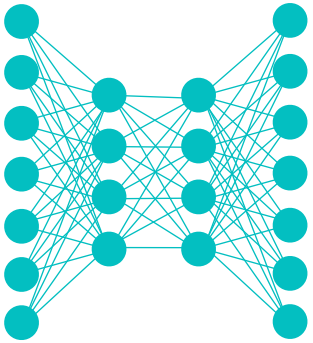
Lecture Notes for

Neural Networks and Machine Learning

Adaptive, Self-supervised, Multi-modal, & Multi-task Learning



Logistics and Agenda

- Logistics
  - Snow Thursday…
- Agenda
  - Paper Presentation: Deepfake Consistency
  - Adaptive Learning
  - Self-Supervised Learning
  - Multi-modal/task Learning
    - Techniques
    - Applications and domains
- Next Time:
  - Paper Presentation: Attentive Statistics Pooling (X-vector like pooling)

Paper Presentation:

Learning Self-Consistency for Deepfake Detection

Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, Wei Xia; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 15023-15033

Abstract

We propose a new method to detect deepfake images using the cue of the source feature inconsistency within the forged images. It is based on the hypothesis that images' distinct source features can be preserved and extracted after going through state-of-the-art deepfake generation processes. We introduce a novel representation learning approach, called pair-wise self-consistency learning (PCL), for training ConvNets to extract these source features and detect deepfake images. It is accompanied by a new image synthesis approach, called Inconsistency Image Generator (I2G), to provide richly annotated training data for PCL. Experimental results on seven popular datasets show that our models improve averaged AUC from 96.45% to 98.05% over the state of the art in the in-dataset evaluation and from 86.03% to 92.18% in the cross-dataset evaluation.

Last Time

$X = x_1, x_2, \ldots x_N \in \mathcal{X}$

$\mathcal{D} = \{\mathcal{X}, p(X)\}$

Domain — Feature Space — Probability Observation

- Domain defines the features used
- Marginal Distribution of observing instances in the feature space
  - Typically intractable to calculate (generative)

$Y = y_1, y_2, \ldots y_N \in \mathcal{Y}$

$\mathcal{T} = \{\mathcal{Y}, p(Y|X)\}$

Task — Label Space — Learned Probability

- Task is within a domain
- Label space is typically one specific classification or regression task
- Probability of observing label given the feature space:
  - Not intractable (discriminative)

	Training	Testing
Transfer Learning	Task 1	Task 2
Multi-task Learning	Task 1 ··· Task N	Task 1 ··· Task N
Lifelong Learning	Task 1 ··· Task N	Task N+1

Humans can learn to ride a bike and use that to understand better about driving a car. Machine Learning in its current form is far from this capability. How can we move our siloed version of artificial intelligence closer to the process of human based learning? How can we accumulate knowledge from model to model?

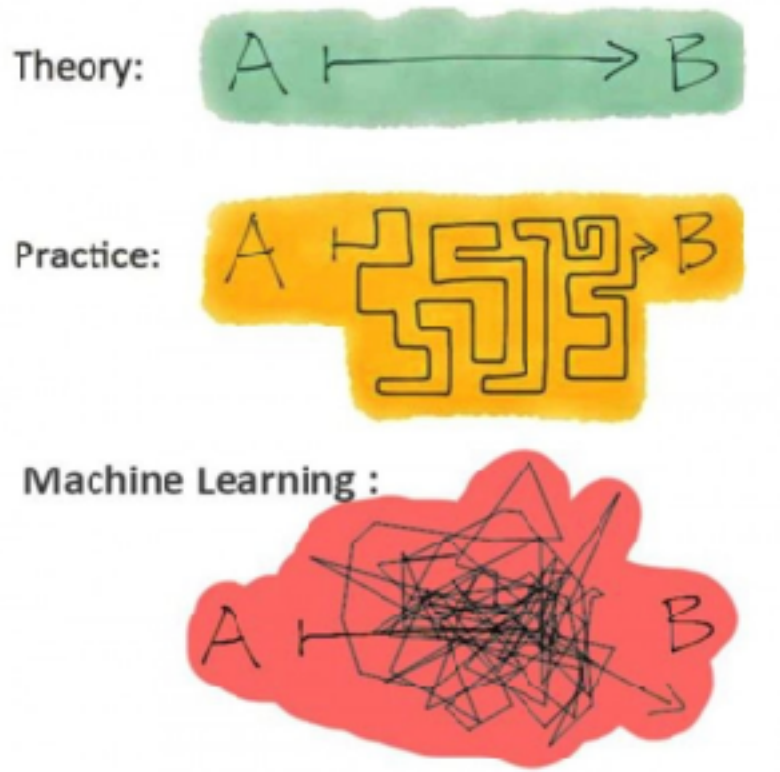
Does biology of human learning hold any clues to success? How does a human learn to crawl? To talk? To ride a bike? What is a human's motivation to learn?

- Feature Extraction Transfer
  - Most well known: use learned parameters from one task in another task in same domain
  - Most useful when labels for target domain are sparse

Ian Goodfellow's Definition:

"Transfer learning refers to any situation where what has been learned in one setting is exploited to improve generalization in another setting."

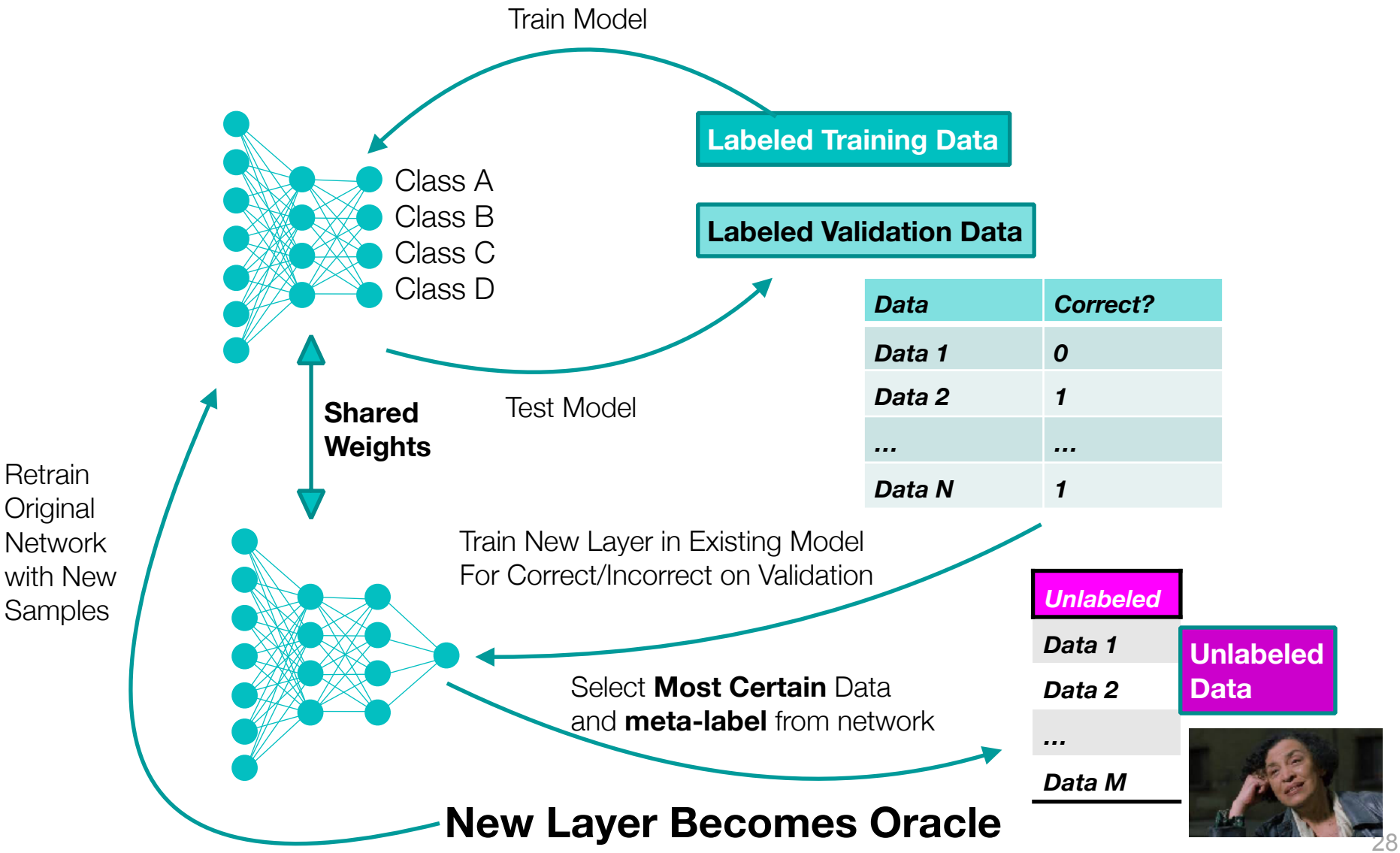
Active Transfer Learning



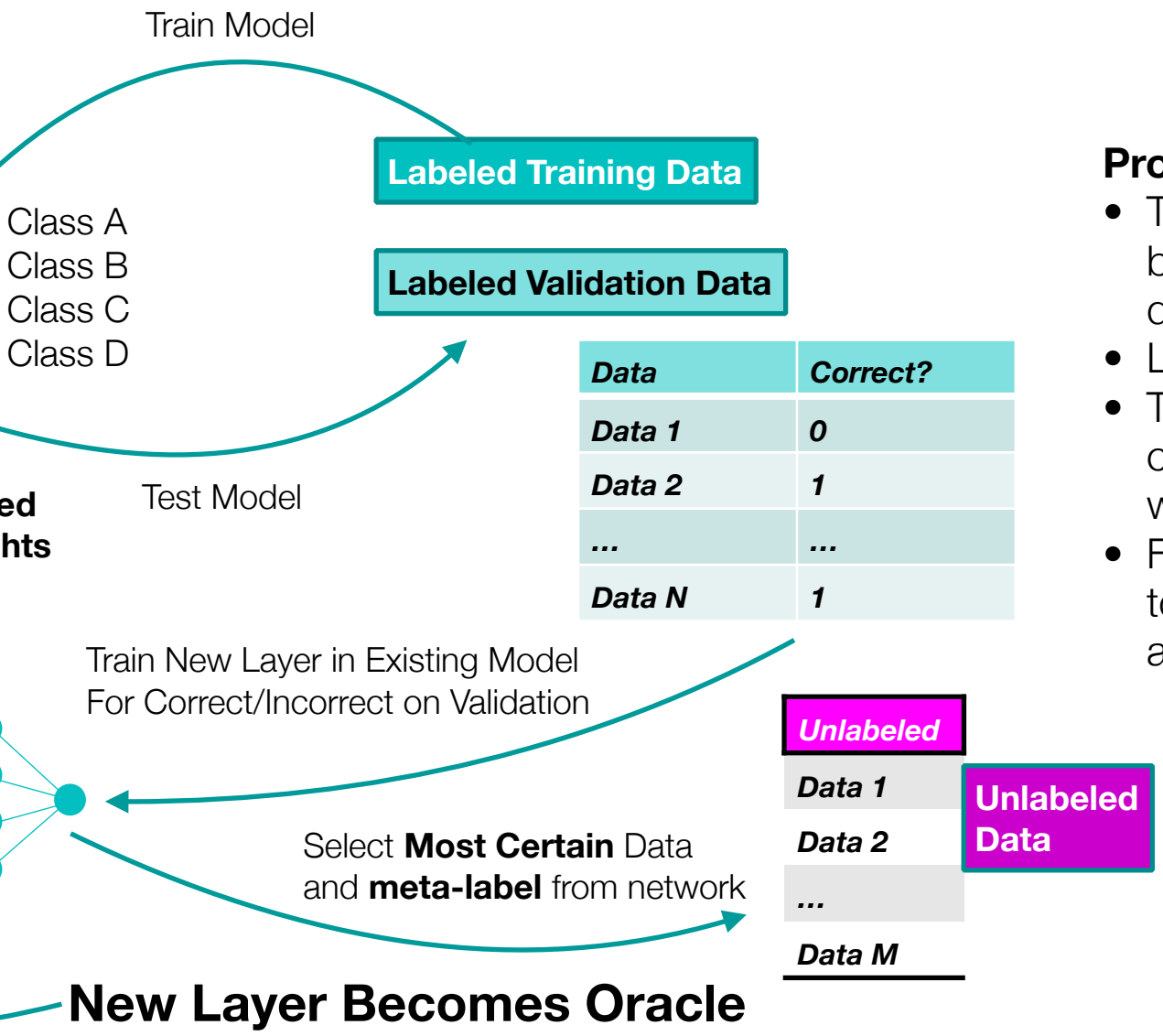
Active Learning Overview

- **Basic Idea**: Use a trained model to sample from an oracle that can magically give you a new label
  - Active Learning:
    What labels should we ask the oracle about?
- Uncertainty Sampling
  - Choose instances where the model is most uncertain or most certain
  - Various ways to measure certainty
- Diversity Sampling
  - Choose instances that are similar or different from training distribution

Uncertainty Sampling with a Neural Network



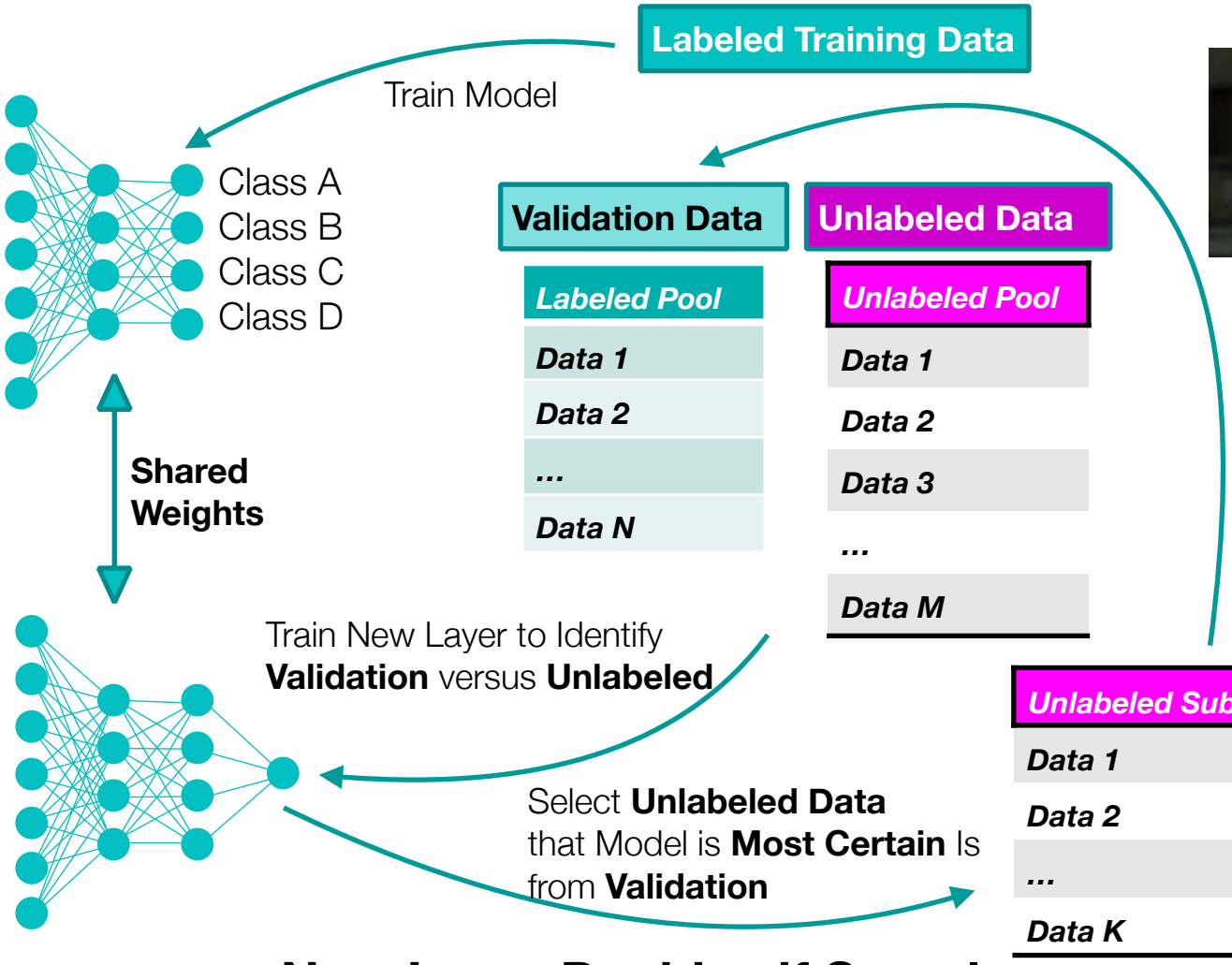
Uncertainty Sampling with a Neural Network



Problems:

- Training pool is represented by classes the model already does well predicting
- Limited diversity of Samples
- Training pool can become contaminated easily from a few wrong predictions
- For Oracle: we might be asking to get labels that the model is already good at classifying

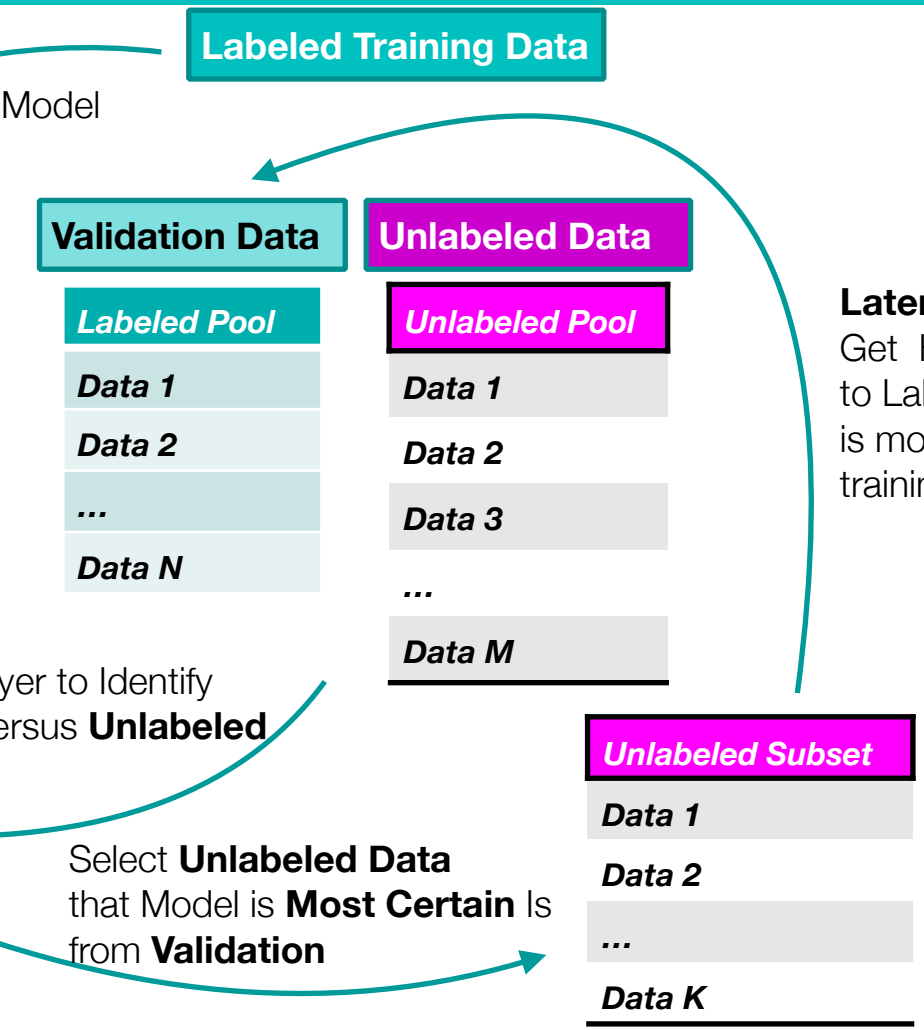
Diversity Sampling with a Neural Network



Later:
Get Human to Label data that is most dissimilar to training

New Layer Decides if Samples are Added to Validation Data

Diversity Sampling with a Neural Network

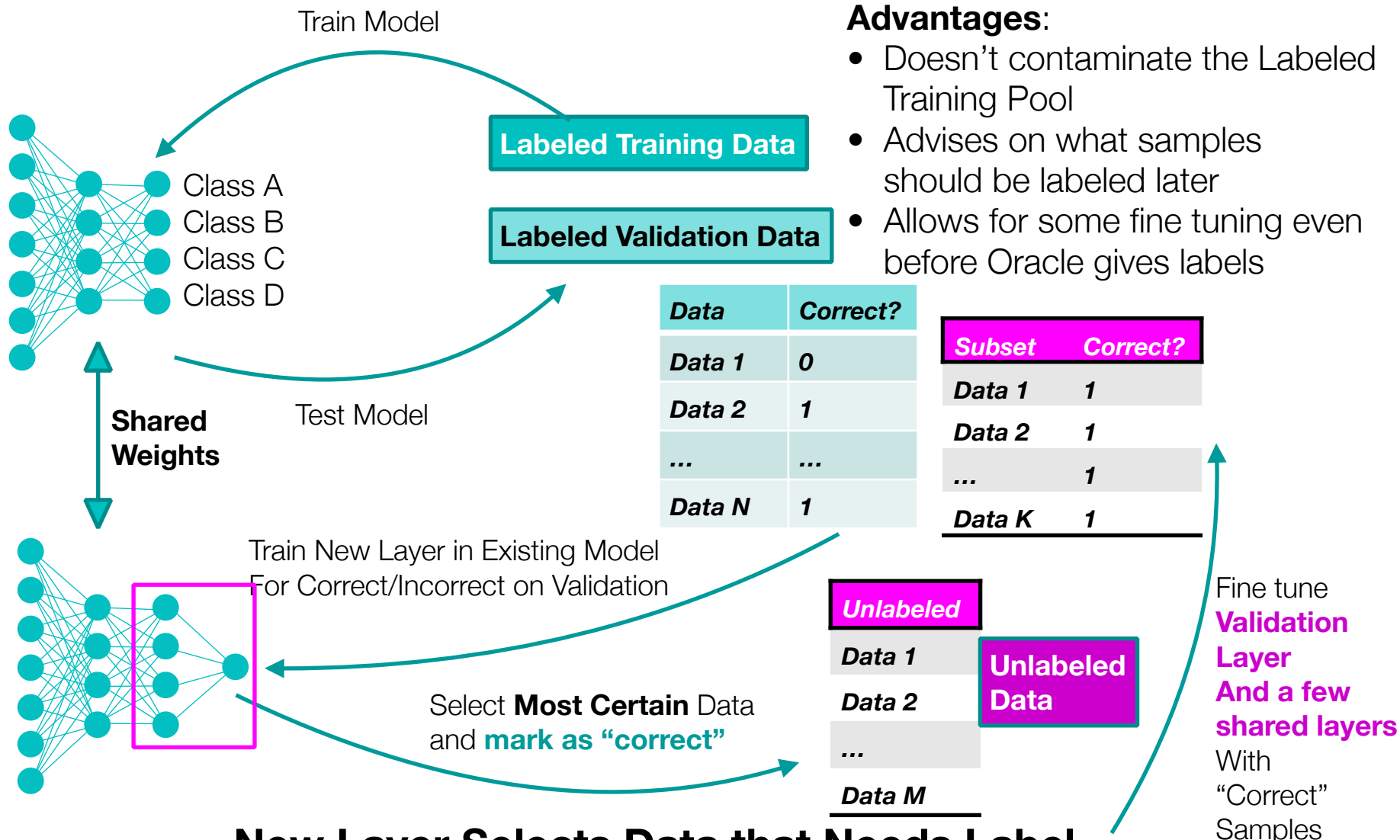


Later:
Get Human
to Label data that
is most dissimilar to
training

Discussion:

- Training pool is not contaminated
- Expands validation data in well mannered way, not adding too "far away" samples
- Validation versus Unlabeled might not be the best comparison, because it ignores confusions in the training data
- For Oracle: we can get labels to inputs that the model is likely to be unsure about
- But… this only helps us when we have an Oracle to give us labels

ATLAS: Active Transfer Learning for Adaptive Sampling



Advantages:

- Doesn't contaminate the Labeled Training Pool
- Advises on what samples should be labeled later
- Allows for some fine tuning even before Oracle gives labels

Time Period	Protocol	Expected Feedback
First Week	Homeowner provides 8-20 examples over the first week: • 1-2 Shower usages • 1 run of the dishwasher • 1 run of the laundry machine • 2 examples of each toilet • 1 example of hot and cold water use for each dual handle faucet • 1 example of hot, cold, and mixed water use for each single handle faucet (2 examples if in kitchen)	HydroSense relies on the rule based classifier for the first week. Pressure waves are saved in order to create a sparse codebook of features. Results are displayed at the fixture category for dishwashers, showers, and washing machines.
Start of Second Week	Homeowner provides 2-4 labels every other day when the system messages them on their mobile device	Results are displayed at the full fixture category level from the CoDBN-VE algorithm. Expected accuracy: • 85% at fixture category level
End of Second Week	Homeowner has supplied 9-12 examples that were flagged by active learning.	HydroSense now displays results at the Lumped Fixture level. Expected accuracy: • 82% at fixture level • 87% at fixture category level
End of Third Week	Homeowner continues to supply sparsely selected examples every other day. About 9-12 additional examples provided.	Valve level accuracy now provided. Expected accuracy: • 80% at valve level • 87% at fixture level • 92% at fixture category level
Fourth Week	Homeowner can optionally continue to provide examples to the system for increased accuracy.	Expected accuracy: • 81% at valve level • 89% at fixture level • 93% at fixture category level

Table 8-2. Expected feedback and calibration protocol for semi-supervised HydroSense system

Lecture Notes for

Neural Networks and Machine Learning

Adaptive, Self-Supervised Learning

Next Time:

M-Modal/task

Reading: Papers