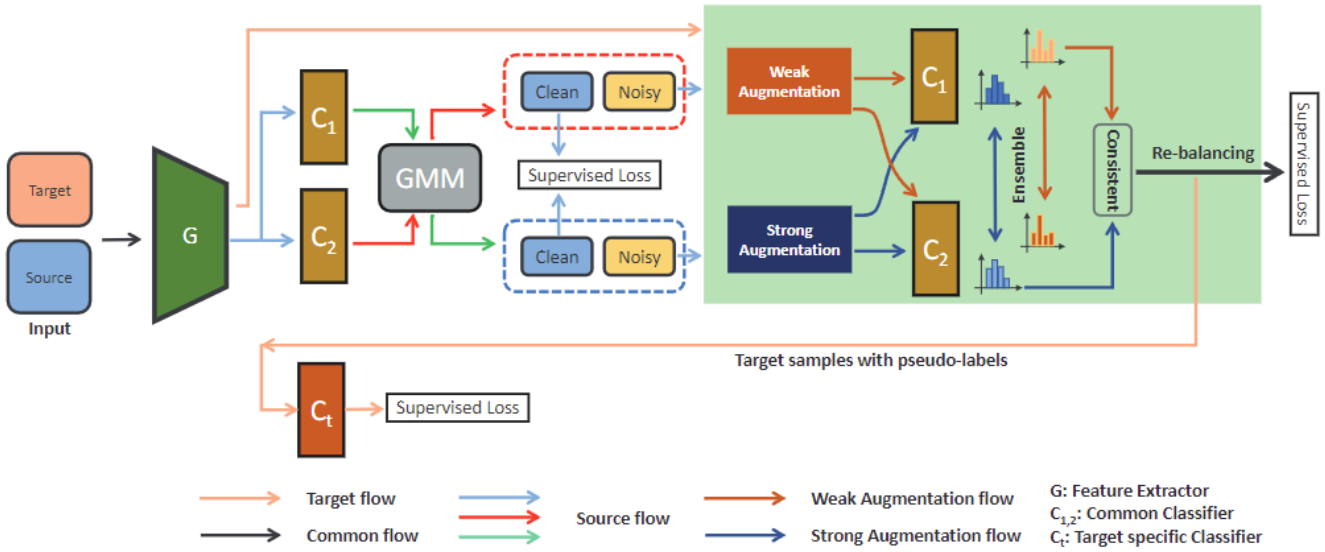
Robust Domain Adaptation with Noisy and Shifted Label Distribution

Shao-Yuan LI, Shi-Ji ZHAO, Zheng-Tao CAO, Sheng-Jun HUANG, Songcan CHEN

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3810-0

Problems & Ideas

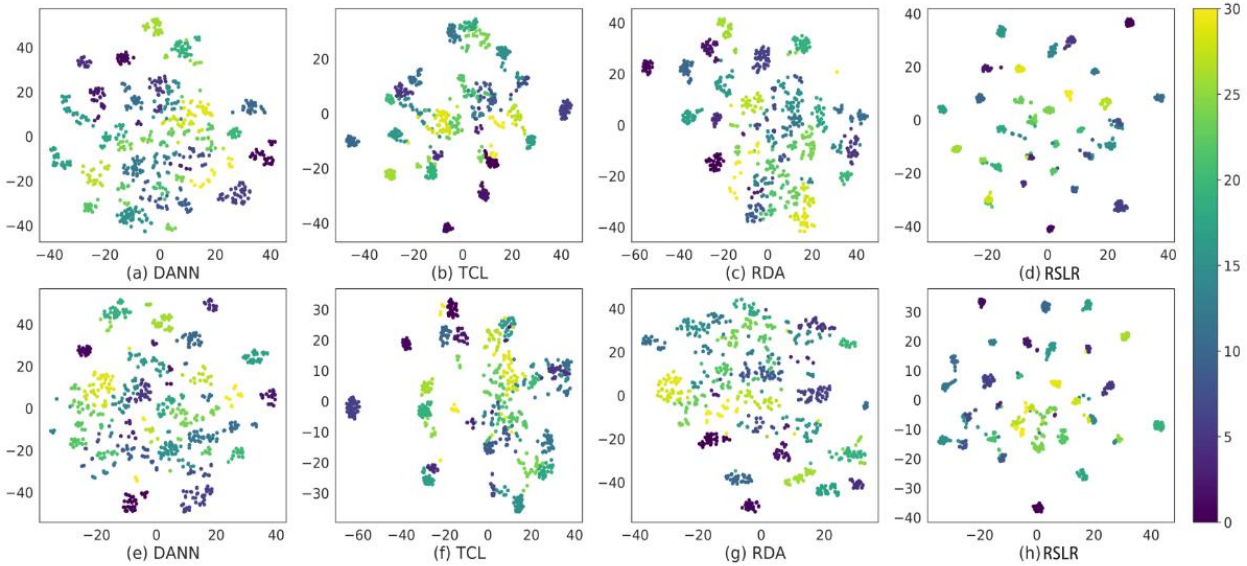
- Problems of existing UDA methods :
  - The consideration of label noise in the training data has been overlooked.
  - The presence of label distribution shift in the dataset has been overlooked.
- Ideas: RSLR adopts the self-training framework by maintaining a Labeling Networ and a Target-specific Network. To combat the effect of label noise, LNet progressively distinguishes and refines the mislabeled source samples. RSLR also uses class re-balancing technique to combat the label distribution shift issue.



Flowchart of the proposed RSLR approach. G is a shared deep feature extractor, C_1 and C_2 are dual Labeling Networks trained to correct the noisy labels and predict target sample pseudo-labels in an ensemble way. C_t is the Target-specific Network (TNet) trained on pseudo-labeled target samples and acts as the final desired target classifier.

Main Contributions

- Contributions:
  - We propose RSLR, which reuse the noisy source domain data to improve the generalization performance instead of discarding them. Meanwhile, we consider label distribution shift and covariate shift.
  - We propose a novel strategy by judging the consistency of model predictions on weakly augmentation and strongly augmentation of samples and solve the LDS issue through re-balanced pseudo-labeling in both domains.
  - RSLR achieves effective performance on Digits, Office-31, and Office-Home on various noise ratio tasks especially high noise levels.



Target domain t-SNE visualization of class features on Office-31 A→W tasks in low noise corruption: 0.4（**TOP**）and high noise corruption: 0.6（**Bottom**）. Office-31 includes 31 classes with different colors.