A Variational Bayesian Approach to Learning Latent Variables for Acoustic Knowledge Transfer

Hu Hu[1], Sabato Marco Siniscalchi[1,2], Chao-Han Huck Yang[1], Chin-Hui Lee[1]
[1]Georgia Institute of Technology, GA, USA [2]University of Enna Kore, Italy

IEEE ICASSP 2022 Paper 2182

Overview

To address acoustic mismatches between training and testing conditions, we propose a **variational Bayesian (VB)** approach to learning distributions of latent variables in deep neural network (DNN) models for cross-domain knowledge transfer. Experimental results on device adaptation in acoustic scene classification show that our proposed approach can obtain good improvements on target devices, and consistently outperforms 13 SOTA knowledge transfer algorithms.

Keywords: Variational inference, Bayesian adaptation, knowledge distillation, latent variable, device mismatch.

Bayesian Adaptive Learning

The classical Bayes' theory
λ: model parameters.
$\mathcal{D}$: data.

$$p(\lambda|\mathcal{D}) = \frac{p(\mathcal{D}|\lambda)p(\lambda)}{p(\mathcal{D})}.$$

For Adaptation Scenarios
S: source domain.
T: target domain.

We have prior knowledge learnt from the source domain to be encoded in the prior distribution:

$$p(\lambda_T) = p(\lambda_S|\mathcal{D}_S).$$

The target domain posterior distribution:

$$p(\lambda_T|\mathcal{D}_T) = \frac{p(\mathcal{D}_T|\lambda_T)p(\lambda_S|\mathcal{D}_S)}{p(\mathcal{D}_T)}.$$

The posterior is usually intractable and difficult to get, thus an approximation is required, such as Maximum a posterior (MAP), variational Bayes (VB), Markov chain Monte Carlo (MCMC), ...
The VB is used in this work.

Variational Bayesian Knowledge Transfer

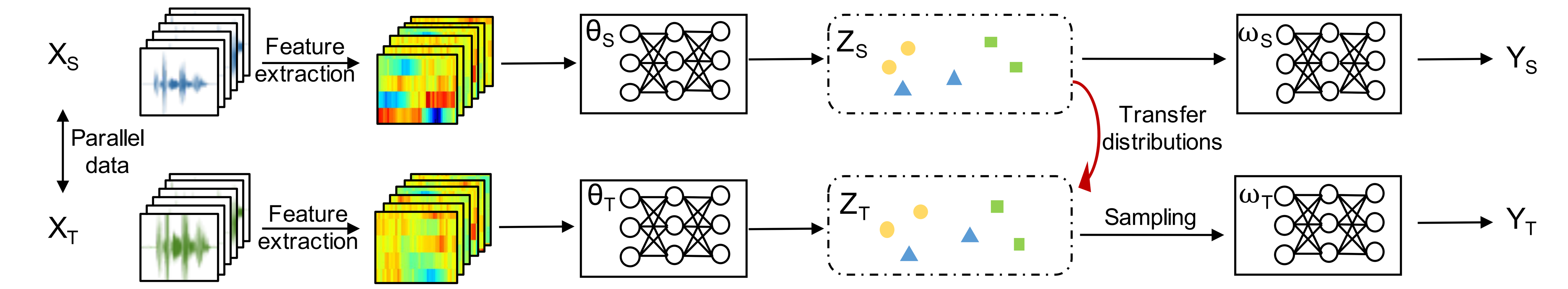


Figure 1: Illustration of the proposed knowledge transfer framework.

Knowledge Transfer of Latent Variables
Deep latent variable: An unobservable representation of data, corresponding to intermediate hidden embedding from a specific layer of DNN.
We propose to perform Bayesian adaptive learing on **deep latent variables** rather than on model weights.
Latent variables are introduced to model parameters λ, as shown in Figure 1:

$$p(\lambda) = p(Z, \theta, \omega) = p(Z|\theta)p(\theta)p(\omega).$$

Posterior of latent variables for target domain:

$$p(\lambda_T|\mathcal{D}_T) = \frac{p(\mathcal{D}_T|\lambda_T)p(\theta_T)p(\omega_T)p(Z_S|\theta_S, \mathcal{D}_S)}{p(\mathcal{D}_T)}.$$

Variational Bayesian Knowledge Transfer
Steps: 1) Set a variational distribution q.
2) Minimize the KLD between variational distribution and real distribution.
3) Take a non-informative prior over θ and ω.
4) Perform Gaussian mean-field variaitonal inference (GMFVI) estimation.
Final learning objective:

$$\mathcal{L}(\lambda_T; \mathcal{D}_T) = \sum_i^{N_T} \mathbb{E}_{z_T^{(i)} \sim \mathcal{N}(\mu_T^{(i)}, \sigma^2)} \log p(y_T^{(i)}|x_T^{(i)}, z_T^{(i)}, \theta_T, \omega_T) - \frac{1}{2\sigma^2}\sum_i^{N_T} \|\mu_T^{(i)} - \mu_S^{(i)}\|_2^2.$$

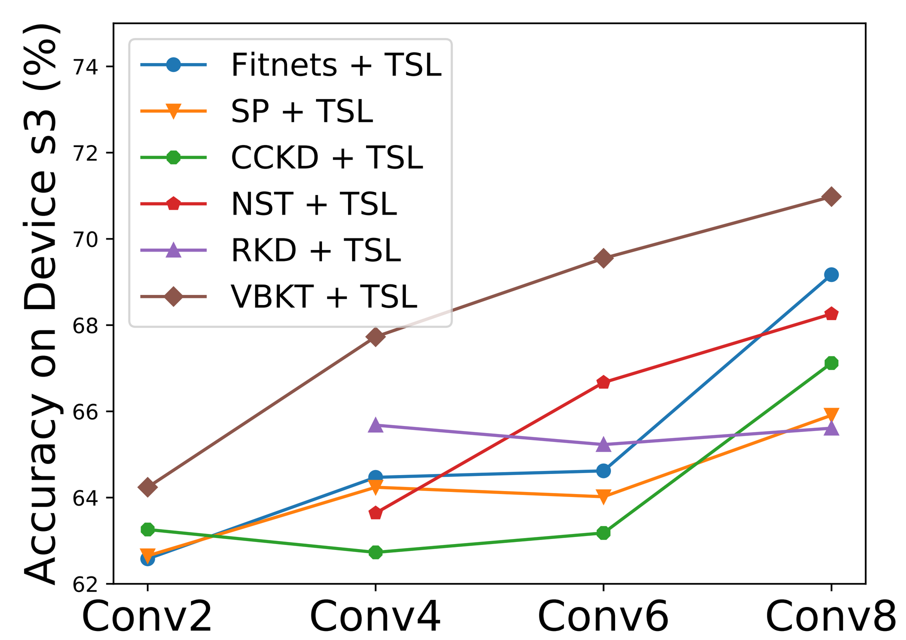
Experimental Evaluations

Selected Evaluation Results
Each method is tested with and without the combination of the TSL. Please refer to our paper for more details.

Method	RESNET avg. (%)	RESNET w/ TSL avg. (%)	FCNN avg. (%)	FCNN w/ TSL avg. (%)
Source.	37.70	-	37.13	-
No trans.	54.29	-	49.97	-
One-hot	63.76	-	64.45	-
TSL	68.04	68.04	66.27	66.27
NLE	65.64	67.76	64.47	64.53
AT	63.73	68.06	64.16	66.35
SP	64.57	68.45	65.74	67.36
RKD	65.28	68.46	65.63	67.27
VBKT	**69.58**	**69.90**	**69.96**	**70.50**

Data set: DCASE 2020 task1a development data set. Device adaptation is performed from 1 source recorder to 8 target recorders.

Effects of Hidden Embedding Depth

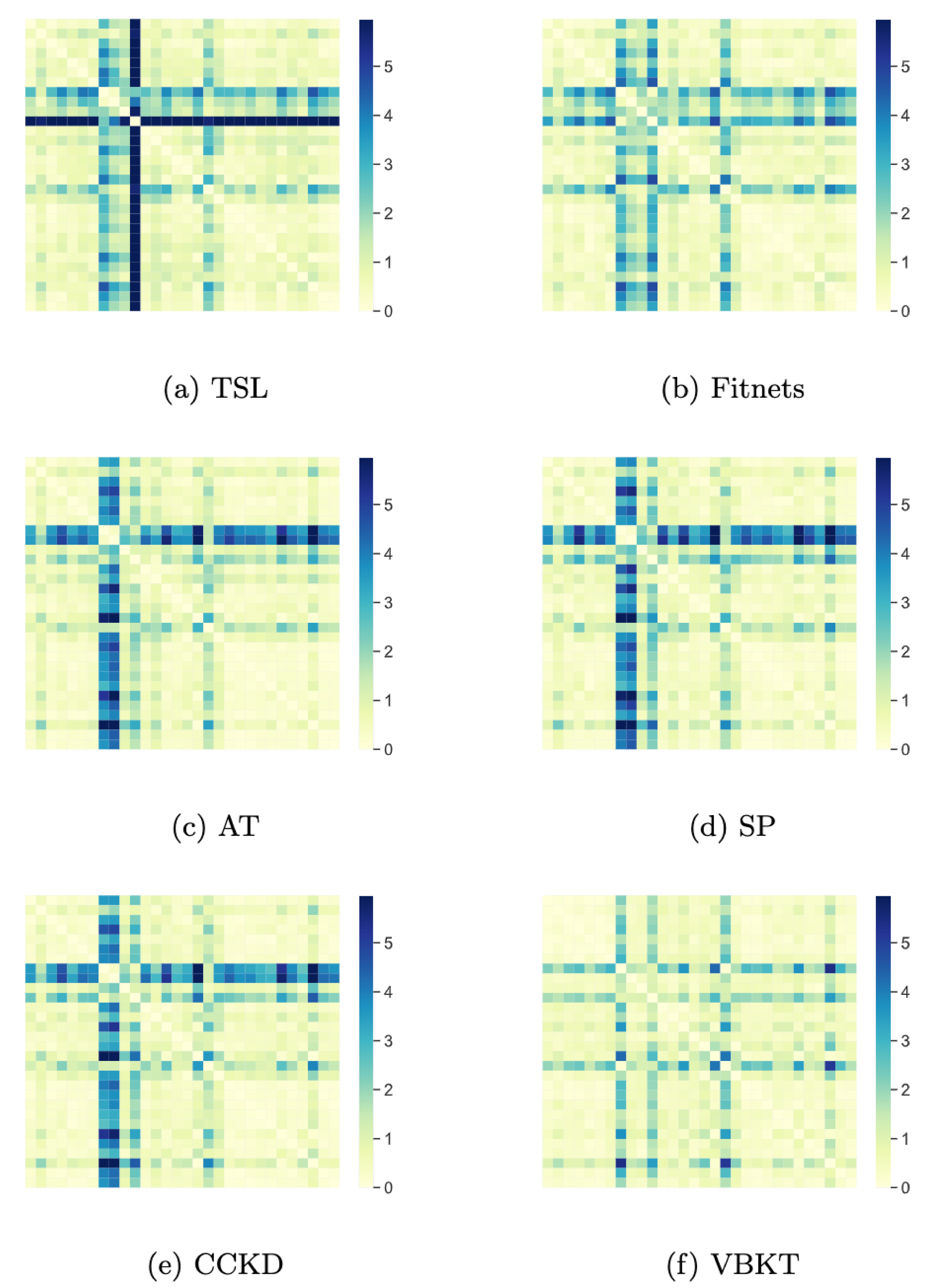


Visualization Analysis

Visualization of Intra-Class Discrepancy

- 30 samples from the same class are randomly selected.
- L2 distance between model outputs are computed and visualized.
- Darker color means bigger intra-class discrepancy.

The proposed method has consistent smaller intra-class discrepancy than others. It has more discriminative information and better cohesion of instances.



(a) TSL (b) Fitnets (c) AT (d) SP (e) CCKD (f) VBKT

More Information

Code available:
https://github.com/MihawkHu/ASC_Knowledge_Transfer