ICML | 2020

Task Understanding From Confusing Multi-task Data

Xin SU
Tsinghua University

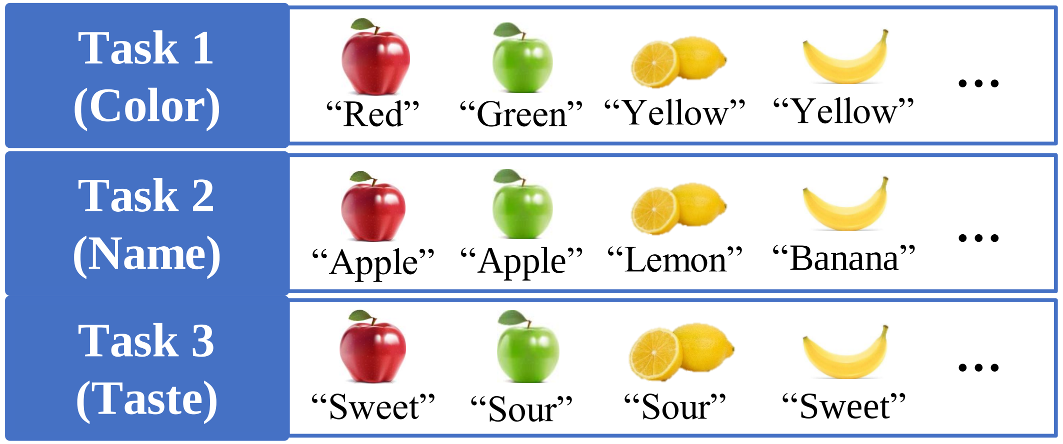
Yizhou JIANG
Tsinghua University

Shangqi GUO
Tsinghua University

Feng CHEN
Tsinghua University

Motivation: From Narrow AI to AGI

- Narrow AI: A specific task in the determined environment.



Multi-Task Learning :
Comprehensive problems in different semantic space

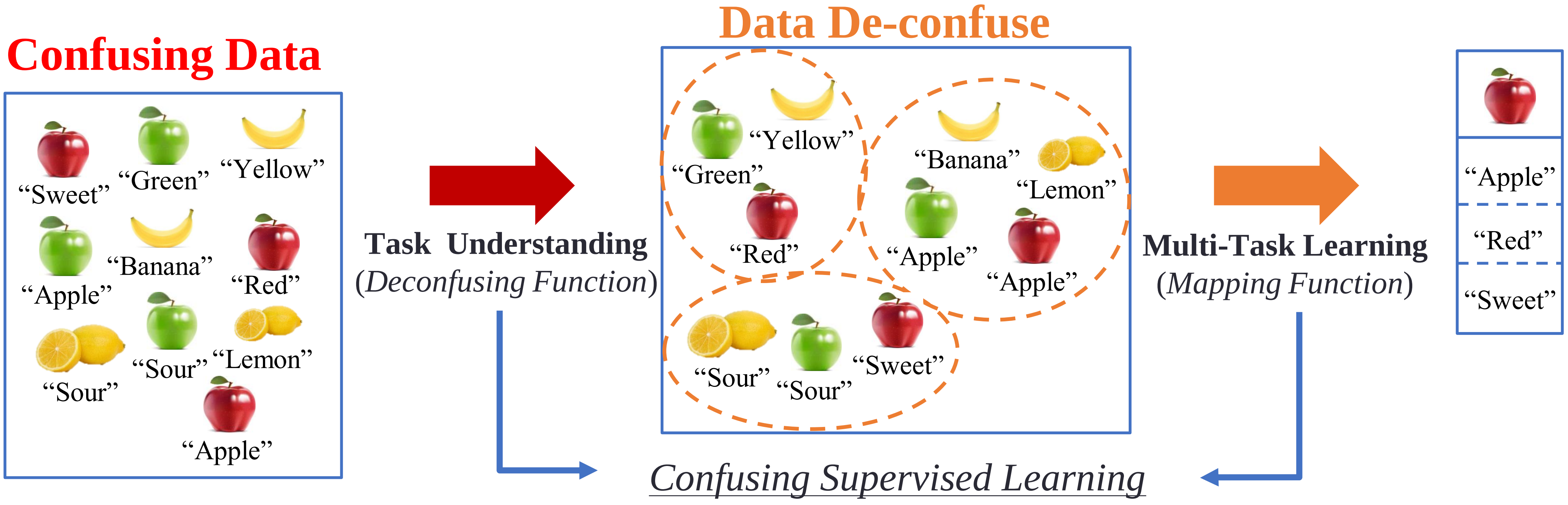
Task Annotation + Label Annotation

Manual Task Definition —— Do not exist in natural raw data

AGI Problem: How can we learn task concept from original raw data?

Confusing Supervised Learning (CSL)

- Without task annotation: Mapping conflicts between multi-task



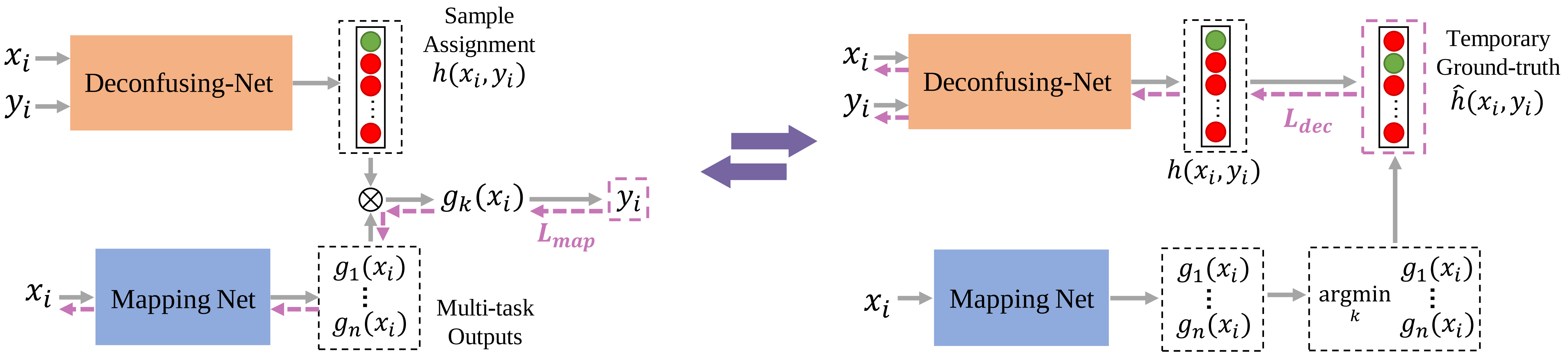
- CSL: Learning task concepts by reducing mapping conflicts

$$R(g,h) = \int_x \sum_{j,k} (f_j(x) - \underbrace{g_k(x)}_{\text{Mapping Function}})^2 \underbrace{h(x, f_j(x), g_k)}_{\text{Deconfusing Function}} p(f_j)p(x)\,\mathrm{d}x$$

Method: CSL-Net

$$\min_{g,h} R_e = \sum_{i=1}^{m} \sum_{k=1}^{n} \left(y_i - g_k\left(x_i\right)\right)^2 \cdot h\left(x_k, y_k; g_k\right)$$

Mapping Net	Deconfusing Net
$\min_{g_k} L_{map}(g_k) = \sum_{i=1}^{m_k} \left\lvert y_i^k - g_k(x_i^k)\right\rvert^2, k = 1, \dots, n$	$\min_{h} L_{dec}(h) = \sum_{i=1}^{m} \left\lvert h(x_i, y_i) - \hat{h}(x_i, y_i)\right\rvert^2$



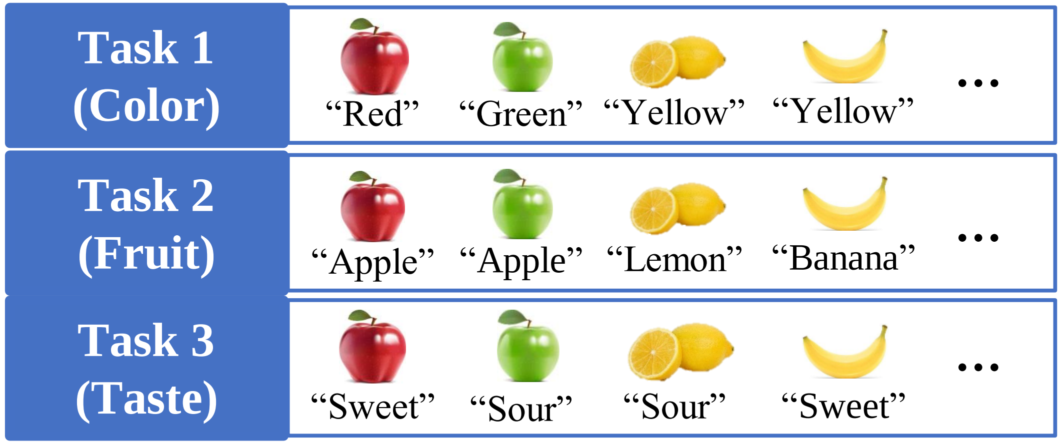
Mapping-Net Training **Deconfusing-Net Training**

Motivation: From Narrow AI to AGI

- AI Success: Exceeded human-level performance on various problems.

- Narrow AI: A specific task in the determined environment.

Motivation: From Narrow AI to AGI



Multi-Task Learning :
Comprehensive problems in different semantic space

Task Annotation + Label Annotation

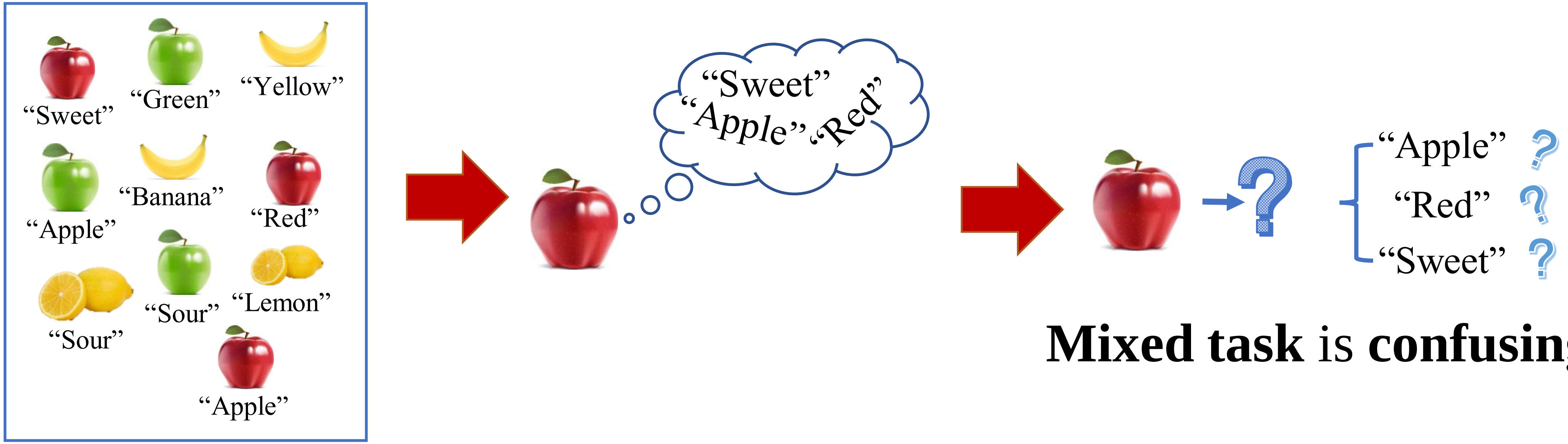
Manual Task Definition —— Do not exist in natural raw data

AGI Problem: How can we learn task concept from original raw data?

Confusing Data

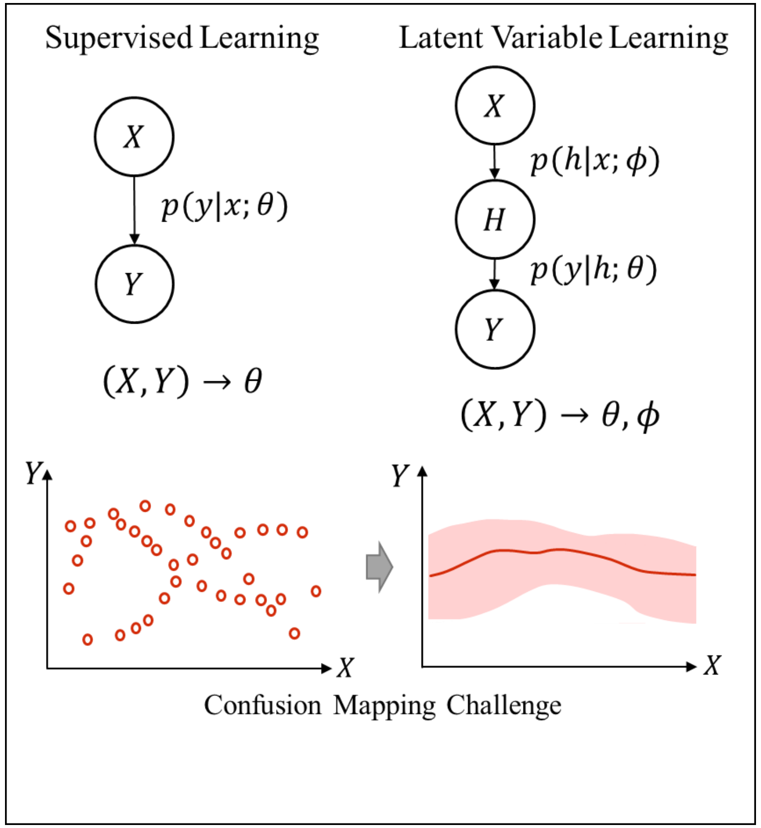
- Multi-tasks cannot be represented by a single mapping function.
- Task understanding is vital for multi-task learning.

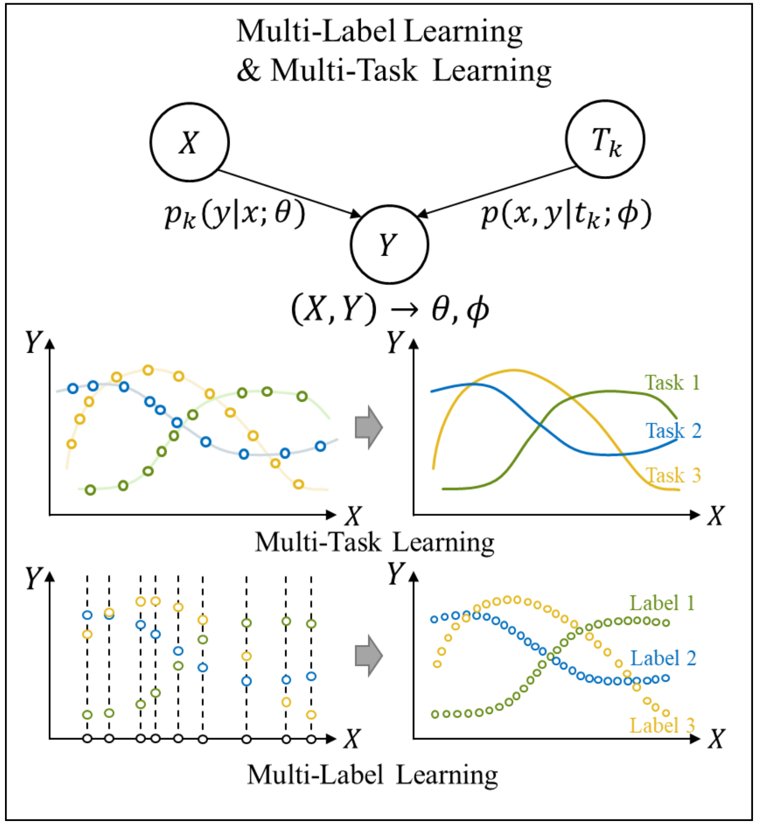
Confusing Data: Multi-task data **without Task Annotation**

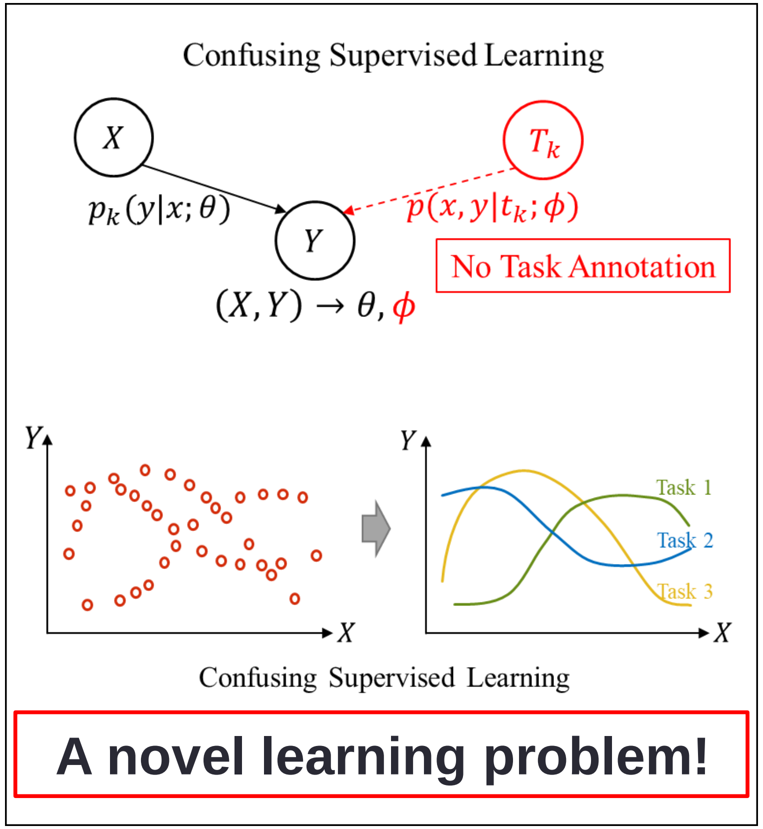


Mixed task is **confusing!**

Comparison of Existing Methods



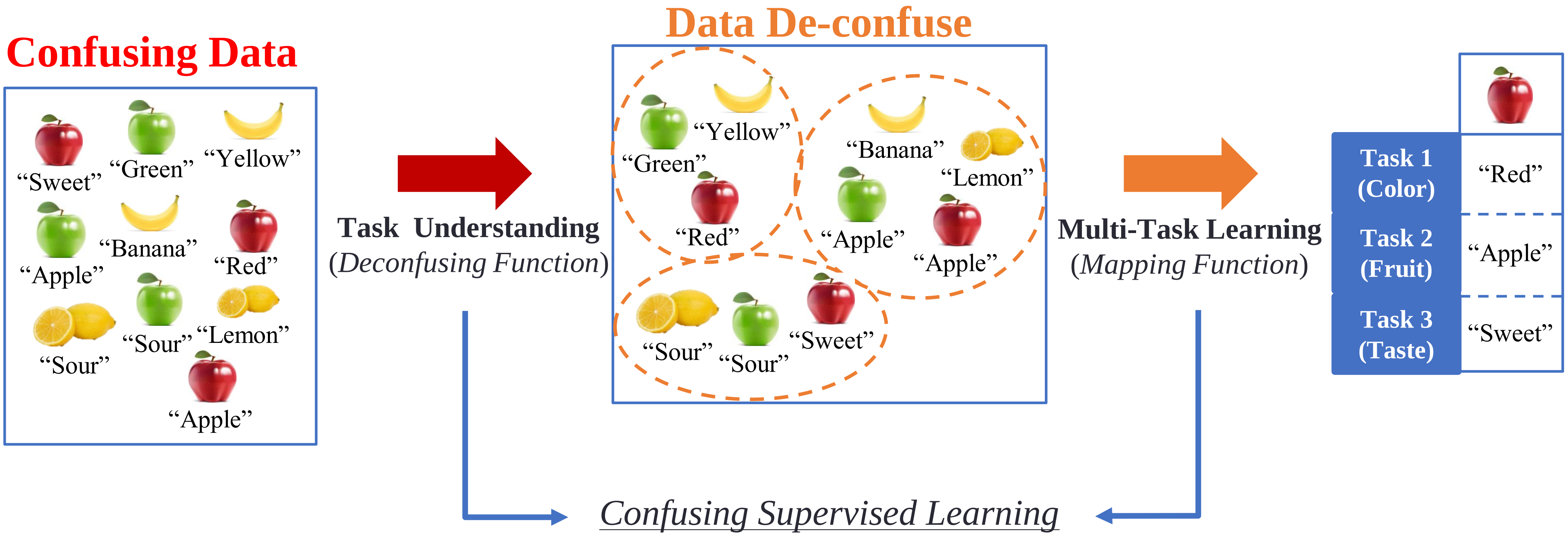




- **Supervised Learning & Latent Variable Learning**: Mapping Confusing.
- **Multi-Task Learning**: Task annotation is needed.
- **Multi-Label Learning**: Multiple labels are allocated.
- Confusing Supervised Learning: No task annotation or samples allocation.

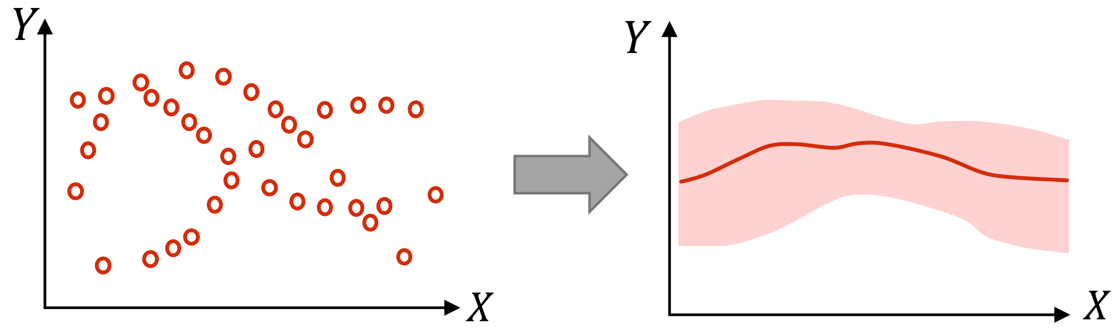
Confusing Supervised Learning (CSL)

- Without task annotation: Mapping conflicts between multi-task

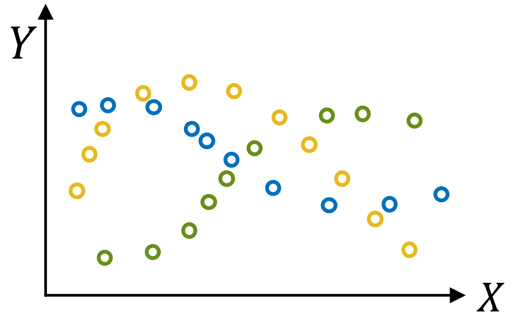


Learning Objective: Risk Functional of CSL

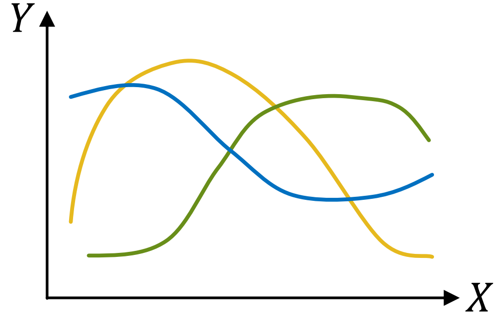
Model	Traditional Supervised Learning	Confusing Supervised Learning
Risk Functional	$R(g) = \int_x \underbrace{\sum_{j=1}^{n} (f_j(x) - g(x))^2 p(f_j)}_{\text{Confusing Multiple Mappings}} \, p(x) \, \mathrm{d}x$	$R(g,h) =$ $\int_x \sum_{j,k} (f_j(x) - \underbrace{g_k(x)}_{\substack{\text{Mapping} \\ \text{Function}}})^2 \underbrace{h(x, f_j(x), g_k)}_{\text{Deconfusing Function}} \, p(f_j) p(x) \, \mathrm{d}x$
Solution	$g^*(x) = \sum_{j=1}^{n} p(f_j) f_j(x) = \bar{f}(x).$ $\min R(g^*) > 0$	$h^*(x, f_j(x), g_k) = I[j = k]$ $g_k^*(x) = f_k(x), \; k = 1, \ldots, n$ $\min R(g^*, h^*) = 0$



Deconfusing $h(x, f, g)$

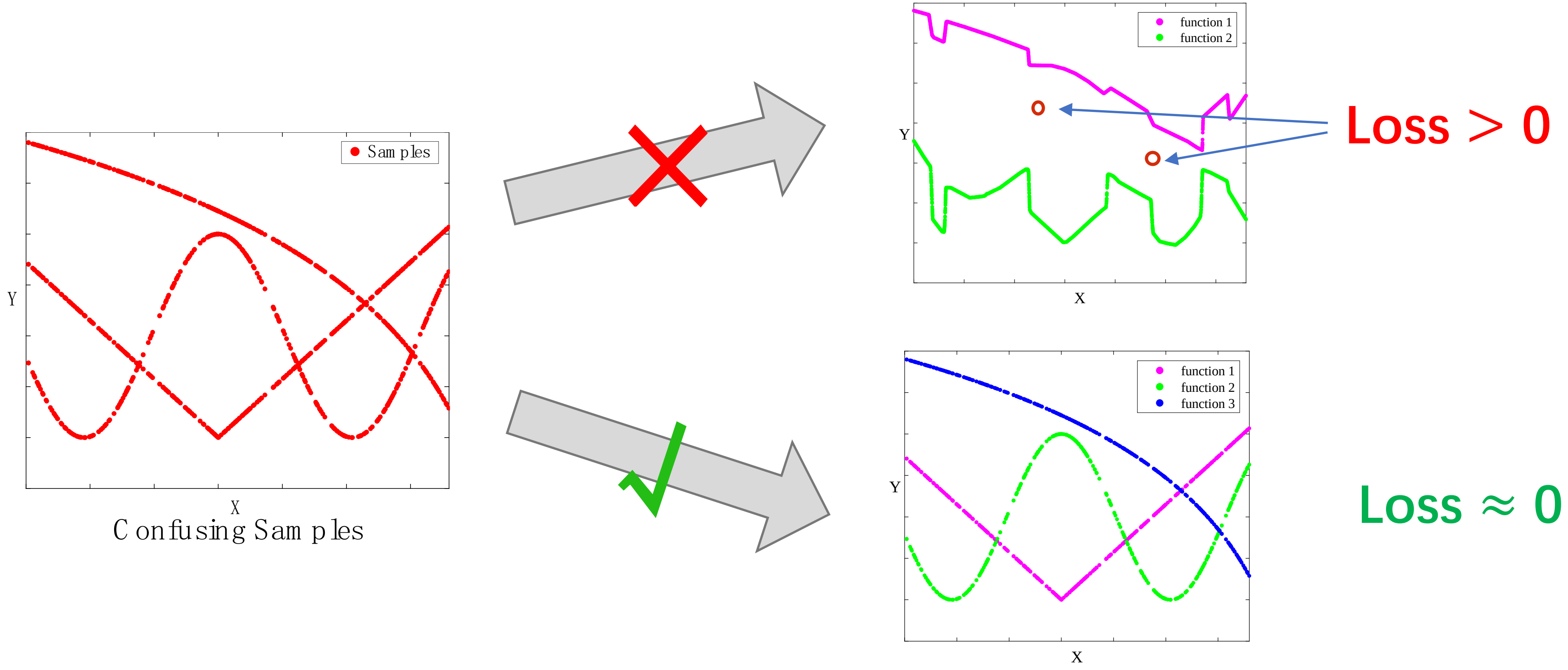


Mapping $g(x)$



Feasibility: Loss → 0

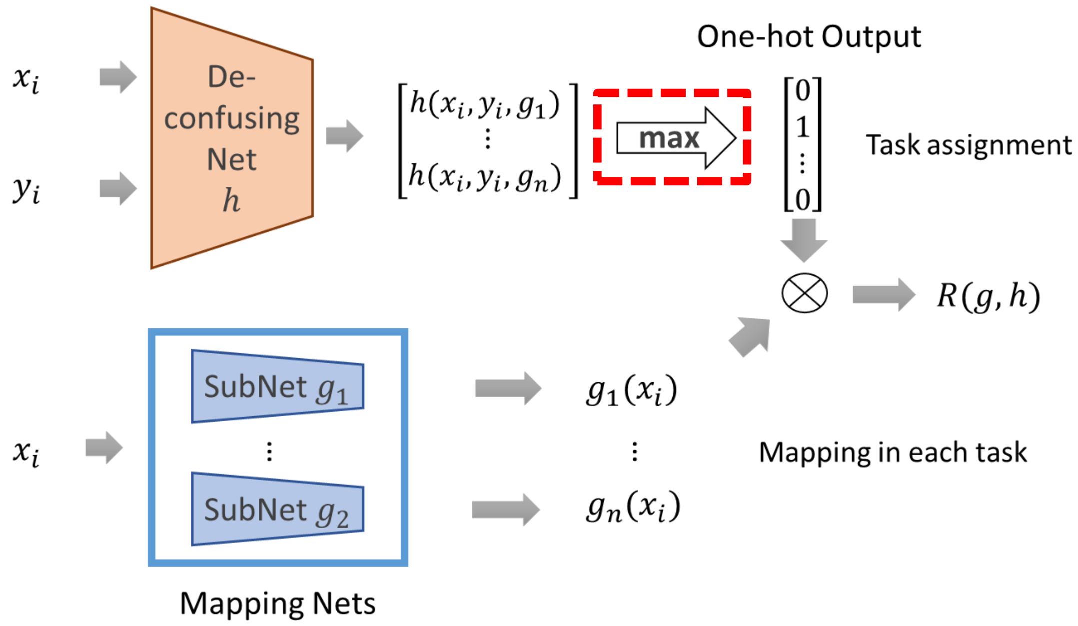
- Wrong allocation of confusing samples leads to unavoidable loss.



- Task concept driven by global loss: Empirical risk should go towards 0!

Training Target & CSL-Net

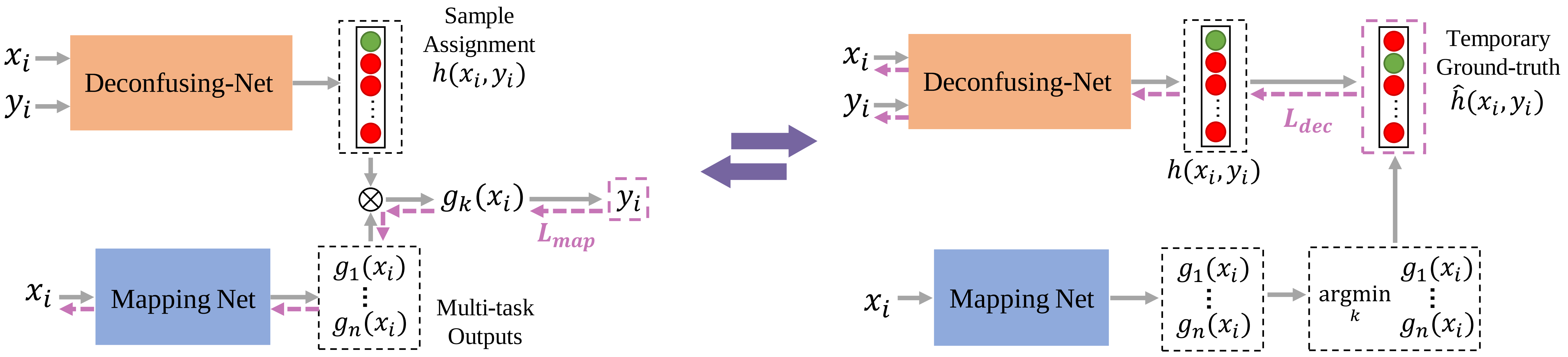
- Optimization Target: $$\min_{g,h} R_e = \sum_{i=1}^{m}\sum_{k=1}^{n}(y_i - g_k(x_i))^2 \cdot h(x_k, y_k; g_k)$$
- Expected Result: $$h^*(x, f_j(x), g_k) = I[j=k] \quad g_k^*(x) = f_k(x),\ k = 1, \ldots, n$$
- Constraint:
  The output of Deconfusing-Net is one-hot!
- Difficulty:
  Approximation of Softmax leads to a trivial solution.
  Joint BP is not available.



Training Algorithm of CSL-Net

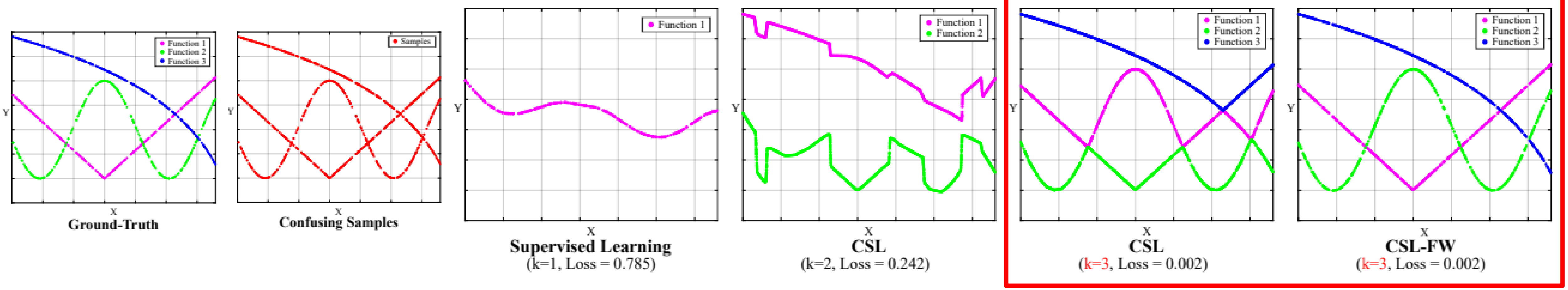
$$\min_{g,h} R_e = \sum_{i=1}^{m}\sum_{k=1}^{n}\left(y_i - g_k\left(x_i\right)\right)^2 \cdot h\left(x_k, y_k; g_k\right)$$

Training of Mapping Net	Training of Deconfusing Net
$\min_{g_k} L_{map}(g_k) = \sum_{i=1}^{m_k} \left\lvert y_i^k - g_k(x_i^k) \right\rvert^2, k = 1, \ldots, n$	$\min_{h} L_{dec}(h) = \sum_{i=1}^{m} \left\lvert h(x_i, y_i) - \hat{h}(x_i, y_i) \right\rvert^2$

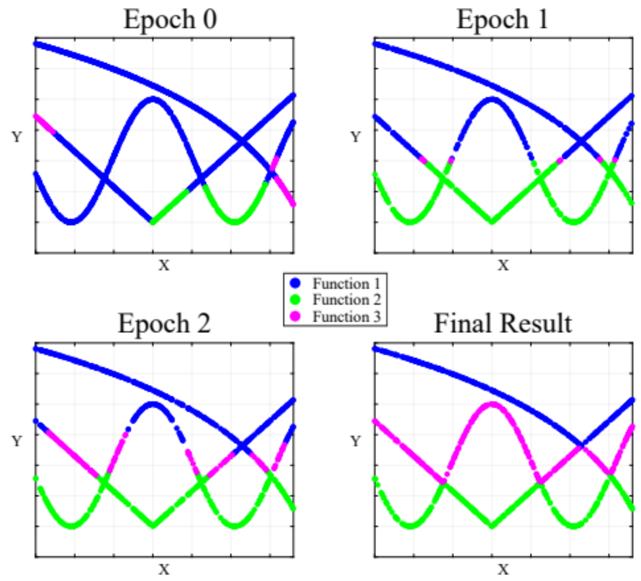


Mapping-Net Training **Deconfusing-Net Training**

Experiment: Function Regression



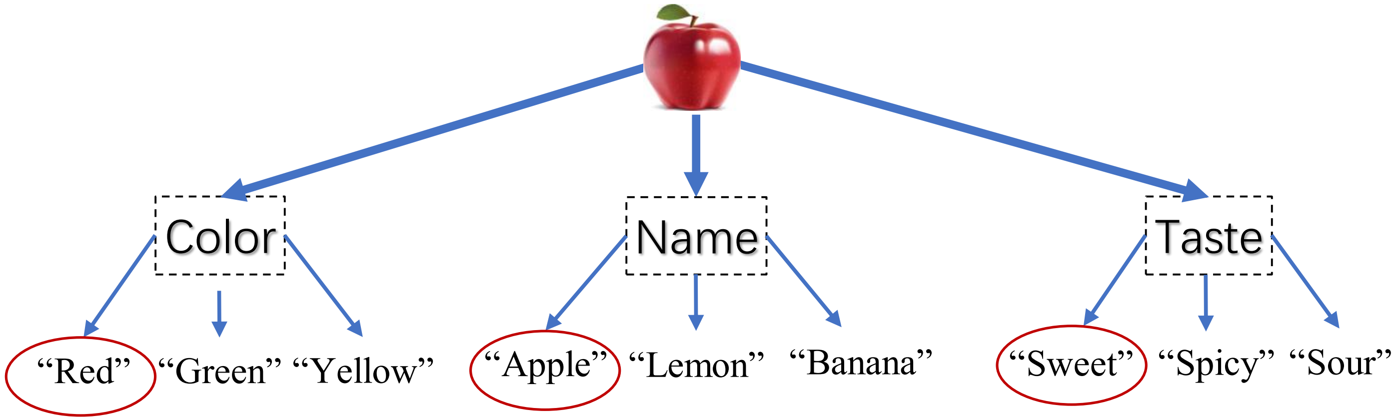
- Supervised learning fails to fit multiple functions.
- Incorrect task number leads to confusing fitting results.
- **CSL-Net learns reasonable task concepts and complete multi-task mapping.**



Results in the training process

Experiment: Pattern Recognition

- Each sample represents the classification result of only one task.



- Two Learning Goal:
  - Task Understanding
  - Classification of Multi-Task
- Two Evaluation Metrics:
  - Task Understanding

$$\alpha_T(j) = \max_k \frac{1}{m} \sum_{i=1}^{m} I[h(x_i, y_i; f_k), \tilde{h}(x_i, y_i; f_j)]$$

  - Classification of Multi-Task

$$\alpha_L(j) = \max_k \frac{1}{m} \sum_{i=1}^{m} 1 - \frac{|g_k(x_i) - f_j(x_i)|}{|f_i(x_i)|}$$

Experiment: Pattern Recognition

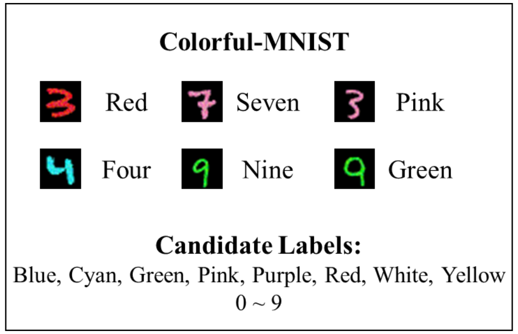
- Results on two confusing supervised datasets.

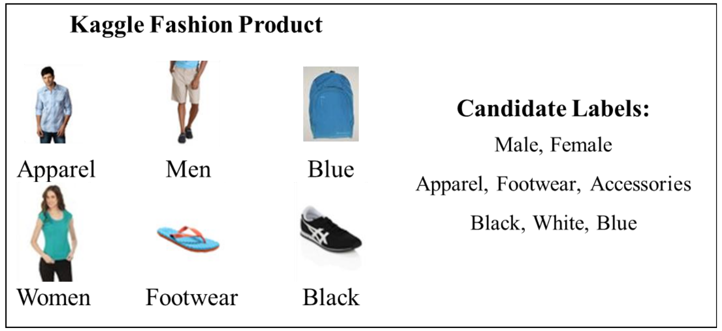
Table 1. Accuracy of Pattern Recognition Experiments.

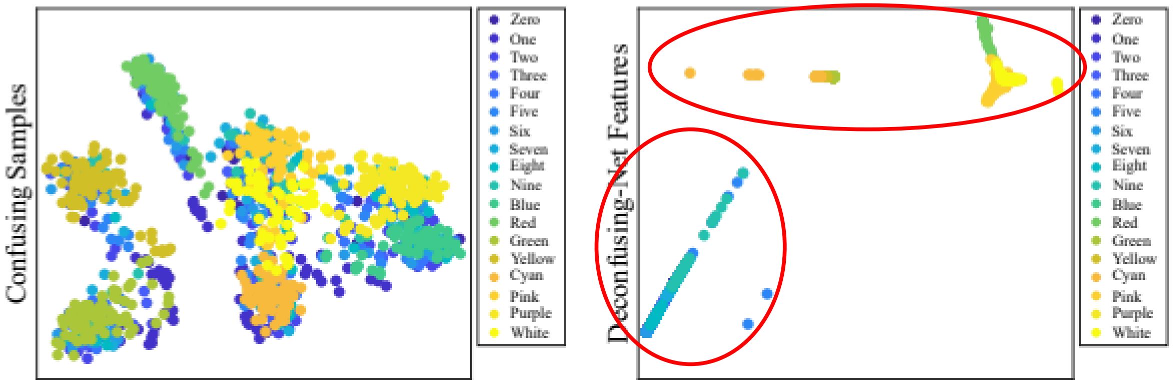
Learning Methods		Colorful-MNIST				Kaggle Fashion Product					
		$\alpha_T(1)$ (Cor)	$\alpha_T(2)$ (Num)	$\alpha_L(1)$ (Cor)	$\alpha_L(2)$ (Num)	$\alpha_T(1)$ (Gen)	$\alpha_T(2)$ (Cat)	$\alpha_T(3)$ (Cor)	$\alpha_L(1)$ (Gen)	$\alpha_L(2)$ (Cat)	$\alpha_L(3)$ (Cor)
Confusing Data	Trad SL	/	/	39.25	52.50	/	/	/	23.59	42.64	29.17
	Pseudo-Label	/	/	36.57	50.01	/	/	/	20.74	33.41	26.30
	SMiLE	/	/	12.94	19.98	/	/	/	16.04	32.74	18.41
	CSL	**98.24**	**99.02**	**99.32**	**97.18**	**98.42**	**99.16**	**98.90**	**93.25**	**97.87**	**90.84**
Task Annotated	Trad MT	99.48	99.61	99.24	98.15	99.01	99.43	99.17	92.91	97.82	91.64
	ML-LOC	99.57	99.58	99.66	98.62	99.12	98.92	99.25	94.54	98.63	94.12

Experiment: Pattern Recognition

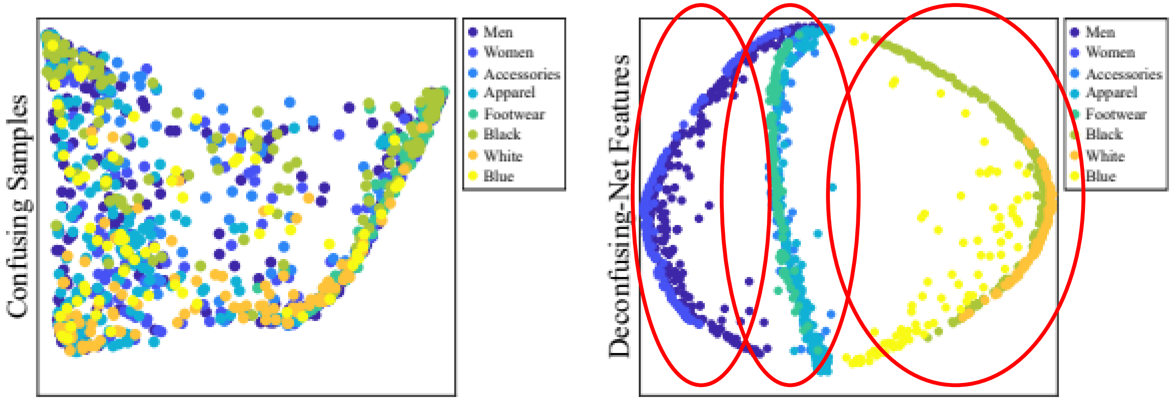
- Feature Visualization of Deconfusing Net.







Before After



Before After

- **Deconfusing Net could separate confusing samples to reasonable task groups.**

Conclusion

- **A novel learning problem** for general raw data:
  - Task annotation is unknown in natural raw data.
  - Understanding task concept from raw data (confusing data).
- **A novel learning paradigm**: Confusing Supervised Learning
  - Deconfusing Function: Samples allocation for tasks
  - Mapping Function: Multi-task mappings.
  - Global Risk Functional: Over all risk of representation for raw data.
- **A novel network**: CSL-Net
  - Algorithm of alternating two-stage training to realize the task constraint.
- **A novel application:** learning system towards general intelligence.
  - The agent autonomously defines task concepts and learns multi-task mapping without manual task annotation.

ICML | 2020

Thanks!

Xin Su, Tsinghua University

suxin16@mails.tsinghua.edu.cn