MT-SLVR: Multi-Task Self-Supervised Learning for Transformation In(Variant) Representations

THE UNIVERSITY of EDINBURGH
Institute for Digital Communications

Calum Heggan (s1529508@ed.ac.uk)

Supervised by Mehrdad Yaghoobi (UOE), Sam Budgett(Thales UK RTI) & Tim Hospedales (UOE, Samsung Research)

INTERSPEECH 2023
20th-24th August 2023 - Dublin, Ireland

1. Motivation & Questions

- Self-supervised methods are incredibly powerful.
- The the most successful methods use augmentation pipelines in order to learn some inductive bias for downstream tasks.
- This inductive bias is not known apriori, meaning that it can either help or hurt performance downstream depending on the task. i.e. rotation invariance is great for object detection but poor for pose estimation
- This often leads to specific upstream models needing pre-trained
- Are we able to learn conflicting inductive biases within a single model?

2. MT-SLVR Part I

- We choose to approach this problem through multi-task learning between contrastive (invariance learning) and predictive (variance learning) algorithms
- Our pipeline has 3 main parts:
  - Augmentation pipeline + processing
  - Task specific parameters + adapters
  - Multi-task loss
- We augment input samples into two correlated views as is done in SimCLR (our contrastive algorithm of choice). Applied augmentations per sample are tracked to learn our predictive model (Multi-Label Augmentation Prediction)
- These correlated views are passed through both base network and task specific parameters

3. MT-SLVR Part II

- The use of task specific parameters (implemented using adapters) is one of the key insights that allows this approaches success
- By giving the base model some parameters that are only updated through either predictive or contrastive gradient updates, task specific information (ideally augmentation invariance/variances) from both objectives can be stored in a single model

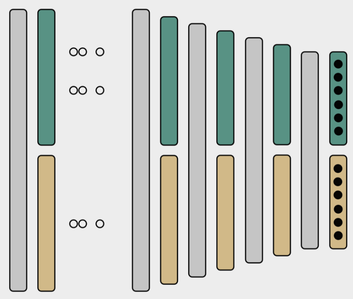

Figure 1:Diagram of how our backbone feature extractor looks with task-specific adapters throughout it

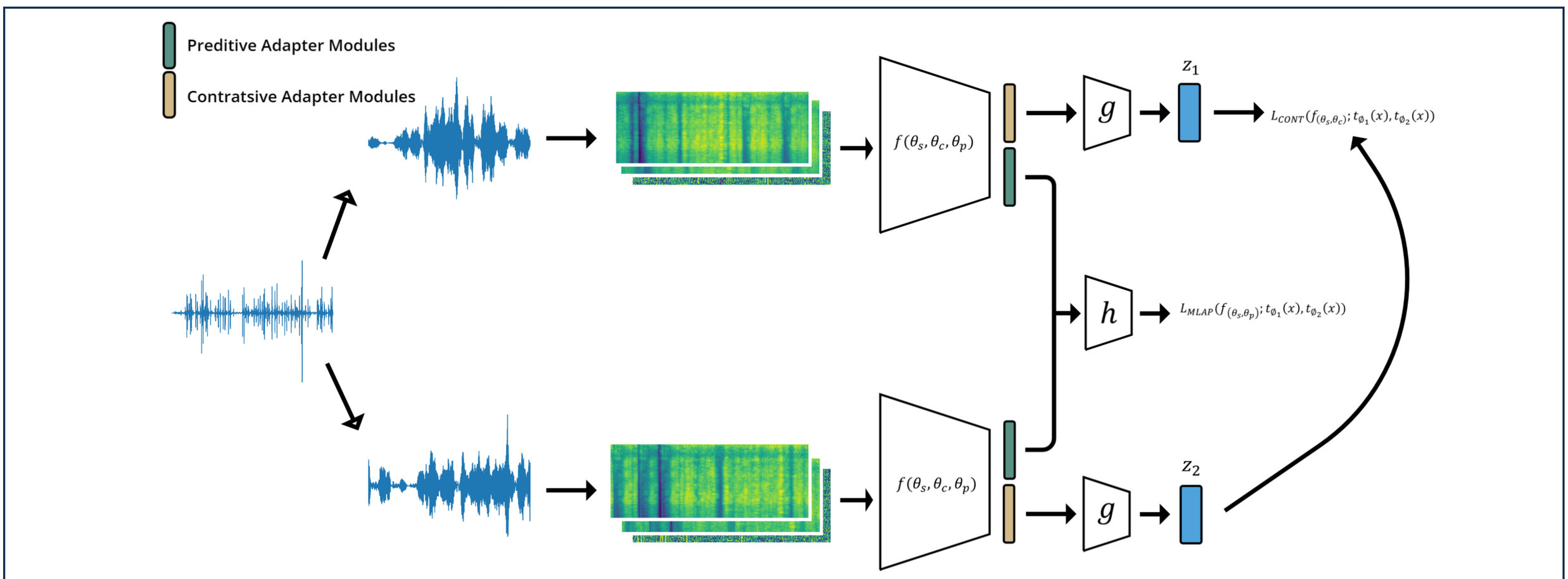

Figure 2: Main pipeline schematic for the MT-SLVR algorithm. The approach multi-task learns both a predictive (Multi-Label Augmentation Prediction) and contrastive (SimCLR) objectives. The main insight of our work is that by using lightweight neural adapters, we are able to learn both augmentation invariance and variance within a single model , allowing for better performance on a diverse set of downstream tasks.

4. Few-Shot Classification

- Although not limited to this, we evaluate on few-shot audio classification tasks spanning 10 datasets

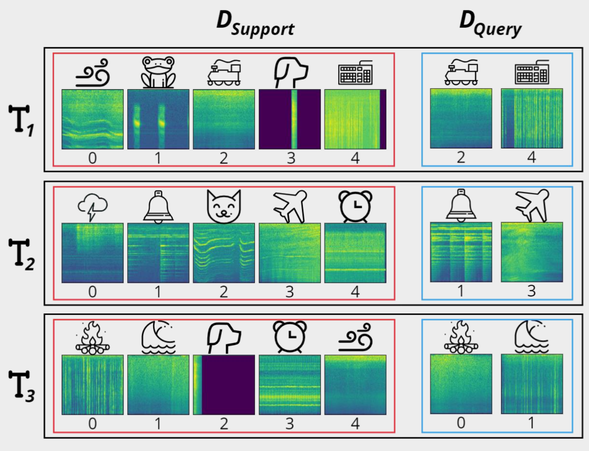

Figure 3: Example of 3 5-way 1-shot audio tasks

- We evaluate our models in their frozen state, with a new linear classifier assigned to each sampled task

5. Experimental Results

- We evaluate our approach using either SimCLR or SimSiam

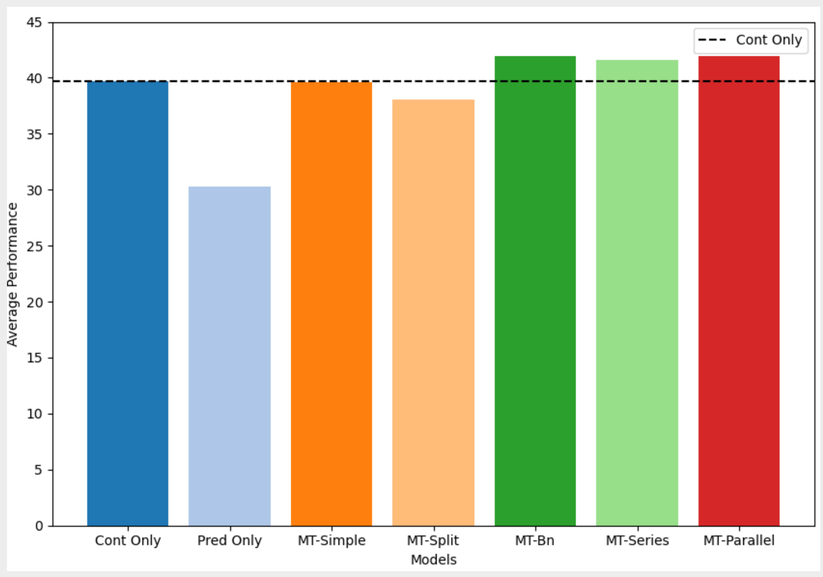

Figure 4: Average experimental results for SimCLR based approaches on 5-way 1-shot tasks. Averages are over 10 diverse datasets

- All adapter based approaches beat baselines and simple multi-task

6. Takeaways & Future Work

- We want to learn a model which can capture both side of an inductive bias to help aid downstream tasks
- Our approach (MT-SLVR) utilised lightweight adapters in order to achieve this
- MT-SLVR performed better than simple multi-task approaches as well as baselines across a wide variety of test sets
- Future work will include an expanded version of MT-SLVR as well as learning how to specialise inductive biases

UDRC THALES