PCL: Partitioned Continual Learning via Unsupervised Latent Experts for Audio Classification

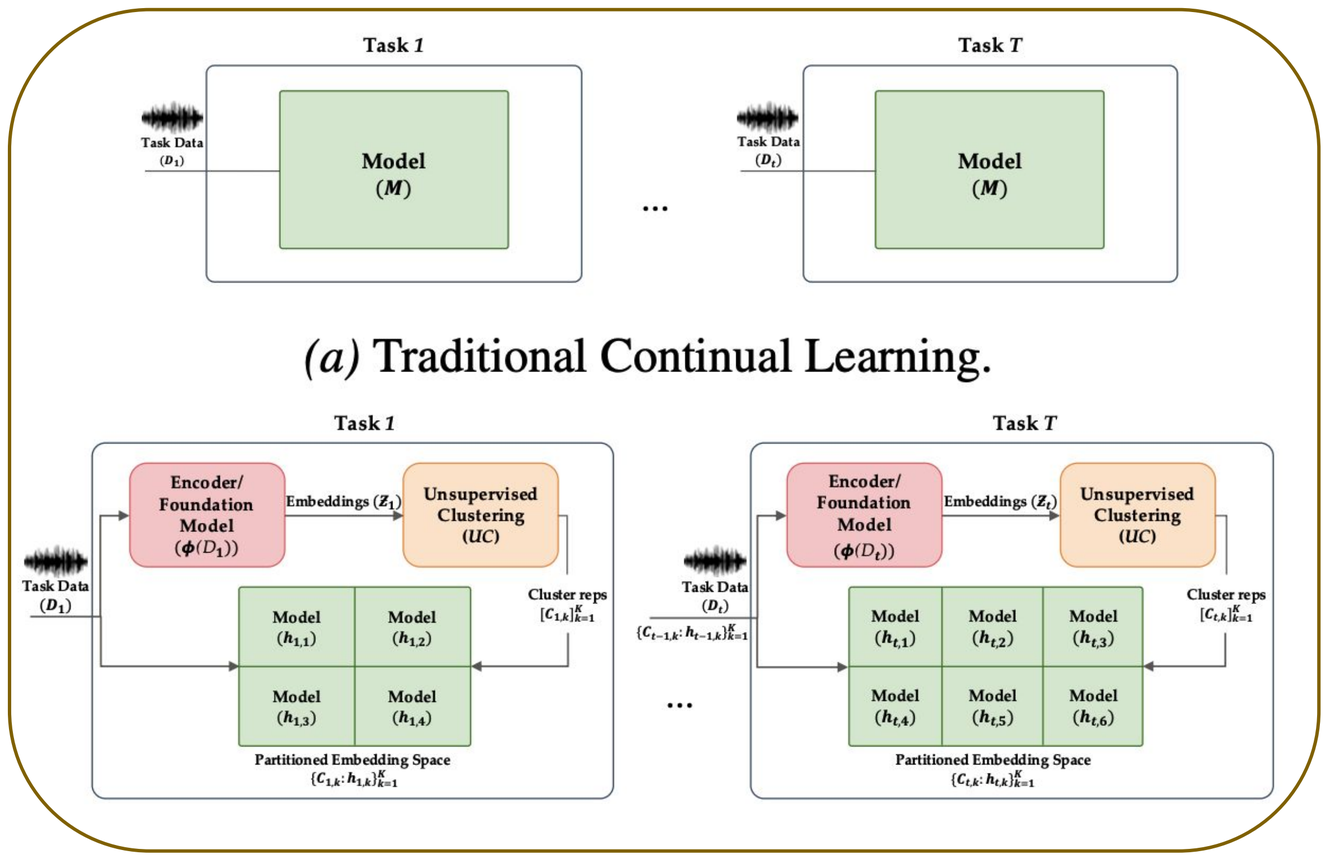
Gautham Krishna Gudur[1*], Mohit Malu[2*], Tanmay Khandait[2],
Reza Rahimi Azghan[2], Anirudh Rayas[2], Pavan Turaga[2], Joydeep Ghosh[1],
Hassan Ghasemzadeh[2], Edison Thomaz[1], Giulia Pedrielli[2]
The University of Texas at Austin[1], Arizona State University[2]

ICML International Conference On Machine Learning

PROBLEM STATEMENT

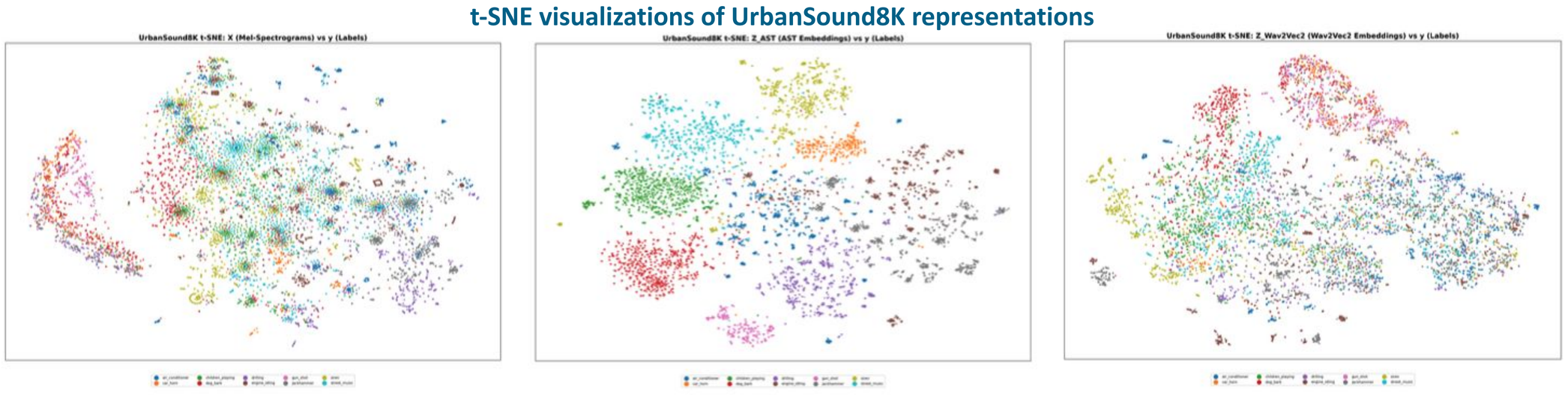
- Real-world audio streams in continual learning (CL): new sound classes, devices, and acoustic environments evolve.
- Naive fine-tuning adapts to new tasks but catastrophically forgets previously learned sounds.
- Full retraining during CL is costly and impractical.
- Monolithic CL methods (e.g. EWC) update one shared model → limited specialization for heterogeneous audio.

Can latent-space partitioning of audio foundation model embeddings improve continual learning?



(a) Traditional Continual Learning.

t-SNE visualizations of UrbanSound8K representations



(a) X (Mel-spectrograms) vs Labels (y) *(b)* Z-space (AST) vs Labels (y) *(c)* Z-space (Wav2Vec2) vs Labels (y)

PCL: Partitioned Continual Learning

Audio foundation model embeddings (CLAP, AST, Wav2Vec2) reveal structured latent regions that can be partitioned into specialized continual learning experts.

- ***Frozen Latent Representation****:* Freeze an audio foundation model and map audio to latent space.
- ***Memory-Aware Weighted Clustering****:* Cluster embeddings into locally homogeneous regions using weighted k-medoids.
- ***One-to-One Expert Assignment****:* Reuse matched experts for old regions; instantiate new experts for novel latent regions using representation similarity.
- ***Partitioned Continual Adaptation****:* Train each lightweight expert only on its assigned partition – reduces interference.

Inference: Route each test sample to its nearest latent cluster and activate only the corresponding expert.

EXPERIMENTS AND RESULTS

Table 1. Class-incremental results on ESC-50 and UrbanSound8K using CNN-2-layer (CNN-2L) and CNN-4-layer (CNN-4L) models with different CL methods and z-spaces. Higher ACC and BWT indicate better performance (↑).

z-space	CL Technique	ESC-50						UrbanSound8K					
		CNN-2L			CNN-4L			CNN-2L			CNN-4L		
		ACC↑	BWT↑	Clusters	ACC↑	BWT↑	Clusters	ACC↑	BWT↑	Clusters	ACC↑	BWT↑	Clusters
–	Finetune	23.254	-0.295	–	23.848	-0.286	–	23.240	-0.379	–	23.850	-0.368	–
	EWC	37.682	-0.252	–	38.416	-0.249	–	27.948	-0.351	–	28.420	-0.348	–
–	LwF	38.275	-0.238	–	38.142	-0.235	–	28.726	-0.345	–	29.285	-0.341	–
–	Joint (Upper bound)	**68.750**	–	–	**69.284**	–	–	**86.320**	–	–	**87.204**	–	–
	EWC	50.846	-0.082	24	52.392	-0.076	24	64.850	-0.124	8	66.420	-0.119	8
CLAP	LwF	52.518	-0.073	24	52.174	-0.073	24	65.370	-0.118	8	67.187	-0.118	8
	EWC	**54.924**	**-0.066**	28	**55.318**	**-0.061**	28	67.940	-0.118	9	68.524	-0.116	9
AST	LwF	54.285	-0.064	28	54.436	-0.065	28	**68.634**	**-0.115**	9	**70.249**	**-0.114**	9
	EWC	42.713	-0.198	15	42.286	-0.187	15	53.180	-0.187	11	55.260	-0.181	11
Wav2Vec2	LwF	43.052	-0.191	15	44.917	-0.204	15	54.057	-0.184	11	54.420	-0.184	11

Paper

SCAN ME