INTERMIX: AN INTERFERENCE-BASED DATA AUGMENTATION AND REGULARIZATION TECHNIQUE FOR AUTOMATIC DEEP SOUND CLASSIFICATION

Ramit Sawhney [1] Atula Tejaswi Neerkaje [2]

[1]Scaler, India [2]Manipal Institute of Technology, India

Code At

Introduction

- Deep Learning models with broad hypothesis space overfit if the size of the training set is not large enough.
- Interpolation-based augmentation and regularization techniques like Mixup have achieved state-of-the-art performances on various tasks.

Contributions

- Building on existing work, we present **InterMix**, an interference-based data augmentation technique for automatic sound classification.
- We compare InterMix against other mixup strategies and highlight the effectiveness of InterMix that uses an interference formula, while simultaneously providing improved privacy safeguards.

InterMix

InterMix - How it Works

- InterMix first introduces **phase shifts** to two randomly sampled sounds.
- Next, InterMix mixes the representations of these two sounds in the m-th layer using an **interference-based formula**.

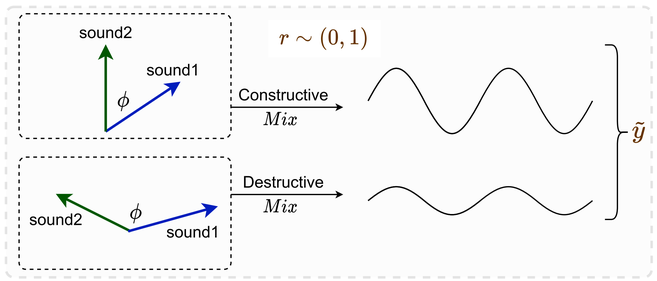

Figure 1. Interference formula based mixing of sounds is controlled by a mixing ratio r, with a phase difference ϕ.

InterMix - Why it Works

- Interference-based mixup provides augmented samples by mixing with a **varied phase differences**.
- Creates representations of varied feature intensities while maintaining the implicit properties of the mixed sounds.
- InterMix creates large number of varied training signals - effect of sensitive crowd-sourced data while training is **minimized**.

Performance Comparison

Model	Learning	Error Rates (%)		
		ESC-50	ESC-10	UrbanSound8K
M18 [21]	Standard	31.5±0.5	18.2±0.5	28.8
	BC Learning	26.7±0.1	14.2±0.9	26.5
	Speechmix	**24.3±0.2**	**12.4±0.5**	25.1
	InterMix (Ours)	25.4±0.5	12.6±0.5	**25.1**
SoundNet5 [7]	Standard	33.8±0.2	16.4±0.8	33.3
	BC Learning	27.4±0.3	13.9±0.4	30.2
	Speechmix	25.6±0.2	11.6±0.3	27.4
	InterMix (Ours)	**25.1±0.3**	**10.6±0.3**	**26.5**
EnvNet [24]	Standard	29.2±0.1	12.8±0.4	33.7
	BC Learning	24.1±0.2	11.3±0.6	28.9
	Speechmix	22.5±0.3	9.3±0.4	**26.5**
	InterMix (Ours)	**22.5±0.3**	**9.1±0.2**	26.8
PiczakCNN [22]	Standard	27.6±0.2	13.2±0.4	25.3
	BC Learning	23.1±0.3	9.4±0.4	23.5
	Speechmix	22.1±0.3	8.4±0.2	22.1
	InterMix (Ours)	**21.9±0.2**	**8.3±0.4**	**21.1**
EnvNet-v2 [6]	Standard	25.6±0.3	14.2±0.8	30.9
	BC Learning	18.2±0.2	10.6±0.6	23.4
	Speechmix	16.2±0.3	8.5±0.4	21.6
	InterMix (Ours)	**15.8±0.4**	**8.2±0.4**	**21.4**
EnvNet-v2 +Augmentation	Standard	21.2±0.3	10.9±0.6	24.9
	BC Learning	15.1±0.2	8.6±0.1	21.7
	Speechmix	13.1±0.2	**7.1±0.1**	20.8
	InterMix (Ours)	**12.9±0.4**	7.2±0.1	**20.5**
Human		18.7	4.3	

Figure 2. We present a comparison between Standard Learning, BC Learning [2], Speechmix [1], and InterMix

- InterMix generally outperforms other techniques across the given models and datasets.
- The best-performing model is the augmented EnvNet-v2.
- We observe relative improvements of 14.6%, 16.3%, and 5.5% with respect to BC learning on ESC-50, ESC-10, and UrbanSound8K respectively

References

[1] Amit Jindal, Narayanan Elavathur Ranganatha, Aniket Didolkar, Arijit Ghosh Chowdhury, Di Jin, Ramit Sawhney, and Rajiv Ratn Shah. Speechmix-augmenting deep sound recognition using hidden space interpolations. In *INTERSPEECH*, pages 861–865, 2020.

[2] Yuji Tokozume, Yoshitaka Ushiku, and Tatsuya Harada. Learning from between-class examples for deep sound recognition. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

Ablation

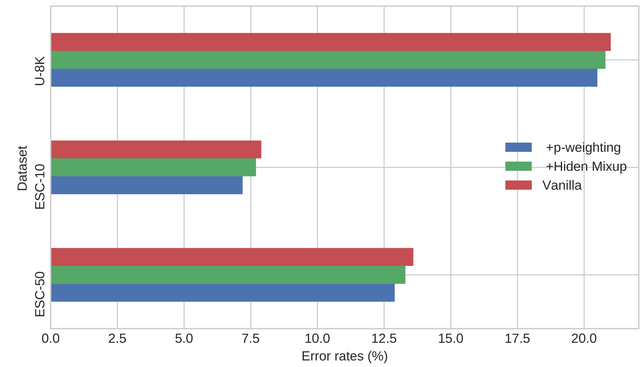

Figure 3. Ablation study of InterMix components.

- We observe **significant** performance improvements on introducing both hidden space mixing and p-weighting.
- Suggests the effectiveness of mixing in the hidden space, and also considering the difference in sound pressure levels.

Adversarial Robustness

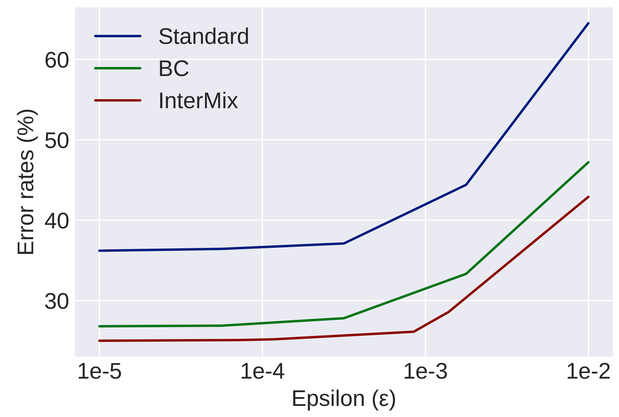

Figure 4. Error rates over adversarial examples with increasing values of Epsilon (ϵ).

- InterMix provides better privacy safeguards through an **improved regularizing effect**.
- InterMix reduces the reliance on sensitive training data by using virtual samples.