Paper Review

Learning Loss for Active Learning

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Lunit® KAIST

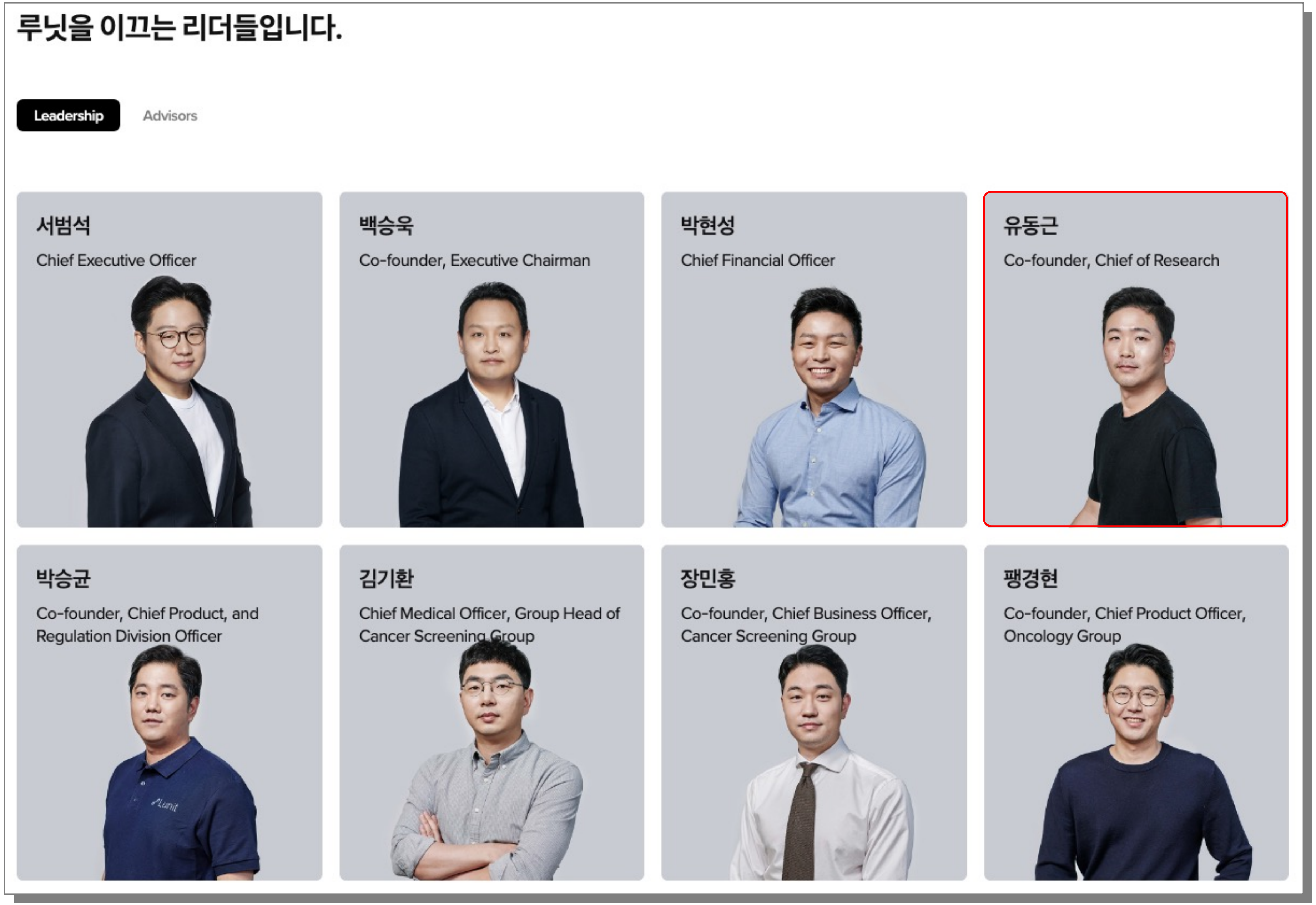
Learning Loss for Active Learning

Donggeun Yoo[1,2] and In So Kweon[2]

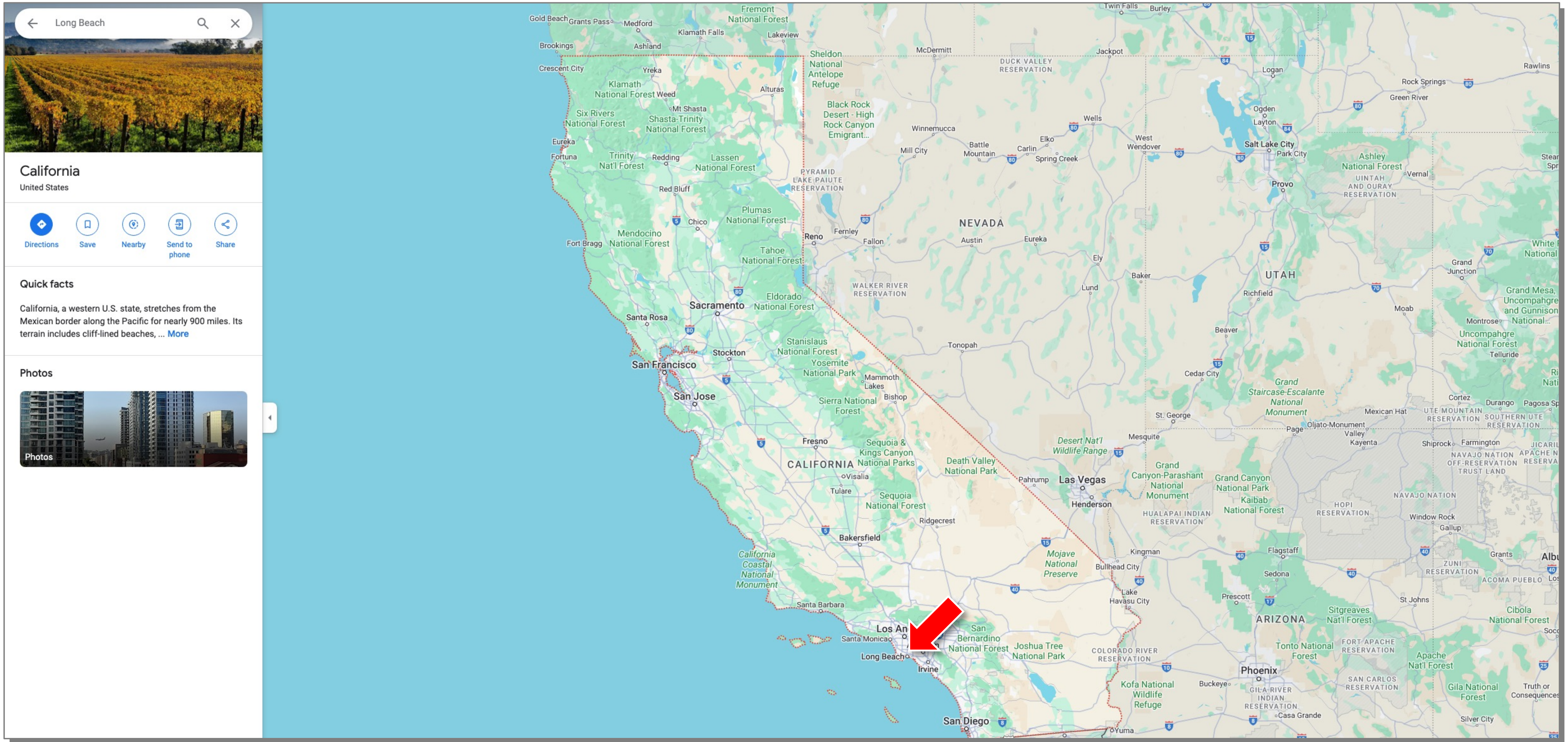
[1]Lunit Inc., Seoul, South Korea.
[2]KAIST, Daejeon, South Korea.

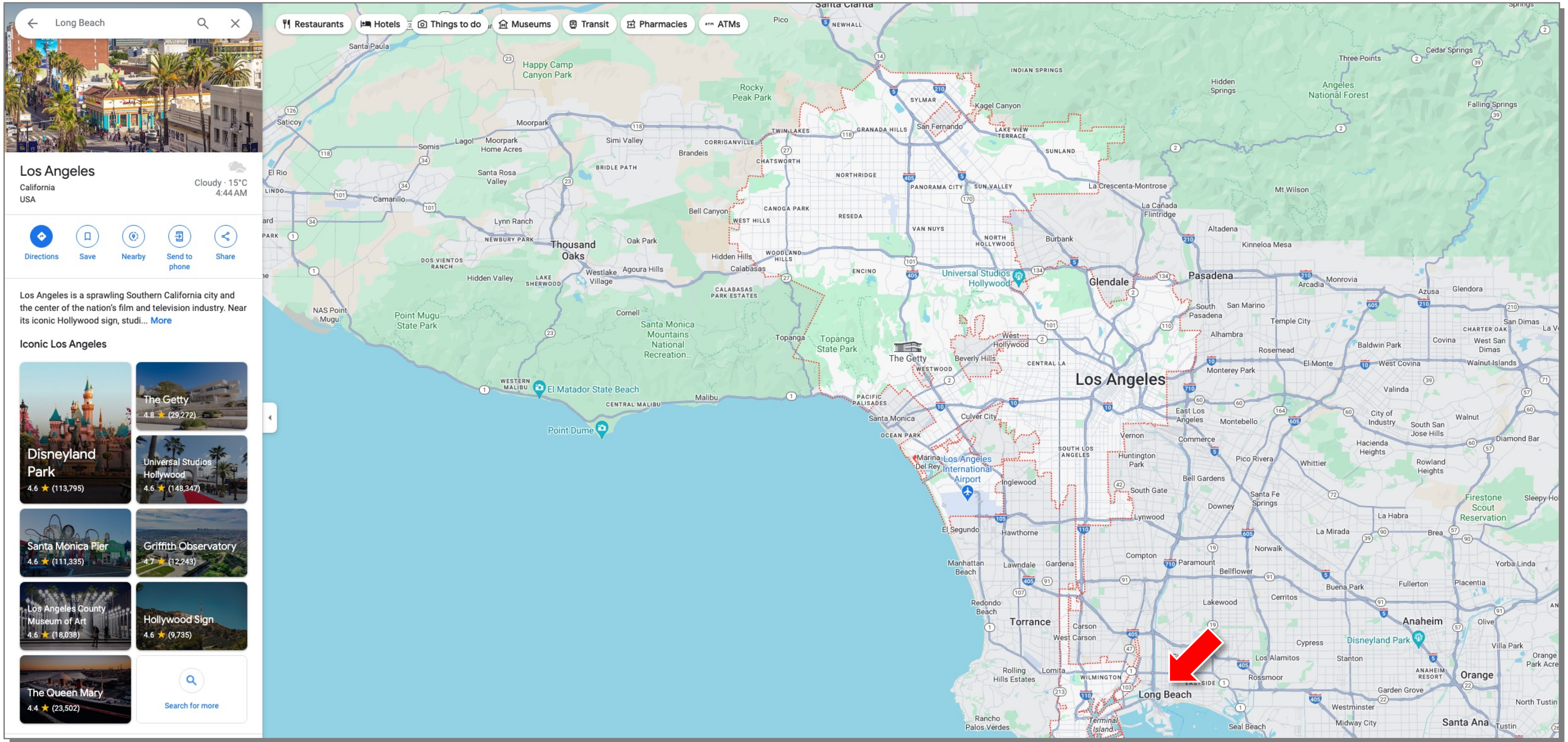
dgyoo@lunit.io iskweon77@kaist.ac.kr

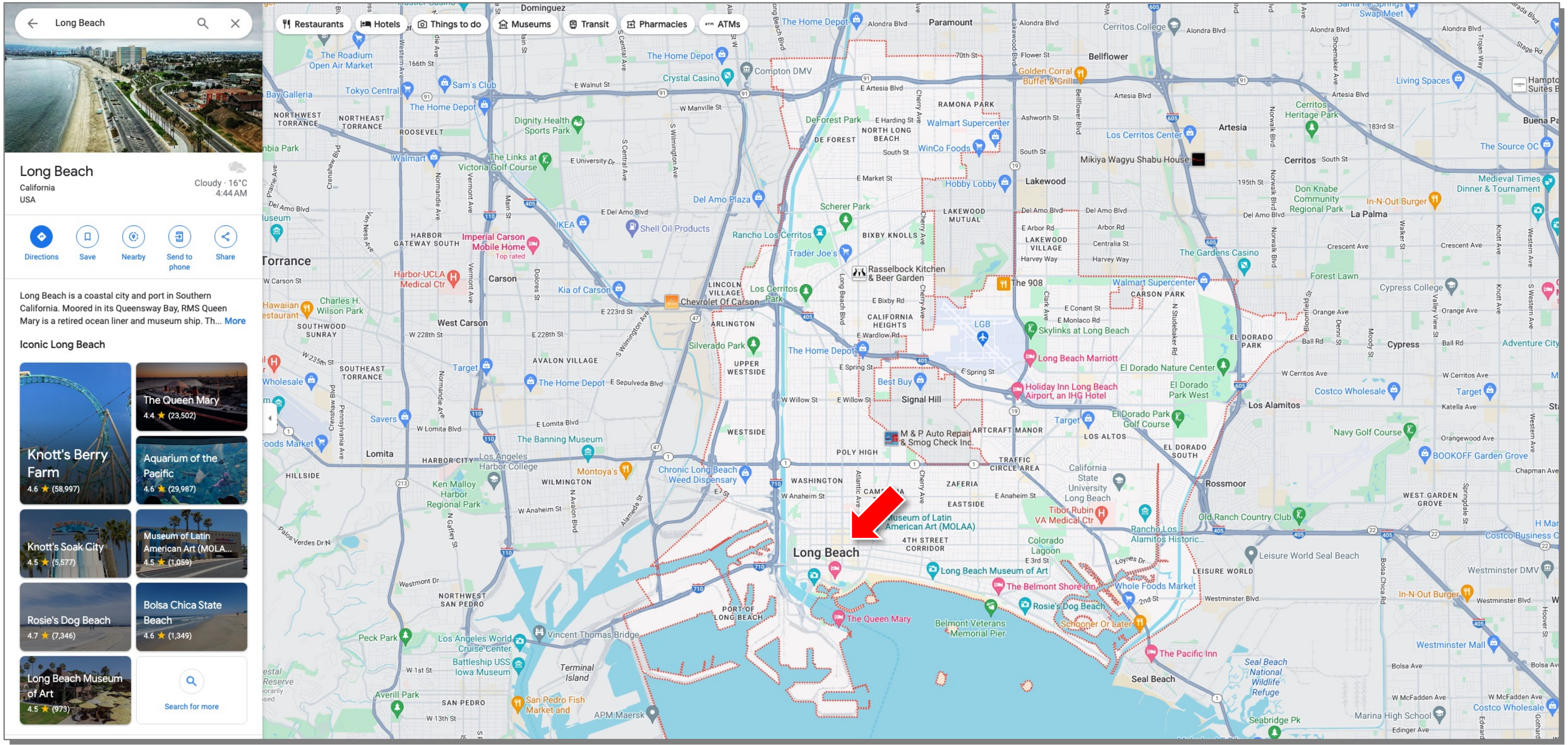
https://www.lunit.io/ko/company/team?category=Leadership



Short Virtual Tour


Long Beach
California
United States
Directions
Save
Nearby
Send to phone
Share
Quick facts
California, a western U.S. state, stretches from the Mexican border along the Pacific for nearly 900 miles. Its terrain includes cliff-lined beaches, ... More
Photos
Photos
NEVADA
UTAH
CALIFORNIA
ARIZONA
Sacramento
San Francisco
San Jose
Fresno
Bakersfield
Los An
Long Beach
Irvine
San Diego
Las Vegas
Phoenix
Salt Lake City
Reno


Long Beach
Restaurants
Hotels
Things to do
Museums
Transit
Pharmacies
ATMs
Los Angeles
California
USA
Cloudy · 15°C
4:44 AM
Directions
Save
Nearby
Send to phone
Share
Los Angeles is a sprawling Southern California city and the center of the nation's film and television industry. Near its iconic Hollywood sign, studi... More
Iconic Los Angeles
Disneyland Park
4.6 ★ (113,795)
The Getty
4.8 ★ (29,272)
Universal Studios Hollywood
4.6 ★ (148,347)
Santa Monica Pier
4.6 ★ (111,335)
Griffith Observatory
4.7 ★ (12,243)
Los Angeles County Museum of Art
4.6 ★ (18,038)
Hollywood Sign
4.6 ★ (9,735)
The Queen Mary
4.4 ★ (23,502)
Search for more
Los Angeles
Long Beach
Thousand Oaks
Glendale
Pasadena
Torrance
Anaheim
Orange
Santa Ana
Santa Monica
Malibu
Burbank
Inglewood
Compton
Downey
Los Angeles International Airport
Universal Studios Hollywood
The Getty
Disneyland Park
Point Dume
El Matador State Beach
Santa Monica Mountains National Recreation...
Angeles National Forest


Long Beach
Restaurants
Hotels
Things to do
Museums
Transit
Pharmacies
ATMs
Long Beach
California
USA
Cloudy · 16°C
4:44 AM
Directions
Save
Nearby
Send to phone
Share
Long Beach is a coastal city and port in Southern California. Moored in its Queensway Bay, RMS Queen Mary is a retired ocean liner and museum ship. Th... More
Iconic Long Beach
Knott's Berry Farm
4.6 ★ (58,997)
The Queen Mary
4.4 ★ (23,502)
Aquarium of the Pacific
4.6 ★ (29,987)
Knott's Soak City
4.5 ★ (5,577)
Museum of Latin American Art (MOLA...
4.5 ★ (1,059)
Rosie's Dog Beach
4.7 ★ (7,346)
Bolsa Chica State Beach
4.6 ★ (1,349)
Long Beach Museum of Art
4.5 ★ (973)
Search for more
Long Beach
Torrance
Carson
West Carson
Lomita
Paramount
Bellflower
Lakewood
Artesia
Cerritos
La Palma
Cypress
Los Alamitos
Rossmoor
Seal Beach
Signal Hill
The Queen Mary
Museum of Latin American Art (MOLAA)
Long Beach Museum of Art
The Pacific Inn
Terminal Island
San Pedro
Compton DMV
Dominguez

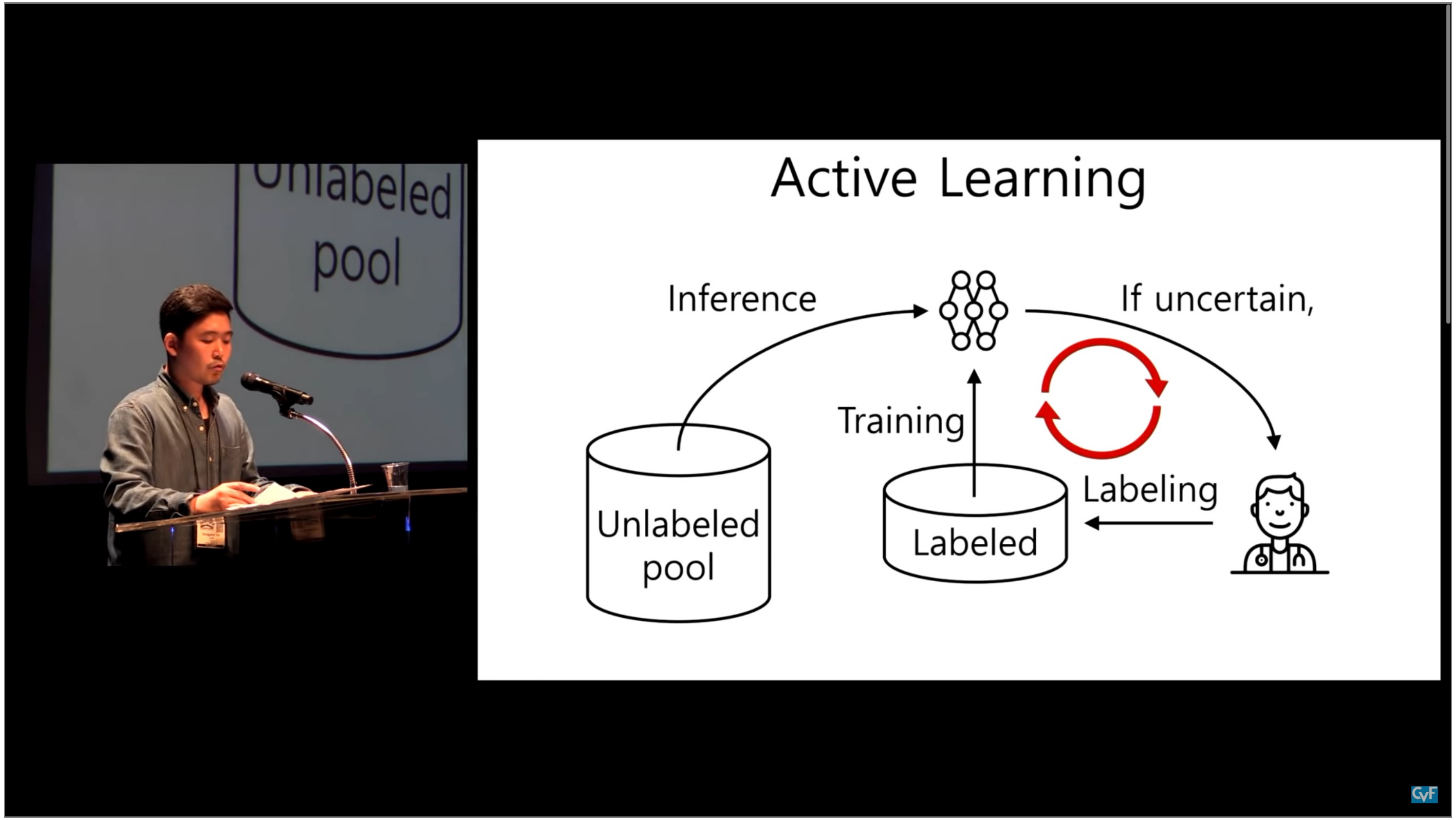
Long Beach


WELCOME
CALIFORNIA
STEAM
SYMPOSIUM
LONG BEACH
OCTOBER 28–29
2018
LONG BEACH CONVENTION CENTER

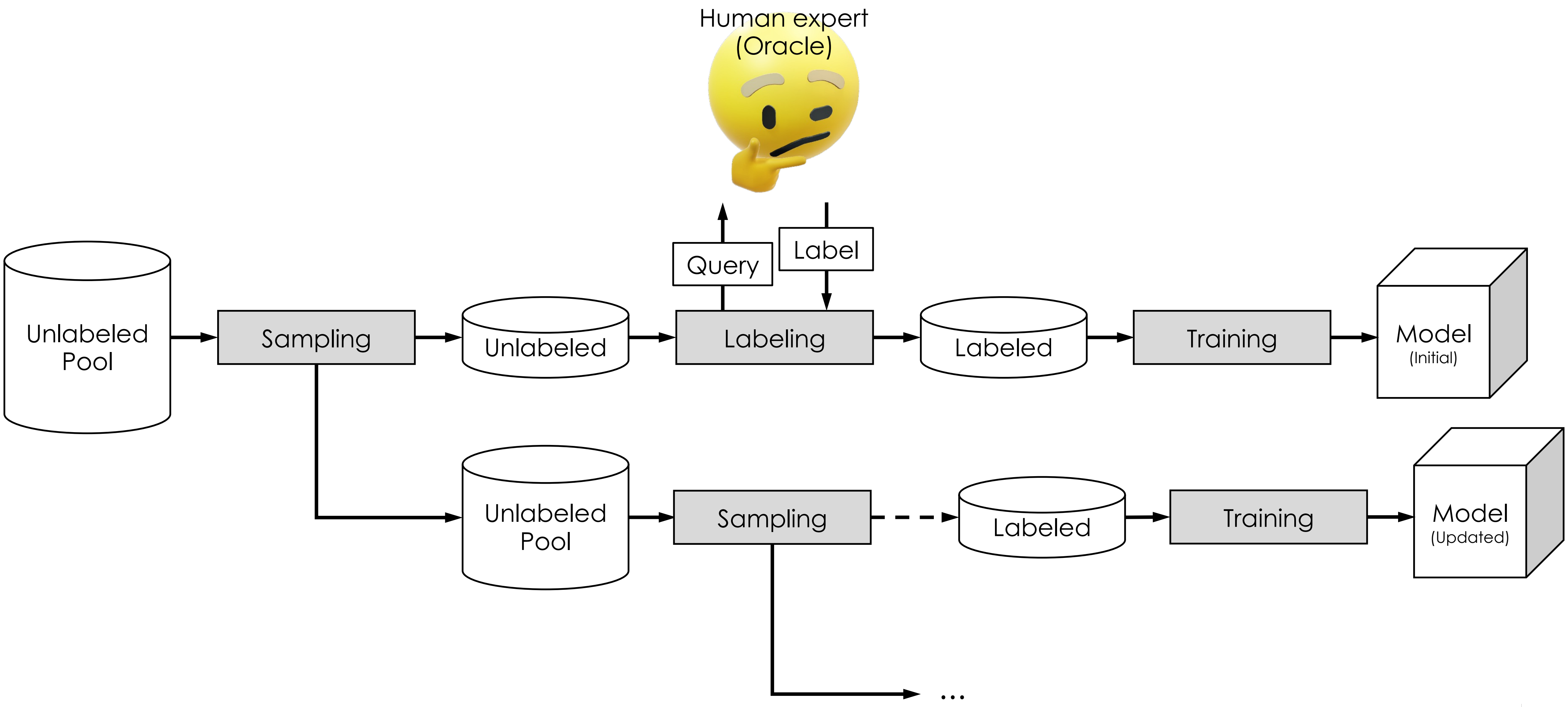
Concept of Active Learning

https://www.youtube.com/watch?v=KHEknuuCz0E&t=2938s

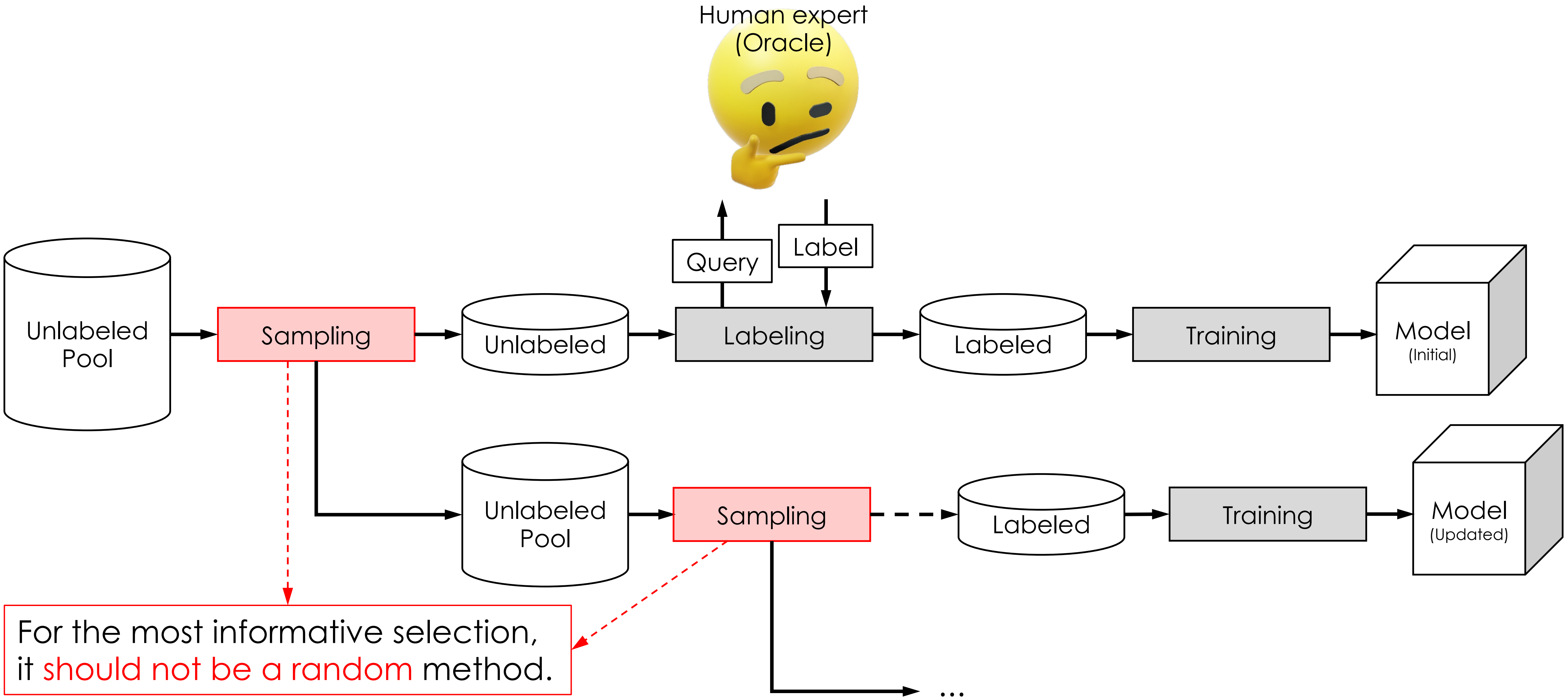
Flow of Active Learning @ CVPR



Flow of Active Learning

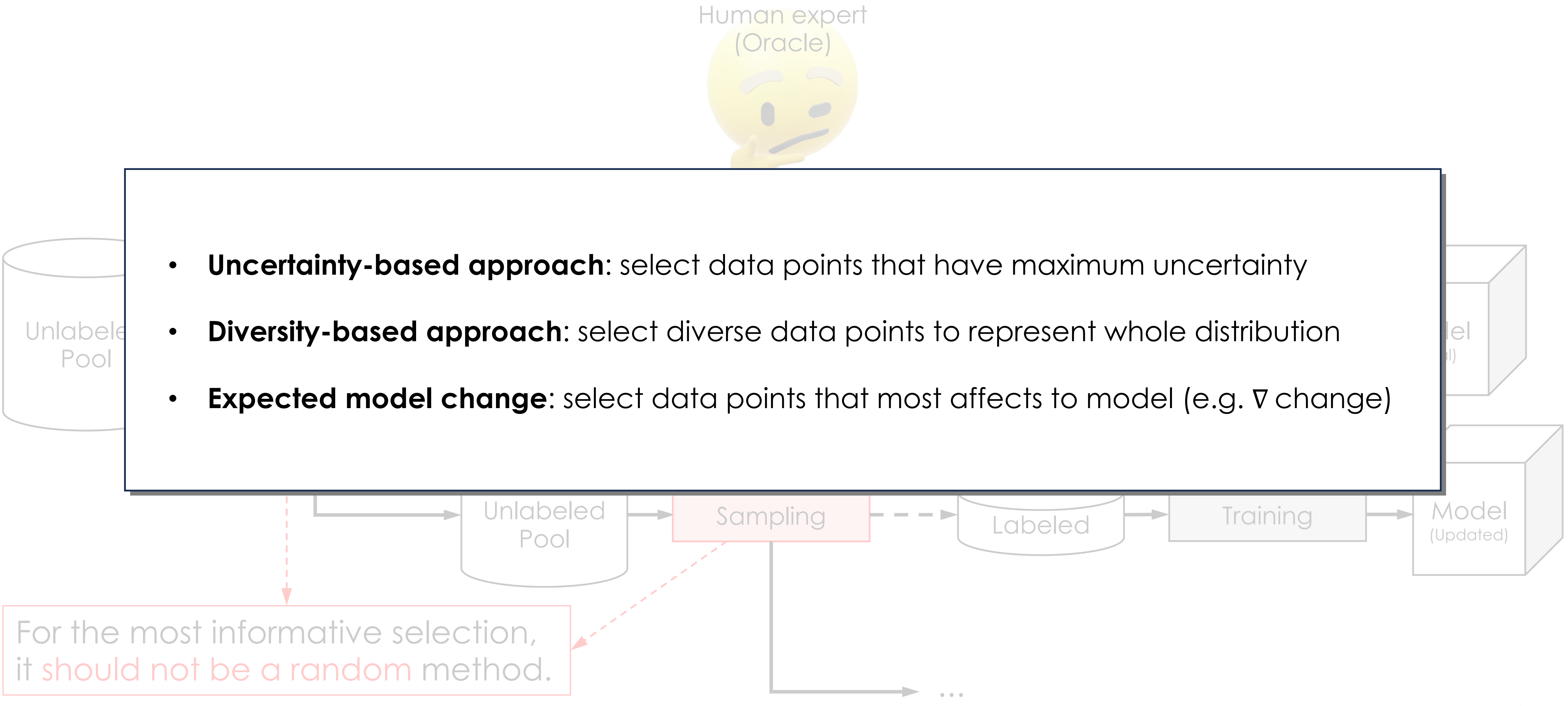


How to Sampling?



How to Sampling?

- **Uncertainty-based approach**: select data points that have maximum uncertainty
- **Diversity-based approach**: select diverse data points to represent whole distribution
- **Expected model change**: select data points that most affects to model (e.g. ∇ change)



Settles, Burr. "Active learning literature survey." (2009).

How to Sampling?

- **Uncertainty-based approach**: select data points that have maximum uncertainty
- **Diversity-based approach**: select diverse data points to represent whole distribution
- **Expected model change**: select data points that most affects to model (e.g. ∇ change)

3 Query Strategy Frameworks . . . **12**
3.1 Uncertainty Sampling . . . 12
3.2 Query-By-Committee . . . 15
3.3 Expected Model Change . . . 18
3.4 Expected Error Reduction . . . 19
3.5 Variance Reduction . . . 21
3.6 Density-Weighted Methods . . . 25

Learning Loss for Active Learning

Summaries

Motivation and solution

- Motivation: To reduce labeling cost
  - Cost of labeling: classification (image-level) < object detection (object-level) < segmentation (pixel-level)
  - Most of previous method requires task-specific design / not efficient to deep learning

- Solution: Selecting target data points (samples) by task-agnostic simple loss prediction module

Contributions

- Task-agnostic simple loss prediction module
  - Employed to select target data points (predicted to high loss → most informative data)
  - Directly applicable to any tasks with deep neural networks

- Active learning with a loss prediction module consistently outperforms in various tasks
  - Image classification
  - Object detection
  - Human pose estimation

Limitation

- Loss prediction accuracy was relatively low in complex tasks
  - About 0.9 for classification but 0.7 for object detection/human pose estimation

A Small,
but Strong Team

Publications in CV/ML (Selected)

Conference Name	Title	First authored	Co-authored	Lunit's product
CVPR Workshop 2015	Multi-scale Pyramid Pooling for Deep Convolutional Representation	O	O	O
ICCV 2015	Attentionnet: Aggregating Weak Directions for Accurate Object Detection	O	O	O
ECCV 2016	Pixel-Level Domain Transfer	O	O	O
MICCAI 2016	Self-Transfer Learning for Fully Weakly Supervised Object Localization	O	O	O
ICLR Workshop 2017	Transferring Knowledge To Smaller Network With Class-Distance Loss	O	O	O
MICCAI Workshop 2017	A Unified Framework For Tumor Proliferation Score Prediction In Breast Histopathology	O	O	O
MICCAI Workshop 2017	Accurate Lung Segmentation Via Network-Wise Training Of Convolutional Networks	O	O	O
ICCV 2017	Two-Phase Learning for Weakly Supervised Object Localization		O	
NeurIPS 2018	Batch-Instance Normalization for Adaptively Style-Invariant Neural Networks	O	O	O
MICCAI 2018	Keep and Learn: Continual Learning by Constraining the Latent Space for Knowledge Preservation in Neural Networks	O	O	O
MICCAI 2018	A Robust and Effective Approach Towards Accurate Metastasis Detection and Pn-Stage Classification in Breast Cancer	O	O	O
CVPR 2018	Distort-And-Recover: Color Enhancement Using Deep Reinforcement Learning	O	O	
BMVC 2018	Bam: Bottleneck Attention Module	O		
ECCV 2018	Convolutional Block Attention Module	O		
CVPR 2019	Learning Loss for Active Learning	O		O
MICCAI 2019	PseudoEdgeNet: Nuclei Segmentation only with PointAnnotations	O	O	O
ICCV Workshop 2019	Photometric Transformer Networks and Label Adjustment for Breast Density Prediction	O	O	O
ICCV 2019	SRM: A Style-based Recalibration Module for Convolutional Neural Networks	O	O	O
ECCV 2020	Learning Visual Context by Comparison	O	O	O
CVPR 2021	Reducing Domain Gap by Reducing Style Bias	O	O	O
CVPR 2021	Weakly-Supervised Physically Unconstrained Gaze Estimation		O	
CVPR 2021	Polygonal Point Set Tracking	O		
CVPR 2022	Interactive Multi-Class Tiny-Object Detection	O	O	O
CVPR 2022	Stereo Depth From Events Cameras: Concentrate and Focus on the Future		O	
ECCV 2022	PT4AL: Using Self-Supervised Pretext Tasks for Active Learning		O	
ECCV Workshop 2022	Variability Matters : Evaluating inter-rater variability in histopathology for robust cell detection	O	O	O
ICLR 2022	Sparse DETR: Efficient End-to-End Object Detection with Learnable Sparsity		O	
MICCAI 2022	Did You Get What You Paid For? Rethinking Annotation Cost of Deep Learning Based Computer Aided Detection in Chest Radiographs	O	O	O
MICCAI 2022	Intra-class Contrastive Learning Improves Computer Aided Diagnosis of Breast Cancer in Mammography	O	O	O
MICCAI Workshop 2022	OOOE: Only-One-Object-Exists Assumption to Find Very Small Objects in Chest Radiographs	O	O	O
MICCAI Workshop 2022	Efficient Neighbor Context-aware Breast Cancer Classification in Digital Breast Tomosynthesis using Transformers	O	O	O
WACV 2023	Improving Multi-Fidelity Optimization With a Recurring Learning Rate for Hyperparameter Tuning	O	O	O
CVPR 2023	OCELOT: Overlapped Cell on Tissue Dataset for Histopathology	O	O	O
CVPR 2023	Benchmarking Self-Supervised Learning on Diverse Pathology Datasets	O	O	O
ICCV 2023	Bayesian Optimization Meets Self-Distillation	O	O	O
MICCAI 2023	Enhancing Breast Cancer Risk Prediction by Incorporating Prior Images	O	O	O

Publications in Medicine (Selected)

Journal Name	Impact Factor	Publications
Journal of Clinical Oncology	50.7	[Park et al., 2022] [Park et al., 2023]
Lancet Digital Health	30.8	[Dembrower et al., 2020] [Kim et al., 2020]
JAMA Oncology	33.0	[Salim et al., 2020]
Radiology	29.2	[Park et al., 2018] [Hwang et al., 2019] [Jang et al., 2020] [Lee et al., 2020] [Nam et al., 2020] [Hwang et al., 2021] [Hong et al., 2022] [Nam et al., 2023] [Lee et al., 2023]
Clinical Cancer Research	13.8	[Jung et al., 2022]
JAMA Network Open	13.8	[Hwang et al., 2020] [Schaffter et al., 2020]
European Journal of Cancer	10.0	[Choi et al., 2022]

A Small, but Strong Team

Publications in CV/ML (Selected)

Conference Name	Title	First authored	Co-authored	Lunit's product
CVPR Workshop 2015	Multi-scale Pyramid Pooling for Deep Convolutional Representation	O	O	O
ICCV 2015	Attentionnet: Aggregating Weak Directions for Accurate Object Detection	O	O	O
ECCV 2016	Pixel-Level Domain Transfer	O	O	O
MICCAI 2016	Self-Transfer Learning for Fully Weakly Supervised Object Localization	O	O	O
ICLR Workshop 2017	Transferring Knowledge To Smaller Network With Class-Distance Loss	O	O	O
MICCAI Workshop 2017	A Unified Framework For Tumor Proliferation Score Prediction In Breast Histopathology	O	O	O
MICCAI Workshop 2017	Accurate Lung Segmentation Via Network-Wise Training Of Convolutional Networks	O	O	O
ICCV 2017	Two-Phase Learning for Weakly Supervised Object Localization		O	
NeurIPS 2018	Batch-Instance Normalization for Adaptively Style-Invariant Neural Networks	O	O	O
MICCAI 2018	Keep and Learn: Continual Learning by Constraining the Latent Space for Knowledge Preservation in Neural Networks	O	O	O
MICCAI 2018	A Robust and Effective Approach Towards Accurate Metastasis Detection and Pn-Stage Classification in Breast Cancer	O	O	O
CVPR 2018	Distort-And-Recover: Color Enhancement Using Deep Reinforcement Learning	O	O	
BMVC 2018	Bam: Bottleneck Attention Module	O		
ECCV 2018	Convolutional Block Attention Module	O		
CVPR 2019	Learning Loss for Active Learning	O		O
MICCAI 2019	PseudoEdgeNet: Nuclei Segmentation only with PointAnnotations	O	O	O
ICCV Workshop 2019	Photometric Transformer Networks and Label Adjustment for Breast Density Prediction	O	O	O
ICCV 2019	SRM: A Style-based Recalibration Module for Convolutional Neural Networks	O	O	O
ECCV 2020	Learning Visual Context by Comparison	O	O	O
ICLR 2022	Sparse DETR: Efficient End-to-End Object Detection with Learnable Sparsity		O	
MICCAI 2022	Did You Get What You Paid For? Rethinking Annotation Cost of Deep Learning Based Computer Aided Detection in Chest Radiographs	O	O	O
MICCAI 2022	Intra-class Contrastive Learning Improves Computer Aided Diagnosis of Breast Cancer in Mammography	O	O	O
MICCAI Workshop 2022	OOOE: Only-One-Object-Exists Assumption to Find Very Small Objects in Chest Radiographs	O	O	O
MICCAI Workshop 2022	Efficient Neighbor Context-aware Breast Cancer Classification in Digital Breast Tomosynthesis using Transformers	O	O	O
WACV 2023	Improving Multi-Fidelity Optimization With a Recurring Learning Rate for Hyperparameter Tuning	O	O	O
CVPR 2023	OCELOT: Overlapped Cell on Tissue Dataset for Histopathology	O	O	O
CVPR 2023	Benchmarking Self-Supervised Learning on Diverse Pathology Datasets	O	O	O
ICCV 2023	Bayesian Optimization Meets Self-Distillation	O	O	O
MICCAI 2023	Enhancing Breast Cancer Risk Prediction by Incorporating Prior Images	O	O	O

Conference Name	Title	First authored	Co-authored	Lunit's product
CVPR 2019	Learning Loss for Active Learning	O		O

Publications in Medicine (Selected)

Journal Name	Impact Factor	Publications
Journal of Clinical Oncology	50.7	[Park et al., 2022] [Park et al., 2023]
Lancet Digital Health	30.8	[Dembrower et al., 2020] [Kim et al., 2020]
JAMA Oncology	33.0	[Salim et al., 2020]
Radiology	29.2	[Park et al., 2018] [Hwang et al., 2019] [Jang et al., 2020] [Lee et al., 2020] [Nam et al., 2020] [Hwang et al., 2021] [Hong et al., 2022] [Nam et al., 2023] [Lee et al., 2023]
Clinical Cancer Research	13.8	[Jung et al., 2022]
JAMA Network Open	13.8	[Hwang et al., 2020] [Schaffter et al., 2020]
European Journal of Cancer	10.0	[Choi et al., 2022]

Overview

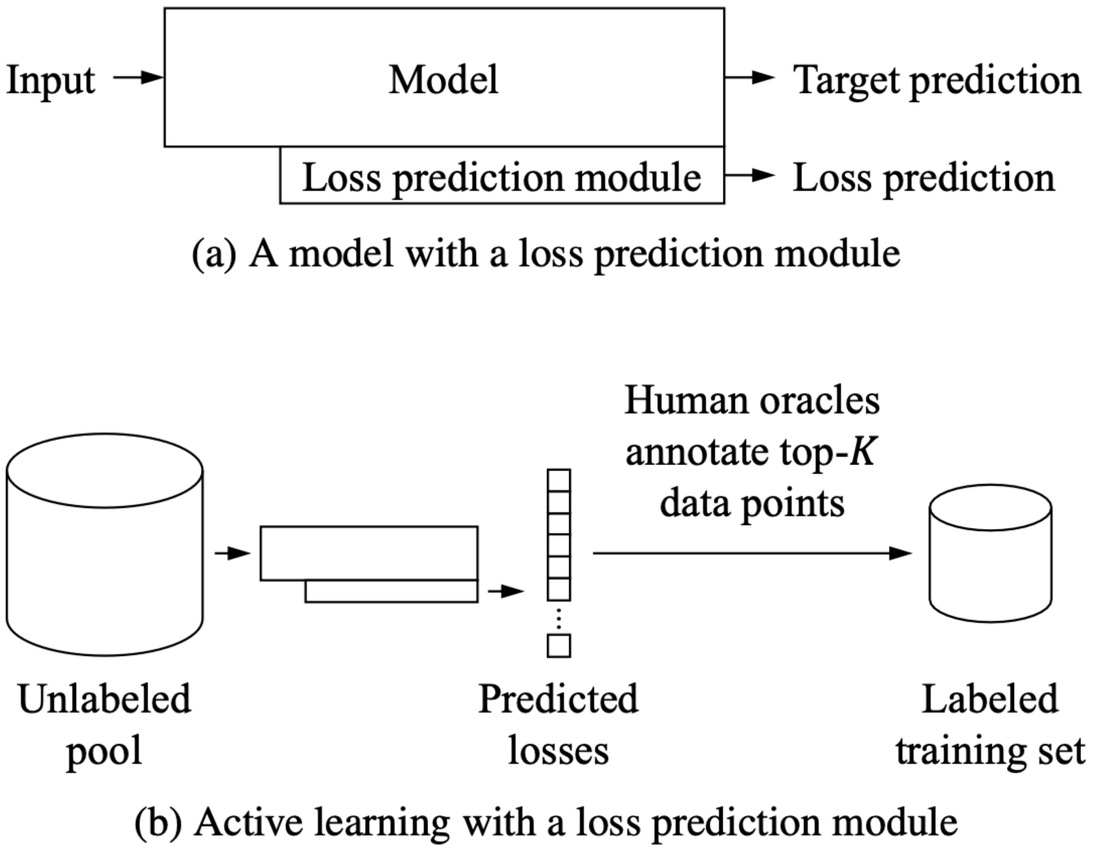
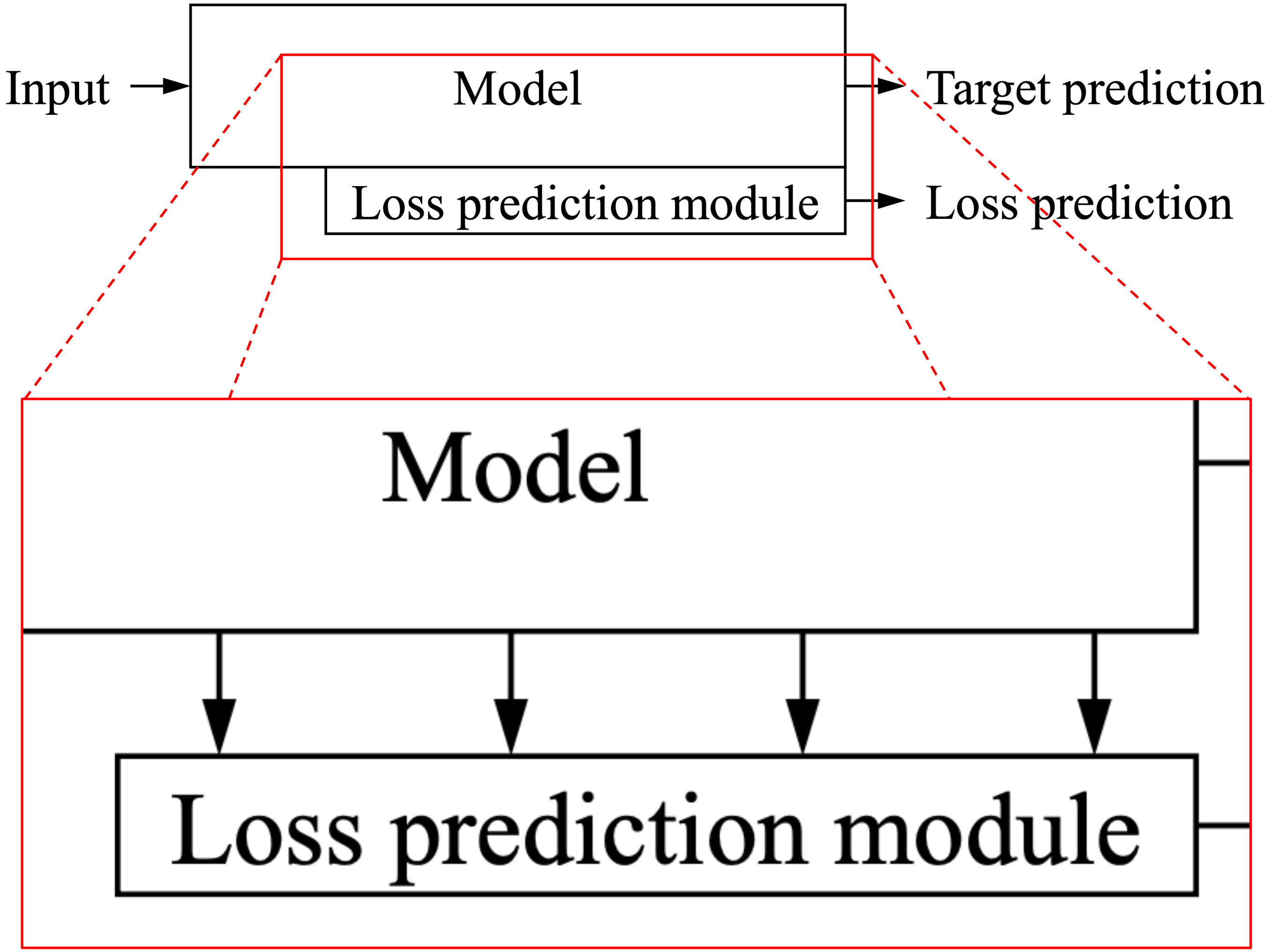
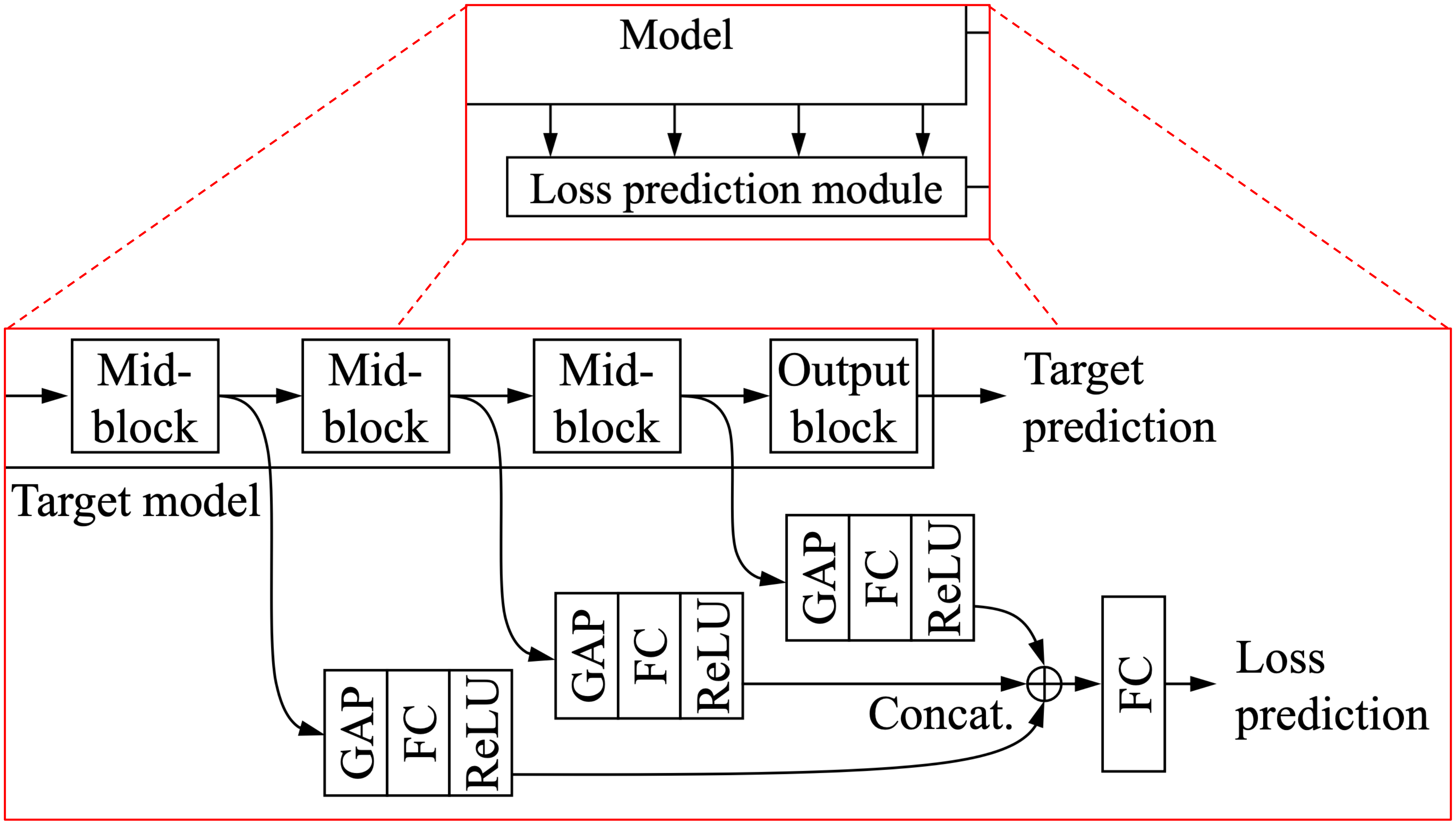


Figure 1. A novel active learning method with a loss prediction module. (a) A loss prediction module attached to a target model predicts the loss value from an input without its label. (b) All data points in an unlabeled pool are evaluated by the loss prediction module. The data points with the top-K predicted losses are labeled and added to a labeled training set.

Loss Prediction Module

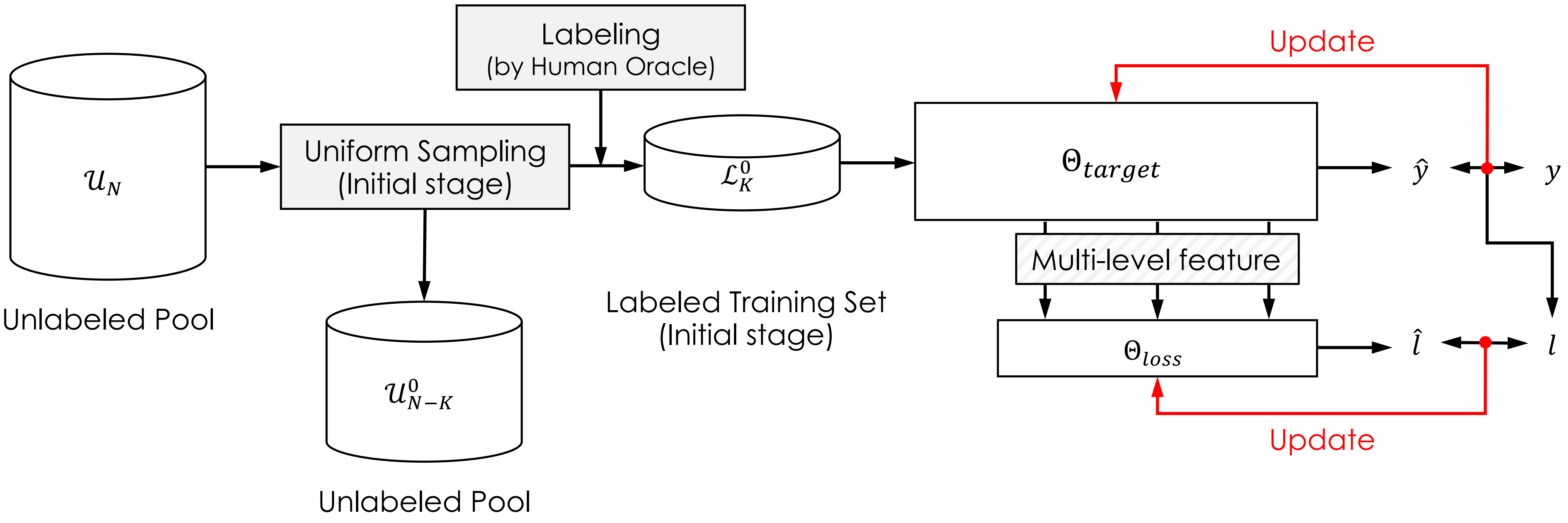


Loss Prediction Module

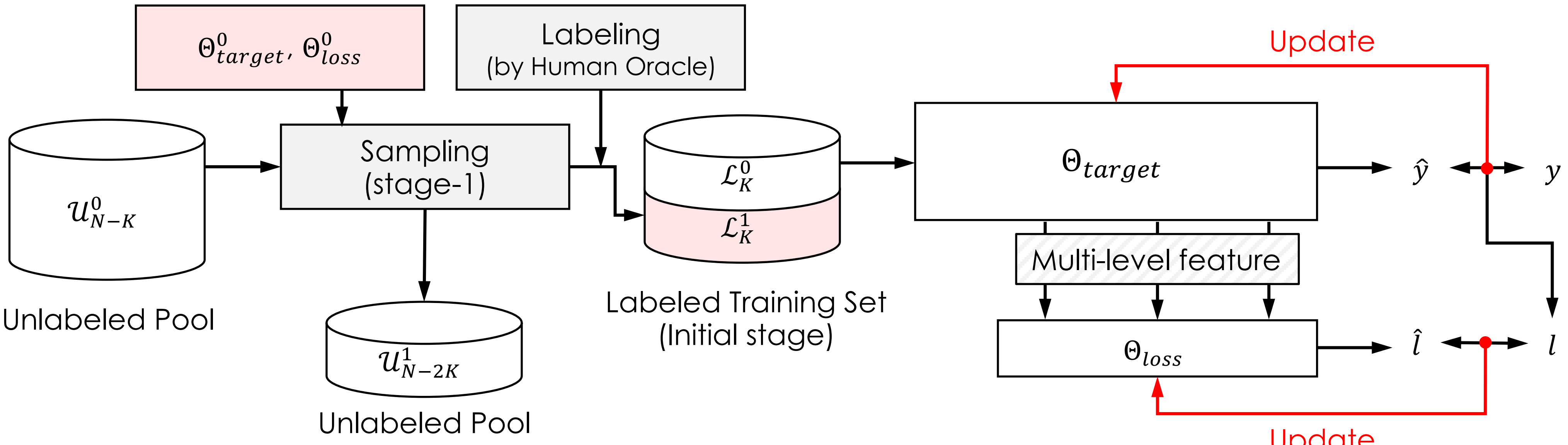


Aggregation of fixed equal dimensional features

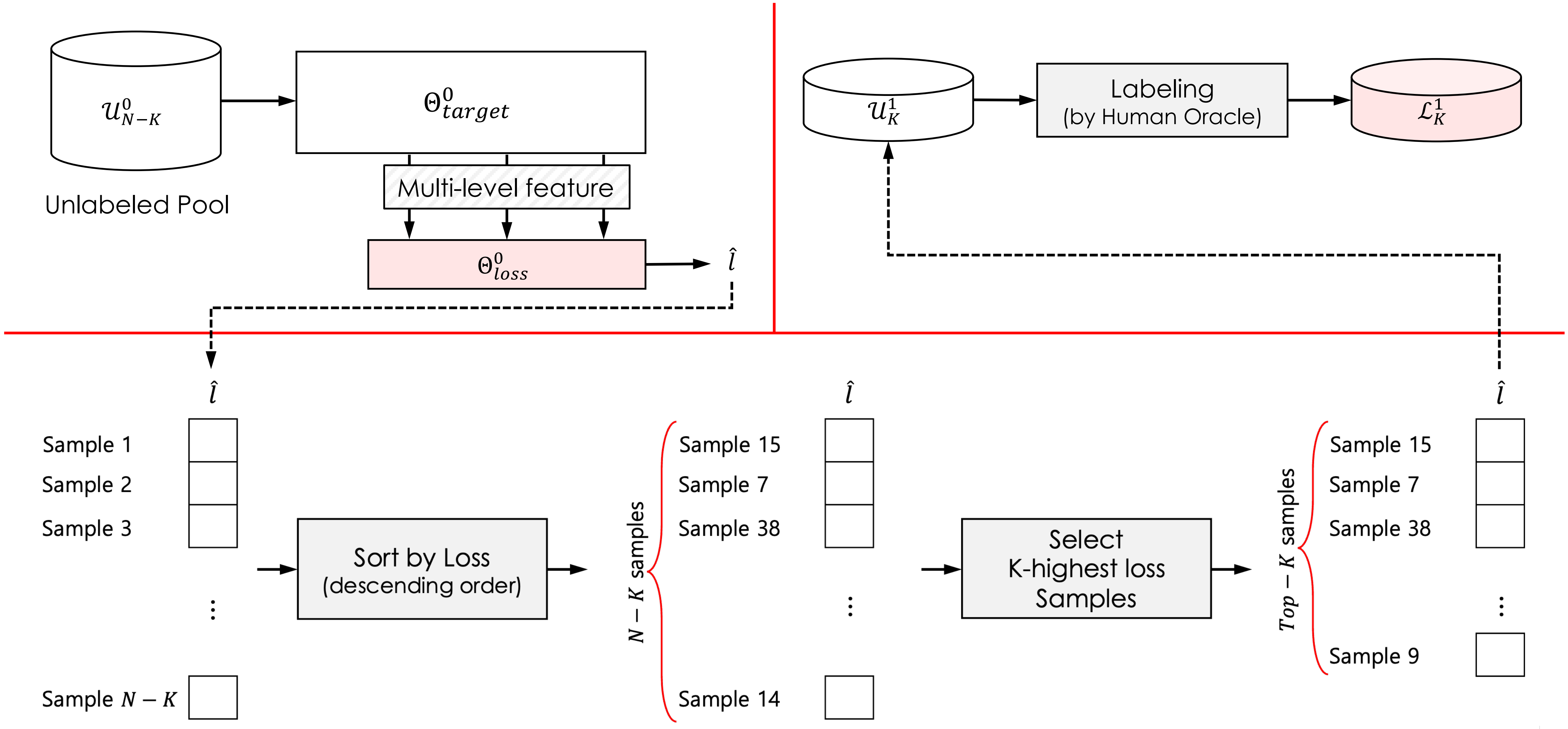
Training – Initial Stage



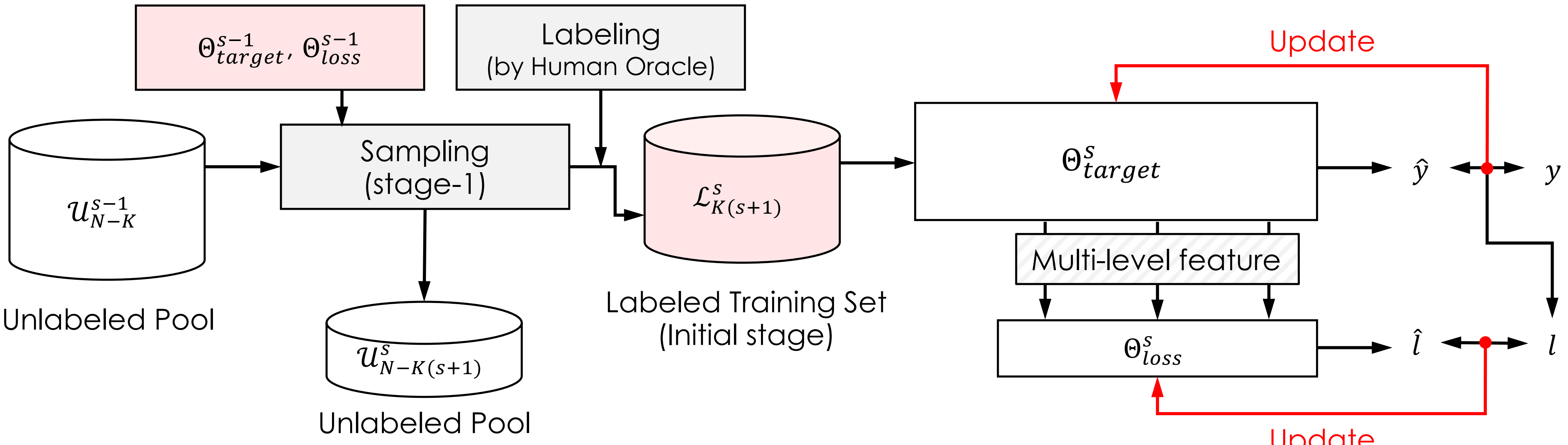
Training – Stage-1



Training – Sampling



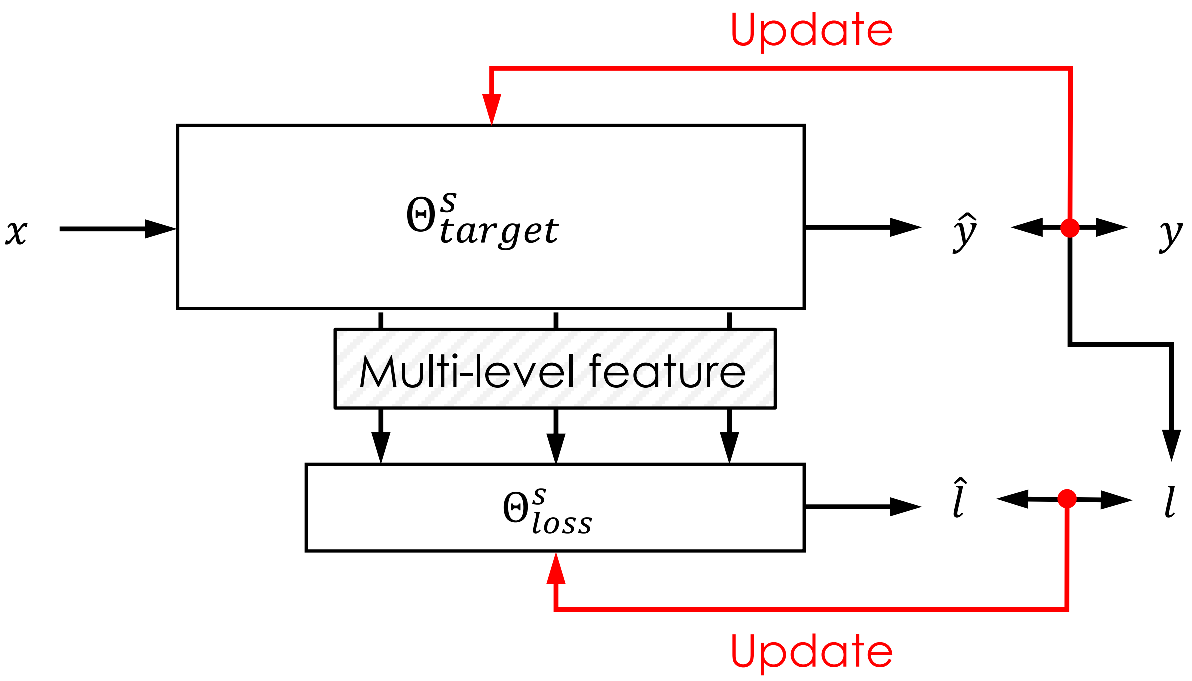
Training – Generalized Form



Effective Trick for Learning Loss

(Only for Training)

Model Update



$$L_{target}(\hat{y}, y) + \lambda L_{loss}(\hat{l}, l)$$

When using MSE as a L_{target},
What if the L_{loss} also uses MSE?

Limitation of classical L_{loss}

When using MSE as a L_{target}, what if the L_{loss} also uses MSE?

The loss prediction module, Θ_{loss}^{s}, will be highly affected by the initial training set

- L_{loss} will be high in initial (strong update)
- L_{loss} decreases through the stages (weak update)
- Not proper to learn new data
- Negative effect on selecting next data points

Ranking Loss

$$L_{loss}(\hat{l}^p, l^p) = \max(0, -\mathbb{I}(l_i, l_j) \cdot (\hat{l}_i - \hat{l}_j) + \zeta)$$

$$s.t.\ \mathbb{I}(l_i, l_j) = \begin{cases} +1, if\ l_i > l_j \\ -1, if\ l_i \leq l_j \end{cases}$$

a.k.a. marginal ranking loss

Case of $\boldsymbol{l_i > l_j}$ $\rightarrow L_{loss}(\hat{l}^p, l^p) = \max(0, -\hat{l}_i + \hat{l}_j + \zeta)$

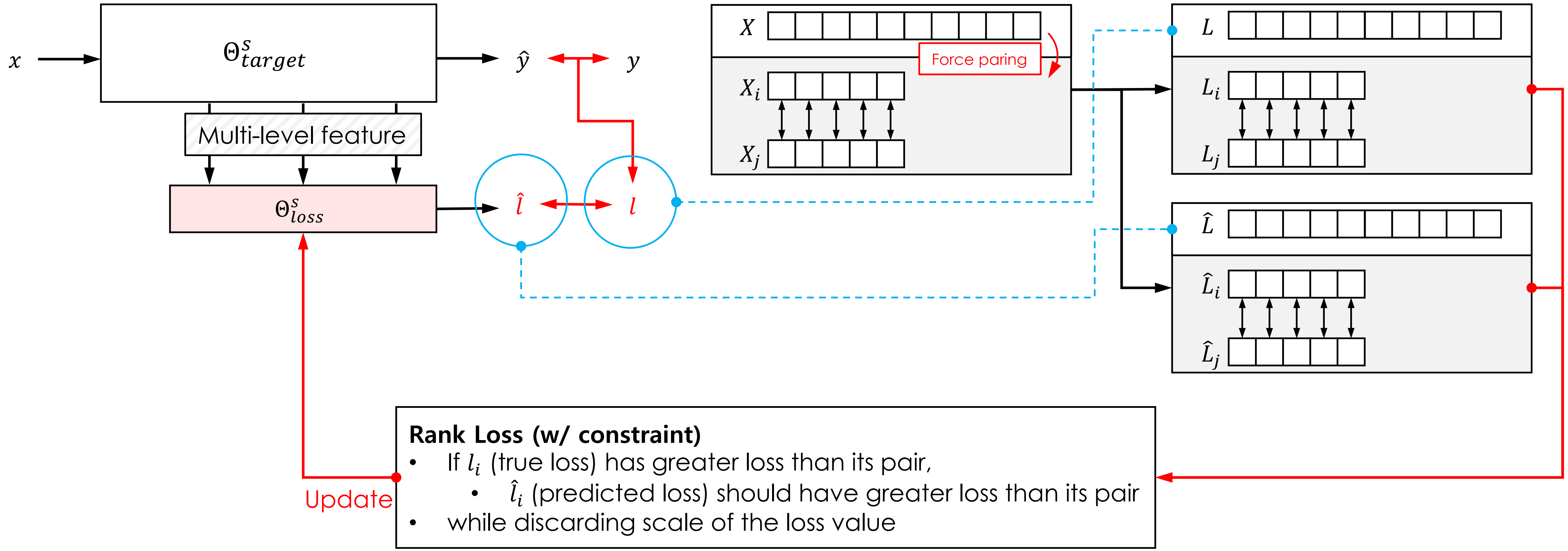
Loss is given when $\underline{abs(\hat{l}_i) < \hat{l}_j + \zeta}$ to reduce the gap between $\hat{l}_i$ and $\hat{l}_j$

Miss prediction of rank

Case of $\boldsymbol{l_i \leq l_j}$ $\rightarrow L_{loss}(\hat{l}^p, l^p) = \max(0, \hat{l}_i - \hat{l}_j + \zeta)$

Loss is given when $\underline{abs(\hat{l}_j) < \hat{l}_i + \zeta}$ to reduce the gap between $\hat{l}_i$ and $\hat{l}_j$

Rank Loss for Loss Prediction Model



Experimental Results

- Solid line: average of 5 trials
- Dashed line: minimum or maximum of trials

Image Classification

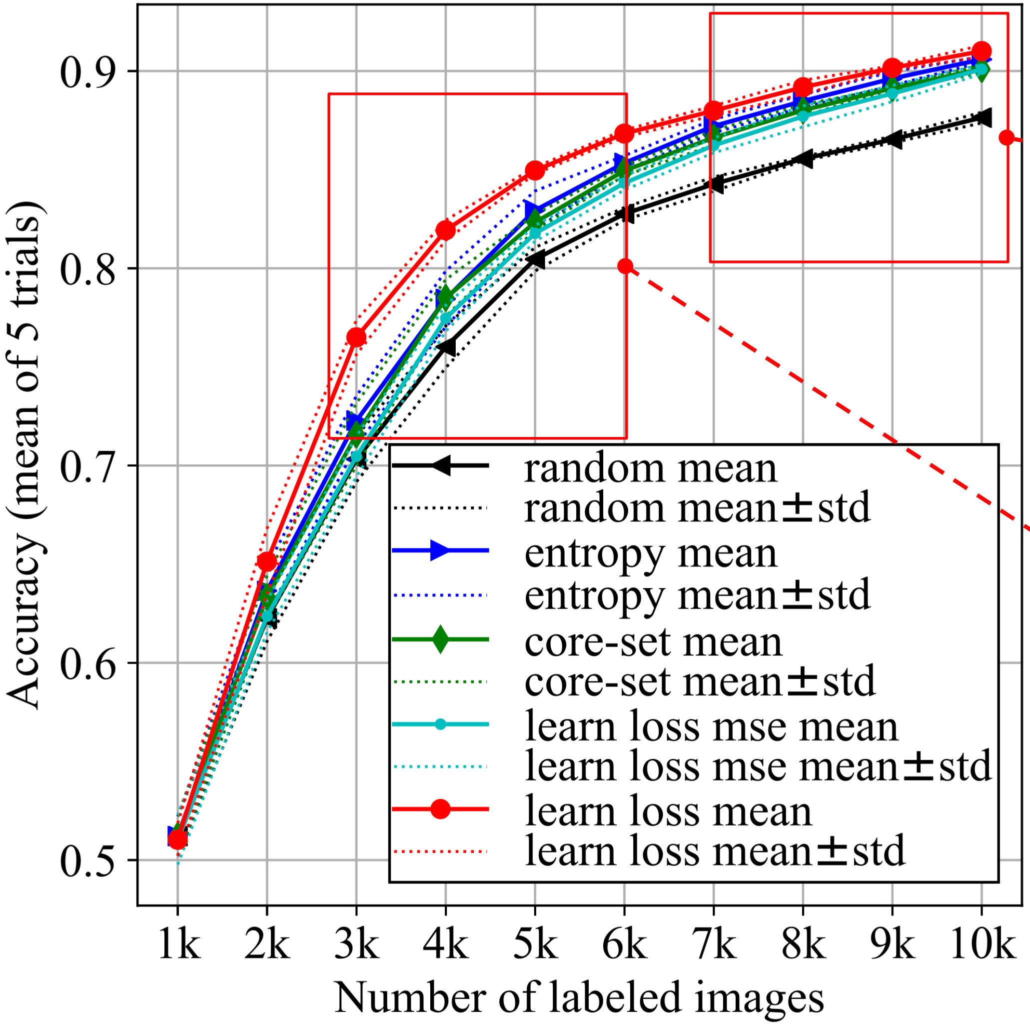
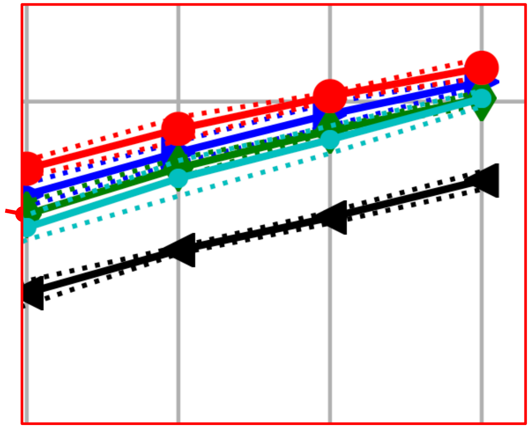
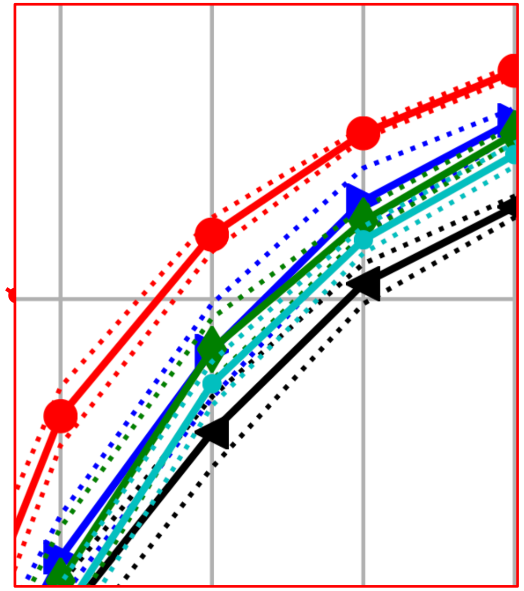


Figure 4. Active learning results of image classification over CIFAR-10.



Slightly better performance on large training set



Better performance on small training set

- Solid line: average of 3 trials
- Dashed line: minimum or maximum of trials

Object Detection & Human Pose Estimation

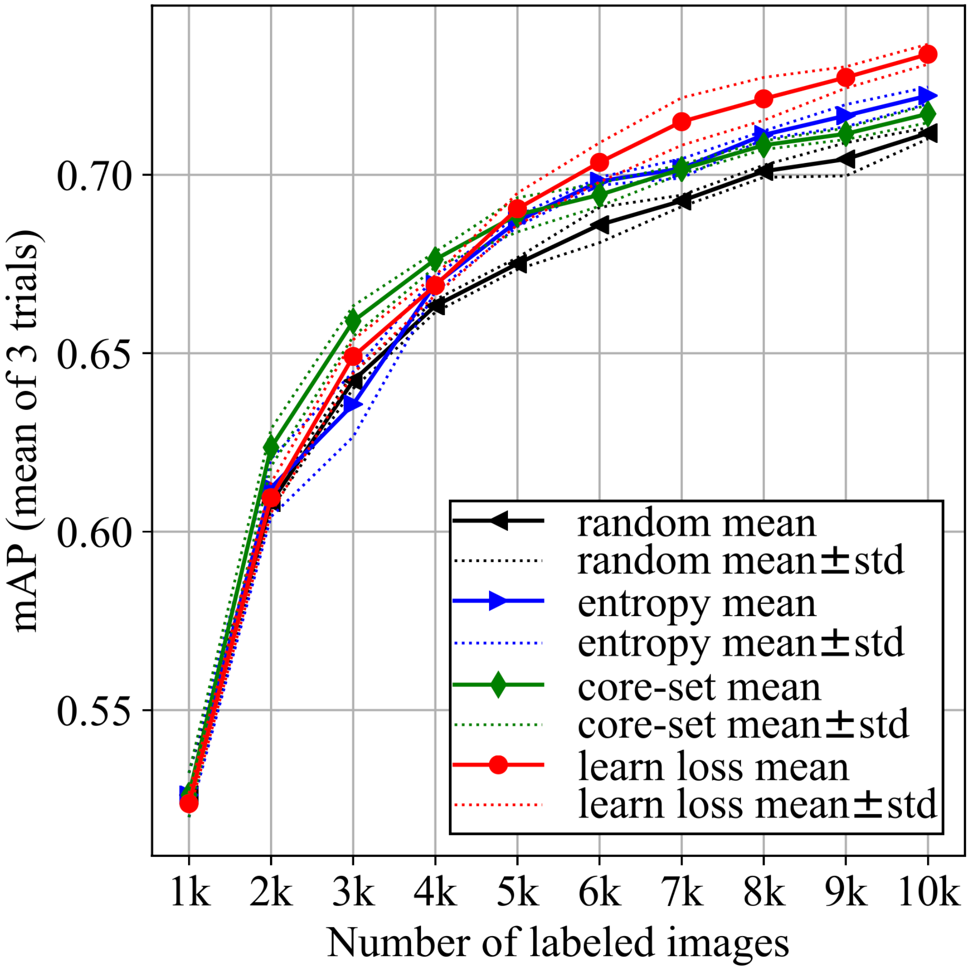


Figure 6. Active learning results of object detection over PASCAL VOC 2007+2012.

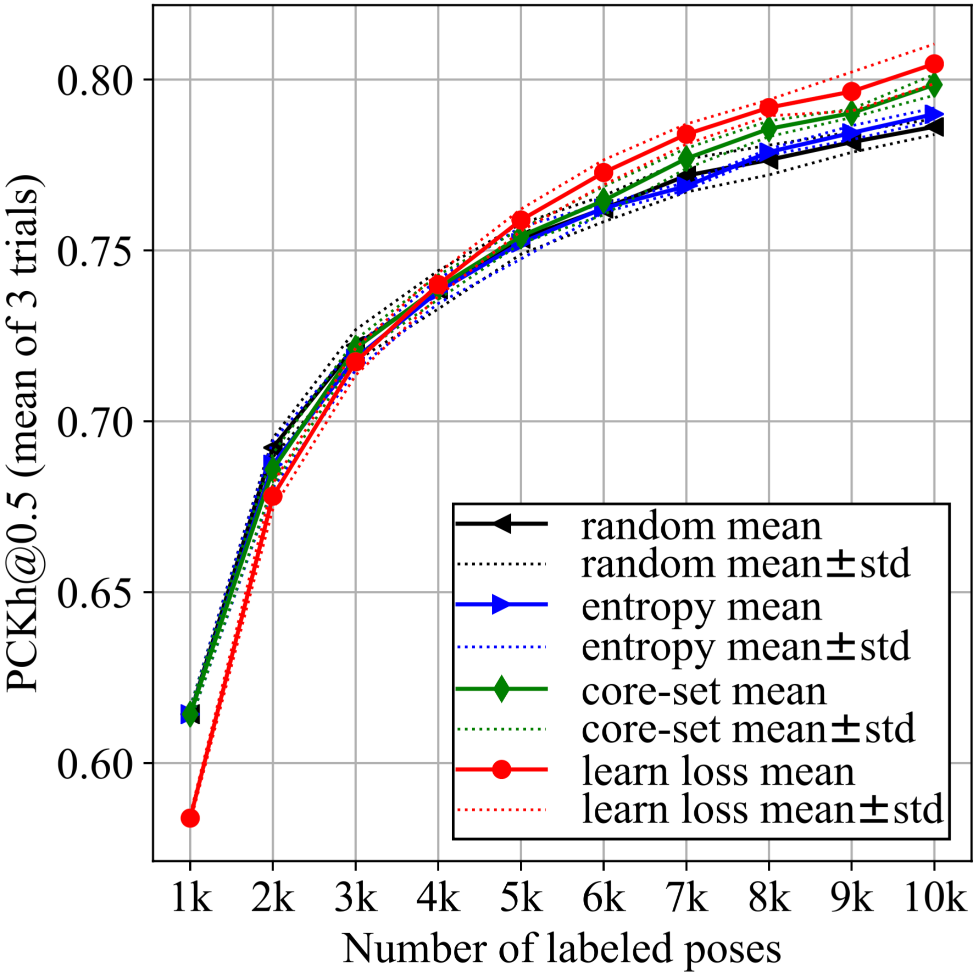


Figure 7. Active learning results of human pose estimation over MPII.

Loss Prediction Accuracy

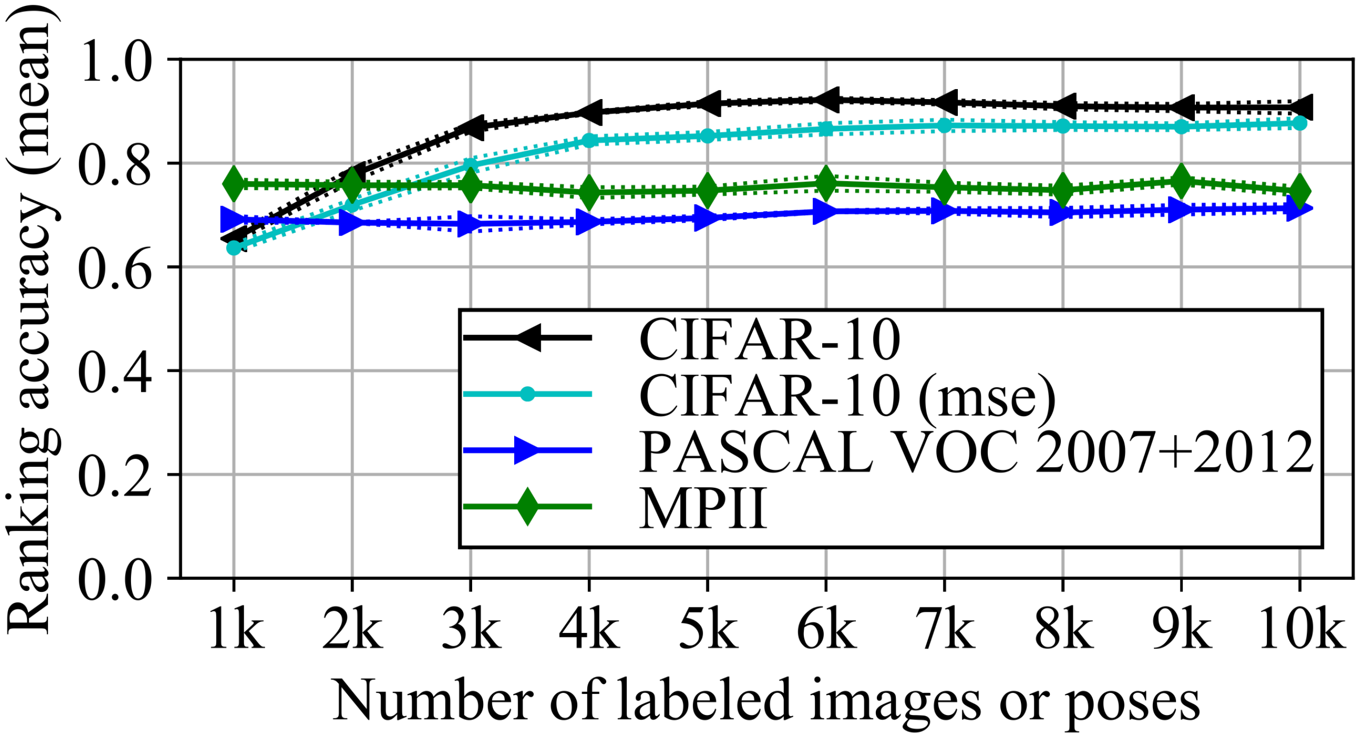


Figure 5. Loss-prediction accuracy of the loss prediction module.

Summaries

Motivation and solution

- Motivation: To reduce labeling cost
  - Cost of labeling: classification (image-level) < object detection (object-level) < segmentation (pixel-level)
  - Most of previous method requires task-specific design / not efficient to deep learning

- Solution: Selecting target data points (samples) by task-agnostic simple loss prediction module

Contributions

- Task-agnostic simple loss prediction module
  - Employed to select target data points (predicted to high loss → most informative data)
  - Directly applicable to any tasks with deep neural networks

- Active learning with a loss prediction module consistently outperforms in various tasks
  - Image classification
  - Object detection
  - Human pose estimation

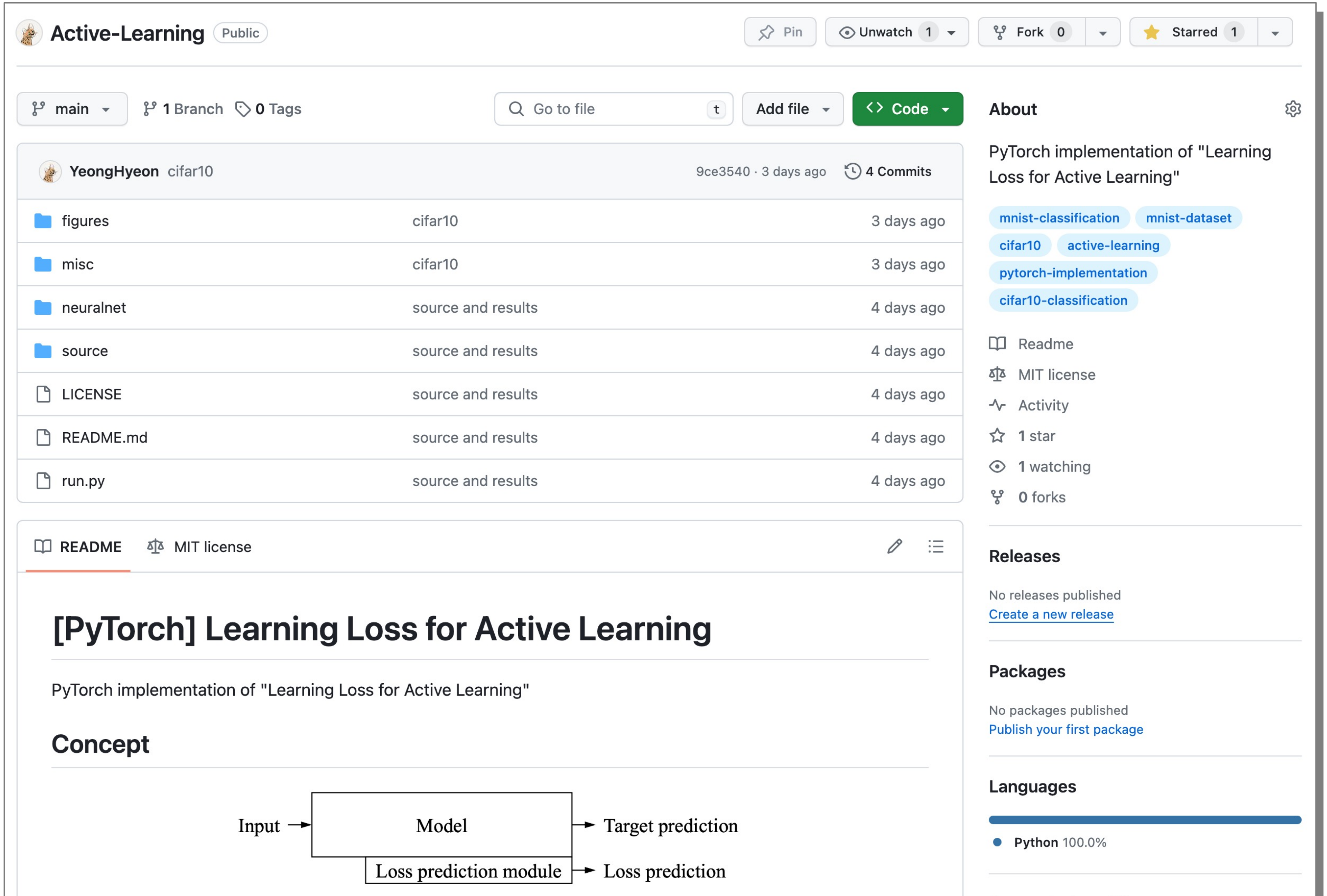
Limitation

- Loss prediction accuracy was relatively low in complex tasks
  - About 0.9 for classification but 0.7 for object detection/human pose estimation

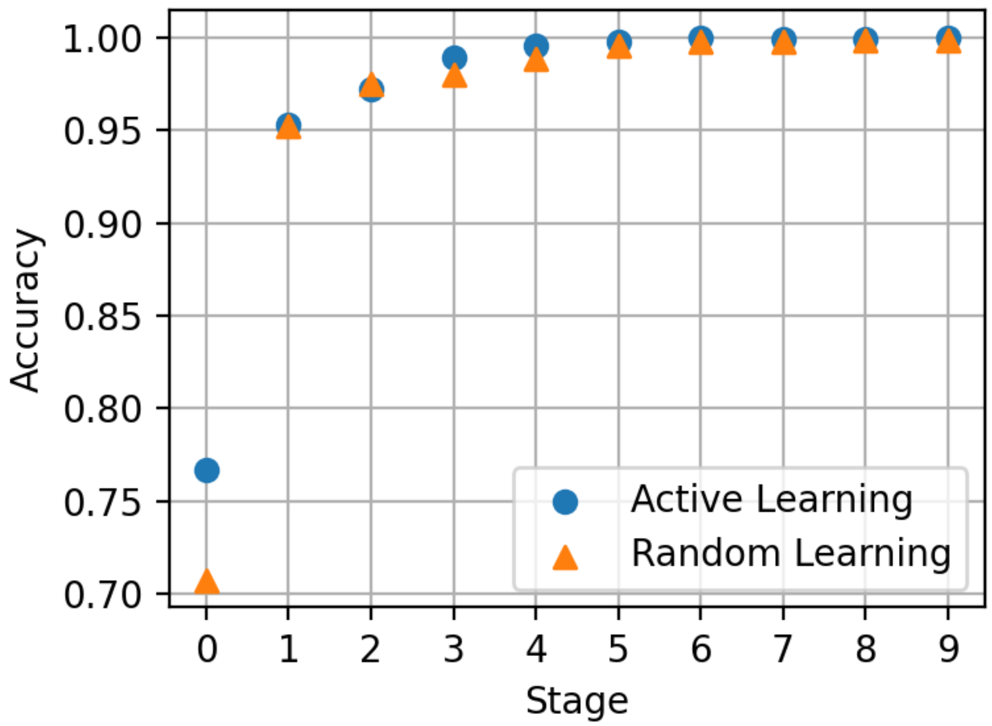
Appendix

https://github.com/YeongHyeon/Active-Learning

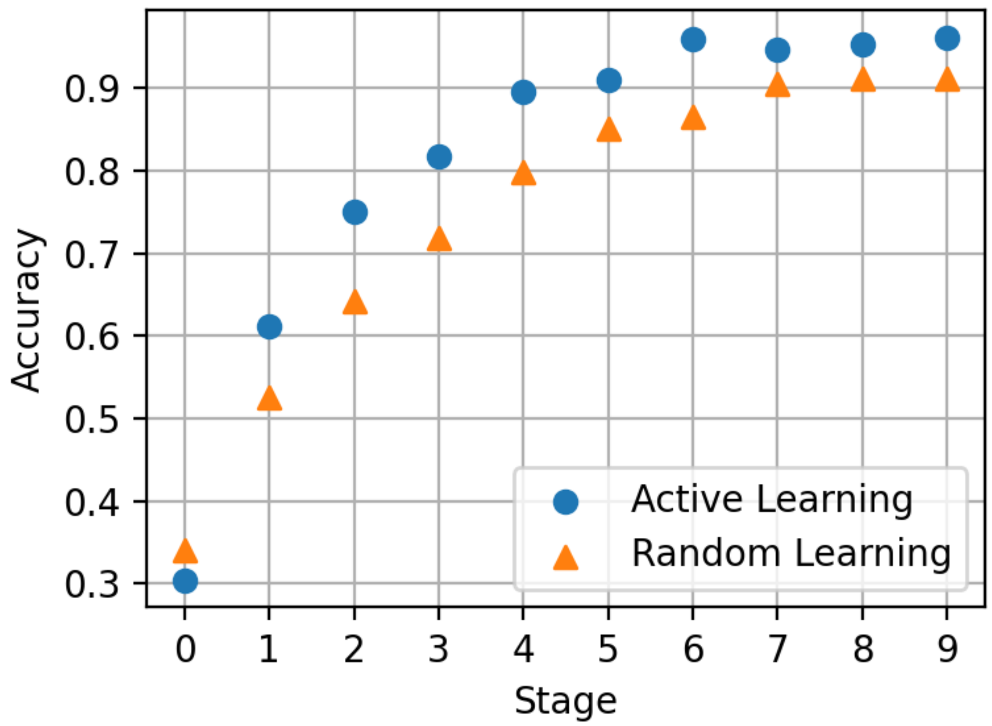
Reproduction



Validation



Classification Accuracy
(w/ MNIST dataset)



Classification Accuracy
(w/ CIFAR10 dataset)