Supplementary to "Distilling Universal and Joint Knowledge for Cross-Domain Model Compression on Time Series Data"

1 Teacher and Student Model Architecture

The teacher and student model consist of three CNN blocks as the backbone and single fully connected layer as the classifier. Each CNN block contains a 1D convolutional layer, followed by a BatchNorm layer, an activation layer (ReLU) and a 1D Max-Pooling layer. Table 1 presents the details of teacher and student model. 'Conv1D' represents the 1D convolutional layer and the first variable in the bracket represents the number of input channels and the second one represents the number of output channels. 'BN' is a BatchNorm layer. 'FC' represents a fully connected layer. 'NC' represents the number of classes.

Table 1: Details of Teacher and Student Model

Blocks	Teacher	Student
#1	Conv1D (INPUT, 64) + BN+ReLU+MaxPool1d	Conv1D (INPUT, 16)+BN+ReLU+MaxPool1d
#2	Conv1D (64, 128)+BN+ReLU+MaxPool1d	Conv1D (16, 32)+BN+ReLU+MaxPool1d
#3	Conv1D (128, 128)+BN+ReLU+MaxPool1d	Conv1D (32, 32)+BN+ReLU+MaxPool1d
#4	Adaptive Average Pooling	Adaptive Average Pooling
#5	FC(128, NC)	FC(32, NC)

2 Details of Benchmark Approaches

We first compare the adversarial feature KD with some commonly-used feature distillation approaches:

- Fitnet: minimizes the L2 distance between teacher and student's feature maps;
- PKT: matches the probability distribution of the data instead of actual feature representations.
- AT: forces the student to mimic teacher's attention maps.

- IEKD: formulates the feature distillation loss into two parts: a similarity loss and a dis-similarity loss. The total loss encourages the student to not only inherit teacher's knowledge but also explore undiscovered knowledge.
- ICKD: minimizes the Mean Squared Error (MSE) of inter-channel correlation matrices between teacher and student.

We also compare our method with some advanced UDA approaches and SOTA UDA and KD combined approaches as listed:

- deep domain confusion (DDC): minimizes the MMD distance between source and target domains.
- minimum discrepancy for domain adaptation (MDDA): integrates MMD and CORAL for cross-domain feature adaptation.
- higher-order moment matching (HoMM): matches the higher-order moments between source and target domain.
- convolutional deep domain adaptation for time series data (CoDATS): utilize adversarial learning scheme for single-source domain adaptation.
- conditional adversarial domain adaptation (CDAN): leverages the discriminant information to enable the alignment of multi-modal distributions via conditional adversarial learning paradigm.
- decision-boundary iterative refinement training with a teacher (DIRT-T): utilizes a virtual adversarial DA model as the initialization and iteratively refines it with natural gradient to minimize the cluster assumption violation.
- joint knowledge distillation and unsupervised domain adaptation (JKU): jointly trains a teacher and student to simultaneously address the KD and UDA.
- adversarial adaptation with distillation (AAD): is built upon the ADDA and employs a teacher pre-trained on source domain for knowledge distillation.
- MobileDA: employs a model pre-trained on source domain as the teacher for KD and integrates the CORAL to minimize domain discrepancy.

3 Effectiveness of Adversarial Feature Distillation

To further validate the effectiveness of the adversarial feature distillation, we include more t-SNE visualization results from other three datasets (*i.e.*, **UCI HAR**,**FD** and **SSC**). As illustrated in Fig.1, Fig.2 and Fig.3, the features learned from our proposed method are more concentrated and all classes are well separated without overlapping. This observation demonstrates that the adversarial feature KD scheme could efficiently transfer the universal feature-level knowledge to the student for cross-domain scenario.

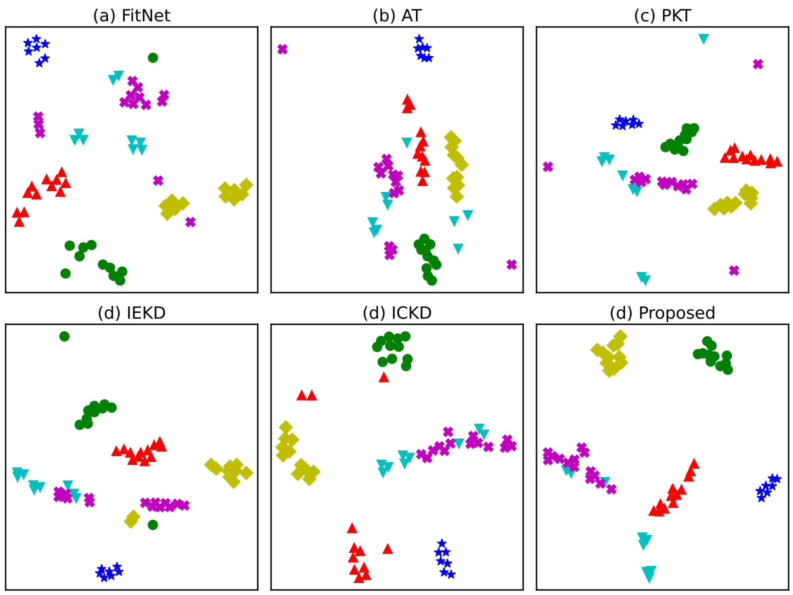

Figure 1: t_SNE of different feature distillation methods on UCI HAR.

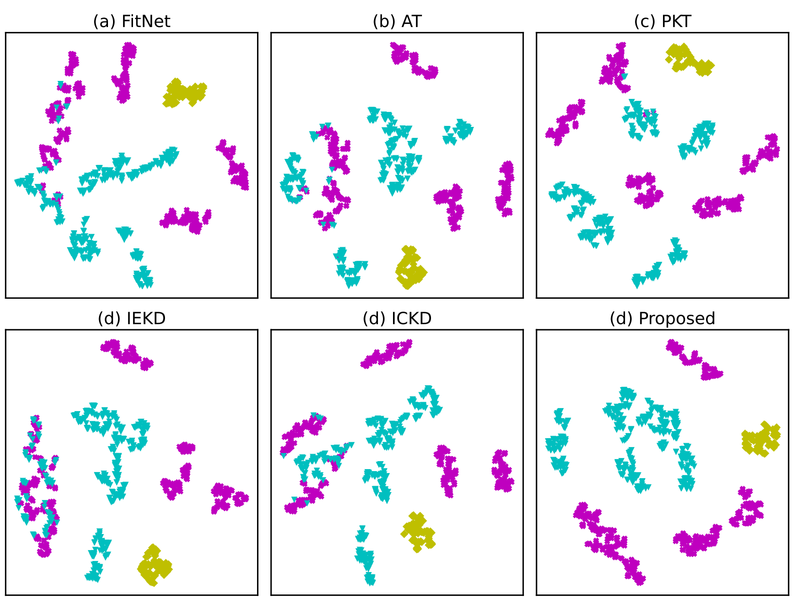

Figure 2: t_SNE of different feature distillation methods on FD.

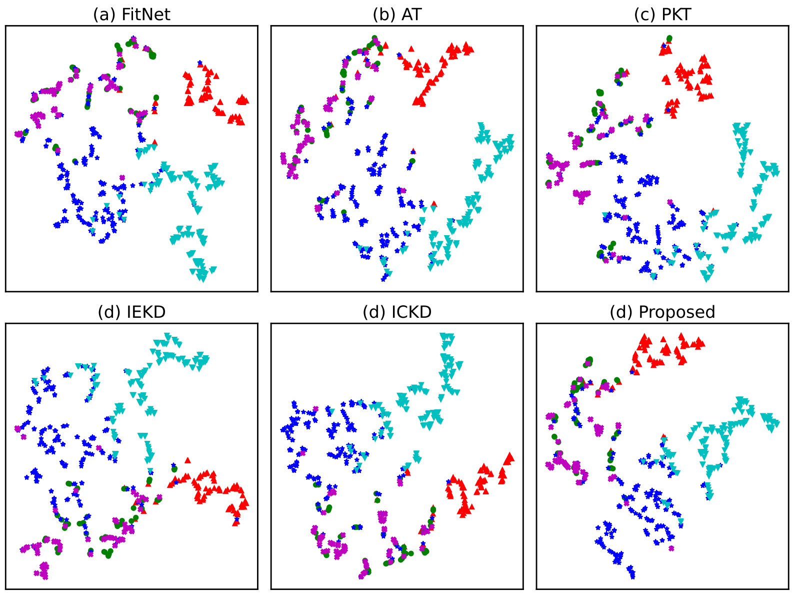

Figure 3: t_SNE of different feature distillation methods on SSC.

4 Additional Transfer Scenarios

To further investigate the effectiveness of our proposed approach, we select five additional transfer scenarios from four datasets and report the results as Table 2 and Table 3 show. We can see that our proposed method consistently outperforms other benchmark methods on these additional transfer scenarios. These results further prove that our proposed approach could significantly enhance the performance of a compact model with the knowledge from a complex teacher for cross-domain scenario compared with other SOTA methods.

Table 4 summaries the averaged Macro F1-score and standard deviation over all ten transfer scenarios for four datasets. We can observe that our approach achieve better performance than other methods.

Table 2: Marco F1-score on UCI HAR and HHAR across three independent runs on five additional scenarios

Methods	UCI HAR Transfer Scenario						HHAR Transfer Scenario					
	$18 \rightarrow 27$	$20 \rightarrow 5$	$24 \rightarrow 8$	$28 \rightarrow 27$	$30 \rightarrow 20$	Avg	$0 \rightarrow 2$	$5 \rightarrow 0$	$6 \rightarrow 1$	$7 \rightarrow 4$	$8 \rightarrow 3$	Avg
Teacher	100.00	94.19	99.54	100.00	91.28	97.00	65.59	42.20	95.10	95.00	96.62	78.90
Student (src-only)	96.44	49.23	74.66	71.41	71.82	72.71	51.85	32.65	78.69	78.13	83.07	64.88
DDC	92.74	66.59	84.74	96.41	78.77	83.85	61.68	33.98	74.93	83.47	80.72	66.96
MDDA	**100.00**	84.16	68.44	95.85	83.05	86.30	64.10	33.35	91.15	94.18	96.26	75.81
HoMM	98.61	75.04	94.95	96.73	78.12	88.69	61.62	29.18	85.96	85.72	92.11	70.92
CoDATS	95.06	91.61	89.39	91.57	63.03	86.13	61.89	34.97	91.67	94.90	85.24	73.73
CDAN	**100.00**	87.19	94.53	**100.00**	84.88	93.32	65.36	33.11	85.47	91.79	93.88	73.92
DIRT-T	**100.00**	93.15	97.92	95.32	85.91	94.46	**67.93**	33.64	90.46	93.80	95.41	76.25
JKU	97.97	63.25	92.62	96.30	74.50	84.93	62.24	35.29	82.38	78.42	91.10	69.89
AAD	90.27	66.88	86.09	94.73	84.82	84.56	58.23	37.24	69.10	81.99	83.61	66.03
MobileDA	92.86	84.96	90.45	79.12	77.56	84.99	50.27	30.83	76.12	89.70	79.25	65.23
Proposed	**100.00**	**94.42**	**100.00**	92.26	**87.10**	**94.76**	62.33	**39.01**	**92.89**	**96.90**	**96.52**	**77.53**

Table 3: Marco F1-score on Bearing FD and SSC across three independent runs on five additional scenarios

Methods	FD Transfer Scenario						SSC Transfer Scenario					
	$1 \rightarrow 0$	$1 \rightarrow 3$	$3 \rightarrow 0$	$3 \rightarrow 1$	$3 \rightarrow 2$	Avg	$3 \rightarrow 19$	$5 \rightarrow 15$	$6 \rightarrow 2$	$13 \rightarrow 17$	$18 \rightarrow 12$	Avg
Teacher	62.46	100.00	59.87	100.00	92.60	82.99	71.43	71.62	76.44	52.03	58.43	65.99
Student (src-only)	45.34	95.89	44.76	99.54	72.11	71.53	59.57	58.58	70.47	37.44	32.98	51.81
DDC	43.32	96.81	53.18	97.84	74.52	73.13	63.70	60.39	**72.61**	44.60	46.14	57.49
MDDA	40.04	**99.97**	42.85	97.94	79.14	71.99	68.72	65.78	70.49	48.77	49.01	60.55
HoMM	42.96	92.34	48.05	99.26	79.04	72.33	57.19	65.07	63.14	43.81	38.92	53.63
CoDATS	46.89	99.04	49.62	99.89	**87.54**	76.60	60.76	63.35	71.27	45.88	45.90	57.43
CDAN	37.78	97.80	43.00	99.62	86.84	73.01	**69.28**	63.52	69.59	38.56	35.55	55.30
DIRT-T	53.73	99.64	51.79	99.51	85.17	77.97	66.89	65.96	68.38	46.39	44.40	58.40
JKU	44.39	97.01	48.82	99.48	78.72	73.68	59.54	67.29	70.48	43.72	48.42	57.89
AAD	46.42	94.65	52.09	98.65	87.11	75.78	62.75	64.81	71.78	44.52	49.18	58.61
MobileDA	51.71	94.92	51.17	99.86	78.51	75.23	64.16	67.67	56.74	47.50	**56.56**	58.53
Proposed	**60.91**	**99.97**	**61.00**	**100.00**	87.08	**81.79**	66.84	**70.76**	65.70	**50.19**	49.77	**60.65**

Table 4: Averaged Marco F1-score and standard deviation over all ten transfer scenarios

Methods	UCI HAR	HHAR	FD	SSC
Student (src-only)	65.48±11.32	61.18±7.92	68.79±4.66	51.49±4.66
DDC	80.18±5.48	65.14±8.76	72.41±2.59	58.25±3.26
MDDA	83.88±6.73	76.61±3.30	76.89±3.79	59.27±**2.53**
HoMM	85.43±2.83	72.16±5.74	73.43±2.67	55.79±5.43
CoDATS	81.60±3.93	70.97±5.84	77.07±**1.48**	57.32±8.34
CDAN	87.54±2.58	73.64±2.73	78.80±6.03	56.49±6.23
DIRT-T	88.76±1.99	77.24±3.45	83.37±1.79	59.14±3.93
JKU	81.32±9.21	70.58±3.85	71.37±4.15	56.87±5.22
AAD	83.78±5.79	70.66±6.55	76.81±4.90	58.69±4.67
MobileDA	84.33±5.06	69.36±3.33	71.87±2.97	58.59±6.43
Proposed	**90.80±1.72**	**78.43±1.60**	**85.93**±1.84	**61.41**±3.20