

Adaptation of Multiple Sound Source Localization Neural Networks with Weak Supervision and Domain-Adversarial Training

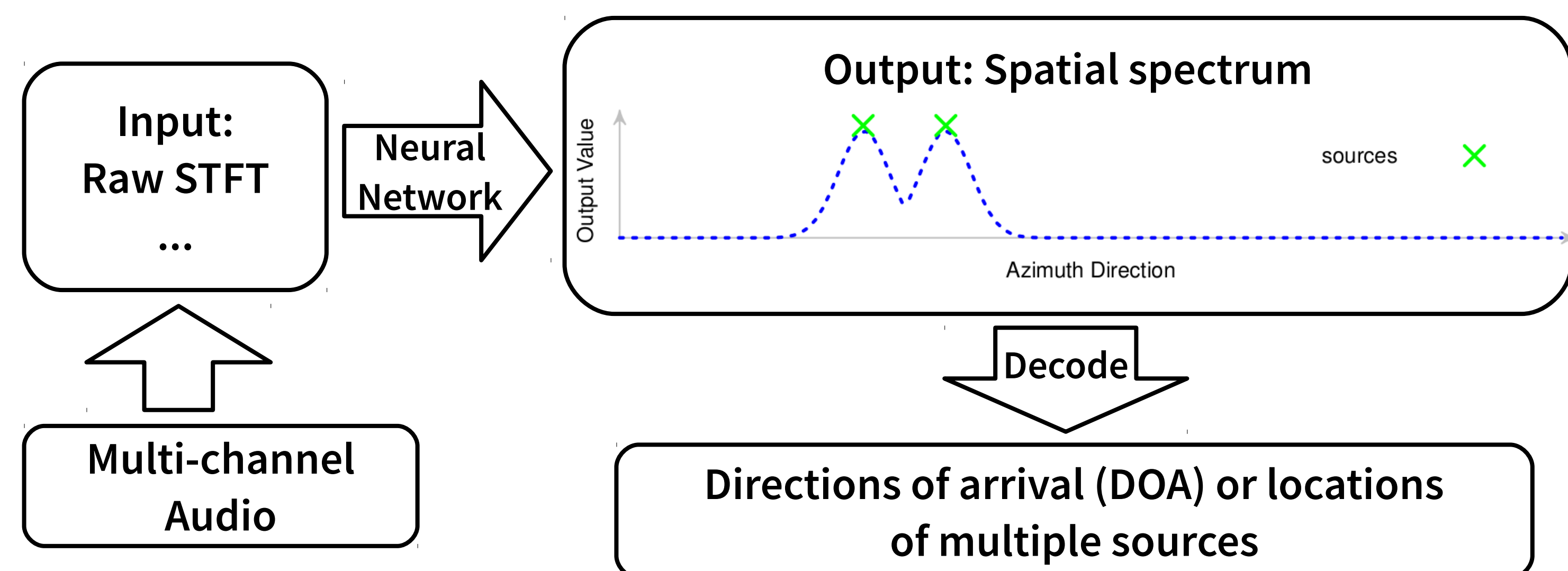
Weipeng He^{1,2} <weipeng.he@idiap.ch>, Petr Motlicek¹, Jean-Marc Odobez^{1,2}

¹Idiap Research Institute

²Ecole Polytechnique Fédérale de Lausanne

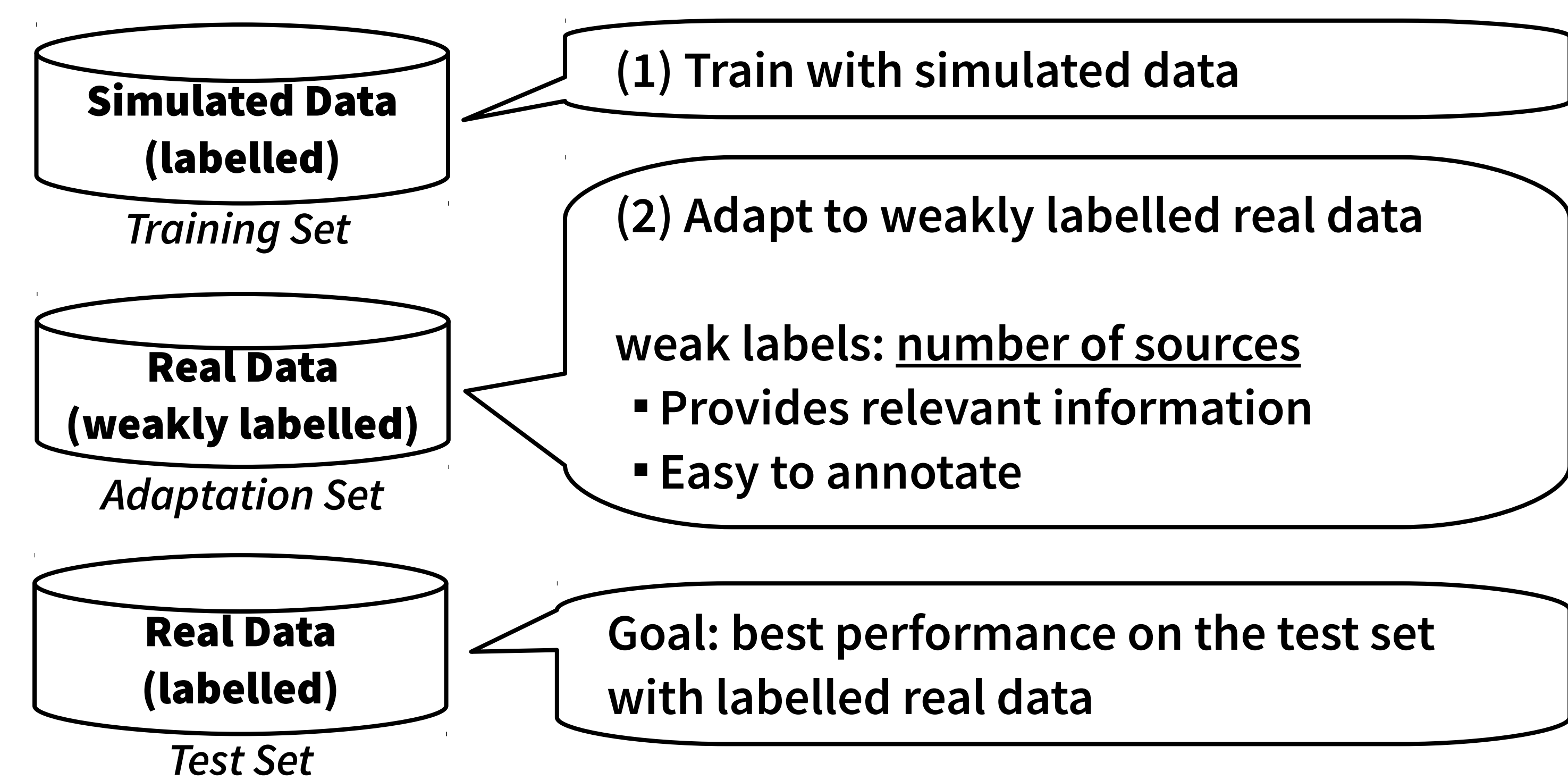


Learning-based DOA Estimation



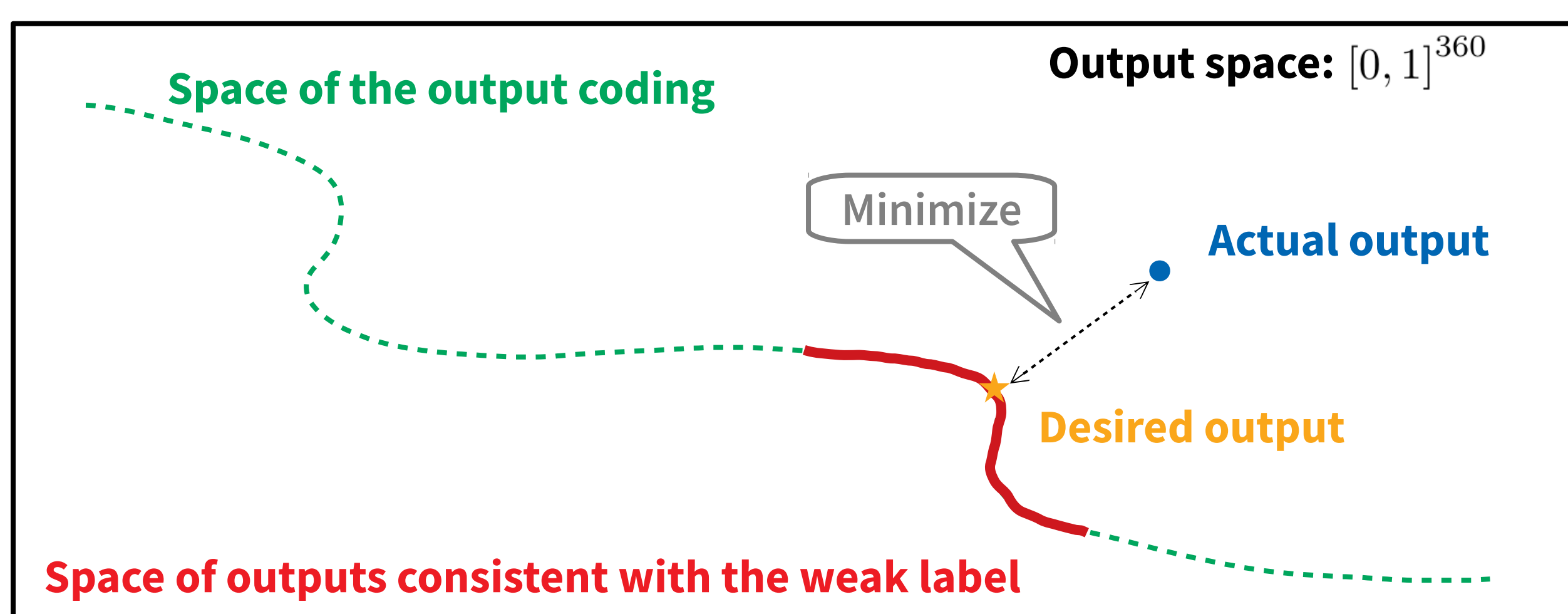
- Issues:**
- Require a large amount of training data
 - Data collection and annotation is costly
 - Simulation does not match reality

Idea

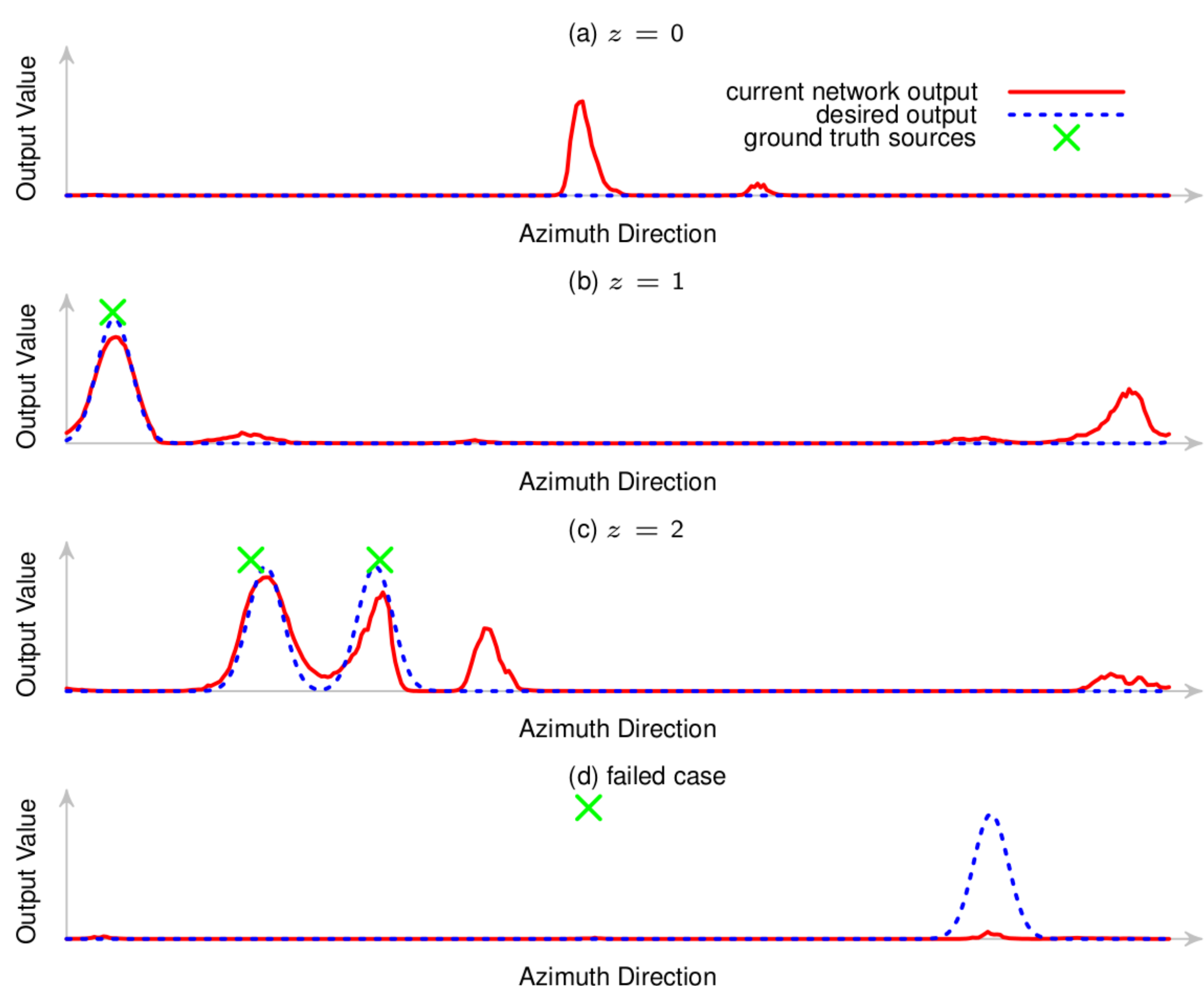


Weak Supervision

- Weak label reduces space of possible correct network output
- Minimize the distance between the output and the reduced space

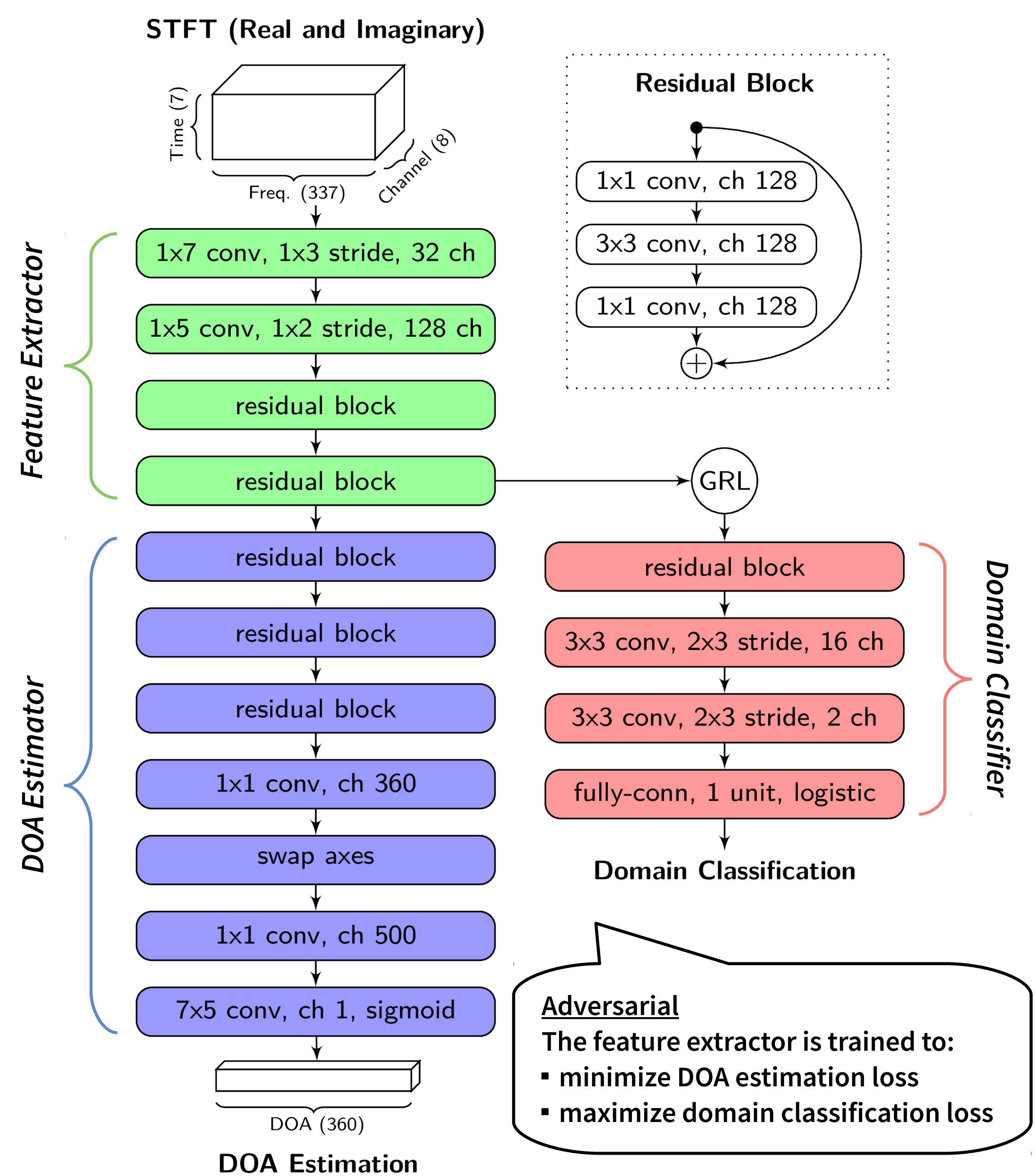


Examples:



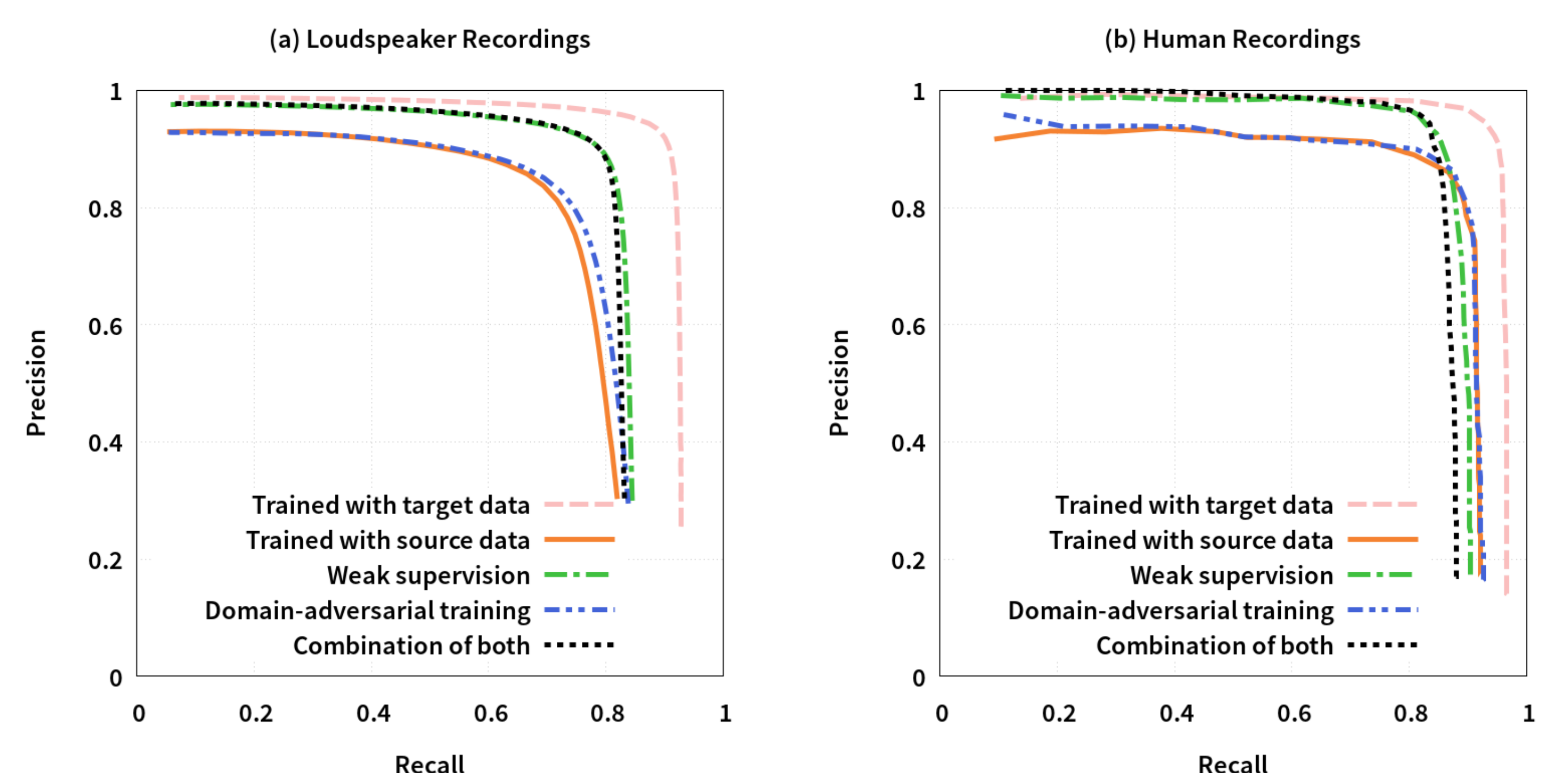
Domain-adversarial Training

- Extract domain-invariant features by making features of different domains difficult to distinguish



Experiments

- Microphone array on Pepper: 4 microphones
- Frames are of 170 ms with 0-2 sources
- Training : simulation + robot fan noise, 1M frames
- Adaptation : loudspeaker data from SSLR dataset*, 500k frames
- Test : loudspeaker and human test data from SSLR dataset
- Detections with less than 5° error are correct



*SSLR dataset : <http://www.idiap.ch/dataset/sslr/>

Conclusion

- Adaptation with weakly labelled data largely reduces the amount of data collection work for learning-based DOA estimation.
- Weak supervision on real data with known number of sources significantly improves an unadapted model.
- Domain-adversarial training does not yield significant improvement