
Leveraging Multi-task Learning for Unambiguous and Flexible Deep Neural Network Watermarking

Fang-Qi Li₁, Lei Yang₁, Shi-Lin Wang₁,
Alan Wee-Chung Liew₂

₁ School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University,

₂ School of Information and Communication Technology, Griffith University.

February, 2022

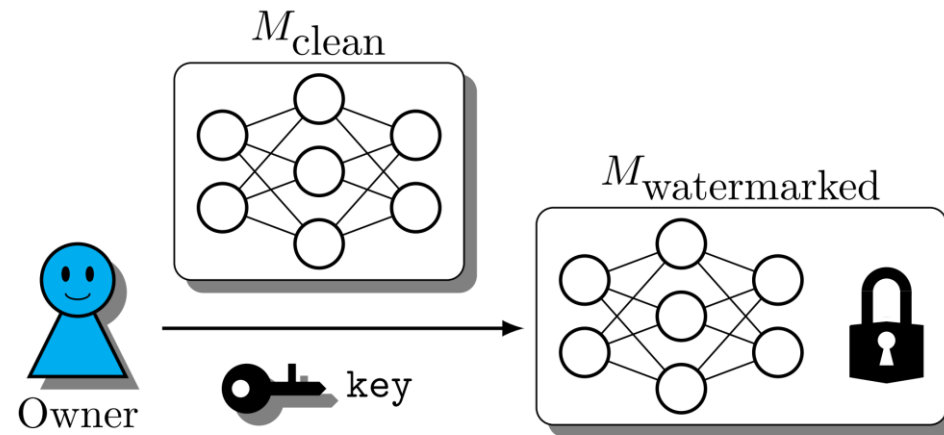
SafeAI2022@AAAI2022

饮水思源 · 爱国荣校

The scenario: ownership verification of DNN models

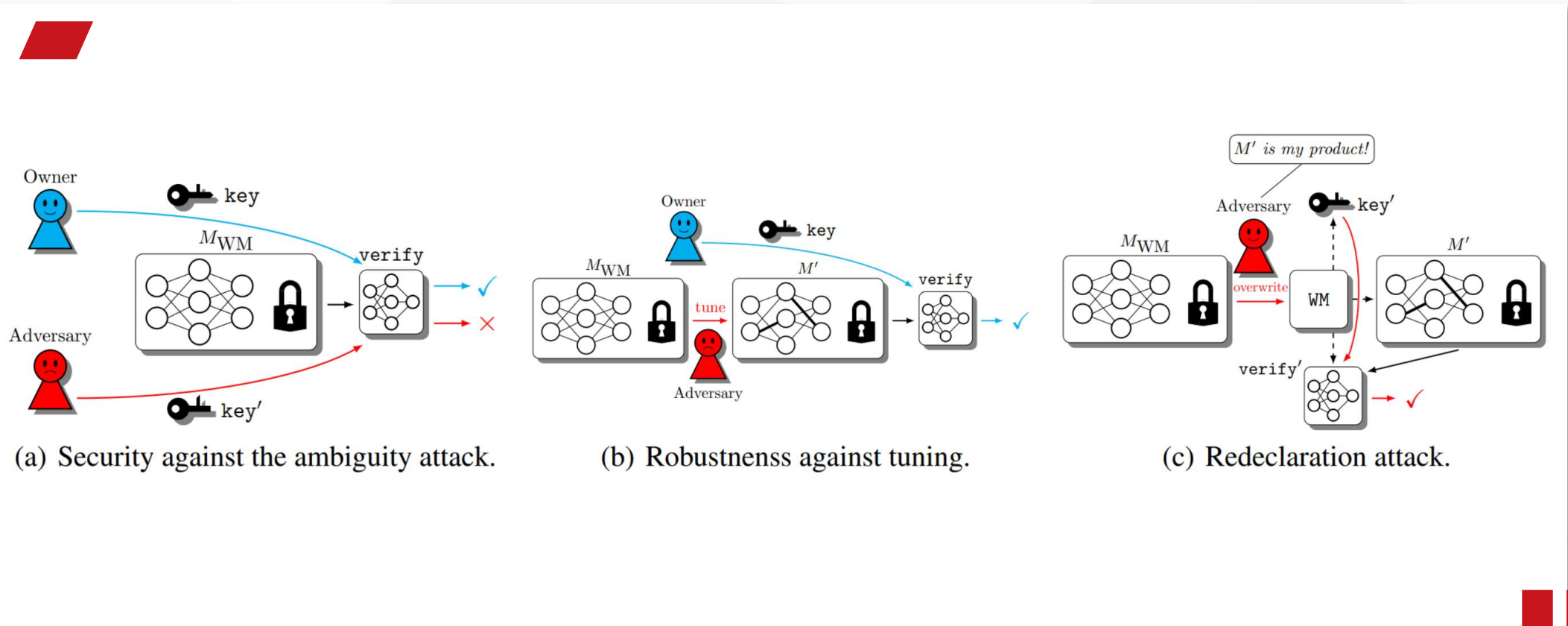
The background:

To protect deep neural networks as intellectual properties, watermarking schemes have been widely adopted.






Targets: unambiguity, robustness, flexibility, etc.





Targets: unambiguity, robustness, flexibility, etc.

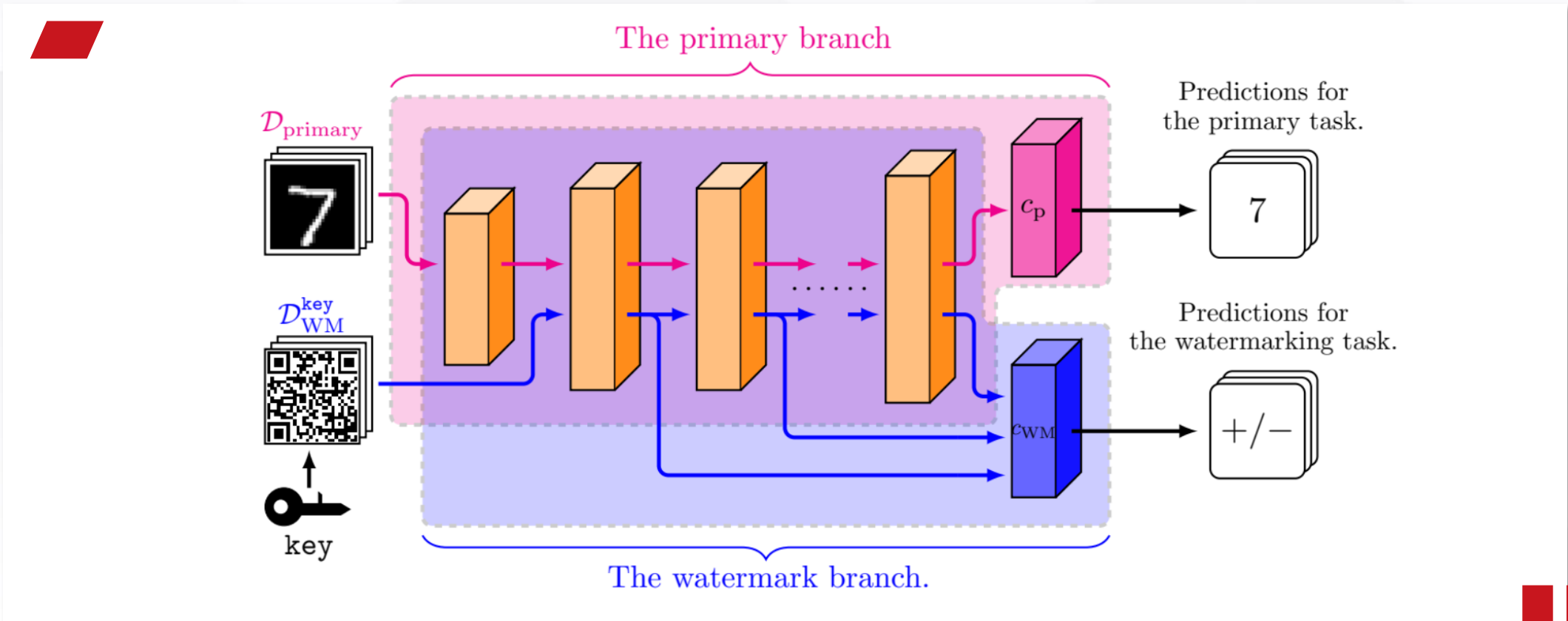


Scheme	Type	Unambiguity	Functionality-preserving	Robustness against tuning	Flexibility
(Uchida et al. 2017)	White-box	×	✓	×	✓
(Darvish, Chen, and Koushanfar 2019)	White-box	✓	✓	✓	×
(Li et al. 2019a)	Black-box	✓	✓	✓	×
(Zhu et al. 2020)	Black-box	✓	✓	✓	×
(Guan et al. 2020)	White-box	✓	✓	×	×
(Le Merrer, Perez, and Trédan 2020)	Black-box	×	✓	✓	✓
(Ong et al. 2021)	Black-box	×	✓	✓	×
(Fan et al. 2021)	Black-box	✓	✓	✓	×
(Liu, Weng, and Zhu 2021)	White-box	×	✓	✓	×
Ours.	White-box	✓	✓	✓	✓



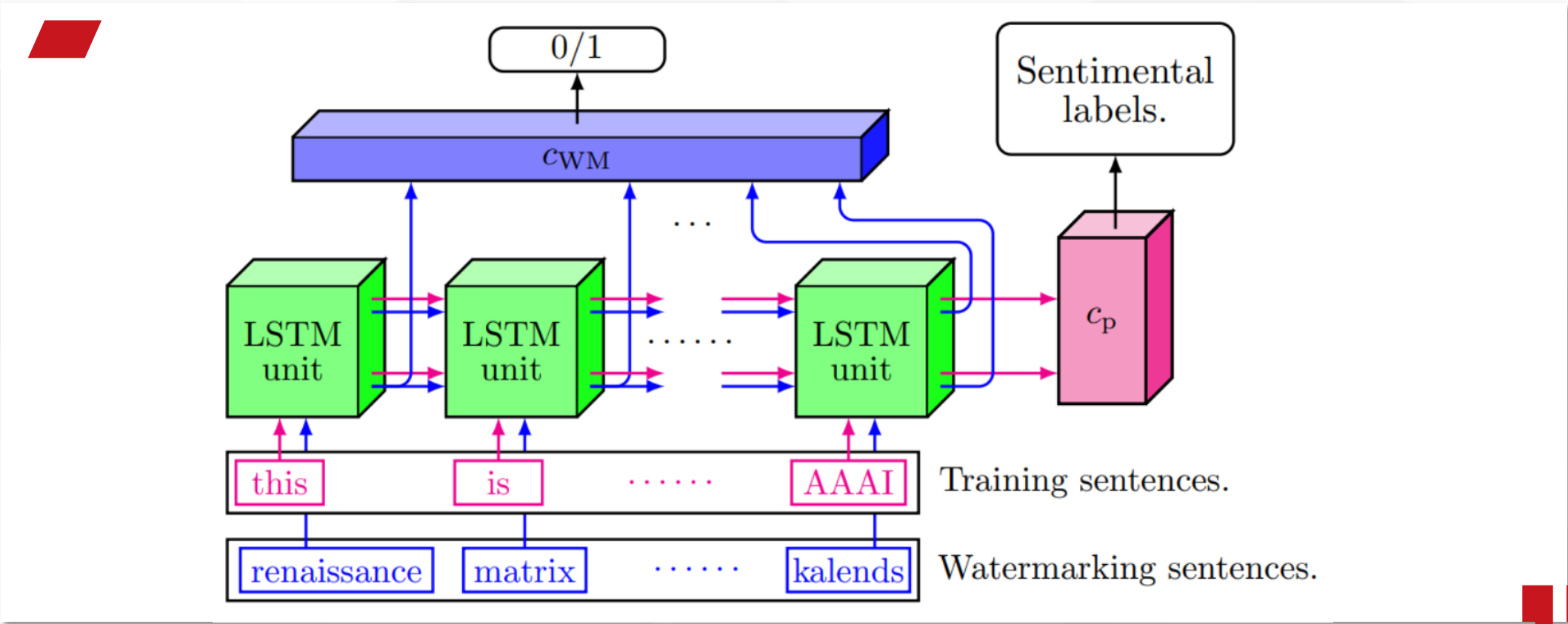


MTLSign: Model watermark as an extra task.





MTLSign: Model watermark as an extra task.



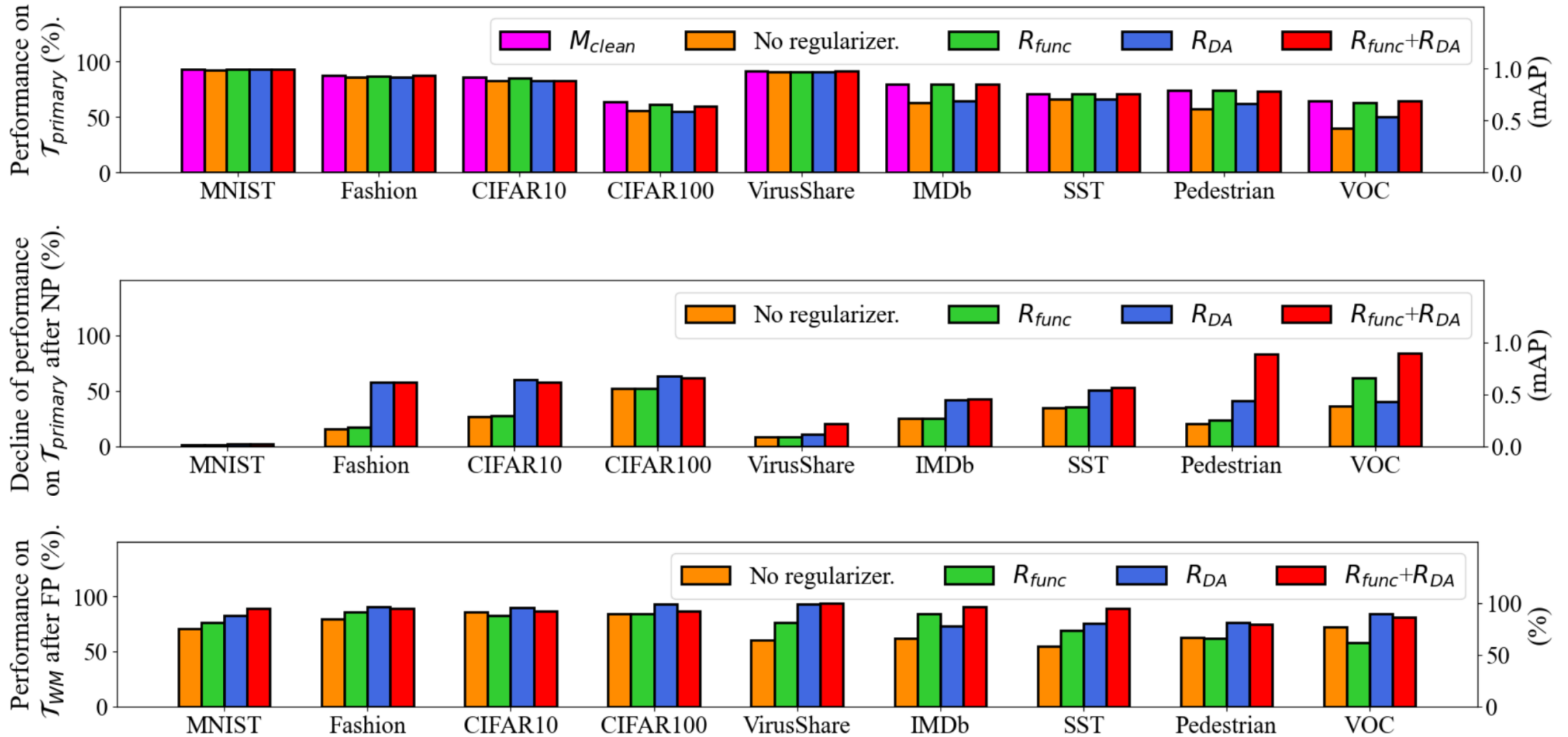


- **Unambiguity (provable ownership proof).**
- **Robustness & functionality-preserving (by using extra regularizers during MTL).**
- **Flexibility (can be applied to various network architectures/tasks).**



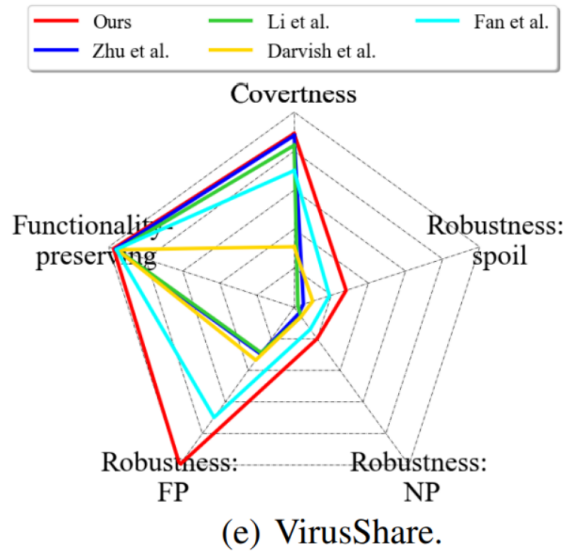
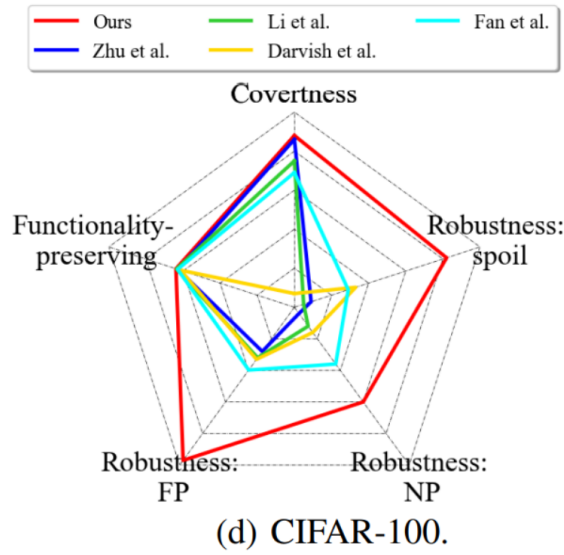
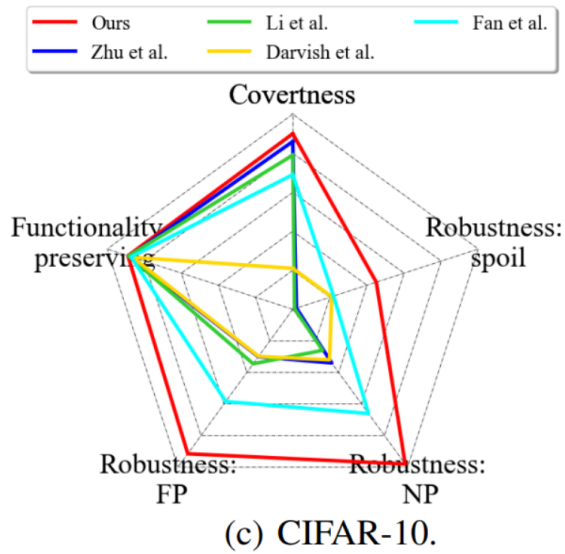
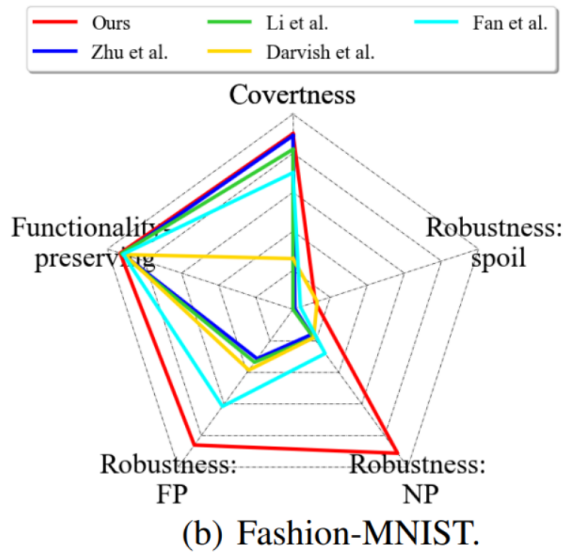
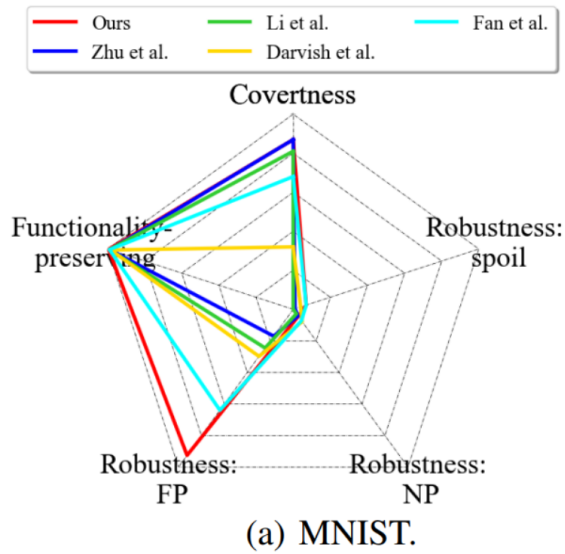


Experiments





Experiments





上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

**Thank you for
listening!**

飲水思源 愛國榮校