

Learning Multi-Tasks with Inconsistent Labels by Using Auxiliary Big Task

Quan FENG, Songcan CHEN

Frontiers of Computer Science, DOI: [10.1007/s11704-022-2251-x](https://doi.org/10.1007/s11704-022-2251-x)

Problems & Ideas

- Multi-task learning in the real world usually has more general problems.:
 - Joint learning among multiple tasks, with fewer training samples for each task and only partially overlapping sets of labels.
 - Joint learning between multiple tasks, with fewer training samples for each task and with label sets that do not partially overlap.
- Ideas: To address this problem, this paper proposes to assist efficient learning of multiple tasks with partially overlapping/non-overlapping label sets by jointly utilizing learned auxiliary large tasks with sufficient information.

Table 5: Testing accuracy of each comparing method on Exp 1, where the optimal performances are bolded. The ranking and average ranking are reported in the corresponding bracket and the last column, respectively.

Methods	Group 1		Group 2		Group 3		Avg Rank
	Office-Caltech	Caltech-256	Amazon	Webcam	Dlsr	Product	
Single-Task	0.76 ± 0.037(6)	0.45 ± 0.037(9)	0.74 ± 0.036(5)	0.63 ± 0.038(9)	0.77 ± 0.031(5)	0.63 ± 0.035(9)	7.16
Multi-task	0.76 ± 0.050(7)	0.53 ± 0.055(5)	0.80 ± 0.046(3)	0.69 ± 0.049(3)	0.80 ± 0.046(3)	0.68 ± 0.049(5)	4.33
Cross-Stich	0.75 ± 0.050(8)	0.53 ± 0.059(6)	0.80 ± 0.046(3)	0.67 ± 0.054(8)	0.79 ± 0.042(4)	0.65 ± 0.056(8)	6.16
NDDR-CNN	0.75 ± 0.054(9)	0.51 ± 0.055(8)	0.79 ± 0.042(4)	0.67 ± 0.044(7)	0.75 ± 0.049(6)	0.67 ± 0.050(6)	6.66
MTAL	0.78 ± 0.054(5)	0.53 ± 0.051(4)	<u>0.81 ± 0.042(2)</u>	<u>0.69 ± 0.048(2)</u>	<u>0.80 ± 0.038(2)</u>	0.66 ± 0.055(7)	3.66
LSSA	<u>0.80 ± 0.058(2)</u>	0.54 ± 0.373(3)	0.74 ± 0.180(6)	0.69 ± 0.155(5)	0.61 ± 0.428(9)	<u>0.73 ± 0.337(2)</u>	4.50
MCN	0.79 ± 0.341(4)	<u>0.59 ± 0.277(2)</u>	0.73 ± 0.450(8)	0.68 ± 0.138(6)	0.63 ± 0.265(7)	0.70 ± 0.272(4)	5.16
MLwSGSU	0.79 ± 0.216(3)	0.53 ± 0.421(7)	0.74 ± 0.285(7)	0.69 ± 0.103(4)	0.63 ± 0.346(8)	0.71 ± 0.050(3)	5.33
ADMTL	0.91 ± 0.019(1)	0.82 ± 0.197(1)	0.90 ± 0.022(1)	0.81 ± 0.133(1)	0.88 ± 0.346(1)	0.92 ± 0.097(1)	1.00

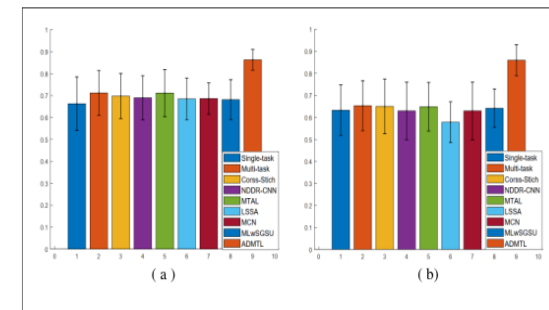


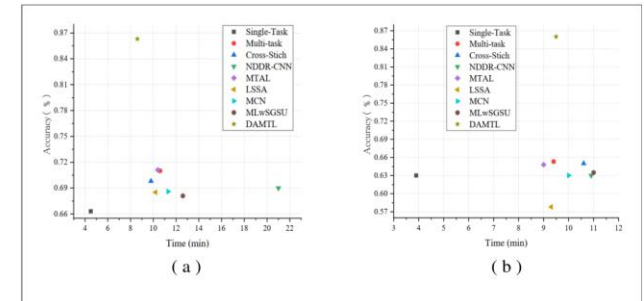
Fig (a) and (b) show the performance comparison of the mean and mean squared error of various methods on Exp 1 and Exp 2, respectively.

Main Contributions

- Contributions:
 - A new framework is proposed to address multi-task learning scenarios with partially overlapping and non-overlapping label sets.
 - A novel knowledge selection strategy is designed for improving the generalization performance of each task.
 - Extensive experiments on multiple datasets with different settings demonstrate the significant competitiveness of our model in comparison with the-state-of-the-art methods.

Table 6: Testing accuracy of each comparing method on Exp 2, where the optimal performances are bolded. The ranking and average ranking are reported in the corresponding bracket and the last column, respectively.

Methods	Group 1		Group 2		Group 3		Avg Rank
	Art	Real World	Office-Caltech	Webcam	Tiny-ImagNet	Amazon	
Single-Task	0.60 ± 0.039(7)	0.51 ± 0.043(7)	0.76 ± 0.037(4)	0.63 ± 0.038(6)	0.51 ± 0.048(7)	0.77 ± 0.031(6)	6.16
Multi-task	0.62 ± 0.055(5)	0.53 ± 0.054(4)	0.78 ± 0.050(3)	0.68 ± 0.050(3)	0.53 ± 0.058(5)	0.78 ± 0.048(5)	4.16
Cross-Stich	0.63 ± 0.057(4)	0.51 ± 0.056(8)	<u>0.78 ± 0.048(2)</u>	0.67 ± 0.050(4)	0.52 ± 0.062(6)	0.79 ± 0.046(4)	4.66
NDDR-CNN	0.59 ± 0.056(8)	0.52 ± 0.056(5)	0.76 ± 0.051(5)	0.66 ± 0.050(5)	0.46 ± 0.461(9)	0.79 ± 0.044(3)	5.83
MTAL	0.61 ± 0.056(6)	0.52 ± 0.058(6)	0.73 ± 0.046(6)	<u>0.69 ± 0.044(2)</u>	0.54 ± 0.067(4)	<u>0.80 ± 0.042(2)</u>	4.33
LSSA	0.69 ± 0.348(3)	0.57 ± 0.712(3)	0.46 ± 0.103(8)	0.50 ± 0.836(8)	0.57 ± 0.564(3)	0.68 ± 0.125(8)	5.50
MCN	0.59 ± 0.056(8)	0.52 ± 0.056(5)	0.76 ± 0.051(5)	0.66 ± 0.050(5)	0.46 ± 0.451(8)	0.79 ± 0.044(3)	5.66
MLwSGSU	<u>0.70 ± 0.149(2)</u>	<u>0.58 ± 0.521(2)</u>	0.58 ± 0.028(7)	0.63 ± 0.493(7)	<u>0.57 ± 0.169(2)</u>	0.75 ± 0.651(7)	4.50
ADMTL	0.88 ± 0.026(1)	0.74 ± 0.414(1)	0.93 ± 0.043(1)	0.83 ± 0.389(1)	0.91 ± 0.029(1)	0.90 ± 0.065(1)	1.00



Compare the time cost (in minutes) of MTL on Exp 1 (a) and Exp 2 (b).

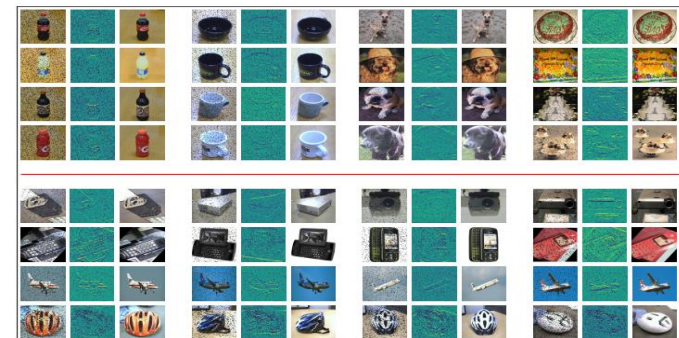


Fig. 6: Visual illustration of knowledge extraction. The top four rows of the red line are visualized in Exp 1 by the mask extraction of knowledge, while the bottom four rows are visualized in Exp 2 by the mask extraction of knowledge. Where in each 4 × 3 square rows 1 and 2 are masked auxiliary tasks and row 3 are the learning tasks.