

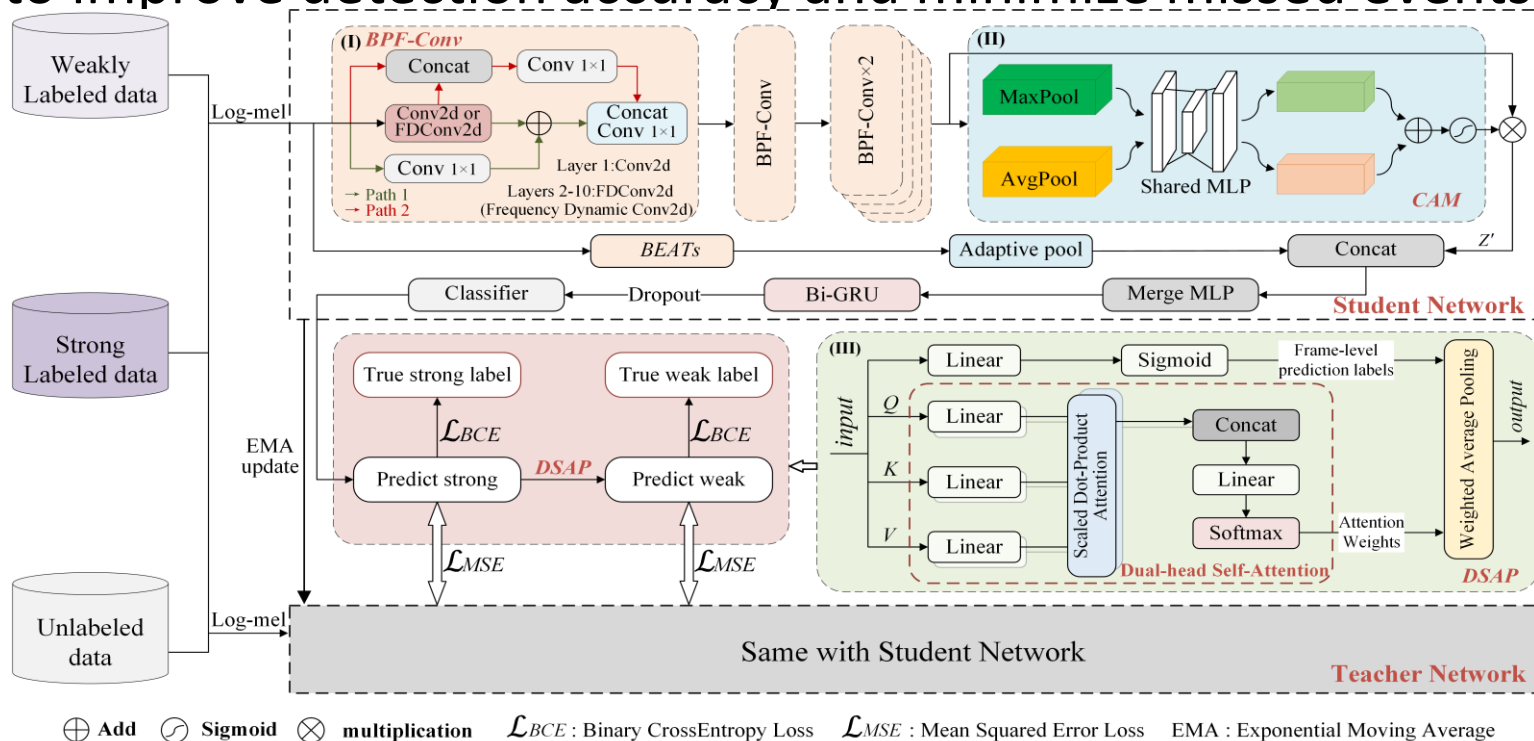
Improving sound event detection through enhanced feature extraction and attention mechanisms

**Dongping ZHANG, Siyi WU, Zhanhong LU,
Zhehao ZHANG, Haimiao HU, Jiabin YU**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-41108-7](https://doi.org/10.1007/s11704-025-41108-7)

Problems & Ideas

- Challenges of sound event detection in home environments:
 - Overlapping sound events: Difficulty in separating events from background noise, leading to detection errors and missed events.
 - Insufficient Feature Extraction: Limited capability to extract relevant features and difficulty in identifying key channels.
- Ideas: Enhance feature extraction and prioritize key channels to improve detection accuracy and minimize missed events.



This diagram illustrates a semi-supervised learning framework for improving audio event detection, integrating key modules.

Main Contributions

- Contributions:
 - A BPF-Conv module with a Bi-path structure that enhances feature extraction and improves feature representation and transfer.
 - A DSAP function designed to aggregate frame-level predictions from weakly labeled samples, thereby enhancing the model’s prediction accuracy for such data.
 - Implementation of CAM to selectively emphasize critical features in the audio feature map while filtering out irrelevant information.

Table1 Comparison with different models on validation dataset

year	models	PSDS1↑	PSDS2↑	CB-F1↑
2019	Baseline CRNN	0.362	0.561	0.423
2020	DY [3]	0.423	0.637	0.508
2022	FDY [4]	0.431	0.645	0.520
2023	MFD-CMT [5]	0.460	0.678	0.539
2023	Li-USTC-task4a-7 [6]	0.451	0.673	-
2023	AST-GRU [7]	0.434	-	0.552
-	Ours	0.489	0.771	0.567

Table1 demonstrates that the proposed model outperforms other methods, highlighting its effectiveness in audio event detection. Table2 compares the performance of different model configurations, showing the improvements achieved by the proposed approach over the baseline.

Table2 Evaluation of the proposed model on the validation and public evaluation dataset

	Method	validation dataset					evaluation dataset	
		Pretrained↑	PSDS1↑	PSDS2↑	CB-F1↑	IB-F1↑	PSDS1↑	PSDS2↑
(a)	Baseline	-	0.362	0.561	0.423	0.645	0.408	0.602
(b)	Baseline	✓	0.453	0.737	0.554	0.788	0.504	0.778
(c)	+BPF-C	✓	0.478	0.753	0.561	0.799	0.509	0.790
(d)	+DSAP	✓	0.462	0.752	0.560	0.792	0.507	0.784
(e)	+BPF-C+DSAP	✓	0.481	0.758	0.565	0.801	0.516	0.794
(f)	+BPF-C+DSAP+CAM	-	0.439	0.681	0.530	0.741	-	-
(g)	+BPF-C+DSAP+CAM	✓	0.489	0.771	0.567	0.803	0.522	0.802