Learning Real-World Label Noise from Multiple Labellings: A Multi-Task Learning Approach

Presented by
Jiaming (Jazlyn) Lin
Bachelor of Advanced Studies (Honours)

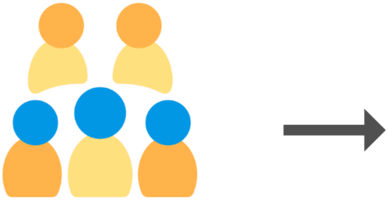
Supervised by
Dr. Tongliang Liu

School of Computer Science, The University of Sydney

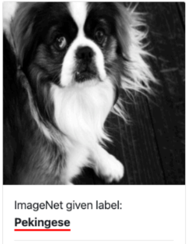
July 3, 2022

THE UNIVERSITY OF SYDNEY

Deep Learning with Label Noise [1]

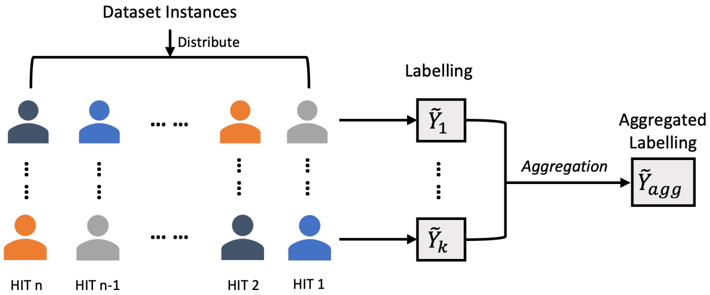


Crowdsourcing

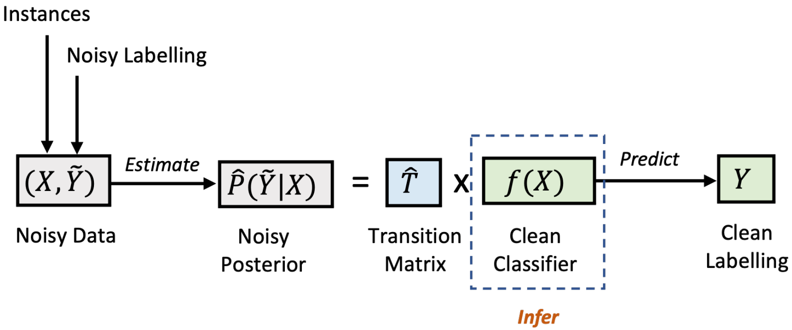


Pekingese VS Japanese Chin

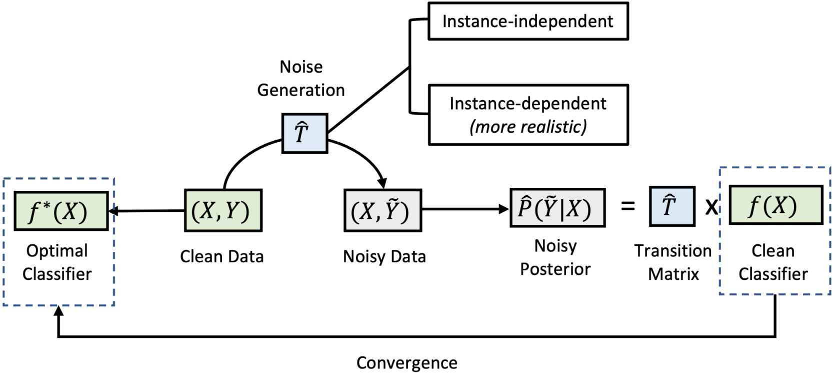
Crowdsourcing Field [2]



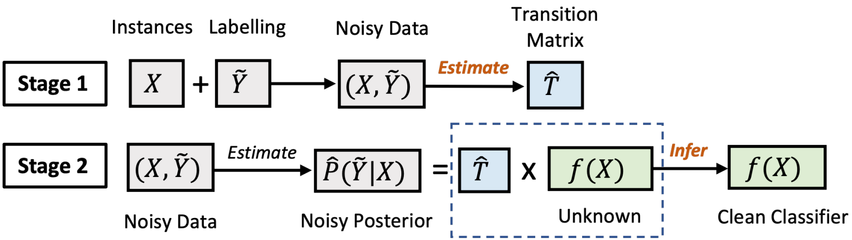
Label-Noise Learning Field [1]



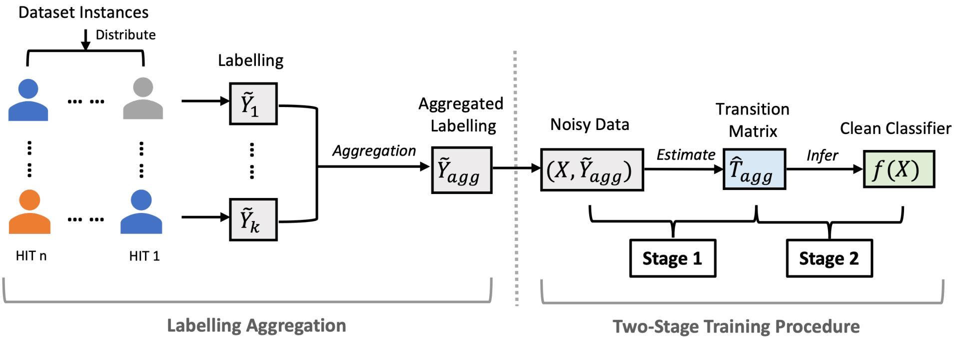
Transition Matrix and its Estimation [3]



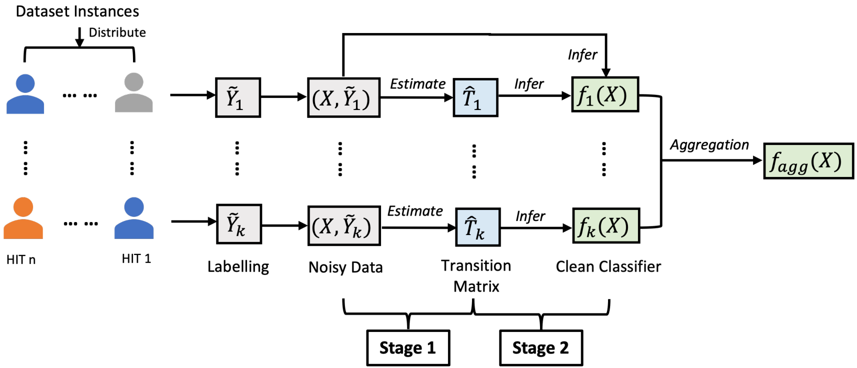
Two-Stage Training Procedure [4]



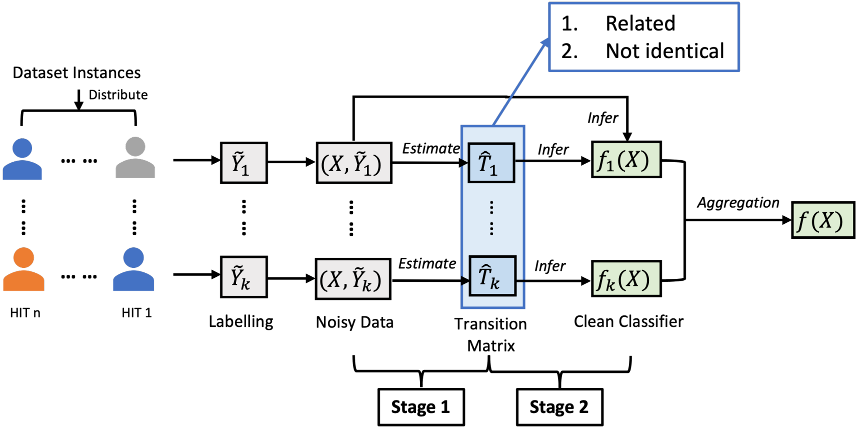
Recent Work [5]: AGG-T



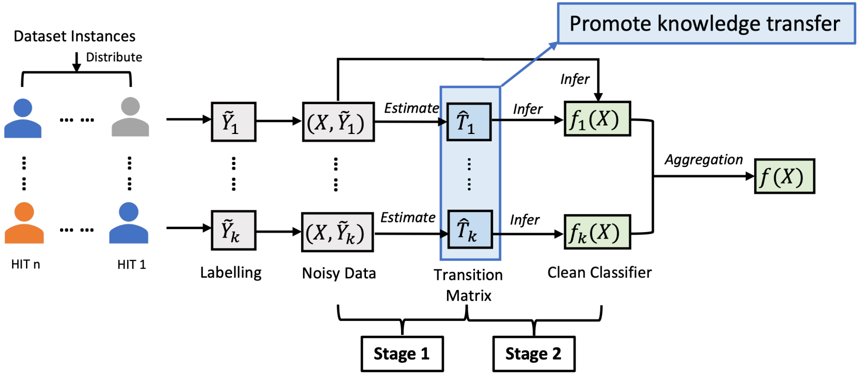
Naive Approach: STL-T [6]



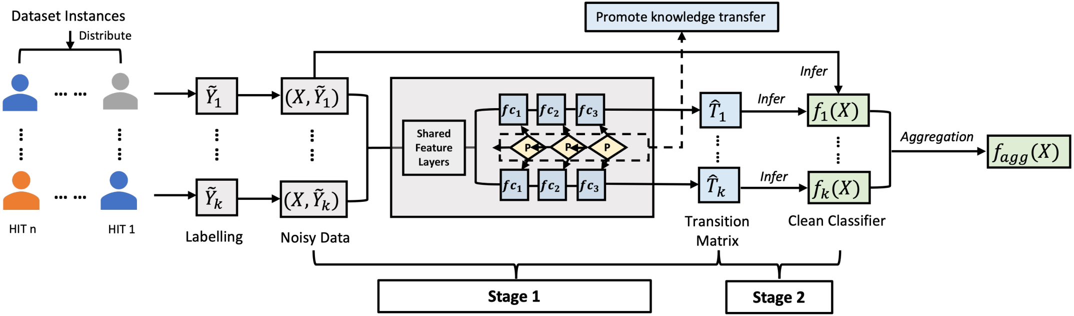
Conditions for MTL [6]



Proposed MTL-T

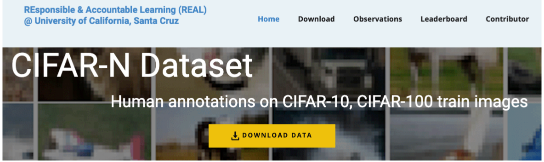


Proposed MTL-T [7][8]



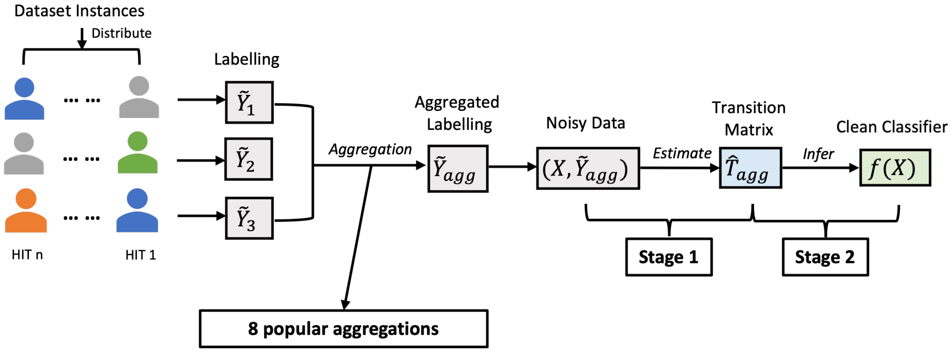
- Adapt Bayes-Label-T [7] to MTL
- Adapt MRN model [8] to Label-Noise Learning

Dataset

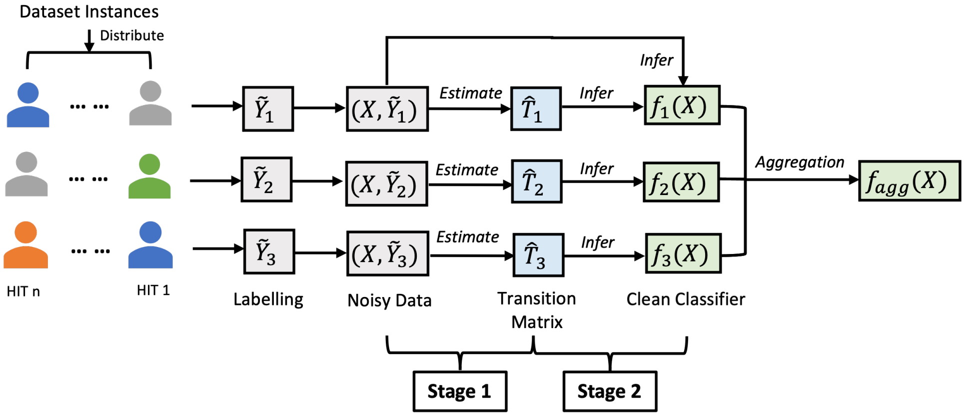


- CIFAR-10N [5], a re-annotation of CIFAR-10 with 3 labellings
- The only relevant dataset, which is
  1. Image-related
  2. Multi-class
  3. Having multiple labellings
  4. Large enough for learning accurate transition matrices

Baselines: AGG-T [2]



Baseline: STL-T



Experimental Research Questions

- ***STL-T vs MTL-T***

  (Q1) Does knowledge transfer occur between transition matrices?

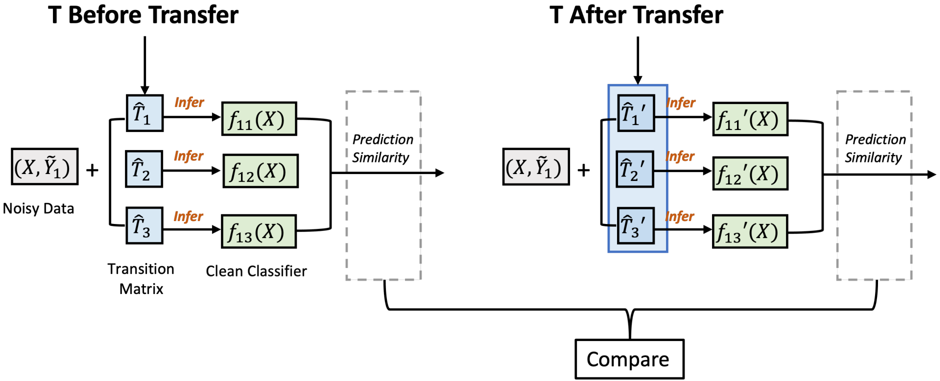
  (Q2) Does knowledge transfer improve the transition matrix estimation?

  (Q3) Does knowledge transfer ultimately improve the final classifier accuracy?

- ***AGG-T vs MTL-T***

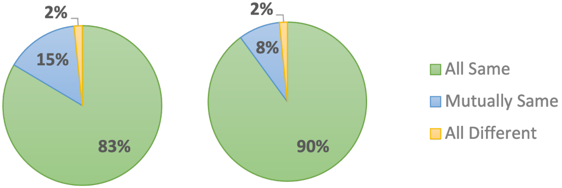
  (Q4) Does MTL-T outperform any baselines in AGG-T?

(Q1) Does Knowledge Transfer Occur?



(Q1) Does Knowledge Transfer Occur?

Fixed $(X, \tilde{Y}_1)$ ⇨



Before Transfer After Transfer

(A1) Knowledge transfer occurs

(Q2) Does Knowledge Transfer Improves T Estimation?

	Test Classification Accuracy (%)		
	$f_1(X)$	$f_2(X)$	$f_3(X)$
STL-T (Without Transfer)	89.34±0.13	89.33±0.11	89.41±0.15
MTL-T (With Transfer)	**89.91±0.23**	**89.97±0.17**	**89.94±0.19**

(A2) Knowledge transfer improves the estimation of each transition matrix

(Q3) Does Knowledge Transfer Improves Final Classifier Accuracy?

	Test Classification Accuracy (%)
	$f_{agg}(X)$
STL-T (Without Transfer)	90.74±0.18
MTL-T (With Transfer)	**91.28±0.21**

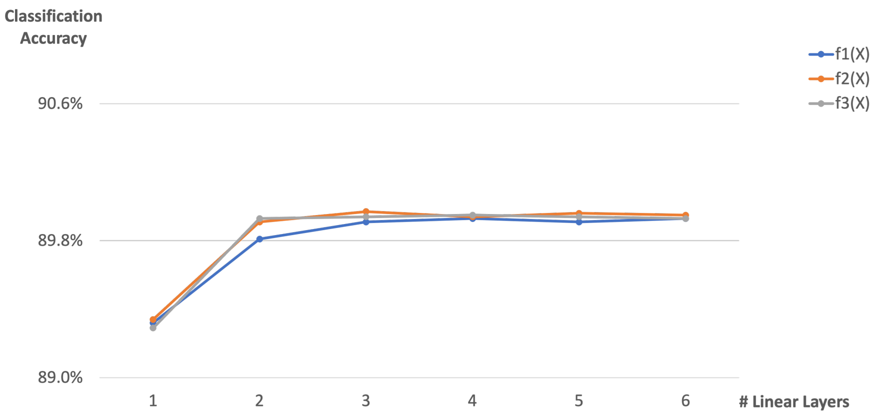
(A3) Knowledge transfer improves the final classifier accuracy (0.5% uplift)

(Q4) Does MTL-T outperform any baselines in AGG-T?

Approach	Aggregation Algorithms	Test Classification Accuracy (%)
AGG-T	MV	89.49±0.12
	ZC (Demartini et al., 2012)	89.91±0.16
	GLAD (Whitehill et al., 2009)	89.89±0.13
	CATD (Li et al., 2014a)	89.92±0.11
	PM (Li et al., 2014b)	89.73±0.13
	D&S (Dawid and Skene, 1979)	90.14±0.09
	BCC (Kim and Ghahramani, 2012)	90.21±0.07
	LFC (Raykar et al., 2010)	90.25±0.10
MTL-T		**91.28±0.21**

(A4) MTL-T outperforms all AGG-T baselines (1.03% to 1.79% uplift)

Ablation Study



Limitations and Future Work

Limitations

1. Only one dataset
2. Assumed distribution of parameter tensor
3. Only popular label aggregation techniques

Future Work

1. Reproduce the experiment when more datasets are available
2. Investigate more alternatives with no assumptions
3. Compare against more aggregation techniques

Conclusion

- Proposed MTL-T which adapts MTL to simultaneously learn from multiple labellings
- Conducted thorough evaluations using CIFAR-10N dataset, showing
  1. Occurrence of knowledge transfer
  2. Improved transition matrix estimation
  3. Improved clean classifier accuracy
  4. Better than popular label aggregation baselines

References I

[1] Bo Han, Quanming Yao, Tongliang Liu, Gang Niu, Ivor W Tsang, James T Kwok, and Masashi Sugiyama. A survey of label-noise representation learning: Past, present and future. *arXiv preprint arXiv:2011.04406*, 2020.

[2] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. Truth inference in crowdsourcing: Is the problem solved? *Proceedings of the VLDB Endowment*, 10(5):541–552, 2017.

[3] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning? *Advances in Neural Information Processing Systems*, 32:6838–6849, 2019.

[4] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1944–1952, 2017.

[5] Jiaheng Wei, Zhaowei Zhu, Hao Cheng, Tongliang Liu, Gang Niu, and Yang Liu. Learning with noisy labels revisited: A study using real-world human annotations. In *International Conference on Learning Representations*, 2021.

[6] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Multi-task Learning*, page 126–140. Cambridge University Press, 2020.

References II

[7] Shuo Yang, Erkun Yang, Bo Han, Yang Liu, Min Xu, Gang Niu, and Tongliang Liu. Estimating instance-dependent bayes-label transition matrix using a deep neural network. In *ICML*, 2022.

[8] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Philip S Yu. Learning multiple tasks with multilinear relationship networks. *Advances in neural information processing systems*, 30, 2017.

Appendix: Key Formulations

$$
\begin{aligned}
p(\mathcal{W}|\mathcal{X}^*, \mathcal{Y}^*, \widetilde{\mathcal{Y}}) &\propto p(\mathcal{W}) \cdot p(\widetilde{\mathcal{Y}}|\mathcal{X}^*, \mathcal{Y}^*, \mathcal{W}) \\
&= \underbrace{\prod_{\ell \in \mathcal{L}} p(\mathcal{W}^\ell)}_{\text{Prior}} \cdot \underbrace{\prod_{j=1}^{k} \prod_{i=1}^{N^*} p(\widetilde{y}_i^j | \hat{y}_i^*, x_i^*, \mathcal{W})}_{\text{MLE}} .
\end{aligned}
$$

$$
\min_{\mathcal{W}^\ell, \Sigma^\ell | \ell \in \mathcal{L}} \underbrace{\sum_{j=1}^{k} \sum_{i=1}^{N^*} \widetilde{y}_i^t \log\left(\hat{y}_i^* T_j^*(x_i^*; \mathcal{W})\right)}_{L_{cel} \text{ from MLE}} + \underbrace{\sum_{\ell \in \mathcal{L}} \lambda_\ell \left[vec(\mathcal{W}^\ell)^T (\Sigma^\ell)^{-1} vec(\mathcal{W}^\ell) + D_1^\ell D_2^\ell \log(|\Sigma^\ell|) \right]}_{L_{reg} \text{ from Prior}}, \tag{3.3}
$$

Appendix: Aggregation Algorithm

Algorithm 1: *Solution Framework*

Input: workers' answers V

Output: inferred truth v_i^* $(1 \leq i \leq n)$, worker quality q^w $(w \in \mathcal{W})$

Initialize all workers' qualities (q^w for $w \in \mathcal{W}$);

while true **do**

- *// Step 1: Inferring the Truth*
- **for** $1 \leq i \leq n$ **do**
  - Inferring the truth v_i^* based on V and $\{q^w \mid w \in \mathcal{W}\}$;
- *// Step 2: Estimating Worker Quality*
- **for** $w \in \mathcal{W}$ **do**
  - Estimating the quality q^w based on V and $\{v_i^* \mid 1 \leq i \leq n\}$;
- *// Check for Convergence*
- **if** Converged **then**
  - **break**;

return v_i^* for $1 \leq i \leq n$ and q^w for $w \in \mathcal{W}$;

Questions?