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/ICASSP48485.2024.10447883

ROBUST DOA ESTIMATION FROM DEEP ACOUSTIC IMAGING

Adrian S. Roman[1]* *Iran R. Roman*[2] *Juan P. Bello*[2]

[1] Viterbi School of Engineering, University of Southern California, California, USA
[2] Music and Audio Research Laboratory, New York University, New York, USA

ABSTRACT

Direction of arrival estimation (DoAE) aims at tracking a sound in azimuth and elevation. Recent advancements include data-driven models with inputs derived from ambisonics intensity vectors or correlations between channels in a microphone array. A spherical intensity map (SIM), or acoustic image, is an alternative input representation that remains underexplored. SIMs benefit from high-resolution microphone arrays, yet most DoAE datasets use low-resolution ones. Therefore, we first propose a super-resolution method to upsample low-resolution microphones. Next, we benchmark DoAE models that use SIMs as input. We arrive to a model that uses SIMs for DoAE estimation and outperforms a baseline and a state-of-the-art model. Our study highlights the relevance of acoustic imaging for DoAE tasks.

***Index Terms*—** spherical intensity maps, acoustic imaging, sound event localization, super-resolution.

1. INTRODUCTION

The release of large-scale sound event localization and detection (SELD) datasets [1, 2, 3] allowed for deep learning approaches for direction of arrival estimation (DoAE) in favor of traditional beamforming [4]. Today, the best-performing models use a combination of two inputs: 1) the intensity vectors of first order ambisonics (FOA) and 2) the correlations between channels in a tetrahedral microphone (4 channels) [5, 6]. However, a spherical intensity map (SIM) is an alternative input representation that remains unexplored.

A SIM is computed using delay-and-sum beamforming (DASB) [7]. DeepWave, an *acoustic imaging* model [8], builds upon DASB by applying sparse coding [9] to deblur a cleaner representation of spatial sound activity (see Figure 1). We evaluated whether we could use DeepWave's SIM to localize sound on the entire LOCATA dataset [2], and we found that it is possible to achieve great localization performance (see Table 1, top row). However, DeepWave assumes high-resolution inputs, such as those captured by the EigenMike (32 channels) [10] or Pyramic (48 channels) [11] arrays.

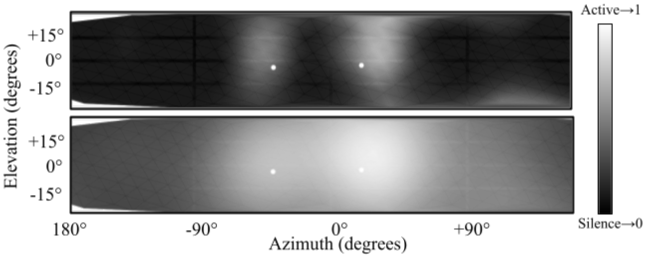

Fig. 1. The acoustic image by DeepWave (top) and DASB (bottom) for two sound sources. Dots denote ground truth.

In this study we investigate whether high-resolution SIMs can be used to carry out DoAE. Since most SELD datasets use low-resolution arrays, we first demonstrate it is possible to upsample them using a super-resolution model. Next, we benchmark models that use SIMs as input to carry out DoAE, and compare against a baseline and open-source state-of-the-art (SoTA) model. Our code and data are openly-available.[1] In summary, our contributions are:

1. A method to upsample the covariance between channels in a low-resolution microphone to high-resolution.
2. A benchmark of models that use SIMs for DoAE, evaluated on the LOCATA [2] and STARSS23 [1] datasets.

2. RELATED WORK

Current DoAE models process the intensity vectors of FOA and correlations between channels in a tetrahedral microphone [5, 6]. Significant ones include the "Sound event localization and detection network" (SELDnet) [12], considered to be a baseline for the task, as well as SoTA models that use multi-head self-attention [5] and multi-task outputs [6].

A SIM is an acoustic image indicating the location of a sound source [13]. SIMs have not been assessed quantitatively as input for DoAE deep-learning models. The DASB algorithm [7] is the most commonly-used method to compute SIMs, but resulting images exhibit poor angular resolution [8]. DeepWave achieves an equivalent SIM using a *backprojection* operation, and further denoises it using *deblurring*,

$$x^{(\ell)} = \tanh\Big(\underbrace{[\bar{\mathbf{B}} \circ \mathbf{B}]^H \mathrm{vec}(\mathbf{\Sigma})}_{backprojection} + \underbrace{\mathbf{P}_\theta(\mathbf{L}) x^{(\ell-1)}}_{deblurring} - \tau\Big), \quad (1)$$

*corresponding author email: romanguz@usc.edu

[1] https://github.com/adrianSRoman/DeepWaveDOA

Authorized licensed use limited to: New York University. Downloaded on May 24,2024 at 13:56:14 UTC from IEEE Xplore. Restrictions apply.

where $\mathbf{\Sigma} \in \mathbb{C}^{M \times M}$ is the instantaneous covariance matrix of a microphone array with M channels. *Backprojection* is equivalent to DASB (sec. 5.2 in [14]). It maps $\mathbf{\Sigma} \in \mathbb{C}^{M \times M}$ onto a uniformly-tiled SIM (i.e. tesselated [15]), representing azimuth and elevation. $\text{vec}(\mathbf{\Sigma}) \in \mathbb{C}^{M^2}$ is a column-wise matrix-flattening operator, $\mathbf{B} \in \mathbb{C}^{M \times N}$ is a trainable matrix, and $\bar{\mathbf{B}}$ is its complex conjugate. $\circ$ is the Khatri-Rao product and H denotes a Hermitian matrix. *Deblurring* iterates over an initial random spherical map $x^{(0)} \in \mathbb{R}^N$ using graphical convolutions that clean it. $\mathbf{P}_\theta(\mathbf{L})$ is a polynomial of the graph Laplacian $\mathbf{L} \in \mathbf{R}^{NxN}$, and θ are graph convolution filters.

DeepWave applies these operations L times with tanh for sparsity. The trainable parameters are θ, $\mathbf{B}$ and τ (a bias term), and are shared across the L iterations. In practice, F of these operations happen in parallel, one for each frequency band. DeepWave's output acoustic image (or SIM) has N pixels, proportional to the number of microphone channels.

3. APPROACH

3.1. DoAE from acoustic images via K-means clustering

First we optimized a DeepWave model to get a sense of maximum performance on the LOCATA dataset (see Methods). This model processes a high-resolution microphone. We apply K-means clustering over DeepWave's SIMs to calculate the DoA. For a 32ch input with $F = 9$ and $L = 5$, DeepWave generates a $N = 242$ acoustic image with values ranging between 0 and 1. After arranging the N pixels in a 2D space $N = A \times E$ for azimuth (A) and elevation (E) we obtain $\mathbf{I} \in \mathbb{R}_{[0,1)}^{F \times A \times E}$. Next, we apply a 1D Tukey window along the elevation axis with a 0.8 tapering factor to remove artifacts closer to the poles. Finally, we run K-means clustering with $K = 3$ on the 15 pixels with the maximal intensity (all others are clipped to zero). This yields three centroids that represent DoA coordinates. To make this algorithm robust, we apply two post-processing heuristics: (1) ignore clusters where points are separated by more than 15° from the centroid, and (2) merge clusters with centroids within 15° of each other.

3.2. Upsampling $\mathbf{\Sigma}$ using super-resolution

DeepWave's resolution is proportional to the number of microphone channels. Lower-resolution microphones, such as 4 channel, are commonly used in SELD datasets. We introduce a channel upsampling method derived from the Deep Back Projection Network (DBPN) by Haris et al. [16], a computer vision super-resolution model. Our complex-valued DBPN (CDBPN) upsamples $\mathbf{\Sigma} \in \mathbb{C}^{4 \times 4} \rightarrow \mathbf{\Sigma} \in \mathbb{C}^{32 \times 32}$ (i.e. from a 4ch to a 32ch array; a factor of 8).

3.3. DeepWave SIMs as inputs for DoAE

The original DeepWave processes 100ms at a time and lacks temporal memory [8]. We also study the effect of adding a gated recurrent unit (GRU) on top. In other words, DeepWave's SIM becomes the GRU input (Figure 2). We train models in an end-to-end fashion using the ADPIT loss, which makes this type of model have the multi-ACCDOA representation to localize overlapping sound sources [17].

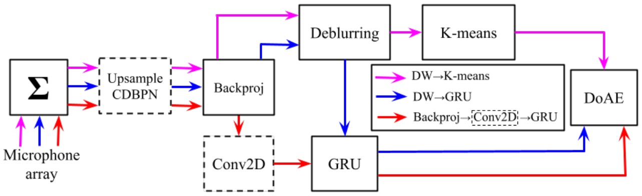

Fig. 2. Signal pipeline in the models we study.

4. METHODOLOGY

4.1. Datasets

Evaluations with real recordings. We evaluate models on LOCATA because it was recorded with an EigenMike (32ch) in a room where human actors moved while speaking (the microphone also moves in some recordings) [2]. To apply CDBPN upsampling we simulate a tetrahedral microphone (4ch) using the 6th, 10th, 22nd, and 26th EigenMike channels (following Adavane et al.'s [18] approach). Metadata does not indicate a speaker's gender, so annotations only have the instantaneous DoA (every 100ms). Our study focuses on DoAE, so these annotations are all we need (i.e. we do not classify sound events into categories). We also evaluate on STARSS23 [1], which contains more rooms and uses a 4ch microphone. We only use its DoA annotations, as we do not classify events into specific classes.

Simulated recordings for training. To train models we simulate 32ch recordings. Our simulated dataset was generated using the SpatialScaper library [19]. We integrated two RIR databases that use the EigenMike. The first one is METU-SPARG, collected at 240 points on a cubic grid in a classroom [20], and the second one is ARNI [21], collected in the "variable acoustics room" at Aalto University. We simulate soundscapes with speech from the FSD50K dataset [22] spatialized in either of the two rooms. We generate a total of 50min of data in the METU room and 10min in ARNI.

Models evaluated on STARSS23 are trained using the companion dataset with simulated 4ch recordings using RIRs from 12 rooms [3]. We refer to it as "DCASE-sim".

4.2. Models, ablations, and baselines

The two main models we study are DeepWave+K-means and DeepWave+GRU. We carry out an ablation study of the DeepWave+GRU model[2] to understand the role of the deblurring

[2] the original DeepWave paper has an ablation study without the GRU [8]

Authorized licensed use limited to: New York University. Downloaded on May 24,2024 at 13:56:14 UTC from IEEE Xplore. Restrictions apply.

Input	**Model / operations**	$LE \downarrow$ (std.)	$LR \uparrow$ (std.)
32ch	Backproj → Deblurring → K-means	**14.8^o** **(± 0.00)**	**99.20** **(± 0.00)**
	Backproj → Deblurring → GRU	20.9^o (± 3.63)	69.36 (± 2.15)
	Backproj → GRU	20.0^o (± 3.01)	70.46 (± 4.14)
	Backproj → Conv2D → GRU	15.96^o (± 2.51)	76.36 (± 9.21)
4ch → CDBPN → 32ch	Backproj → Deblurring → K-means	27.10^o (± 0.00)	**99.20** (± 0.00)
	Backproj → Deblurring → GRU	20.30^o (± 2.76)	72.33 ($\pm$ 0.04)
	Backproj → GRU	18.06^o (± 3.53)	70.96 (± 3.82)
	Backproj → Conv2D → GRU	**17.42^o** (± 0.73)	82.40 (± 5.76)
4ch	SELDnet23	16.8^o (± 0.45)	77.13 (± 3.84)
FOA	SELDnet23	22.43^o (± 3.80)	80.73 (± 10.98)
FOA + 4ch	EINV2	19.83^o (± 0.75)	80.66 (± 7.26)

Table 1. Localization error (LE) and recall (LR) on LOCATA. Scores reflect the average across three experimental replications. Note that the EINV2 model uses both tetrahedral microphone (4ch) and first order ambisonics (FOA) inputs.

operation given the additional GRU. Therefore we have three variants: 1) Backproj → Deblur → GRU, 2) Backproj → GRU, and 3) Backproj → Conv2D → GRU (i.e. we replace deblurring with a 2D convolution layer)[3].

Our optimizations are gradual steps to enable DoAE using a DeepWave backbone. DeepWave+K-means enables DoAE on SIMs, provided high-resolution array data. CDBPN introduces low-resolution array upsampling. A GRU on top of a DeepWave backbone allows end-to-end training using multi-ACCDOA representations. In each variant DeepWave operations are key for DoAE.

Across datasets we compare performance against two models: SELDnet23, a baseline for DoAE, and EINV2, the highest-ranked open-source model[4]. Note these models were designed to carry out SELD (i.e. classification in addition to DoAE). Also, EINV2 assumes data augmentation, but the authors did not make this code available on their repository[5]. Therefore we use the EINV2 without data augmentation.

4.3. Training procedure and evaluation metrics

CDBPN: We train CDPBN to upsample from 4ch to 32ch. We use METU soundscapes for training and ARNI for validation. The input and target are the instantaneous (100ms) 4ch $\mathbf{\Sigma} \in \mathbb{C}^{4\times4\times F}$ and 32ch $\mathbf{\Sigma} \in \mathbb{C}^{32\times32\times F}$, respectively. We use the original DBPN code, but we use F channels instead of RGB, and train two models, one for the real part and another for the imaginary part of $\mathbf{\Sigma}$. We freeze the optimal CDBPN and use it in experiments where we upsample from 4ch to 32ch.

LOCATA experiments: We train models using the generated METU soundscapes, and we validate using the ARNI ones. The final evaluation is carried out across the entire LOCATA dataset. Frequency bands are nine total ($F = 9$) and linearly spaced from 1.5kHz to 4.5kHz. We evaluate DeepWave models with and without the CDBPN upsampling.

STARSS23 experiments: STARSS23 comes divided in four directories ("dev-train-tau", "dev-train-sony", "dev-test-tau" and "dev-test-sony"), each with a unique set of rooms, and each room has its own sound sources (i.e. a set of "actors" and sounding objects such as a guitar or household blender). We train models using "DCASE-sim", "dev-train-tau", and "dev-train-sony", and we validate using "dev-test-tau". The final evaluation is carried out on the "unseen" rooms in "dev-test-sony". Note that the optimal DeepWave model that we evaluate on the STARSS23 dataset uses CDBPN since the STARSS23 and "DCASE-sim" datasets only have 4ch.

Metrics. We evaluate Localization Error (LE), the radial difference between predicted and true location per sound event, and Localization Recall (LR), the true positive rate of instantaneous detections out of the total annotated sound event instances [23]. We do not measure sound classification performance since we focus on sound localization.

5. RESULTS

Table 1 shows performance by the model variants on the LOCATA dataset and compares against SELDnet23 and EINV2. DeepWave+K-means with 32ch a input outperforms all models with low LE and high LR. We optimized this model to give a sense of "ceiling" performance. Its counterpart using 4ch upsampled with CDBPN shows a deteriorated LE (see Figure 3 bottom), but LR is same. We enhanced CDBPN for upsampling $\mathbf{\Sigma}$; however, this does not encompass SIM generation, thus posing challenges for K-means post-processing.

The models with Backprojection, Deblurring, and GRU show LE around 20^o, independent of whether the input is

[3] The variant that passes $\mathbf{\Sigma}$ directly to the GRU does not learn, independent of whether CDBPN upsampling is used.

[4] acording to the "DCASE" 2023 SELD challenge

[5] https://github.com/Jinbo-Hu/DCASE2022-TASK3

Authorized licensed use limited to: New York University. Downloaded on May 24,2024 at 13:56:14 UTC from IEEE Xplore. Restrictions apply.

Input	**Model**	*LE* ↓	*LR* ↑
up32ch	Backproj → Conv → GRU*	**20.5^o**	70.1
4ch	SELDnet23	23.3^o	82.3
FOA	SELDnet23	21.9^o	83.0
FOA + 4ch	EINV2	24.0^o	**84.2**

Table 2. Performance of SELDnet23, EINV2 and our best model on the STARSS23 "dev-test-sony" split. *Plus two MHSA layers after the GRU.

32ch or upsampled from 4ch using CDBPN. This indicates that the GRU is resilient to CDBPN distortions, and can improve LE by ~7^o compared to DeepWave+K-means.

The models that only use Backprojection and GRU further improve LE, indicating that Deblurring is unnecessary given the GRU. This is consistent with the result we discussed in the previous paragraph, where we note that the GRU is resilient to CDBPN distortions. In other words, DeepWave's Deblurring becomes unnecessary when the GRU is added. Adding a Conv2D layer between Backprojection and the GRU further improves performance. This makes sense, as the Conv2D layer operates across the F frequency "channels", and passes the GRU a representation with encoded information across frequency bands. In other words, it saves the GRU the step of having to aggregate information across the frequency axis.

Note that all GRU models show a deteriorated LR compared to the ones using K-means. This could be caused by the temporal memory that the GRU adds, which may come with a time constant to "react" in response to sounds appearing and disappearing from the scene. In contrast, DeepWave+K-means operated on single frames of audio without memory, and was optimized to "quickly" react to sound activity.

Next we evaluated the optimal model (CDBPN → Backproj → Conv2D → GRU) on STARSS23. Compared to LOCATA, STARSS23 contains more diverse sound sources and acoustic conditions. We found optimal performance by adding two multi-headed self-attention (MHSA) layers after the GRU [24]. Our model outperforms the LE performance of SELDnet trained with FOA, as well as SELDnet trained with 4ch and EINV2. Reasons for the deteriorated EINV2 performance compared to SELDnet23 include the smaller validation set we used ("dev-test-tau"), and the test set being smaller and more challenging ("dev-test-sony"). Note that EINV2 is known to have a female speech LE "a lot higher than average" [6].

When it comes to CDBPN, we analyzed its performance in terms of the magnitude and phase error. Figure 3 breaks down these results. The top matrices show the input $\mathbf{\Sigma}_4 \in \mathbb{C}^{4\times4\times F}$, CDBPN output $\hat{\mathbf{\Sigma}}_{32} \in \mathbb{C}^{32\times32\times F}$ and target $\mathbf{\Sigma}_{32} \in \mathbb{C}^{32\times32\times F}$ (averaged across the frequency axis) for magnitude (first row) and phase (second row). The bottom plot shows the magnitude and phase error between avg($\hat{\mathbf{\Sigma}}_{32}$) and avg($\mathbf{\Sigma}_{32}$) as a function of frequency (16 logaritmically-spaced bands). Results indicate that upsampling $\mathbf{\Sigma}$ at higher-frequency bands is more challenging for CDBPN.

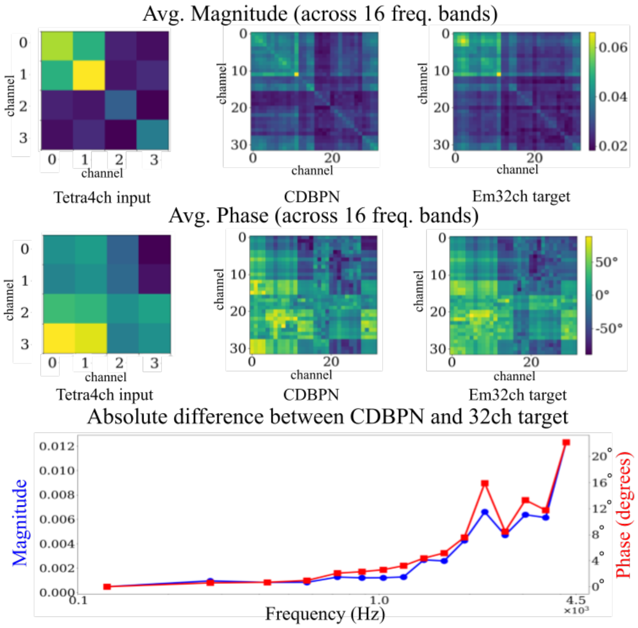

Fig. 3. CDBPN upsampling of a single reverberant white noise source directly facing the front of the microphone.

6. CONCLUSION AND FUTURE WORK

We have studied SIMs for DoAE. We focused on DeepWave because of its high-resolution SIM. We proposed a CDBPN model to upsample microphone array information from 4ch to 32ch. This allows our optimal model (4ch → CDBPN → 32ch → Backproj → Conv2D → GRU) to process existing SELD data with 4ch audio. The model can be interpreted as the combination of DeepWave (Backprojection to generate a SIM) and SELDnet23 (Conv2D + GRU) operations. Our systematic benchmark against the SELDnet and EINV2 models on the LOCATA and STARSS dataset is an indicator that our models can be generalized to other DoAE tasks.

Future work includes developing our model to also carry out sound event classification. Furthermore, generalizing CDBPN work with arbitrary microphone array shapes will be an important endeavor. For the time being, our results demonstrate the advantages of using SIMs for DoAE, which could benefit existing and future DoAE and SELD models.

7. ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation grant no. IIS-1955357. The authors thank the funding source and their grant collaborators.

Authorized licensed use limited to: New York University. Downloaded on May 24,2024 at 13:56:14 UTC from IEEE Xplore. Restrictions apply.

8. REFERENCES

[1] Kazuki Shimada, Archontis Politis, et al., “Starss23: An audio-visual dataset of spatial recordings of real scenes with spatiotemporal annotations of sound events,” *arXiv preprint 2306.09126*, 2023.

[2] Heinrich W Löllmann, Christine Evers, Alexander Schmidt, Heinrich Mellmann, Hendrik Barfuss, Patrick A Naylor, and Walter Kellermann, “The locata challenge data corpus for acoustic source localization and tracking,” in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing*, 2018, pp. 410–414.

[3] Archontis Politis, “[DCASE2022 Task 3] Synthetic SELD mixtures for baseline training,” Apr. 2022.

[4] Pierre-Amaury Grumiaux, Srđan Kitić, Laurent Girin, and Alexandre Guérin, “A survey of sound source localization with deep learning methods,” *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 107–151, 2022.

[5] Qing Wang, Jun Du, Hua-Xin Wu, Jia Pan, Feng Ma, and Chin-Hui Lee, “A four-stage data augmentation approach to resnet-conformer based acoustic modeling for sound event localization and detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1251–1264, 2023.

[6] Jinbo Hu, Yin Cao, Ming Wu, Qiuqiang Kong, Feiran Yang, Mark D. Plumbley, and Jun Yang, “A trackwise ensemble event independent network for polyphonic sound event localization and detection,” in *IEEE ICASSP*, 2022, pp. 9196–9200.

[7] Boaz Rafaely, *Fundamentals of spherical array processing*, vol. 8, Springer, 2015.

[8] Matthieu Simeoni, Sepand Kashani, Paul Hurley, and Martin Vetterli, “Deepwave: a recurrent neural-network for real-time acoustic imaging,” *Advances In Neural Information Processing Systems*, vol. 32, 2019.

[9] Karol Gregor and Yann LeCun, “Learning fast approximations of sparse coding,” in *Proceedings of the 27th intl. conf. on machine learning*, 2010, pp. 399–406.

[10] MH Acoustics, “Eigenmike microphone array release notes (v17. 0),” *25 Summit Ave, Summit, NJ, USA*, 2013.

[11] Robin Scheibler, Juan Azcarreta, René Beuchat, and Corentin Ferry, “Pyramic: Full stack open microphone array architecture and dataset,” in *2018 16th IWAENC*. IEEE, 2018, pp. 226–230.

[12] Sharath Adavanne, Archontis Politis, Joonas Nikunen, and Tuomas Virtanen, “Sound event localization and detection of overlapping sources using convolutional recurrent neural networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 34–48, 2018.

[13] Ning Chu, José Picheral, Ali Mohammad-Djafari, and Nicolas Gac, “A robust super-resolution approach with sparsity constraint in acoustic imaging,” *Applied Acoustics*, vol. 76, pp. 197–208, 2014.

[14] Alle-Jan van der Veen and Stefan J Wijnholds, “Signal processing tools for radio astronomy,” in *Handbook of Signal Processing Systms.*, pp. 421–463. Springer, 2013.

[15] Saksham Singh Kushwaha, Iran R Roman, and Juan P Bello, “Analyzing the effect of equal-angle spatial discretization on sound event localization and detection.,” in *DCASE*, 2022.

[16] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita, “Deep back-projection networks for super-resolution,” in *Proceedings of the IEEE CVPR conference*, 2018, pp. 1664–1673.

[17] Kazuki Shimada, Yuichiro Koyama, Shusuke Takahashi, Naoya Takahashi, Emiru Tsunoo, and Yuki Mitsufuji, “Multi-accdoa: Localizing and detecting overlapping sounds from the same class with auxiliary duplicating permutation invariant training,” in *IEEE ICASSP*. IEEE, 2022, pp. 316–320.

[18] Sharath Adavanne, Archontis Politis, and Tuomas Virtanen, “Localization, detection and tracking of multiple moving sound sources with a convolutional recurrent network,” *arXiv preprint 1904.12769*, 2019.

[19] Iran Roman, Chris Ick, Sivan Ding, Adrian Roman, Brian McFee, and Juan Pablo Bello, “Spatial scaper: a library to simulate and augment soundscapes for sound event localization and detection in realistic rooms,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024.

[20] Orhun Olgun and Huseyin Hacihabiboglu, “METU SPARG Eigenmike em32 Acoustic Impulse Response Dataset v0.1.0,” Apr. 2019.

[21] Thomas McKenzie, Leo McCormack, and Christoph Hold, “Dataset of spatial room impulse responses in a variable acoustics room for six degrees-of-freedom rendering and analysis,” *arXiv preprint 2111.11882*, 2021.

[22] Eduardo Fonseca, Xavier Favory, Jordi Pons, Frederic Font, and Xavier Serra, “Fsd50k: an open dataset of human-labeled sound events,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 829–852, 2021.

[23] Archontis Politis, Annamaria Mesaros, et al., “Overview and evaluation of sound event localization and detection in dcase 2019,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 684–698, 2020.

[24] Parthasaarathy Sudarsanam, Archontis Politis, and Konstantinos Drossos, “Assessment of self-attention on learned features for sound event localization and detection,” *arXiv preprint arXiv:2107.09388*, 2021.

Authorized licensed use limited to: New York University. Downloaded on May 24,2024 at 13:56:14 UTC from IEEE Xplore. Restrictions apply.