDRR, and pNN50, displaying more stable performance across different distances. Also, our performance exhibits greater stability in terms of both mean and median errors. These underscore the robustness of our approach in accurately extracting heartbeat timings.

In summary, our system consistently demonstrates superior performance across varying distances, showcasing its ability to reliably extract intricate HRV-related features.

3.3. Ablation Study

To demonstrate the importance of considering the entire body reflections in radar-based HRV monitoring rather than the single-place body reflection, we first conduct the ablation experiment by comparatively using the cardiac spatio-temporal signals or one-dimensional cardiac signals at a single body surface place described in [9] as the network input. As shown in Table 2, the performance significantly improves when using spatial-temporal signals that consider the entire body reflections.

Type of errors		IBI (ms) ↓	RMSSD (ms) ↓	SDRR (ms) ↓	pNN50 (%) ↓
Single reflection	Mean	42.53	16.44	14.50	10.14
	Median	26.00	11.40	9.10	8.70
Entire body reflections	Mean	**18.24**	**11.66**	**5.80**	**7.78**
	Median	**12.00**	**7.30**	**2.90**	**5.53**

Table 2: Ablation study of cardiac signal representation.

We also analyze the effectiveness of the spatial modeling in the network. Specifically, we conduct the experiment without spatial modeling by averaging transformer input along the spatial dimension to blur the spatial information in X_f. The result shown in Table 3 demonstrates our network design effectively models spatial information and achieves superior performance.

Type of errors		IBI (ms) ↓	RMSSD (ms) ↓	SDRR (ms) ↓	pNN50 (%) ↓
Without spacial modeling	Mean	22.38	13.61	8.50	9.79
	Median	16.00	8.15	4.20	7.35
With spacial modeling	Mean	**18.24**	**11.66**	**5.80**	**7.78**
	Median	**12.00**	**7.30**	**2.90**	**5.53**

Table 3: Ablation study of spatial modeling.

4. CONCLUSIONS

In this paper, we presented a novel system designed to enhance HRV monitoring by efficiently extracting information from the entire torso surface. Leveraging entire body reflections and a deep spatio-temporal network, we have achieved the best results in a fair HRV evaluation approach. Furthermore, we emphasized the significance of using entire body reflections and spatial modeling for HRV analysis, demonstrating its superior performance. This system holds promise for non-contact HRV monitoring, with potential applications in early cardiovascular disease diagnosis and stress assessment.

5. REFERENCES

[1] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, pp. 1031–1051, 2006.

[2] Borejda Xhyheri, Olivia Manfrini, Massimiliano Mazzolini, Carmine Pizzi, and Raffaele Bugiardini, "Heart rate variability today," *Progress in cardiovascular diseases*, vol. 55, no. 3, pp. 321–331, 2012.

[3] Fred Shaffer and Jay P Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[4] Jinbo Chen, Dongheng Zhang, Zhi Wu, Fang Zhou, Qibin Sun, and Yan Chen, "Contactless electrocardiogram monitoring with millimeter wave radar," *IEEE Transactions on Mobile Computing*, 2022.

[5] Unsoo Ha, Sohrab Madani, and Fadel Adib, "Wistress: Contactless stress monitoring using wireless signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–37, 2021.

[6] Vladimir L Petrović, Milica M Janković, Anita V Lupšić, Veljko R Mihajlović, and Jelena S Popović-Božović, "High-accuracy real-time monitoring of heart rate variability using 24 ghz continuous-wave doppler radar," *Ieee Access*, vol. 7, pp. 74721–74733, 2019.

[7] Fengyu Wang, Xiaolu Zeng, Chenshu Wu, Beibei Wang, and KJ Ray Liu, "mmhrv: Contactless heart rate variability monitoring using millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 22, pp. 16623–16636, 2021.

[8] Mingmin Zhao, Fadel Adib, and Dina Katabi, "Emotion recognition using wireless signals," in *Proceedings of the 22nd annual international conference on mobile computing and networking*, 2016, pp. 95–108.

[9] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo, "Can we obtain fine-grained heartbeat waveform via contact-free rf-sensing?," in *IEEE INFOCOM 2022-IEEE conference on computer communications*. IEEE, 2022, pp. 1759–1768.

[10] You Ran, Dongheng Zhang, Jinbo Chen, Yang Hu, and Yan Chen, "Contactless blood pressure monitoring with mmwave radar," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 541–546.

[11] Dongheng Zhang, Yang Hu, Yan Chen, and Bing Zeng, "Breathtrack: Tracking indoor human breath status via commodity wifi," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3899–3911, 2019.

[12] Jinbo Chen, Dongheng Zhang, Dong Zhang, Qibin Sun, and Yan Chen, "Mmcamera: an imaging modality for future rf-based physiological sensing," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 894–896.

[13] Sinh Huynh, Rajesh Krishna Balan, JeongGil Ko, and Youngki Lee, "Vitamon: measuring heart rate variability using smartphone front camera," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 1–14.

[14] Dongheng Zhang, Yang Hu, and Yan Chen, "Mtrack: Tracking multiperson moving trajectories and vital signs with radio signals," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3904–3914, 2020.

[15] Colin Lea, Rene Vidal, Austin Reiter, and Gregory D Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 47–54.

[16] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[19] Huazhang Hu, Sixun Dong, Yiqun Zhao, Dongze Lian, Zhengxin Li, and Shenghua Gao, "Transrac: Encoding multi-scale temporal correlation with transformers for repetitive action counting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19013–19022.