Supplementary Material:
Joint-Label Learning by Dual Augmentation for Time Series Classification

Qianli Ma[1,2], **Zhenjing Zheng**[1], **Jiawei Zheng**[1], **Sen Li**[1], **Wanqing Zhuang**[1], **Garrison W. Cottrell**[3]

[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou
[2]Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education
[3]Department of Computer Science and Engineering, University of California, San Diego, CA, USA
qianlima@scut.edu.cn, 982360227@qq.com

A. Introduction to the datasets and baselines

In subsection Comparison with State-of-the-art Methods of the main text, we report the experimental results of JobDA on 85 UCR[1] (Chen et al. 2015) datasets. Here, we show the statistics of these 85 datasets in Table 5.

The proposed method is compared with three SOTA deep learning-based time series classification methods (Wang, Yan, and Oates 2017): Multilayer Perceptron (MLP), Fully Convolutional Network (FCN), and Residual Network (ResNet). The introduction to these three baselines are as follow:

- **MLP**: The MLP stacks three fully-connected layers with 500 neurons for each layer, and the softmax layer is used to obtain the classification results. The MLP is trained with Adadelta (Zeiler 2012) with learning rate 1.0, $\rho = 0.95$ and $\epsilon = 1e-6$.
- **FCN**: The FCN stacks three convolution blocks with 128, 256, and 128 filters in each block, where the filter size in each block is 8, 5, and 3. the convolutional results are fed into a global average pooling layer and a softmax layer to get the classification results. The FCN is trained with Adam (Kingma and Ba 2014) with the learning rate 0.001, $\beta1 = 0.9, \beta2 = 0.999$ and $\epsilon = 1e-8$.
- **ResNet**: The ResNet stacks three residual blocks, each of which contains three convolution blocks. The number of filters in the three residual blocks are 64, 128, and 128, respectively. The global average pooling layer and a softmax layer are used to obtain classification results. The ResNet is trained with Adam (Kingma and Ba 2014) with the learning rate 0.001, $\beta1 = 0.9, \beta2 = 0.999$ and $\epsilon = 1e-8$.

B. Exploration of Different Time-series Transformations

In this paper, we propose a novel time-series transformation called Time-Series Warping (TSW). TSW alternately compresses and expands different subsequences of the time series by using downsampling and upsampling operation while

[1]https://www.cs.ucr.edu/~eamonn/time_series_data/

Table 4: Classification accuracy (%) of four time-series transformations on 44 UCR datasets. The best accuracy is indicated as bold.

Dataset	Flipping	Jittering	Scaling	TSW
ArrowHead	0.840	0.842	0.806	**0.857**
Beef	0.744	0.722	0.744	**0.760**
BeetleFly	0.883	**0.900**	0.817	0.890
BirdChicken	**0.883**	**0.883**	0.800	0.860
Car	0.911	**0.939**	0.889	0.927
CBF	**0.994**	0.933	0.983	0.981
CinCECGTorso	**0.838**	0.807	0.819	0.835
Coffee	**1.000**	**1.000**	0.988	**1.000**
DiatomSizeR	0.947	0.960	0.938	**0.975**
DistPhaxAgeGrp	0.791	**0.797**	0.783	0.786
DistPhaxTW	0.759	**0.767**	0.754	0.759
Earthquakes	0.713	0.746	**0.764**	0.754
ECG200	0.887	0.857	0.890	**0.894**
ECGFiveDays	**0.998**	0.979	0.994	0.993
FaceFour	**0.955**	**0.955**	0.936	**0.955**
FacesUCR	**0.956**	0.922	0.926	**0.956**
Fish	0.987	0.981	0.985	**0.991**
GunPoint	0.998	0.993	0.989	**1.000**
Ham	0.762	0.743	0.743	**0.764**
Haptics	0.500	0.526	0.526	**0.555**
Herring	0.625	0.583	**0.646**	0.619
InlineSkate	0.334	0.363	0.346	**0.398**
ItalyPower	0.962	0.942	**0.964**	0.958
Lightning2	0.781	0.765	0.760	**0.787**
Lightning7	**0.831**	0.808	0.795	0.781
Mallat	0.965	0.964	0.951	**0.968**
Meat	**0.967**	0.956	0.922	0.940
MoteStrain	**0.904**	0.900	0.862	0.900
MidPhaxAgeGrp	**0.743**	**0.743**	0.738	0.740
MidPhaxTW	**0.611**	0.592	0.587	0.586
OliveOil	0.789	0.856	0.844	**0.860**
OSULeaf	0.983	**0.989**	0.949	0.984
Plane	**1.000**	**1.000**	**1.000**	**1.000**
ShapeletSim	**1.000**	0.826	0.946	**1.000**
SonyAIBORobot1	0.971	0.946	0.904	**0.978**
SonyAIBORobot2	0.954	0.952	0.876	**0.970**
Symbols	0.924	0.946	0.899	**0.959**
ToeSegmentation1	**0.978**	0.966	0.942	0.933
ToeSegmentation2	0.900	0.879	0.903	**0.934**
Trace	**1.000**	**1.000**	**1.000**	**1.000**
TwoLeadECG	**1.000**	0.996	0.999	**1.000**
Wine	0.716	0.605	0.722	**0.737**
Worms	0.617	**0.630**	0.599	0.629
WormsTwoClass	0.705	**0.716**	0.713	0.692

Table 5: Statistics of 85 UCR datasets.

Datasets	#Train	#Test	#Class	Length	Dataset type	Meat	60	60	3	448	Spectro
Adiac	390	391	37	176	Image	MedicalImages	381	760	10	99	Image
ArrowHead	36	175	3	251	Image	MidPhxAgeGrp	154	400	3	80	Image
Beef	30	30	5	470	Spectro	MidPhxCorr	291	600	2	80	Image
BeetleFly	20	20	2	512	Image	MidPhxTW	154	399	6	80	Image
BirdChicken	20	20	2	512	Image	MoteStrain	20	1252	2	84	Sensor
Car	60	60	4	577	Sensor	NonInv_Thor1	1800	1965	42	750	ECG
CBF	30	900	3	128	Simulated	NonInv_Thor2	1800	1965	42	750	ECG
Chlorine	467	3840	3	166	Sensor	OliveOil	30	30	4	570	Spectro
CinCECGTorso	40	1380	4	1639	Sensor	OSULeaf	200	242	6	427	Image
Coffee	28	28	2	286	Spectro	PhalCorr	1800	858	2	80	Image
Computers	250	250	2	720	Device	Phoneme	214	1896	39	1024	Sensor
CricketX	390	390	12	300	Motion	Plane	105	105	7	144	Sensor
CricketY	390	390	12	300	Motion	ProxPhxAgeGp	400	205	3	80	Image
CricketZ	390	390	12	300	Motion	ProxPhxCorr	600	291	2	80	Image
DiatomSizeR	16	306	4	345	Image	ProxPhxTW	205	400	6	80	Image
DisPhxAgeGp	139	400	3	80	Image	RefrigerationDevices	375	375	3	720	Device
DisPhxCorr	276	600	2	80	Image	ScreenType	375	375	3	720	Device
DisPhxTW	139	400	6	80	Image	ShapeletSim	20	180	2	500	Simulated
Earthquakes	139	322	2	512	Sensor	ShapesAll	600	600	60	512	Image
ECG200	100	100	2	96	ECG	SmlKitApp	375	375	3	720	Device
ECG5000	500	4500	5	140	ECG	SonyAIBORobot1	20	601	2	70	Sensor
ECGFiveDays	23	861	2	136	ECG	SonyAIBORobot2	27	953	2	65	Sensor
ElectricDevices	8926	7711	7	96	Device	StarLightCurves	1000	8236	3	1024	Sensor
FaceAll	560	1690	14	131	Image	Strawberry	370	613	2	235	Spectro
FaceFour	24	88	4	350	Image	SwedishLeaf	500	625	15	128	Image
FacesUCR	200	2050	14	131	Image	Symbols	25	995	6	398	Image
FiftyWords	450	455	50	270	Image	SyntheticControl	300	300	6	60	Simulated
Fish	175	175	7	463	Image	ToeSegmentation1	40	228	2	277	Motion
FordA	1320	3601	2	500	Sensor	ToeSegmentation2	36	130	2	343	Motion
FordB	810	3636	2	500	Sensor	Trace	100	100	4	275	Sensor
GunPoint	50	150	2	150	Motion	TwoLeadECG	23	1139	2	82	ECG
Ham	109	105	2	431	Spectro	TwoPatterns	1000	4000	4	128	Simulated
HandOutlines	370	1000	2	2709	Image	UWaveGestAll	896	3582	8	945	Motion
Haptics	155	308	5	1092	Motion	UWaveGest_X	896	3582	8	315	Motion
Herring	64	64	2	512	Image	UWaveGest_Y	896	3582	8	315	Motion
InlineSkate	100	550	7	1882	Motion	UWaveGest_Z	896	3582	8	315	Motion
InsectWing	220	1980	11	256	Sensor	Wafer	1000	6164	2	152	Sensor
ItalyPower	67	1029	2	24	Sensor	Wine	57	54	2	234	Spectro
LrgKitApp	375	375	3	720	Device	WordSynonyms	267	638	25	270	Image
Lightning2	60	61	2	637	Sensor	Worms	77	181	5	900	Motion
Lightning7	70	73	7	319	Sensor	WormsTwoClass	77	181	2	900	Motion
Mallat	55	2345	8	1024	Simulated	Yoga	300	3000	2	426	Image

keeping the length of the time series unchanged. In addition to TSW, we have tried three additional time-series transformations, and their introduction are as follow: 1. Flipping: Flipping the sign of original time series; 2. Jittering: Adding zero-mean Gaussian noise ($\mu = 0$, $\sigma = 0.03$); 3. Scaling: Multiply all elements by a scalar drawn from a Gaussian ($\mu = 1$, $\sigma = 0.1$). Based on Time-Series Warping ($N = 2$) and these three transformations, we analyze the impact of different time-series transformations on model performance from the perspective of dataset type, sequence length, and number of categories. The results of four time-series transformations (Use ResNet as the classifier) on 44 UCR datasets are shown in Table 4.

Dataset Type. UCR time-series classification archive includes many different types of datasets, such as Image, Spectro, Sensor, etc. There are usually huge differences between different types of datasets, so it is necessary to explore the performance of different transformations on different types of datasets. The statistical results of Time-Series Warping ($N = 2$) and these three time-series transformations on several datasets are shown in Table 6.

From Table 6, we can draw two conclusions. First, each transformation may prefer a certain type of dataset (e.g., Jittering works better on Image). Second, TSW is suitable for a wider range of tasks, which performs better on different types of datasets.

Sequence Length. As shown in Tabel 5, the sequence length of different datasets varies greatly. Therefore, we do further exploration on the sequence length. We divide the sequence length into 3 intervals and report the number of datasets with the best results for different transformations. The statistical results of four time-series transfor-

Table 6: Statistical results of four time-series transformations on different types of datasets.

Type	Image (# of best)	Spectro (# of best)	Sensor (# of best)	Simulated (# of best)	ECG (# of best)	Motion (# of best)	# of Total_Best	Avg_rank
ResNet_JL_Flipping	5 / 14	2 / 6	5 / 11	2 / 3	2 / 3	1 / 7	17 / 44	2.239
ResNet_JL_Jittering	7 / 14	1 / 6	3 / 11	0 / 3	0 / 3	2 / 7	13 / 44	2.625
ResNet_JL_Scaling	1 / 14	0 / 6	4 / 11	0 / 3	0 / 3	0 / 7	5 / 44	3.148
ResNet_JL_TSW ($N = 2$)	6 / 14	5 / 6	5 / 11	2 / 3	2 / 3	4 / 7	**24 / 44**	**1.989**

Table 7: Statistical results of four time-series transformations on different sequence length of dataset.

Length	$1 - 300$ (# of best)	$301 - 600$ (# of best)	> 600 (# of best)
Flipping	11 / 20	5 / 17	1 / 7
Jittering	6 / 20	5 / 17	2 / 7
Scaling	3 / 20	2 / 17	0 / 7
TSW ($N = 2$)	11 / 20	9 / 17	4 / 7

Table 8: Statistical results of four time-series transformations on different number of categorie of datasets.

Number of categorie	$1 - 4$ (# of best)	> 4 (# of best)
Flipping	13 / 31	4 / 13
Jittering	9 / 31	4 / 13
Scaling	4 / 31	1 / 13
TSW ($N = 2$)	16 / 31	8 / 13

mations on several datasets are shown in Table 7. We can see that all transformations prefer datasets with shorter sequence lengths. In addition, TSW still has relatively good performance in datasets of different sequence lengths.

Number of Categories. Finally, we do further exploration on the number of categories. We divide the sequence length into 2 intervals and report the number of datasets with the best results for different transformations. The statistical results of four time-series transformations on several datasets are shown in Table 8. For different number of categories of datasets, we can see that Flipping and Scaling prefer datasets with small number of categories. TSW achieves relatively good performance in datasets of different number of categories similar to the result of sequence length.

In general, for each time-series transformation, it will perform better on a specific dataset (e.g., datasets with shorter sequence lengths or small number of categories). To illustrate our approach, we just used one transformation for this paper. In addition, establishing a frame work for understanding which transformations are useful for which datasets is interesting. We leave this question for future work.

C. Comparison with 1NN-DTW

For smaller UCR time series datasets, traditional time series classification methods have achieved good performance. For example, KNN and dynamic time warping has been hard to beat for some time now. To further verify the performance of JobDA, we compare ResNet_SL with 1NN-DTW on 44 UCR datasets with training set sizes of 200 or less.

To make a fair comparison, we compared our ResNet_JL with 1NN-DTW (trained on the original dataset), 1NN-DTW_Trans (trained on the augmented datasets using Flipping, Jittering and Scaling), and 1NN-DTW_TSW (trained

Table 9: Classification accuracy (%) of ResNet_JL and three 1NN-DTW-based methods on 44 UCR datasets. The best accuracy is indicated as bold.

Dataset	1NN-DTW	1NN-DTW_Trans	1NN-DTW_TSW	ResNet_JL
ArrowHead	0.680	0.686	0.686	**0.850**
Beef	0.567	0.567	0.567	**0.833**
BeetleFly	0.700	0.700	0.700	**0.940**
BirdChicken	0.750	0.750	0.750	**0.900**
Car	0.750	0.767	0.733	**0.943**
CBF	**1.000**	**1.000**	**1.000**	0.990
CinCECGTorso	0.691	0.736	0.703	**0.841**
Coffee	0.964	0.964	0.964	**1.000**
DiatomSizeR	0.961	0.948	0.961	**0.965**
DistPhaxAgeGrp	0.795	0.795	0.795	**0.798**
DistPhaxTW	0.728	0.720	0.728	**0.774**
Earthquakes	0.730	0.705	0.730	**0.767**
ECG200	0.800	0.800	0.800	**0.878**
ECGFiveDays	0.775	0.782	0.805	**0.995**
FaceFour	0.841	0.818	0.864	**0.957**
FacesUCR	0.934	0.935	**0.938**	0.831
Fish	0.863	0.846	0.863	**0.986**
GunPoint	0.880	0.887	0.887	**1.000**
Ham	0.562	0.533	0.562	**0.779**
Haptics	0.364	0.351	0.364	**0.564**
Herring	0.547	0.516	0.547	**0.581**
InlineSkate	0.375	0.378	0.375	**0.395**
ItalyPower	0.946	0.945	0.941	**0.963**
Lightning2	0.803	0.820	**0.836**	0.810
Lightning7	**0.767**	0.726	0.726	**0.767**
Mallat	0.914	0.917	0.909	**0.972**
Meat	**0.933**	**0.933**	**0.933**	0.917
MoteStrain	0.891	**0.895**	0.883	0.879
MidPhaxAgeGrp	0.745	0.743	0.745	**0.756**
MidPhaxTW	0.576	0.579	0.576	**0.589**
OliveOil	0.833	0.833	0.833	**0.867**
OSULeaf	0.636	0.653	0.632	**0.988**
Plane	**1.000**	0.990	**1.000**	**1.000**
ShapeletSim	0.756	0.661	0.794	**1.000**
SonyAIBORobot1	0.712	0.719	0.730	**0.966**
SonyAIBORobot2	0.843	0.837	0.846	**0.961**
Symbols	0.953	0.948	**0.956**	**0.956**
ToeSegmentation1	0.798	0.776	0.811	**0.975**
ToeSegmentation2	0.846	0.823	0.831	**0.929**
Trace	0.990	0.990	0.990	**1.000**
TwoLeadECG	0.931	0.937	0.928	**1.000**
Wine	0.593	0.593	0.593	**0.889**
Worms	0.464	0.475	0.481	**0.634**
WormsTwoClass	**0.685**	0.669	0.657	0.650
#Best	5	3	6	**38**
Avg_rank	2.875	3.011	2.682	**1.432**

Table 10: Classification accuracy (%) and standard deviation between parentheses of three deep learning-based methods, three methods using single-label learning and our joint-label learning method (ResNet) on 85 UCR datasets. The best accuracy is indicated as bold.

Dataset	MLP	FCN	ResNet	MLP_SL	FCN_SL	ResNet_SL	ResNet_JL
Adiac	39.7(1.9)	**84.4(0.7)**	82.9(0.6)	35.0(0.9)	83.1(0.9)	79.5(0.8)	69.0(0.3)
ArrowHead	77.8(1.2)	84.3(1.5)	84.5(1.2)	77.3(0.8)	82.6(0.6)	80.5(0.4)	**85.0(0.9)**
Beef	72.0(2.8)	69.7(4.0)	75.3(4.2)	64.7(2.7)	71.3(5.4)	74.0(2.5)	**83.3(0.0)**
BeetleFly	87.0(2.6)	86.0(9.7)	85.0(2.4)	78.0(2.4)	93.0(2.4)	92.0(2.4)	**94.0(3.7)**
BirdChicken	77.5(3.5)	**95.5(3.7)**	88.5(5.3)	77.0(2.4)	90.0(0.0)	90.0(0.0)	90.0(0.0)
Car	76.7(2.6)	90.5(1.4)	92.5(1.4)	85.7(0.8)	87.0(3.7)	87.7(1.7)	**94.3(0.8)**
CBF	87.2(0.7)	99.4(0.1)	99.5(0.3)	89.6(0.5)	**99.9(0.0)**	**99.9(0.1)**	99.0(0.2)
ChlorineConcentration	80.2(1.1)	81.4(0.9)	**84.4(1.0)**	76.4(1.1)	81.2(0.7)	82.7(1.0)	83.7(0.6)
CinCECGTorso	84.0(1.0)	82.4(1.2)	82.6(2.4)	77.9(1.3)	82.0(0.5)	81.6(1.2)	**84.1(1.8)**
Coffee	99.6(1.1)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	96.4(0.0)	**100.0(0.0)**
Computers	56.3(1.6)	82.2(1.0)	81.5(1.2)	54.7(1.1)	**83.7(0.7)**	82.9(1.2)	79.9(0.7)
CricketX	59.1(1.1)	79.2(0.7)	79.1(0.6)	60.3(0.6)	77.4(0.7)	**82.4(0.8)**	66.4(0.5)
CricketY	60.0(0.8)	78.7(1.2)	80.3(0.8)	58.4(1.1)	75.3(1.3)	**82.5(1.4)**	64.2(1.3)
CricketZ	61.7(0.8)	81.1(1.0)	**81.2(1.4)**	62.3(0.7)	75.5(0.8)	78.0(1.0)	68.4(1.1)
DiatomSizeReduction	91.0(1.4)	31.3(3.6)	30.1(0.2)	94.4(0.5)	96.3(0.7)	**99.0(0.4)**	96.5(0.7)
DistalPhalanxOutlineAgeGroup	65.7(1.1)	71.0(1.3)	71.7(1.3)	**81.2(1.0)**	80.6(0.8)	78.4(0.9)	79.8(0.5)
DistalPhalanxOutlineCorrect	72.6(1.3)	76.0(1.5)	77.1(1.0)	78.2(0.5)	80.3(0.2)	80.0(0.7)	**81.2(1.4)**
DistalPhalanxTW	61.7(1.3)	69.0(2.1)	66.5(1.6)	75.3(0.8)	76.4(0.9)	77.0(0.2)	**77.4(0.7)**
Earthquakes	71.7(1.3)	72.7(1.7)	71.2(2.0)	75.2(0.7)	76.5(0.8)	74.5(1.2)	**76.7(2.0)**
ECG200	**91.6(0.7)**	88.9(1.0)	87.4(1.9)	86.6(1.0)	85.8(0.7)	87.0(2.1)	87.8(1.3)
ECG5000	92.9(0.1)	**94.0(0.1)**	93.4(0.2)	93.1(0.1)	93.8(0.1)	93.4(0.1)	93.8(0.2)
ECGFiveDays	97.0(0.5)	98.7(0.3)	97.5(1.9)	95.2(0.4)	90.7(2.3)	98.2(1.3)	**99.5(0.2)**
ElectricDevices	59.2(1.1)	70.2(1.2)	72.9(0.9)	61.5(0.9)	72.5(0.9)	72.5(0.5)	**73.2(0.4)**
FaceAll	79.3(1.1)	94.5(0.9)	83.9(2.0)	84.6(2.3)	**96.4(0.5)**	84.2(0.3)	71.6(0.2)
FaceFour	84.0(1.4)	92.8(0.9)	95.5(0.0)	84.1(0.7)	91.4(0.6)	91.4(2.4)	**95.7(0.5)**
FacesUCR	83.3(0.3)	94.6(0.2)	**95.5(0.4)**	80.0(0.8)	93.6(0.1)	**95.5(0.1)**	83.1(0.4)
FiftyWords	68.4(7.1)	62.7(6.1)	**74.0(1.5)**	71.9(0.2)	64.4(0.7)	70.4(0.7)	49.8(1.3)
Fish	84.8(0.8)	95.8(0.6)	97.9(0.8)	86.2(0.9)	96.6(0.5)	98.3(0.4)	**98.6(0.3)**
FordA	73.0(0.4)	90.4(0.2)	**92.0(0.4)**	67.1(0.4)	89.5(0.2)	91.0(0.2)	**92.0(0.3)**
FordB	60.3(0.3)	87.8(0.6)	**91.3(0.3)**	56.0(0.5)	87.8(0.3)	88.4(0.6)	89.9(0.2)
GunPoint	92.7(1.1)	**100.0(0.0)**	99.1(0.7)	95.2(0.3)	**100.0(0.0)**	98.8(0.3)	**100.0(0.0)**
Ham	69.1(1.4)	71.8(1.4)	75.7(2.7)	72.2(1.8)	71.6(1.5)	75.4(3.5)	**77.9(1.8)**
HandOutlines	**91.8(0.5)**	80.6(7.9)	91.1(1.4)	81.0(1.0)	76.4(4.1)	85.6(0.8)	86.3(0.9)
Haptics	43.3(1.4)	48.0(2.4)	51.9(1.2)	42.2(1.2)	46.5(1.1)	54.0(0.7)	**56.4(0.9)**
Herring	52.8(3.9)	60.8(7.7)	61.9(3.8)	50.0(2.8)	**64.7(1.9)**	57.8(3.0)	58.1(2.3)
InlineSkate	33.7(1.0)	33.9(0.8)	37.3(0.9)	35.1(1.0)	36.9(0.8)	36.8(1.6)	**39.5(1.1)**
InsectWingbeatSound	**60.7(0.4)**	39.3(0.6)	50.7(0.9)	57.6(0.2)	40.3(0.4)	49.1(0.9)	54.8(0.3)
ItalyPowerDemand	95.4(0.2)	96.1(0.3)	**96.3(0.4)**	94.8(0.3)	95.5(0.3)	94.9(0.5)	**96.3(0.1)**
LargeKitchenAppliances	47.3(0.6)	**90.2(0.4)**	90.0(0.5)	54.9(1.2)	88.4(0.5)	89.9(0.4)	89.4(1.1)
Lightning2	67.0(2.1)	73.9(1.4)	77.0(1.7)	75.7(2.4)	77.4(1.2)	**83.9(2.6)**	81.0(2.9)
Lightning7	63.0(1.7)	82.7(2.3)	**84.5(2.0)**	71.5(1.8)	84.1(2.1)	83.3(1.6)	76.7(1.7)
Mallat	91.8(0.6)	96.7(0.9)	97.2(0.3)	91.5(1.0)	96.2(0.2)	**97.4(0.2)**	97.2(0.1)
Meat	89.7(1.7)	85.3(6.9)	**96.8(2.5)**	91.0(2.0)	85.7(5.9)	92.7(3.1)	91.7(1.1)
MedicalImages	72.1(0.7)	77.9(0.4)	77.0(0.7)	73.4(0.6)	78.2(0.6)	80.3(0.8)	**80.5(0.8)**
MiddlePhalanxOutlineAgeGroup	53.1(1.8)	55.3(1.8)	56.9(2.1)	**79.3(0.0)**	73.7(1.3)	75.1(0.6)	75.6(0.7)
MiddlePhalanxOutlineCorrect	77.0(1.1)	80.1(1.0)	80.9(1.2)	74.7(1.5)	**81.8(0.8)**	80.6(0.4)	79.2(0.8)
MiddlePhalanxTW	53.4(1.6)	51.2(1.8)	48.4(2.0)	**60.5(1.0)**	58.8(0.6)	58.3(0.5)	58.9(1.5)
MoteStrain	85.8(0.9)	**93.7(0.5)**	92.8(0.5)	85.9(0.3)	90.1(0.2)	90.4(0.8)	87.9(0.9)
NonInvasiveFetalECGThorax1	91.6(0.4)	**95.6(0.3)**	94.5(0.3)	91.0(0.5)	95.1(0.3)	93.5(0.3)	91.8(0.2)
NonInvasiveFetalECGThorax2	91.7(0.3)	**95.3(0.3)**	94.6(0.3)	91.8(0.3)	95.0(0.2)	94.1(0.2)	92.5(0.3)
OliveOil	66.7(3.8)	72.3(16.6)	83.0(8.5)	57.3(12.2)	58.0(9.1)	**87.3(1.3)**	86.7(2.1)
OSULeaf	55.7(1.0)	97.7(0.9)	97.9(0.8)	53.6(1.9)	97.0(0.5)	**98.8(0.5)**	**98.8(0.5)**
PhalangesOutlinesCorrect	73.5(2.1)	82.0(0.5)	**83.9(1.2)**	75.2(0.7)	81.6(0.8)	82.3(0.9)	83.2(0.5)
Phoneme	9.6(0.3)	32.5(0.5)	33.4(0.7)	10.4(0.1)	33.0(0.3)	**34.2(0.4)**	32.2(0.7)
Plane	97.8(0.5)	**100.0(0.0)**	**100.0(0.0)**	96.4(0.4)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**
ProximalPhalanxOutlineAgeGroup	85.6(0.5)	83.1(1.3)	85.3(0.8)	85.5(1.3)	84.9(1.2)	**87.0(0.9)**	86.3(0.5)
ProximalPhalanxOutlineCorrect	73.3(1.8)	90.3(0.7)	**92.1(0.6)**	78.5(1.3)	91.3(0.8)	90.5(1.0)	90.9(1.0)
ProximalPhalanxTW	76.7(0.7)	76.7(0.9)	78.0(1.7)	79.9(0.4)	**81.4(0.8)**	80.8(0.7)	79.7(0.7)
RefrigerationDevices	37.9(2.1)	50.8(1.0)	52.5(2.5)	37.5(1.3)	53.9(1.7)	**56.1(1.2)**	55.0(0.8)
ScreenType	40.3(1.0)	62.5(1.6)	62.2(1.4)	37.2(1.1)	**63.6(1.1)**	62.7(1.5)	61.1(1.3)
ShapeletSim	50.3(3.1)	72.4(5.6)	77.9(15.0)	51.1(1.5)	96.9(1.3)	**100.0(0.0)**	**100.0(0.0)**
ShapesAll	77.1(0.5)	89.5(0.4)	**92.1(0.4)**	77.2(0.2)	89.3(0.5)	91.3(0.4)	82.1(0.6)
SmallKitchenAppliances	37.1(1.9)	78.3(1.3)	78.6(0.8)	40.4(1.4)	78.9(0.7)	79.4(0.5)	**80.7(0.8)**
SonyAIBORobotSurface1	67.2(1.3)	96.0(0.7)	95.8(1.3)	68.8(0.9)	**96.6(1.0)**	93.8(1.4)	**96.6(1.0)**
SonyAIBORobotSurface2	83.4(0.7)	**97.9(0.5)**	97.8(0.5)	87.5(1.0)	95.3(0.3)	95.8(0.5)	96.1(0.2)
StarLightCurves	94.9(0.2)	96.1(0.9)	97.2(0.3)	95.1(0.3)	97.2(0.1)	97.4(0.1)	**97.7(0.1)**
Strawberry	96.1(0.5)	97.2(0.3)	**98.1(0.4)**	97.4(0.3)	96.6(0.4)	95.7(0.4)	96.2(0.1)
SwedishLeaf	85.1(0.5)	**96.9(0.5)**	95.6(0.4)	87.2(0.8)	96.4(0.3)	96.1(0.2)	96.6(0.3)
Symbols	83.2(1.0)	95.5(1.0)	90.6(2.3)	83.3(0.3)	91.3(1.3)	93.3(0.8)	**95.6(1.0)**
SyntheticControl	97.6(0.4)	98.5(0.3)	**99.8(0.2)**	98.1(0.5)	99.1(0.2)	99.7(0.1)	99.7(0.2)
ToeSegmentation1	58.3(0.9)	96.1(0.5)	96.3(0.6)	58.1(0.8)	96.6(0.3)	96.0(0.6)	**97.5(0.5)**
ToeSegmentation2	74.5(1.9)	88.0(3.3)	90.6(1.7)	73.5(1.6)	90.8(1.1)	**93.5(0.4)**	92.9(1.0)
Trace	80.7(0.7)	**100.0(0.0)**	**100.0(0.0)**	91.6(0.8)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**
TwoLeadECG	76.2(1.3)	**100.0(0.0)**	**100.0(0.0)**	71.7(0.4)	99.9(0.1)	99.8(0.1)	**100.0(0.0)**
TwoPatterns	94.6(0.3)	87.1(0.3)	**100.0(0.0)**	99.1(0.1)	89.7(0.2)	**100.0(0.0)**	**100.0(0.0)**
UWaveGestureLibraryAll	**95.5(0.2)**	81.7(0.3)	86.0(0.4)	94.7(0.2)	81.7(0.3)	87.5(0.3)	88.2(0.4)
UWaveGestureLibraryX	76.7(0.3)	75.4(0.4)	78.0(0.4)	78.1(0.4)	76.3(0.4)	78.5(0.5)	**79.8(0.3)**
UWaveGestureLibraryY	69.8(0.2)	63.9(0.6)	67.0(0.7)	70.6(0.3)	63.8(0.4)	67.8(0.3)	**71.1(0.3)**
UWaveGestureLibraryZ	69.7(0.2)	72.6(0.5)	75.0(0.4)	70.3(0.5)	71.8(0.4)	74.3(0.1)	**75.4(0.6)**
Wafer	99.6(0.0)	99.7(0.0)	**99.9(0.1)**	99.4(0.0)	99.7(0.0)	99.6(0.1)	**99.9(0.0)**
Wine	56.5(7.1)	58.7(8.3)	74.4(8.5)	51.1(2.2)	73.0(6.9)	73.7(4.7)	**88.9(2.6)**
WordSynonyms	59.8(0.8)	56.4(1.2)	62.2(1.5)	**62.6(0.8)**	54.4(0.7)	60.6(0.7)	62.1(0.8)
Worms	45.7(2.4)	76.5(2.2)	**79.1(2.5)**	36.9(1.5)	62.9(1.2)	63.2(1.0)	63.4(1.5)
WormsTwoClass	60.1(1.5)	72.6(2.7)	74.7(3.3)	60.6(2.2)	**74.9(2.6)**	71.5(1.8)	65.0(1.1)
Yoga	85.5(0.4)	83.9(0.7)	87.0(0.9)	84.8(0.3)	84.9(0.4)	85.5(0.9)	**89.2(0.4)**
#Best	4	14	21	5	13	17	**37**
Avg_rank	5.853	4.000	3.053	5.494	3.771	3.176	**2.653**

Table 11: Classification accuracy (%) and standard deviation between parentheses of three deep learning-based methods, three methods using single-label learning and our joint-label learning method (ResNet and FCN) on 85 UCR datasets. The best accuracy is indicated as bold.

Dataset	MLP	FCN	ResNet	MLP_SL	FCN_SL	ResNet_SL	FCN_JL	ResNet_JL
Adiac	39.7(1.9)	**84.4(0.7)**	82.9(0.6)	35.0(0.9)	83.1(0.9)	79.5(0.8)	72.5(0.1)	69.0(0.3)
ArrowHead	77.8(1.2)	84.3(1.5)	84.5(1.2)	77.3(0.8)	82.6(0.6)	80.5(0.4)	**85.7(1.2)**	85.0(0.9)
Beef	72.0(2.8)	69.7(4.0)	75.3(4.2)	64.7(2.7)	71.3(5.4)	74.0(2.5)	77.3(1.3)	**83.3(0.0)**
BeetleFly	87.0(2.6)	86.0(9.7)	85.0(2.4)	78.0(2.4)	93.0(2.4)	92.0(2.4)	88.0(4.0)	**94.0(3.7)**
BirdChicken	77.5(3.5)	**95.5(3.7)**	88.5(5.3)	77.0(2.4)	90.0(0.0)	90.0(0.0)	92.0(4.0)	90.0(0.0)
Car	76.7(2.6)	90.5(1.4)	92.5(1.4)	85.7(0.8)	87.0(3.7)	87.7(1.7)	90.7(0.8)	**94.3(0.8)**
CBF	87.2(0.7)	99.4(0.1)	99.5(0.3)	89.6(0.5)	**99.9(0.0)**	**99.9(0.1)**	98.6(0.2)	99.0(0.2)
ChlorineConcentration	80.2(1.1)	81.4(0.9)	**84.4(1.0)**	76.4(1.1)	81.2(0.7)	82.7(1.0)	82.1(0.2)	83.7(0.6)
CinCECGTorso	84.0(1.0)	82.4(1.2)	82.6(2.4)	77.9(1.3)	82.0(0.5)	81.6(1.2)	80.3(0.8)	**84.1(1.8)**
Coffee	99.6(1.1)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	96.4(0.0)	**100.0(0.0)**	**100.0(0.0)**
Computers	56.3(1.6)	82.2(1.0)	81.5(1.2)	54.7(1.1)	**83.7(0.7)**	82.9(1.2)	**83.7(0.4)**	79.9(0.7)
CricketX	59.1(1.1)	79.2(0.7)	79.1(0.6)	60.3(0.6)	77.4(0.7)	**82.4(0.8)**	63.5(0.8)	66.4(0.5)
CricketY	60.0(0.8)	78.7(1.2)	80.3(0.8)	58.4(1.1)	75.3(1.3)	**82.5(1.4)**	59.7(1.5)	64.2(1.3)
CricketZ	61.7(0.8)	81.1(1.0)	**81.2(1.4)**	62.3(0.7)	75.5(0.8)	78.0(1.0)	65.5(1.1)	68.4(1.1)
DiatomSizeReduction	91.0(1.4)	31.3(3.6)	30.1(0.2)	94.4(0.5)	96.3(0.7)	**99.0(0.4)**	92.9(0.3)	96.5(0.7)
DistalPhalanxOutlineAgeGroup	65.7(1.1)	71.0(1.3)	71.7(1.3)	81.2(1.0)	80.6(0.8)	78.4(0.9)	**81.7(0.7)**	79.8(0.5)
DistalPhalanxOutlineCorrect	72.6(1.3)	76.0(1.5)	77.1(1.0)	78.2(0.5)	80.3(0.2)	80.0(0.7)	**82.3(0.6)**	81.2(1.4)
DistalPhalanxTW	61.7(1.3)	69.0(2.1)	66.5(1.6)	75.3(0.8)	76.4(0.9)	77.0(0.2)	76.4(0.6)	**77.4(0.7)**
Earthquakes	71.7(1.3)	72.7(1.7)	71.2(2.0)	75.2(0.7)	76.5(0.8)	74.5(1.2)	**78.4(0.9)**	76.7(2.0)
ECG200	**91.6(0.7)**	88.9(1.0)	87.4(1.9)	86.6(1.0)	85.8(0.7)	87.0(2.1)	86.6(1.0)	87.8(1.3)
ECG5000	92.9(0.1)	94.0(0.1)	93.4(0.2)	93.1(0.1)	93.8(0.1)	93.4(0.1)	**94.2(0.1)**	93.8(0.2)
ECGFiveDays	97.0(0.5)	98.7(0.3)	97.5(1.9)	95.2(0.4)	90.7(2.3)	98.2(1.3)	95.1(2.1)	**99.5(0.2)**
ElectricDevices	59.2(1.1)	70.2(1.2)	72.9(0.9)	61.5(0.9)	72.5(0.9)	72.5(0.5)	71.2(1.0)	**73.2(0.4)**
FaceAll	79.3(1.1)	94.5(0.9)	83.9(2.0)	84.6(2.3)	**96.4(0.5)**	84.2(0.3)	83.2(0.9)	71.6(0.2)
FaceFour	84.0(1.4)	92.8(0.9)	95.5(0.0)	84.1(0.7)	91.4(0.6)	91.4(2.4)	93.6(0.6)	**95.7(0.5)**
FacesUCR	83.3(0.3)	94.6(0.2)	**95.5(0.4)**	80.0(0.8)	93.6(0.1)	**95.5(0.1)**	81.3(0.5)	83.1(0.4)
FiftyWords	68.4(7.1)	62.7(6.1)	**74.0(1.5)**	71.9(0.2)	64.4(0.7)	70.4(0.7)	47.1(0.6)	49.8(1.3)
Fish	84.8(0.8)	95.8(0.6)	97.9(0.8)	86.2(0.9)	96.6(0.5)	98.3(0.4)	96.6(0.4)	**98.6(0.3)**
FordA	73.0(0.4)	90.4(0.2)	**92.0(0.4)**	67.1(0.4)	89.5(0.2)	91.0(0.2)	89.6(0.2)	**92.0(0.3)**
FordB	60.3(0.3)	87.8(0.6)	**91.3(0.3)**	56.0(0.5)	87.8(0.3)	88.4(0.6)	88.4(0.1)	89.9(0.2)
GunPoint	92.7(1.1)	**100.0(0.0)**	99.1(0.7)	95.2(0.3)	**100.0(0.0)**	98.8(0.3)	99.6(0.3)	**100.0(0.0)**
Ham	69.1(1.4)	71.8(1.4)	75.7(2.7)	72.2(1.8)	71.6(1.5)	75.4(3.5)	72.2(0.4)	**77.9(1.8)**
HandOutlines	**91.8(0.5)**	80.6(7.9)	91.1(1.4)	81.0(1.0)	76.4(4.1)	85.6(0.8)	81.7(1.5)	86.3(0.9)
Haptics	43.3(1.4)	48.0(2.4)	51.9(1.2)	42.2(1.2)	46.5(1.1)	54.0(0.7)	48.9(1.4)	**56.4(0.9)**
Herring	52.8(3.9)	60.8(7.7)	61.9(3.8)	50.0(2.8)	**64.7(1.9)**	57.8(3.0)	61.3(2.3)	58.1(2.3)
InlineSkate	33.7(1.0)	33.9(0.8)	37.3(0.9)	35.1(1.0)	36.9(0.8)	36.8(1.6)	37.2(1.1)	**39.5(1.1)**
InsectWingbeatSound	**60.7(0.4)**	39.3(0.6)	50.7(0.9)	57.6(0.2)	40.3(0.4)	49.1(0.9)	40.9(0.6)	54.8(0.3)
ItalyPowerDemand	95.4(0.2)	96.1(0.3)	**96.3(0.4)**	94.8(0.3)	95.5(0.3)	94.9(0.5)	95.9(0.1)	**96.3(0.1)**
LargeKitchenAppliances	47.3(0.6)	**90.2(0.4)**	90.0(0.5)	54.9(1.2)	88.4(0.5)	89.9(0.4)	86.8(1.2)	89.4(1.1)
Lightning2	67.0(2.1)	73.9(1.4)	77.0(1.7)	75.7(2.4)	77.4(1.2)	**83.9(2.6)**	77.0(1.5)	81.0(2.9)
Lightning7	63.0(1.7)	82.7(2.3)	**84.5(2.0)**	71.5(1.8)	84.1(2.1)	83.3(1.6)	77.5(2.4)	76.7(1.7)
Mallat	91.8(0.6)	96.7(0.9)	97.2(0.3)	91.5(1.0)	96.2(0.2)	**97.4(0.2)**	96.2(0.2)	97.2(0.1)
Meat	89.7(1.7)	85.3(6.9)	**96.8(2.5)**	91.0(2.0)	85.7(5.9)	92.7(3.1)	90.3(2.2)	91.7(1.1)
MedicalImages	72.1(0.7)	77.9(0.4)	77.0(0.7)	73.4(0.6)	78.2(0.6)	80.3(0.8)	78.7(0.5)	**80.5(0.8)**
MiddlePhalanxOutlineAgeGroup	53.1(1.8)	55.3(1.8)	56.9(2.1)	**79.3(0.0)**	73.7(1.3)	75.1(0.6)	74.4(0.8)	75.6(0.7)
MiddlePhalanxOutlineCorrect	77.0(1.1)	80.1(1.0)	80.9(1.2)	74.7(1.5)	**81.8(0.8)**	80.6(0.4)	81.1(0.4)	79.2(0.8)
MiddlePhalanxTW	53.4(1.6)	51.2(1.8)	48.4(2.0)	**60.5(1.0)**	58.8(0.6)	58.3(0.5)	57.9(0.8)	58.9(1.5)
MoteStrain	85.8(0.9)	**93.7(0.5)**	92.8(0.5)	85.9(0.3)	90.1(0.2)	90.4(0.8)	87.0(0.5)	87.9(0.9)
NonInvasiveFetalECGThorax1	91.6(0.4)	**95.6(0.3)**	94.5(0.3)	91.0(0.5)	95.1(0.3)	93.5(0.3)	93.2(0.3)	91.8(0.2)
NonInvasiveFetalECGThorax2	91.7(0.3)	**95.3(0.3)**	94.6(0.3)	91.8(0.3)	95.0(0.2)	94.1(0.2)	93.0(0.1)	92.5(0.3)
OliveOil	66.7(3.8)	72.3(16.6)	83.0(8.5)	57.3(12.2)	58.0(9.1)	**87.3(1.3)**	83.3(5.2)	86.7(2.1)
OSULeaf	55.7(1.0)	97.7(0.9)	97.9(0.8)	53.6(1.9)	97.0(0.5)	**98.8(0.5)**	97.1(0.4)	**98.8(0.5)**
PhalangesOutlinesCorrect	73.5(2.1)	82.0(0.5)	**83.9(1.2)**	75.2(0.7)	81.6(0.8)	82.3(0.9)	82.9(0.5)	83.2(0.5)
Phoneme	9.6(0.3)	32.5(0.5)	33.4(0.7)	10.4(0.1)	33.0(0.3)	**34.2(0.4)**	31.8(0.7)	32.2(0.7)
Plane	97.8(0.5)	**100.0(0.0)**	**100.0(0.0)**	96.4(0.4)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**
ProximalPhalanxOutlineAgeGroup	85.6(0.5)	83.1(1.3)	85.3(0.8)	85.5(1.3)	84.9(1.2)	**87.0(0.9)**	84.1(0.5)	86.3(0.5)
ProximalPhalanxOutlineCorrect	73.3(1.8)	90.3(0.7)	**92.1(0.6)**	78.5(1.3)	91.3(0.8)	90.5(1.0)	91.3(0.3)	90.9(1.0)
ProximalPhalanxTW	76.7(0.7)	76.7(0.9)	78.0(1.7)	79.9(0.4)	**81.4(0.8)**	80.8(0.7)	**81.4(0.1)**	79.7(0.7)
RefrigerationDevices	37.9(2.1)	50.8(1.0)	52.5(2.5)	37.5(1.3)	53.9(1.7)	**56.1(1.2)**	52.7(1.1)	55.0(0.8)
ScreenType	40.3(1.0)	62.5(1.6)	62.2(1.4)	37.2(1.1)	63.6(1.1)	62.7(1.5)	**65.8(1.6)**	61.1(1.3)
ShapeletSim	50.3(3.1)	72.4(5.6)	77.9(15.0)	51.1(1.5)	96.9(1.3)	**100.0(0.0)**	96.3(1.9)	**100.0(0.0)**
ShapesAll	77.1(0.5)	89.5(0.4)	**92.1(0.4)**	77.2(0.2)	89.3(0.5)	91.3(0.4)	80.0(0.2)	82.1(0.6)
SmallKitchenAppliances	37.1(1.9)	78.3(1.3)	78.6(0.8)	40.4(1.4)	78.9(0.7)	79.4(0.5)	79.4(0.7)	**80.7(0.8)**
SonyAIBORobotSurface1	67.2(1.3)	96.0(0.7)	95.8(1.3)	68.8(0.9)	**96.6(1.0)**	93.8(1.4)	95.0(0.6)	**96.6(1.0)**
SonyAIBORobotSurface2	83.4(0.7)	**97.9(0.5)**	97.8(0.5)	87.5(1.0)	95.3(0.3)	95.8(0.5)	96.9(0.4)	96.1(0.2)
StarLightCurves	94.9(0.2)	96.1(0.9)	97.2(0.3)	95.1(0.3)	97.2(0.1)	97.4(0.1)	97.5(0.1)	**97.7(0.1)**
Strawberry	96.1(0.5)	97.2(0.3)	**98.1(0.4)**	97.4(0.3)	96.6(0.4)	95.7(0.4)	96.7(0.1)	96.2(0.1)
SwedishLeaf	85.1(0.5)	**96.9(0.5)**	95.6(0.4)	87.2(0.8)	96.4(0.3)	96.1(0.2)	96.7(0.3)	96.6(0.3)
Symbols	83.2(1.0)	95.5(1.0)	90.6(2.3)	83.3(0.3)	91.3(1.3)	93.3(0.8)	92.4(0.7)	**95.6(1.0)**
SyntheticControl	97.6(0.4)	98.5(0.3)	**99.8(0.2)**	98.1(0.5)	99.1(0.2)	99.7(0.1)	98.9(0.3)	99.7(0.2)
ToeSegmentation1	58.3(0.9)	96.1(0.5)	96.3(0.6)	58.1(0.8)	96.6(0.3)	96.0(0.6)	96.9(1.1)	**97.5(0.5)**
ToeSegmentation2	74.5(1.9)	88.0(3.3)	90.6(1.7)	73.5(1.6)	90.8(1.1)	**93.5(0.4)**	91.8(0.4)	92.9(1.0)
Trace	80.7(0.7)	**100.0(0.0)**	**100.0(0.0)**	91.6(0.8)	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**	**100.0(0.0)**
TwoLeadECG	76.2(1.3)	**100.0(0.0)**	**100.0(0.0)**	71.7(0.4)	99.9(0.1)	99.8(0.1)	99.9(0.0)	**100.0(0.0)**
TwoPatterns	94.6(0.3)	87.1(0.3)	**100.0(0.0)**	99.1(0.1)	89.7(0.2)	**100.0(0.0)**	88.9(0.3)	**100.0(0.0)**
UWaveGestureLibraryAll	**95.5(0.2)**	81.7(0.3)	86.0(0.4)	94.7(0.2)	81.7(0.3)	87.5(0.3)	82.8(0.4)	88.2(0.4)
UWaveGestureLibraryX	76.7(0.3)	75.4(0.4)	78.0(0.4)	78.1(0.4)	76.3(0.4)	78.5(0.5)	77.0(0.4)	**79.8(0.3)**
UWaveGestureLibraryY	69.8(0.2)	63.9(0.6)	67.0(0.7)	70.6(0.3)	63.8(0.4)	67.8(0.3)	65.4(0.6)	**71.1(0.3)**
UWaveGestureLibraryZ	69.7(0.2)	72.6(0.5)	75.0(0.4)	70.3(0.5)	71.8(0.4)	74.3(0.1)	73.5(0.2)	**75.4(0.6)**
Wafer	99.6(0.0)	99.7(0.0)	**99.9(0.1)**	99.4(0.0)	99.7(0.0)	99.6(0.1)	99.7(0.0)	**99.9(0.0)**
Wine	56.5(7.1)	58.7(8.3)	74.4(8.5)	51.1(2.2)	73.0(6.9)	73.7(4.7)	63.3(13.0)	**88.9(2.6)**
WordSynonyms	59.8(0.8)	56.4(1.2)	62.2(1.5)	**62.6(0.8)**	54.4(0.7)	60.6(0.7)	57.1(0.9)	62.1(0.8)
Worms	45.7(2.4)	76.5(2.2)	**79.1(2.5)**	36.9(1.5)	62.9(1.2)	63.2(1.0)	66.4(1.3)	63.4(1.5)
WormsTwoClass	60.1(1.5)	72.6(2.7)	74.7(3.3)	60.6(2.2)	**74.9(2.6)**	71.5(1.8)	74.0(2.3)	65.0(1.1)
Yoga	85.5(0.4)	83.9(0.7)	87.0(0.9)	84.8(0.3)	84.9(0.4)	85.5(0.9)	86.9(0.5)	**89.2(0.4)**
#Best	4	13	21	4	12	17	11	**34**
Avg_rank	6.700	4.612	3.394	6.300	4.329	3.565	4.171	**2.929**

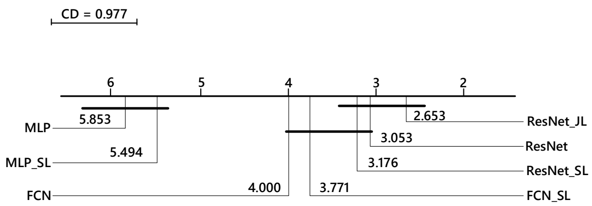


Figure 8: Critical difference diagram of the comparison with deep learning methods. The critical difference is 0.977, which means that two classifiers are not significantly different at $p < 0.05$ level when the rank difference is less than 0.977.

on the augmented datasets using TSW, $N = 2, 4, 8$). From Table 9, two conclusions can be drawn. First, our method achieves the best results on 38 of the 44 datasets and the best average rank of 1.432. Second, TSW can improve the performance of 1NN-DTW. In contrast, the other transformations degrade the performance of 1NN-DTW.

D. Full Results on 85 UCR Datasets

To maintain the integrity of the experiments, we conduct experiments on 85 UCR time-series classification datasets. The full results on 85 UCR datasets are shown in Table 10. ResNet_JL ranks first among all the methods by achieving the best results on 37 datasets.

To explore whether the ResNet_JL is significantly better than the other methods, we conduct the Nemenyi non-parametric statistical test (Demšar 2006). As shown in Figure 8, the critical difference is 0.977, which means that two classifiers are not significantly different at $p < 0.05$ level when the rank difference is less than 0.977. ResNet_JL is significantly superior to MLP-based and FCN-based methods, and slightly superior to ResNet-based methods.

E. Analysis of Transformation Number M

To explore the effect of the transformation number M on the model performance, we evaluate the performance of ResNet_JL with different transformation number M ($M = 2, 3, 4, 5$, including the original time series) on 44 UCR time series datasets.

As shown in Table 12, we see that ResNet_JL with M=2, 3, 4, and 5 achieved the better results (p-value) of 32(0.030), 32(0.006), 35(0.016), and 31(0.005) than ResNet in 44 smaller UCR datasets, respectively. This shows that no matter which value M takes, ResNet_JL can achieve better performance than ResNet.

F. Analysis of Training Set Size

In subsection Analysis of Training Set Size of the main text, we use extremely few training samples, and conduct experiments on 4 UCR datasets. Some people are concerned that it may lead to a smaller batch size and make the training insufficient for ResNet. Therefore, we increase the batch size of ResNet and conduct experiments on 8 UCR datasets.

Table 12: Classification accuracy (%) of ResNet and ResNet_JL with different M on 44 UCR datasets. The best accuracy is indicated as bold.

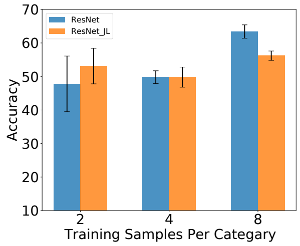
Dataset	ResNet	M=2	M=3	M=4	M=5
ArrowHead	0.845	0.857	**0.867**	0.850	0.845
Beef	0.753	0.760	0.820	0.833	**0.847**
BeetleFly	0.850	0.890	0.920	0.940	**0.990**
BirdChicken	0.885	0.860	**0.900**	**0.900**	**0.900**
Car	0.925	0.927	0.923	**0.943**	0.923
CBF	**0.995**	0.981	0.989	0.990	0.986
CinCECGTorso	0.826	0.835	0.837	**0.841**	0.832
Coffee	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
DiatomSizeR	0.301	**0.975**	0.973	0.965	0.965
DistPhaxAgeGrp	0.717	0.786	0.793	**0.798**	**0.798**
DistPhaxTW	0.665	0.759	0.763	**0.774**	0.760
Earthquakes	0.712	0.754	0.773	0.767	**0.797**
ECG200	0.874	0.894	0.870	0.878	**0.896**
ECGFiveDays	0.975	0.993	0.997	0.995	**0.998**
FaceFour	0.955	0.955	0.955	**0.957**	0.952
FacesUCR	0.955	**0.956**	0.955	0.831	0.947
Fish	0.979	**0.991**	0.987	0.986	0.989
GunPoint	0.991	**1.000**	**1.000**	**1.000**	0.995
Ham	0.757	0.764	**0.794**	0.779	0.777
Haptics	0.519	0.555	0.558	**0.564**	0.553
Herring	0.619	0.619	0.606	0.581	**0.625**
InlineSkate	0.373	**0.398**	0.392	0.395	0.395
ItalyPower	**0.963**	0.958	0.960	**0.963**	0.958
Lightning2	0.770	0.787	0.790	**0.810**	0.807
Lightning7	**0.845**	0.781	0.792	0.767	0.767
Mallat	**0.972**	0.968	0.967	**0.972**	0.963
Meat	**0.968**	0.940	0.903	0.917	0.920
MoteStrain	**0.928**	0.900	0.897	0.879	0.884
MidPhaxAgeGrp	0.569	0.740	0.743	**0.756**	0.749
MidPhaxTW	0.484	0.586	**0.591**	0.589	0.584
OliveOil	0.830	0.860	**0.887**	0.867	0.847
OSULeaf	0.979	0.984	0.983	**0.988**	0.986
Plane	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
ShapeletSim	0.779	**1.000**	0.994	**1.000**	0.996
SonyAIBORobot1	0.958	**0.978**	0.964	0.966	0.971
SonyAIBORobot2	**0.978**	0.970	0.967	0.961	0.964
Symbols	0.906	0.959	**0.960**	0.956	0.959
ToeSegmentation1	0.963	0.933	0.965	**0.975**	0.973
ToeSegmentation2	0.906	**0.934**	0.922	0.929	**0.934**
Trace	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
TwoLeadECG	**1.000**	**1.000**	**1.000**	**1.000**	0.999
Wine	0.744	0.737	0.841	**0.889**	0.848
Worms	**0.791**	0.629	0.645	0.634	0.599
WormsTwoClass	**0.747**	0.692	0.685	0.650	0.693
#Better	-	32	32	35	31
p-value	-	0.030	0.006	0.016	0.005

As shown in Figure 9(a-d), we see that the accuracy of ResNet is indeed improved when the batch size is increased. Despite this, ResNet_JL still performs better overall. In addition, we have added experiments on four additional datasets. As shown in Figure 9(e-h), we see that ResNet_JL also performs better overall.

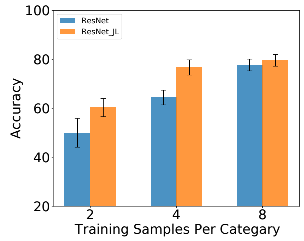
References

Chen, Y.; Keogh, E.; Hu, B.; Begum, N.; Bagnall, A.; Mueen, A.; and Batista, G. 2015. The UCR time series classification archive. www.cs.ucr.edu/~eamonn/time_series_data/.

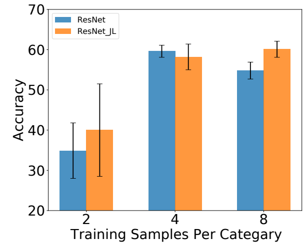
Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7: 1–30.



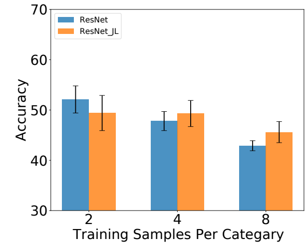
(a) Computers



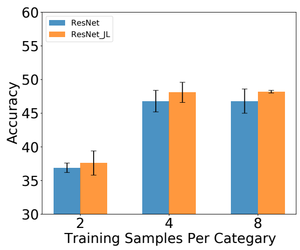
(b) LrgKitApp



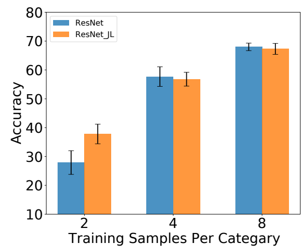
(c) ProxPhxCorr



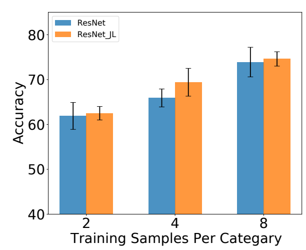
(d) RefrigerationDevices



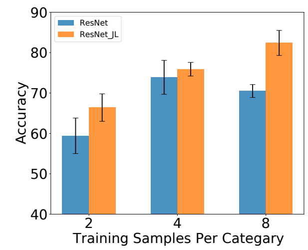
(e) DistalPhalanxOutlineCorrect



(f) DistalPhalanxTW



(g) SmallKitchenAppliances



(h) TwoPatterns

Figure 9: The accuracy vs training set size (only a few training samples per category) on eight UCR datasets.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Wang, Z.; Yan, W.; and Oates, T. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*, 1578–1585. IEEE.

Zeiler, M. D. 2012. ADADELTA: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701* .