Supplementary Material To: Eliminating Quantization Errors in Classification-Based Sound Source Localization

Linfeng Feng[a,b], Xiao-Lei Zhang[a,b,*], Xuelong Li[c]

[a]*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China*
[b]*Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518063, Guangdong, China*
[c]*Institute of Artificial Intelligence (TeleAI), China Telecom Corp Ltd, 31 Jinrong Street, Beijing 100033, P. R. China*

Table 1: Architecture of the PNN. The batch normalization and ReLU activations are not shown in the table.

Layer name	Structure	Output size
Input	—	$1 \times 4 \times 256$
Conv2D-1	2×1, Stride=(1, 1)	$64 \times 3 \times 256$
Conv2D-2	2×1, Stride=(1, 1)	$64 \times 2 \times 256$
Conv2D-3	2×1, Stride=(1, 1)	$64 \times 1 \times 256$
Flatten	—	16384
Dense-1	—	512
Dense-2	—	512
Dense-3	—	$I + 1$

Table 2: Architecture of the PNN-Split. The batch normalization and ReLU activations are not shown in the table.

Layer name	Structure	Output size
Input	—	$1 \times 4 \times 256$
Conv2D-1	2×1, Stride=(1, 1)	$4 \times 3 \times 256$
Conv2D-2	2×3, Stride=(1, 1)	$16 \times 2 \times 256$
Conv2D-3	2×3, Stride=(1, 1)	$32 \times 1 \times 256$
Flatten	—	8192
Dense-1	—	$2I + 2$
BiLSTM	—	$2I + 2$
Dense-2	—	$I + 1$

Table 3: Architecture of the SNet. The batch normalization and ReLU activations are not shown in the table.

Layer name	Structure	Output size
Input	—	$8 \times 7 \times 256$
Conv2D-1	1×7, Stride=(1, 3)	$32 \times 7 \times 84$
Conv2D-2	1×5, Stride=(1, 2)	$128 \times 7 \times 40$
Residual Block	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 128 \end{bmatrix} \times 5$, Stride=(1,1)	$128 \times 7 \times 40$
Conv2D-3	1×1, Stride=(1, 1)	$(I + 1) \times 7 \times 40$
Swap axes	—	$40 \times 7 \times (I + 1)$
Conv2D-4	1×1, Stride=(1, 1)	$500 \times 7 \times (I + 1)$
Conv2D-5	7×5, Stride=(1, 1)	$1 \times 1 \times (I + 1)$
Flatten	—	$I + 1$

Table 4: Architecture of the SNet-Split. The batch normalization and ReLU activations are not shown in the table.

Layer name	Structure	Output size
Input	—	$8 \times 7 \times 256$
Conv2D-1	1×7, Stride=(1, 3)	$32 \times 7 \times 84$
Conv2D-2	1×5, Stride=(1, 2)	$128 \times 7 \times 40$
Residual Block	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 128 \end{bmatrix} \times 5$, Stride=(1,1)	$128 \times 7 \times 40$
Flatten	—	35840
Dense-1	—	$2I + 2$
BiLSTM	—	$2I + 2$
Dense-2	—	$I + 1$

*Corresponding author.
Email addresses: fenglinfeng@mail.nwpu.edu.cn (Linfeng Feng), xiaolei.zhang@nwpu.edu.cn (Xiao-Lei Zhang), li@nwpu.edu.cn (Xuelong Li)

Table 5: Experimental results on office data, where the backbone network is PNN.

One-hot	CE	62.60	3.168	3.181	3.077
	MSE	64.07	3.197	3.242	3.123
	WD	58.36	4.040	4.063	3.963
	NLAE	63.17	3.071	3.087	**2.978**
	MSE (wo)	60.92	3.695	3.753	3.621
GLC	BCE	56.06	4.877	5.282	4.860
	MSE	57.52	3.495	3.795	**3.465**
	NLAE	58.80	3.675	4.036	3.640
	MSE (wo)	57.93	3.518	3.926	3.500
SLD	CE	54.18	7.282	7.797	7.282
	BCE	54.82	6.511	7.029	6.504
	MSE	58.65	4.229	4.631	4.204
	WD	55.02	4.813	5.276	4.802
	NLAE	59.27	4.700	5.026	4.659
	MSE (wo)	60.43	3.184	3.582	**3.163**
ULD	CE	63.26	3.791	3.910	3.752
	BCE	60.12	3.384	3.413	3.290
	MSE	64.25	3.589	3.624	3.501
	WD	62.39	3.521	3.733	3.530
	NLAE	64.09	3.985	4.153	3.964
	MSE (wo)	64.12	3.008	3.116	**2.925**

Table 6: Experimental results on conference room data, where the backbone network is PNN.

One-hot	CE	54.32	6.888	6.905	6.805
	MSE	54.60	5.748	5.762	5.670
	WD	49.63	7.022	7.055	6.983
	NLAE	53.84	5.192	5.268	**5.138**
	MSE (wo)	51.89	6.322	6.668	6.318
GLC	BCE	44.65	11.409	11.673	11.397
	MSE	47.48	9.306	9.516	9.270
	NLAE	50.16	9.719	10.046	9.703
	MSE (wo)	49.90	5.816	6.078	**5.814**
SLD	CE	43.86	14.482	14.835	14.486
	BCE	44.18	14.759	15.084	14.760
	MSE	48.04	11.559	11.816	11.536
	WD	45.08	13.211	13.539	13.237
	NLAE	48.90	9.987	10.253	9.967
	MSE (wo)	53.71	6.052	6.498	**6.052**
ULD	CE	52.97	10.561	10.657	10.503
	BCE	50.87	6.545	6.600	6.457
	MSE	55.90	5.483	5.527	**5.394**
	WD	54.02	6.704	6.849	6.705
	NLAE	54.33	6.654	6.729	6.601
	MSE (wo)	54.18	5.431	5.635	5.420

Table 7: Experimental results on simulated data L1, where the backbone network is SNet.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One-hot	**CE**	74.22	2.305	1.998	**1.970**
	MSE	75.39	2.459	2.114	2.092
	WD	68.54	3.184	2.958	2.936
	NLAE	74.68	2.694	2.353	2.327
	MSE (wo)	75.79	2.612	2.265	2.236
GLC	BCE	66.50	2.458	2.390	2.257
	MSE	68.19	2.779	2.685	2.551
	NLAE	69.38	2.290	2.225	**2.044**
	MSE (wo)	71.44	2.292	2.214	2.055
SLD	CE	63.59	2.911	2.904	2.743
	BCE	66.53	2.794	2.746	2.602
	MSE	67.67	2.641	2.606	2.450
	WD	63.84	3.497	3.399	3.275
	NLAE	68.91	3.262	3.198	3.084
	MSE (wo)	69.93	2.543	2.544	**2.360**
ULD	CE	71.46	2.654	2.285	2.245
	BCE	74.05	2.473	2.116	2.044
	MSE	76.20	2.301	1.912	1.845
	WD	70.90	2.929	2.658	2.583
	NLAE	72.81	2.803	2.454	2.384
	MSE (wo)	76.88	2.176	1.782	**1.696**

Table 8: Experimental results on office data, where the backbone network is SNet.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One-hot	CE	66.72	2.406	2.349	2.349
	MSE	68.24	2.398	2.335	2.305
	WD	64.61	2.798	2.776	2.708
	NLAE	67.59	2.269	2.221	2.193
	MSE (wo)	67.48	2.285	2.197	**2.169**
GLC	BCE	57.30	2.806	3.100	2.751
	MSE	58.47	2.596	2.881	2.531
	NLAE	57.99	2.737	2.946	2.636
	MSE (wo)	61.11	2.474	2.766	**2.413**
SLD	CE	58.88	2.778	3.199	2.768
	BCE	61.93	2.612	3.029	2.627
	MSE	61.70	2.753	3.068	2.708
	WD	62.94	2.933	3.280	2.944
	NLAE	63.26	3.091	3.397	3.036
	MSE (wo)	59.88	2.555	2.894	**2.521**
ULD	CE	66.68	2.375	2.366	2.326
	BCE	66.58	2.352	2.351	2.285
	MSE	66.23	2.290	2.252	2.191
	WD	63.91	2.712	2.753	2.645
	NLAE	68.05	2.417	2.487	2.428
	MSE (wo)	68.96	2.177	2.208	**2.145**

Table 9: Experimental results on conference room data, where the backbone network is SNet.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One-hot	CE	55.79	4.765	4.634	4.629
	MSE	57.53	4.744	4.627	4.599
	WD	52.71	5.310	5.270	5.241
	NLAE	56.67	4.912	4.805	4.791
	MSE (wo)	56.57	4.616	4.488	**4.483**
GLC	BCE	48.85	4.954	5.046	4.866
	MSE	50.38	4.775	4.886	4.687
	NLAE	49.91	4.987	5.034	4.866
	MSE (wo)	53.70	4.337	4.462	**4.257**
SLD	CE	49.45	4.993	5.164	4.934
	BCE	51.28	4.945	5.107	4.914
	MSE	52.56	5.022	5.176	4.948
	WD	53.94	5.434	5.661	5.414
	NLAE	52.34	5.392	5.554	5.323
	MSE (wo)	52.97	4.519	4.696	**4.465**
ULD	CE	56.86	4.749	4.668	4.616
	BCE	54.92	5.380	5.267	5.236
	MSE	56.27	4.657	4.510	4.482
	WD	53.82	4.937	4.836	4.786
	NLAE	58.07	4.820	4.729	4.703
	MSE (wo)	57.18	4.583	4.476	**4.456**

Table 10: Experimental results on simulated data C2, where the backbone network is PNN-Split.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One-hot	CE	69.55	7.981	7.366	7.370
	MSE	69.86	8.169	7.481	7.491
	WD	68.22	7.189	**6.573**	6.576
	NLAE	70.06	7.912	7.296	7.292
	MSE (wo)	75.30	8.105	7.327	7.250
GLC	BCE	73.77	6.012	5.630	5.250
	MSE	75.64	5.999	5.526	5.152
	NLAE	74.13	5.950	5.504	5.143
	MSE (wo)	77.09	5.855	5.391	**5.043**
SLD	CE	47.86	8.725	8.208	8.220
	BCE	57.94	6.934	6.711	6.570
	MSE	64.31	6.466	6.156	6.001
	WD	59.47	6.707	6.410	6.246
	NLAE	59.59	6.787	6.536	6.457
	MSE (wo)	71.88	5.588	5.386	**5.203**
ULD	CE	66.90	7.490	6.799	6.784
	BCE	71.37	6.741	5.991	5.964
	MSE	70.66	7.120	6.369	6.367
	WD	70.61	6.403	5.683	5.666
	NLAE	71.24	6.837	6.048	6.050
	MSE (wo)	79.33	6.089	5.291	**5.114**

Table 11: Experimental results on simulated data L2, where the backbone network is SNet-Split.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One-hot	CE	65.58	6.439	6.022	6.027
	MSE	66.38	6.772	6.324	6.336
	WD	62.56	6.145	5.892	5.896
	NLAE	67.23	5.971	5.515	**5.520**
	MSE (wo)	72.78	6.124	5.687	5.681
GLC	BCE	68.30	4.880	4.577	4.247
	MSE	69.49	5.273	4.915	4.588
	NLAE	67.65	5.101	4.728	4.419
	MSE (wo)	70.74	4.760	4.386	**4.093**
SLD	CE	53.52	6.101	5.967	5.804
	BCE	55.71	5.741	5.651	5.461
	MSE	60.77	5.573	5.435	5.247
	WD	55.37	5.862	5.701	5.538
	NLAE	58.65	5.802	5.810	5.628
	MSE (wo)	70.64	4.798	4.765	**4.495**
ULD	CE	64.39	5.536	4.890	4.873
	BCE	63.49	5.774	5.146	5.139
	MSE	66.60	5.647	4.987	4.997
	WD	63.58	5.456	4.860	4.858
	NLAE	67.31	5.327	4.650	4.651
	MSE (wo)	73.53	5.296	4.554	**4.524**
α**ULD+**$(1-\alpha)$**GLC**	**MSE (wo)**	73.01	4.446	4.008	**3.739**

Table 12: Experimental results on office data, where the backbone network is SNet-Split.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One Hot	CE	60.68	7.357	7.299	7.269
	MSE	62.78	6.677	6.941	6.704
	WD	60.47	6.422	6.309	**6.291**
	NLAE	65.44	7.084	7.276	7.131
	MSE (wo)	58.40	11.682	11.819	11.750
GLC	BCE	62.84	6.419	7.021	6.502
	MSE	62.17	7.052	7.595	7.106
	NLAE	59.74	6.653	7.071	6.634
	MSE (wo)	64.23	6.130	6.756	**6.180**
SLD	CE	48.07	7.382	7.820	7.420
	BCE	50.91	7.152	7.822	7.221
	MSE	59.00	5.764	6.560	**5.832**
	WD	46.98	7.792	8.258	7.781
	NLAE	52.57	7.128	7.868	7.284
	MSE (wo)	63.62	6.204	7.005	6.244
ULD	CE	60.65	6.246	6.371	6.219
	BCE	60.56	6.382	6.536	6.403
	MSE	62.78	5.908	6.017	5.860
	WD	58.81	6.649	6.739	6.622
	NLAE	63.34	5.639	5.751	**5.618**
	MSE (wo)	62.79	6.505	6.500	6.434
α**ULD+**$(1-\alpha)$**GLC**	**MSE (wo)**	65.22	6.107	6.620	**6.081**

Table 13: Experimental results on conference room data, where the backbone network is SNet-Split.

Encoding	Loss	ACC	MAE		
			Top-1	WAD-2	WAD-3
One Hot	CE	46.98	15.048	14.895	14.880
	MSE	48.32	14.003	13.867	13.837
	WD	44.96	12.723	12.552	**12.543**
	NLAE	48.58	13.995	13.845	13.814
	MSE (wo)	49.78	18.446	18.429	18.399
GLC	BCE	52.05	11.680	11.815	11.585
	MSE	51.52	12.005	12.050	11.866
	NLAE	48.69	12.639	12.605	12.424
	MSE (wo)	53.86	11.349	11.442	**11.211**
SLD	CE	37.36	13.191	13.378	13.175
	BCE	38.70	13.209	13.409	13.199
	MSE	46.51	12.050	12.231	12.020
	WD	36.32	13.297	13.481	13.220
	NLAE	39.12	12.847	13.158	12.906
	MSE (wo)	53.63	11.714	11.941	**11.640**
ULD	CE	47.16	13.172	13.007	12.945
	BCE	45.70	14.158	13.985	13.948
	MSE	49.90	12.232	12.086	**12.025**
	WD	42.41	14.342	14.191	14.135
	NLAE	50.28	12.559	12.396	12.342
	MSE (wo)	54.58	12.776	12.655	12.591
α**ULD+**$(1-\alpha)$**GLC**	**MSE (wo)**	53.68	10.519	10.459	**10.229**