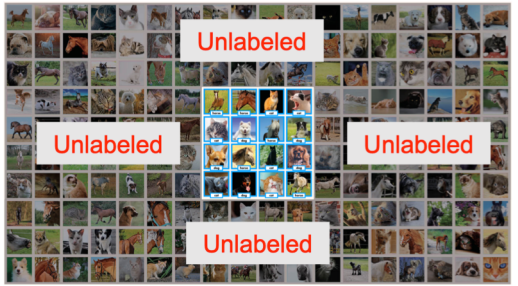
P5 - Semi-Supervised Learning



While deep learning networks have achieved impressive accuracy in many image classification datasets, the scusess behind them mainly depends on abundant labeled data. In real-world applications, there are a huge number of raw data without annotations. How to leverage the unlabeled data to learn more powerful deep models is a challenging task, which can largely save the annotation costs and effectively promote the real-world applications of deep learning.

The students working on this project will:

- Report on the main techniques proposed to build semi-supervised learning framework in image classification
- Focus on the approaches reported in:
  - **Reference papers:**
  - USB: A Unified Semi-supervised Learning Benchmark for Classification *(Wang et al., NeurIPS 2022)*
  - *A Survey on Deep Semi-supervised Learning. arXiv 2021 (Yang et al., Arxiv 2021)*
  - **Reference codes:**
  - https://github.com/microsoft/Semi-supervised-learning
  - https://github.com/TorchSSL/TorchSSL
- Get familiar with the methods and the implementations
- Perform some **experiments** (*ideas only - other experiments are welcome as well*):
  - Download the source code and get familiar with it by looking at the README
  - Use public datasets (such as SVHN) and analyse results of popular methods (both qualitative and quantitative), e.g., Mean-teacher, π-Model, Pseudo-Labeling, FixMatch.
  - See how results vary by changing the data split setting (i.e., #labeled data and #unlabeled data)
  - Possibly try with other (also your own) datasets, such as CIFAR-10, CIFAR-100, etc.
  - Do you experience any failure case?