Active Learning

Junho Park

Vision and Display Systems Lab.

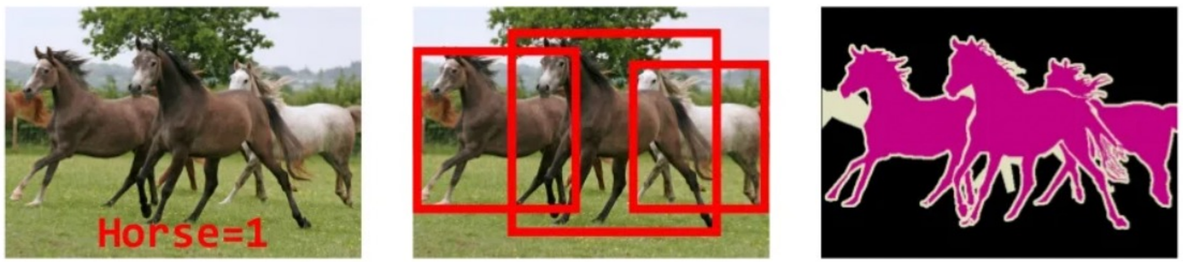
Sogang University

Outline

1. Motivation
2. About Active Learning
3. Recent Work
4. For Hand Pose Estimation
5. Conclusion
6. References

Motivation

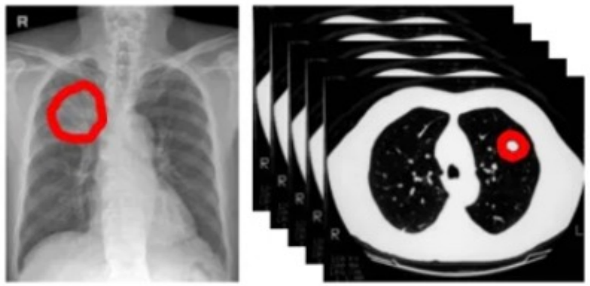
- Success of Supervised Learning
  - A large number of learnable parameters
  - Availability for large-scale annotated data
- Limitations
  - Annotating requires a lot of effort
  - Expensive
  - Time-consuming



$ $$ $$$

Motivation

- Success of Supervised Learning
  - A large number of learnable parameters
  - Availability for large-scale annotated data
- Limitations
  - Annotating requires a lot of effort
  - Expensive
  - Time-consuming



$$$$$

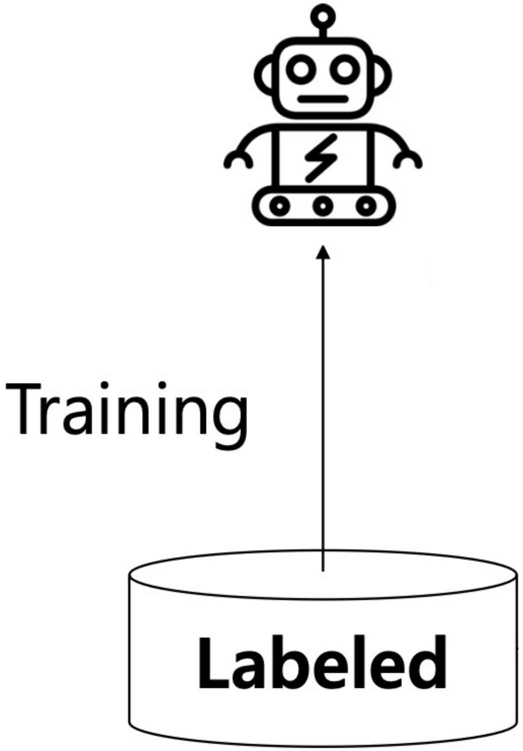
Motivation

- Success of Supervised Learning
  - A large number of learnable parameters
  - Availability for large-scale annotated data
- Limitations
  - Annotating requires a lot of effort
  - Expensive
  - Time-consuming
- Solution
  - Identify the most influencing and discriminative examples to annotate.
  - Efficiently select the most meaningful samples.

➔ ***Active Learning is ABSOLUTELY NEEDED!***

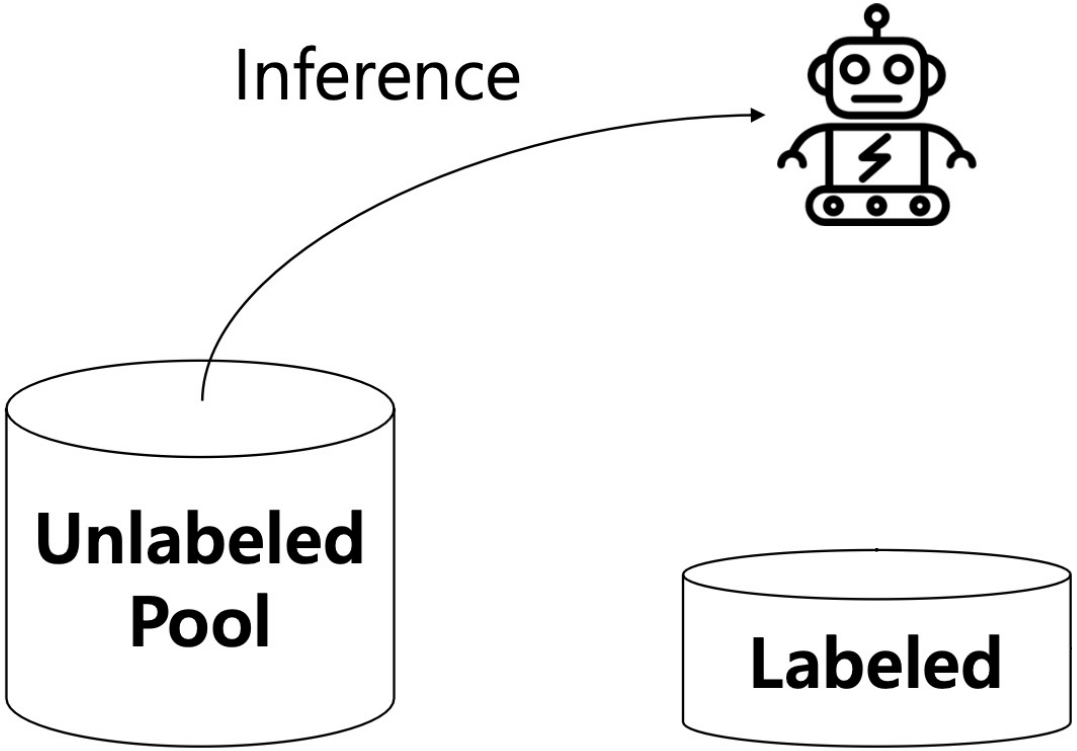
About Active Learning

- Concept



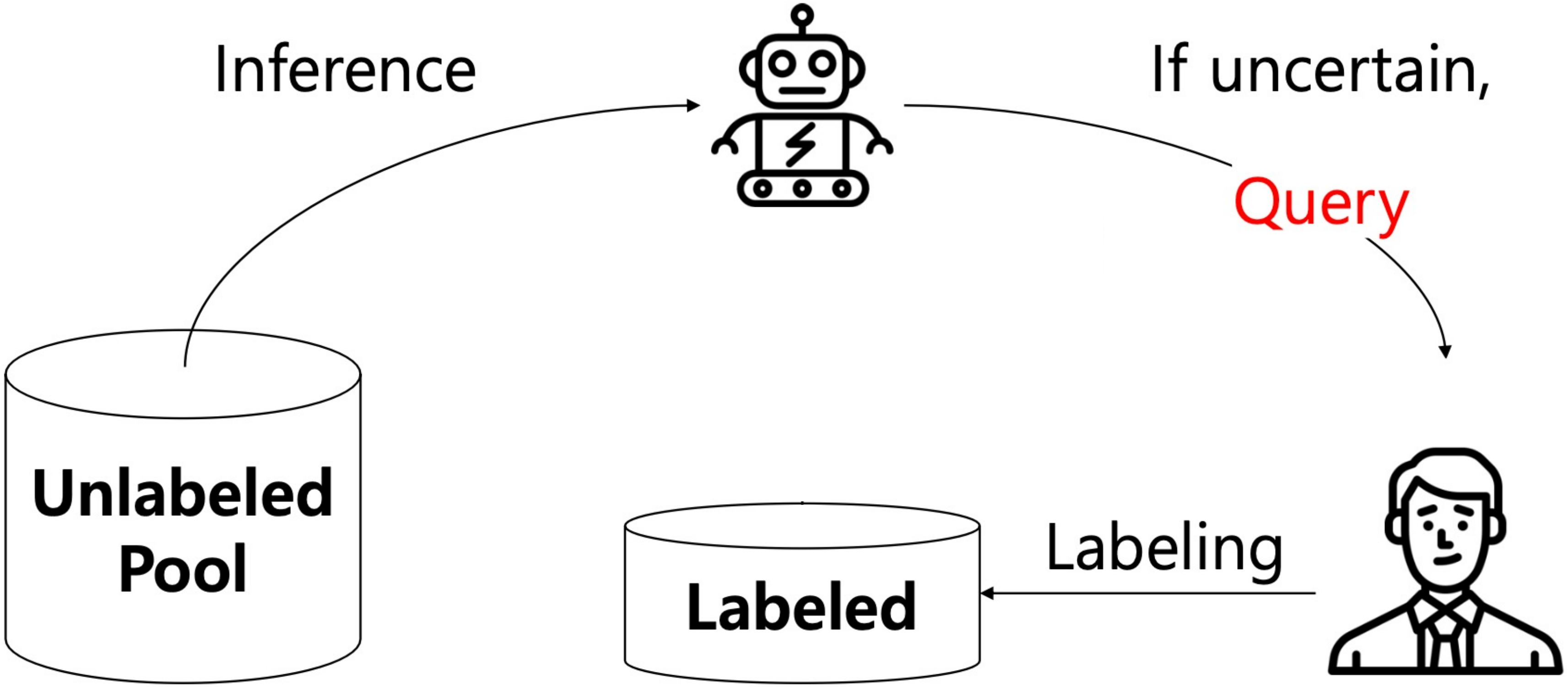
About Active Learning

- Concept



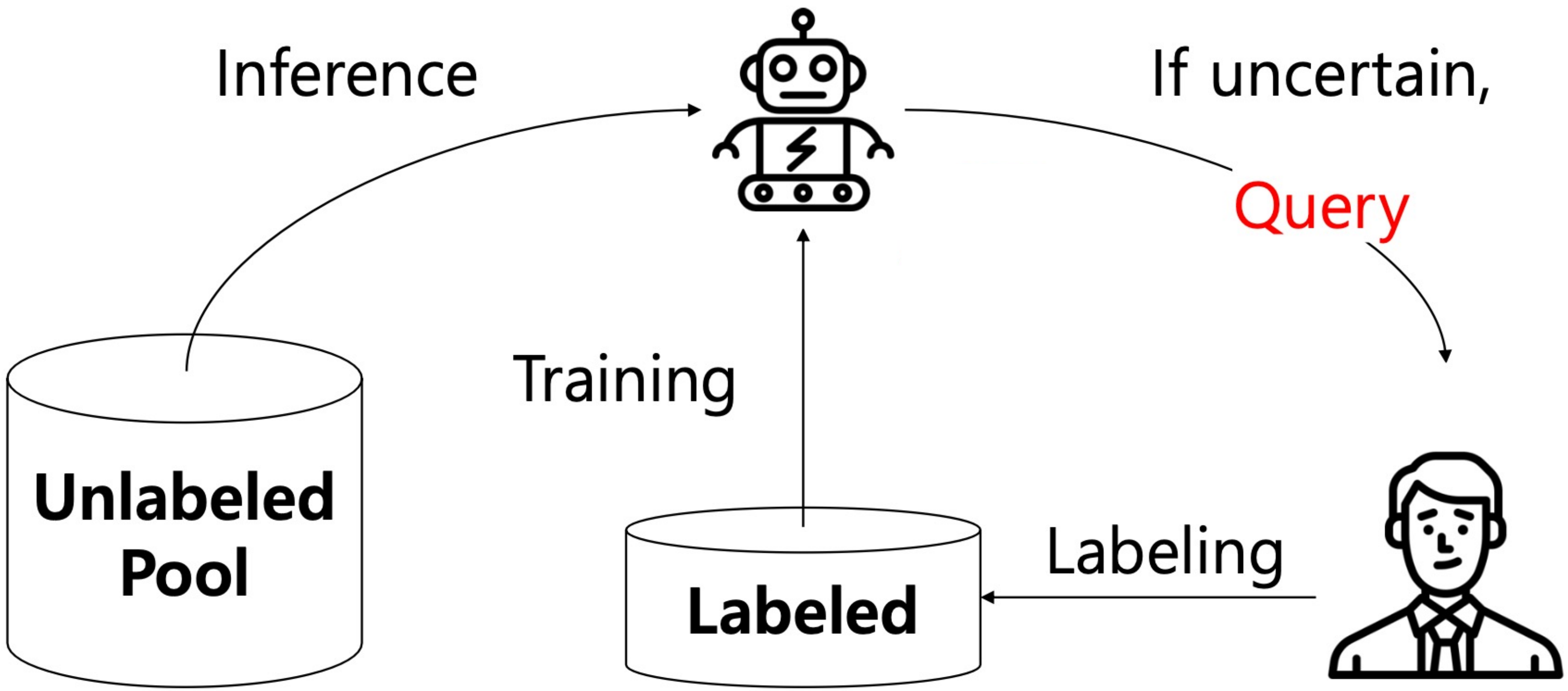
About Active Learning

- Concept



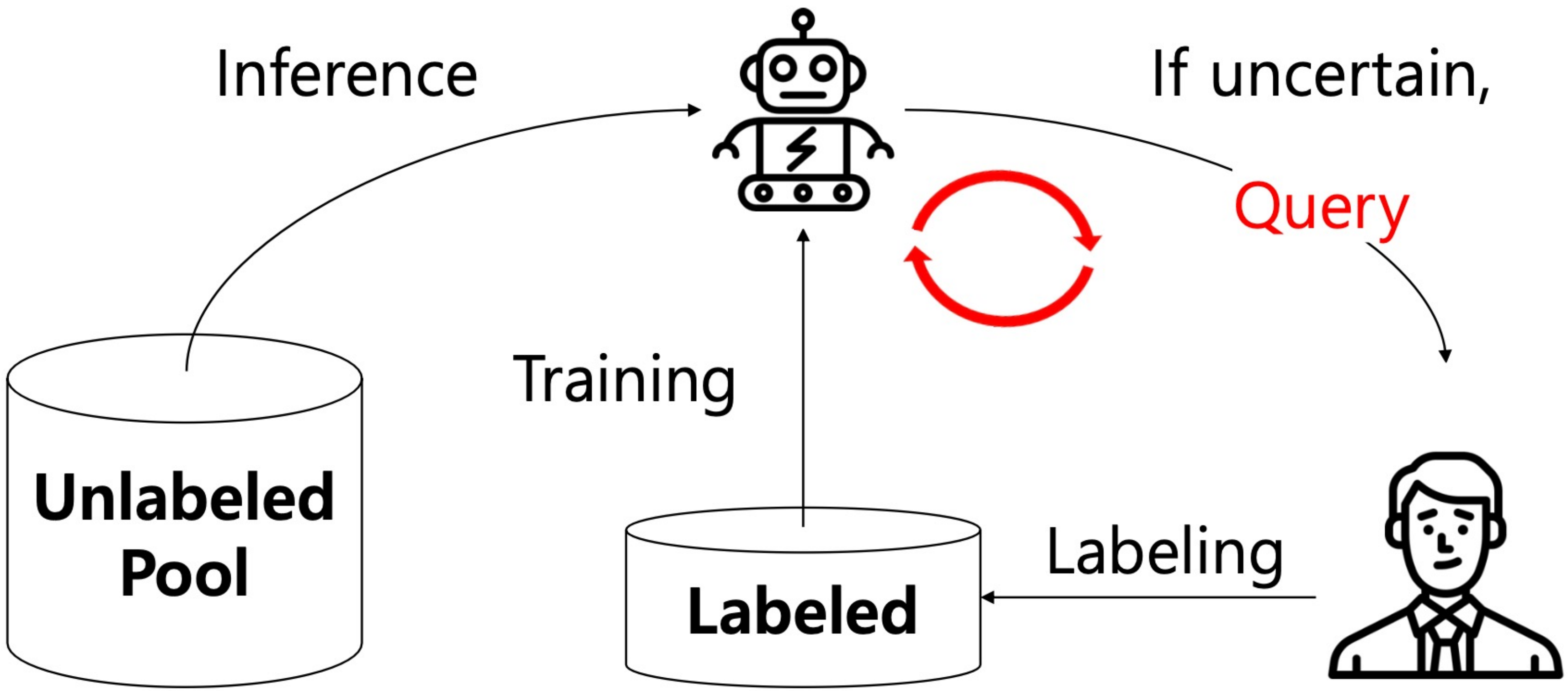
About Active Learning

- Concept



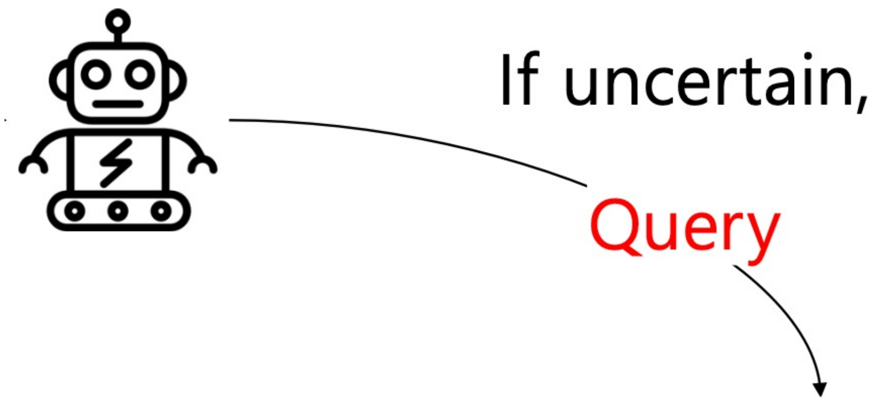
About Active Learning

- Concept



About Active Learning

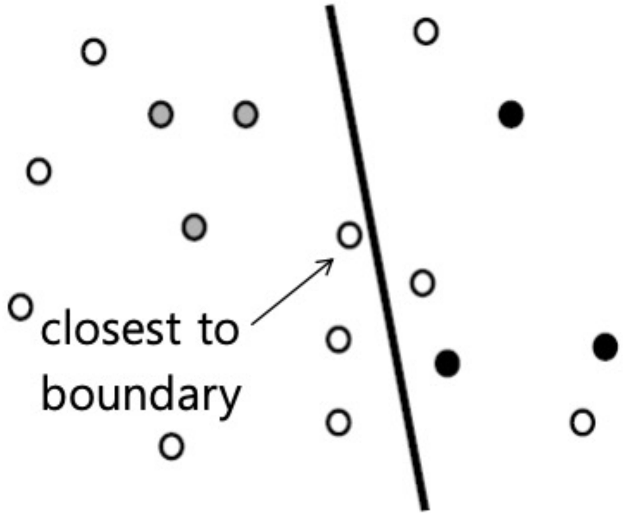
- Concept



The key of active learning:
How to measure the UNCERTAINTY

Uncertainty Sampling

- Goal
  - Choosing the sample closest to the decision boundary.
- Methods
  - Least Confidence
  - Margin Sampling
  - Maximum Entropy



Uncertainty Sampling

- Least Confidence
  - Choose the example where the top label had the smallest probability.

Data ID	Class A	Class B	Class C	Top 1	Query Priority
D1	0.1	0.2	0.7	0.7	4
D2	0.33	0.33	0.34	0.34	1
D3	0.41	0.39	0.2	0.41	3
D4	0.3	0.4	0.3	0.4	2

Uncertainty Sampling

- Margin Sampling
  - The sample with the smallest difference between the top-1 confidence and the top-2 confidence is selected.

Data ID	Class A	Class B	Class C	Top 1 – Top 2	Query Priority
D1	0.1	0.2	0.7	0.5	4
D2	0.33	0.33	0.34	0.01	1
D3	0.41	0.39	0.2	0.02	2
D4	0.3	0.4	0.3	0.1	3

Uncertainty Sampling

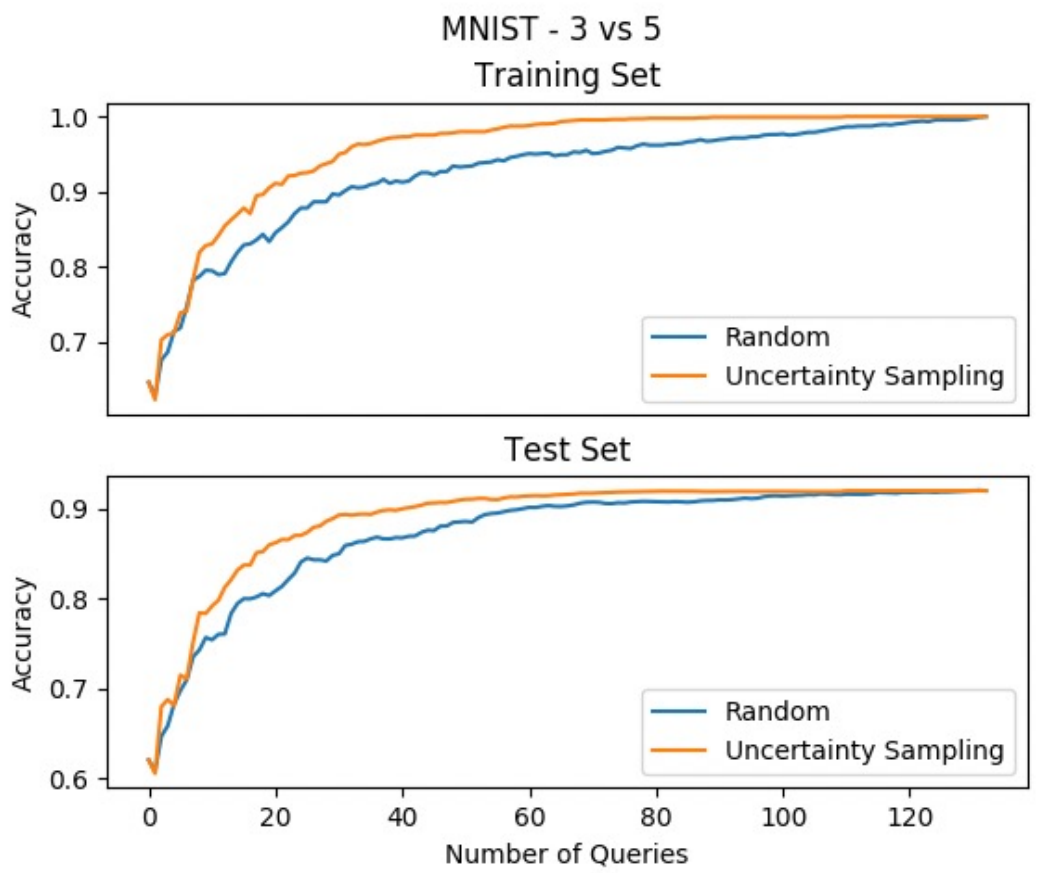
- Maximum Entropy
  - Choose the example which has the highest entropy.
  - Decision Rule:

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log \mathrm{P}(x_i)$$

Data ID	Class A	Class B	Class C	Entropy	Query Priority
D1	0.1	0.2	0.7	0.8018	4
D2	0.33	0.33	0.34	1.0985	1
D3	0.41	0.39	0.2	1.0546	3
D4	0.3	0.4	0.3	1.0888	2

Uncertainty Sampling

- Compared with Random Sampling



Uncertainty Sampling

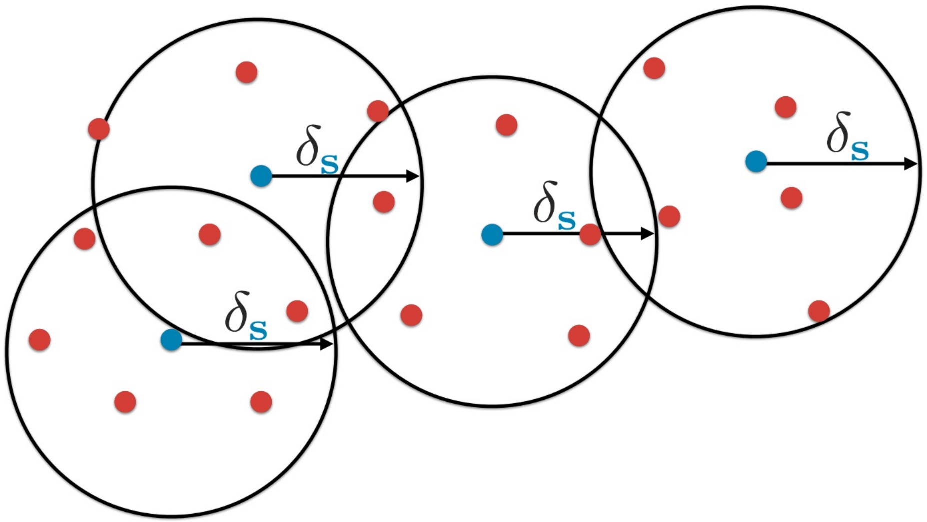
- Advantage
  - Easy to implement
  - High performance
- Disadvantage
  - Suffer from outlier
  - Ignored diversity

Recent Work

- CoreSet
  - Active Learning for Convolutional Neural Networks: A Core-set approach
  - Accepted at ICLR 2018
- Learning Loss
  - Learning Loss for Active Learning
  - Accepted as Oral at CVPR 2019

CoreSet

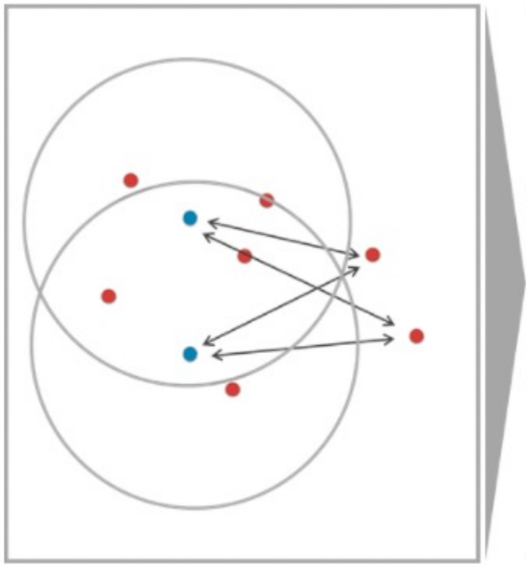
- Concept
  - Using feature in Deep Learning
  - Batch Active Learning
  - Find distinct data points from pre-selected subsets



CoreSet

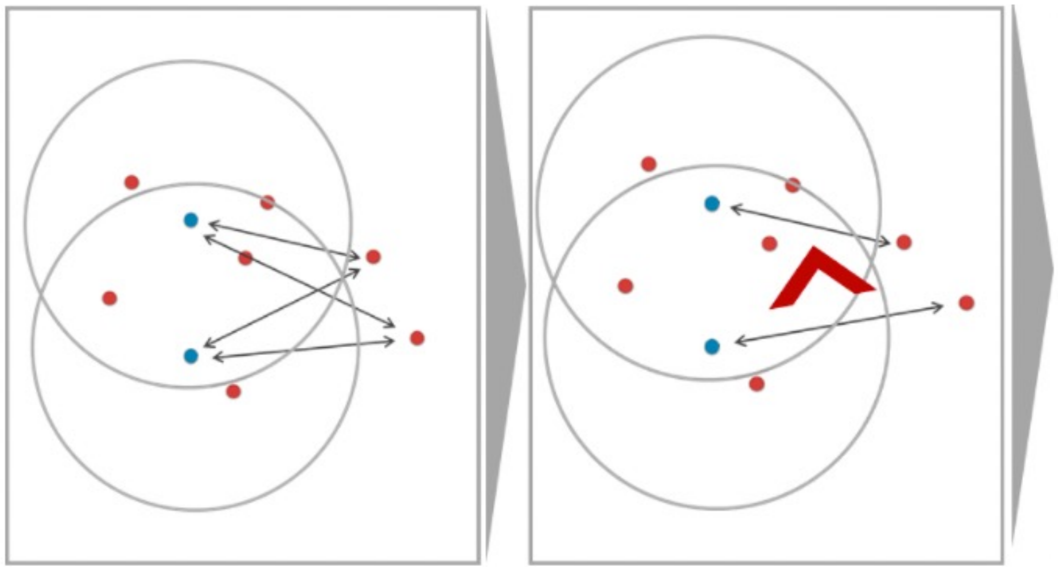
- k-Center-Greedy Algorithm

  1. Find the shortest distance between the data point and the center of the circle.



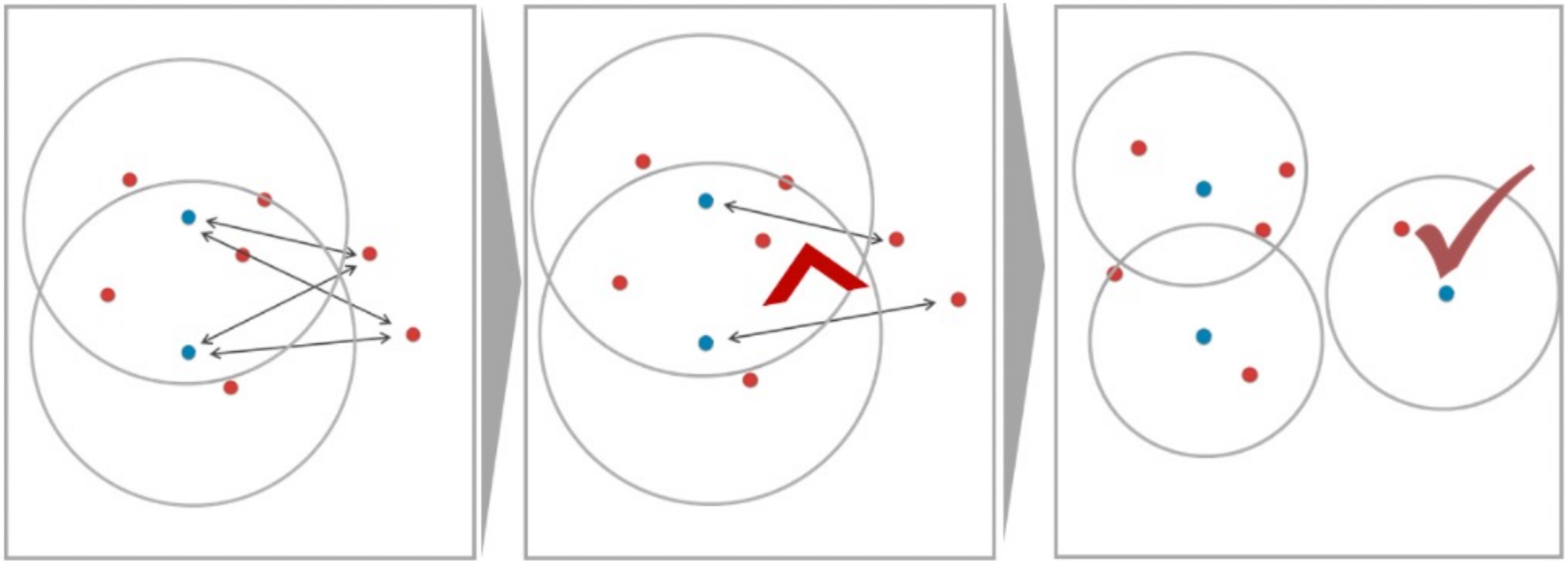
CoreSet

- k-Center-Greedy Algorithm
  1. Find the shortest distance between the data point and the center of the circle.
  2. Select the longest distance among them.



CoreSet

- k-Center-Greedy Algorithm
    1. Find the shortest distance between the data point and the center of the circle.
    2. Select the longest distance among them.
    3. Define new Core-set.

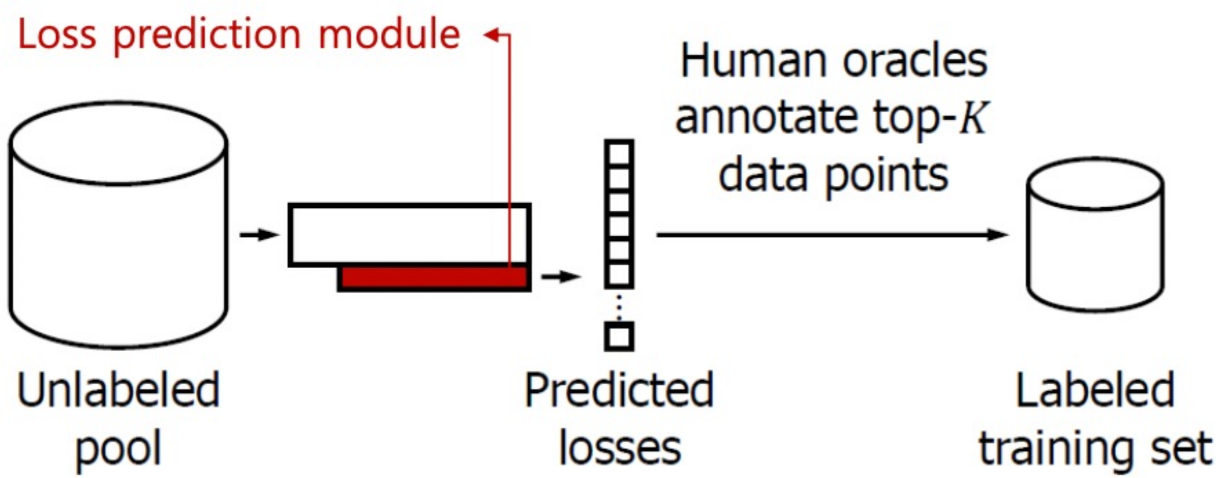


CoreSet

- Advantage
  - Easy to implement.
  - Efficient due to short search time.
- Disadvantage
  - Affected by the density of data.
  - Suffer from outlier.

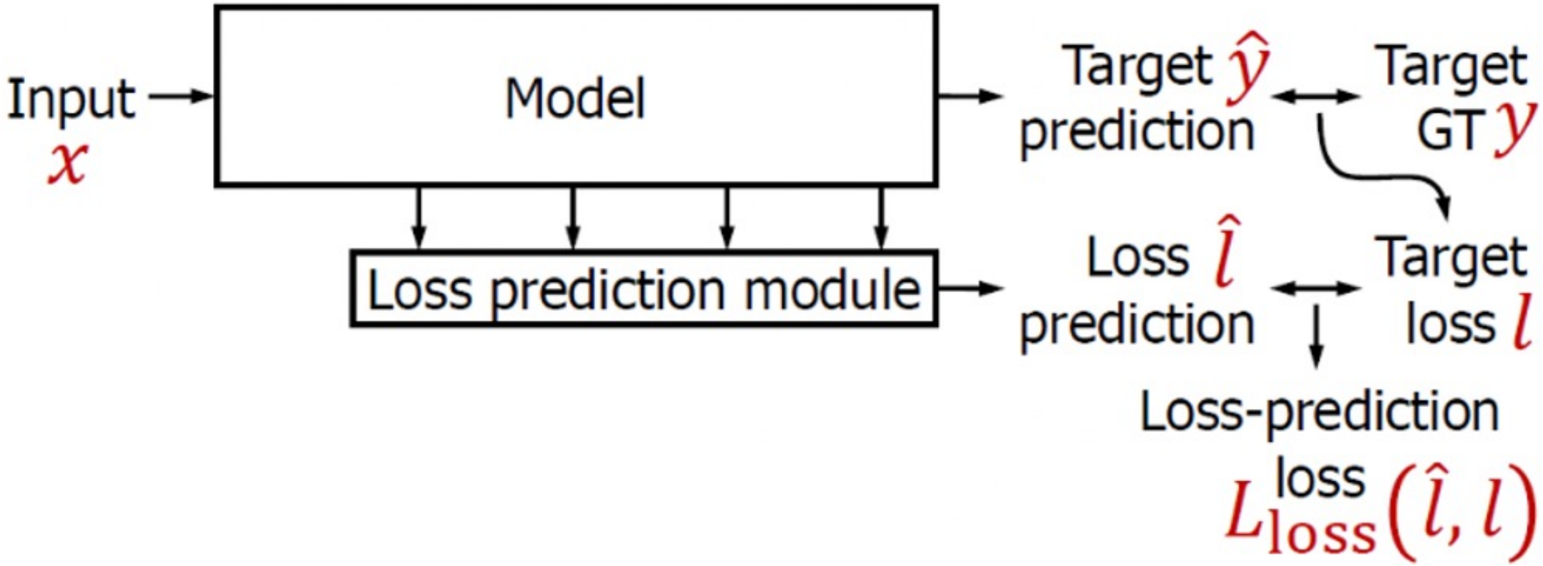
Learning Loss

- Concept
  - Query the sample which has highest loss from unlabeled data.
  - Add loss prediction module to predict top-k data points.



Learning Loss

- Loss function



Learning Loss

- Loss function
  - Constantly decreasing target loss during training.
  - Instead of MSE Loss, using Margin Ranking Loss.
  - MSE Loss:

$$L_{loss}(\hat{l}, l) = \|\hat{l} - l\|^2$$

Scale changes

  - Margin Ranking Loss:

Margin (=1)

$$L_{\text{loss}}(\hat{l}_i, \hat{l}_j, l_i, l_j) = \max\left(0, -\mathbf{1}(l_i, l_j) \cdot (\hat{l}_i - \hat{l}_j) + \xi\right)$$

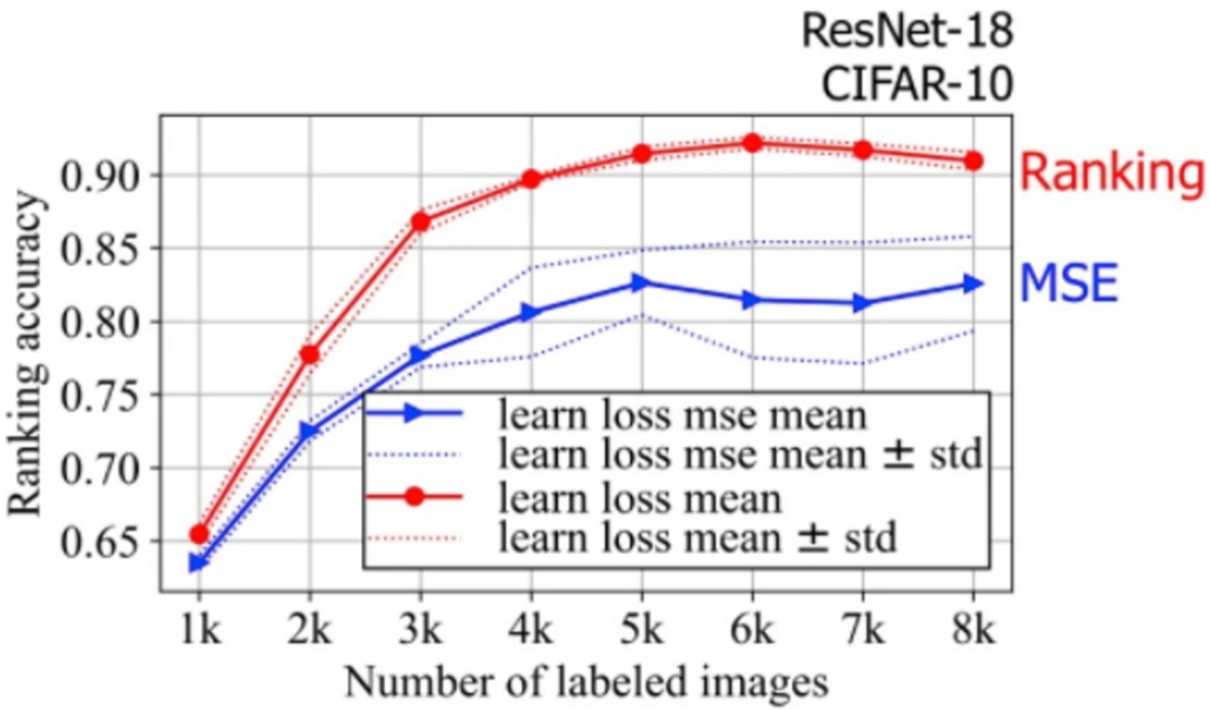
A pair of predicted losses

A pair of real losses

$$\text{where} \quad \mathbf{1}(l_i, l_j) = \begin{cases} +1, & \text{if } l_i > l_j \\ -1, & \text{otherwise} \end{cases}$$

Learning Loss

- Loss function
  - MSE Loss vs. Ranking Loss

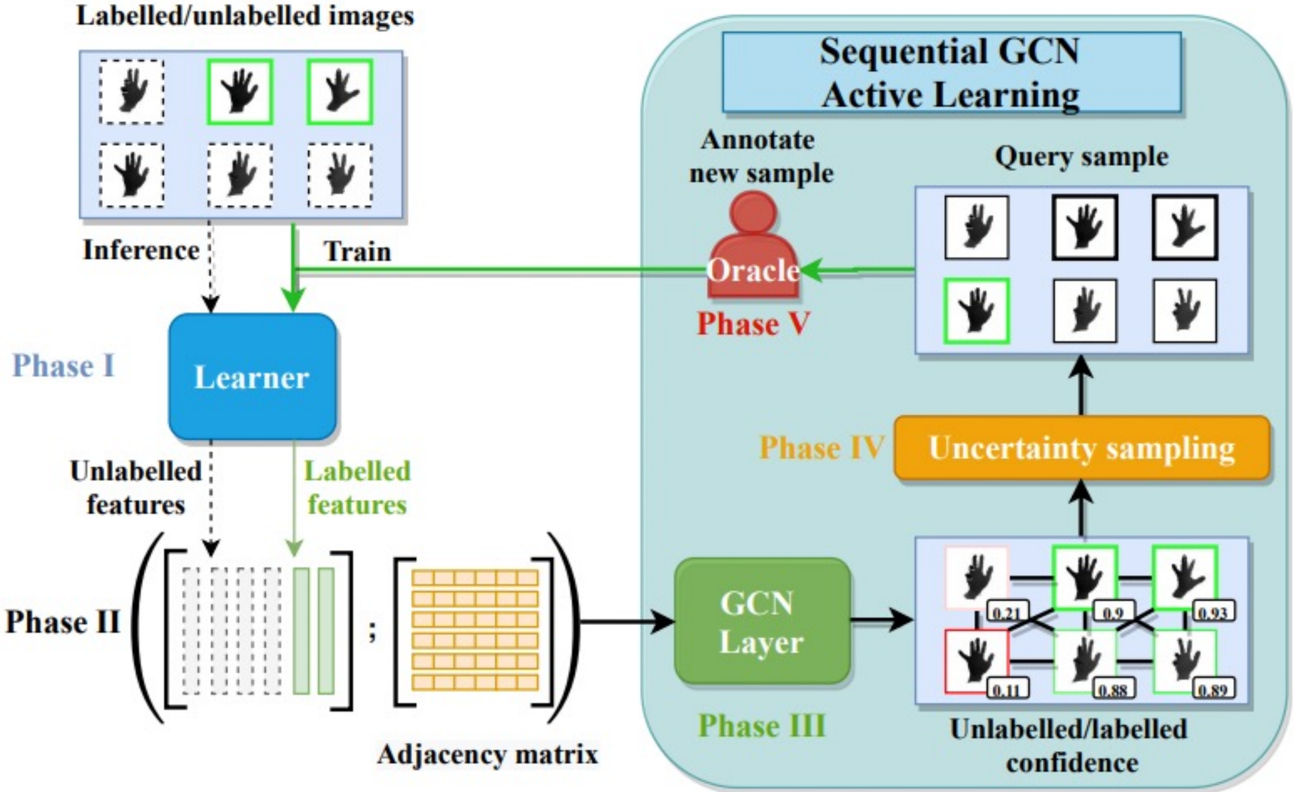


Learning Loss

- Advantage
  - Task-agnostic.
  - End-to-end training.
- Disadvantage
  - Not considered diversity and density.
  - Lack of correlation between the labeled and unlabeled data.

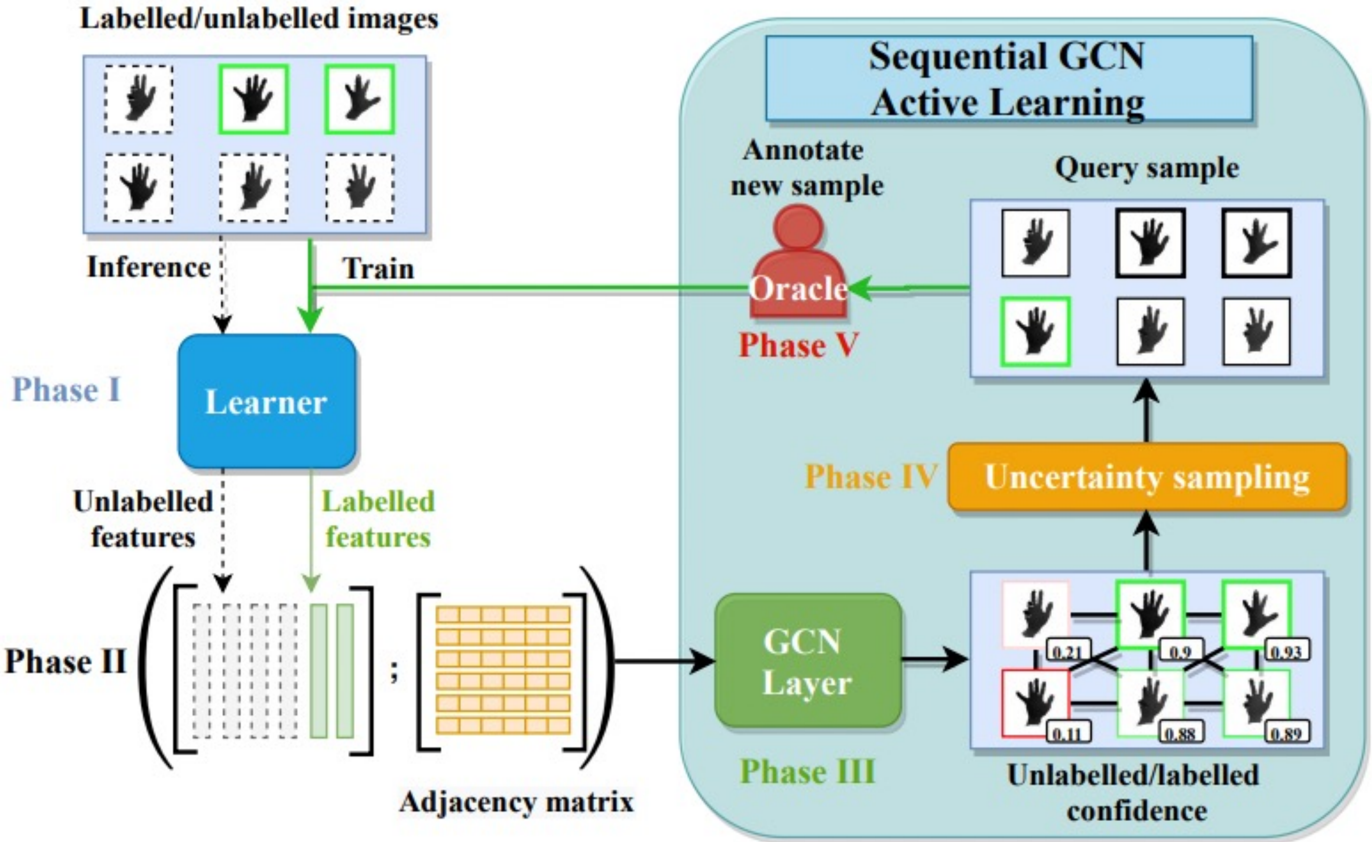
For Hand Pose Estimation

- Sequential Graph Convolutional Network for Active Learning
  - Accepted as Poster at CVPR 2021



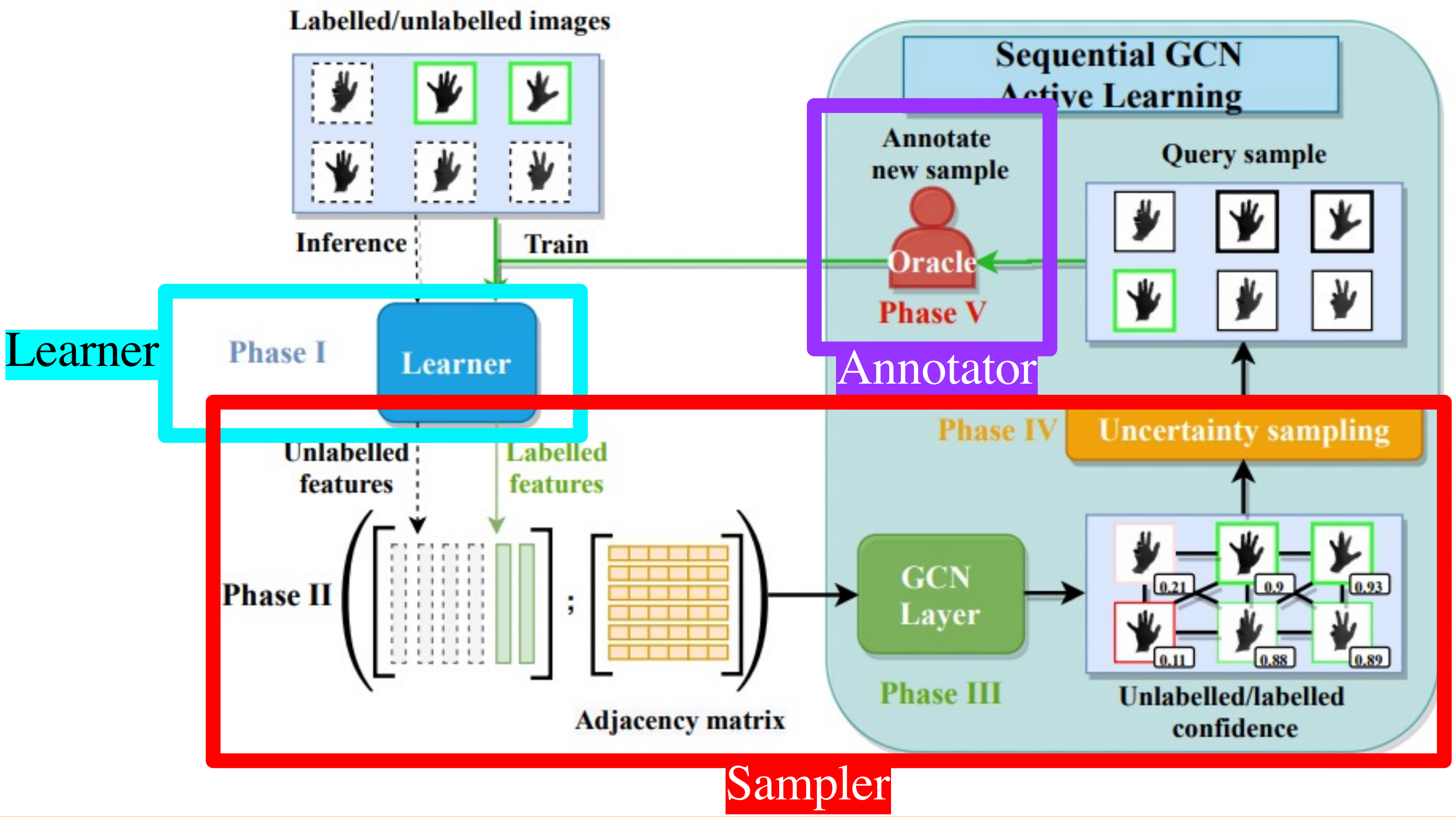
Components of AL framework

- Learner, Sampler, Annotator



Components of AL framework

- Learner, Sampler, Annotator



Learner

- Goal
  - Minimize the objective of target task.
  - Objective may change depending on the task to be dealt with.
- Classification
  - Objective function (cross-entropy):

$$\mathcal{L}^c_{\mathcal{M}}(\mathbf{x}, \mathbf{y}; \theta) = -\frac{1}{N_l} \sum_{i=1}^{N_l} \mathbf{y}_i \log(f(\mathbf{x}_i, \mathbf{y}_i; \theta)),$$

Learner

- Goal
  - Minimize the objective of target task.
  - Objective may change depending on the task to be dealt with.
- Regression
  - Objective function:

$$\mathcal{L}^{r}_{\mathcal{M}}(\mathbf{x}, \mathbf{y}; \theta) = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(\frac{1}{J} \sum_{j=1}^{J} \| \mathbf{y}_{i,j} - f(\mathbf{x}_{i,j}, \mathbf{y}_{i,j}; \theta) \|^2 \right)$$

- Other tasks
  - Just modify the learner.
  - The rest of pipeline remains the same.

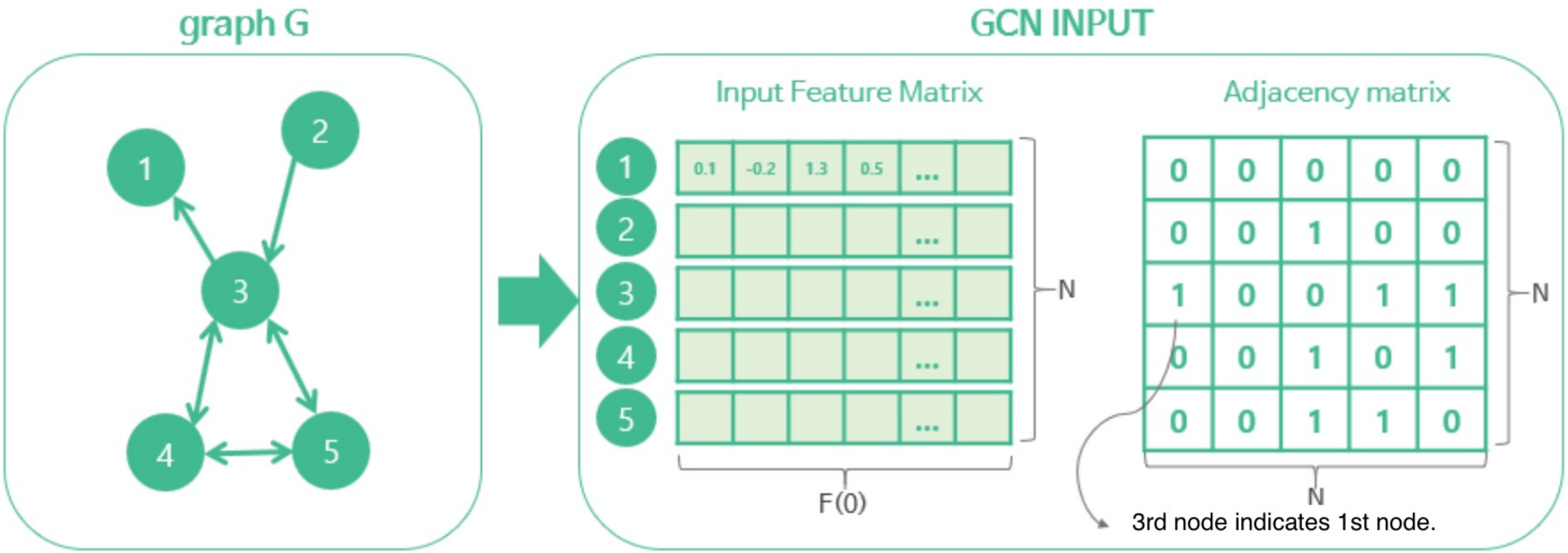
Sampler

- Goal
  - Select the representative unlabeled examples within a fixed budget to deliver the highest performance.
- Mechanism
  - From a pool of unlabeled dataset D_U, randomly select an initial batch for labelling $D_0 \subset D_U$.
  - Achieve minimum loss with the least number of batches D_n.

$$\min_{n} \min_{\mathcal{L}_{\mathcal{M}}} \mathcal{A}(\mathcal{L}_{\mathcal{M}}(\mathbf{x}, \mathbf{y}; \theta) | \mathbf{D}_0 \subset \cdots \subset \mathbf{D}_n \subset \mathbf{D}_U).$$

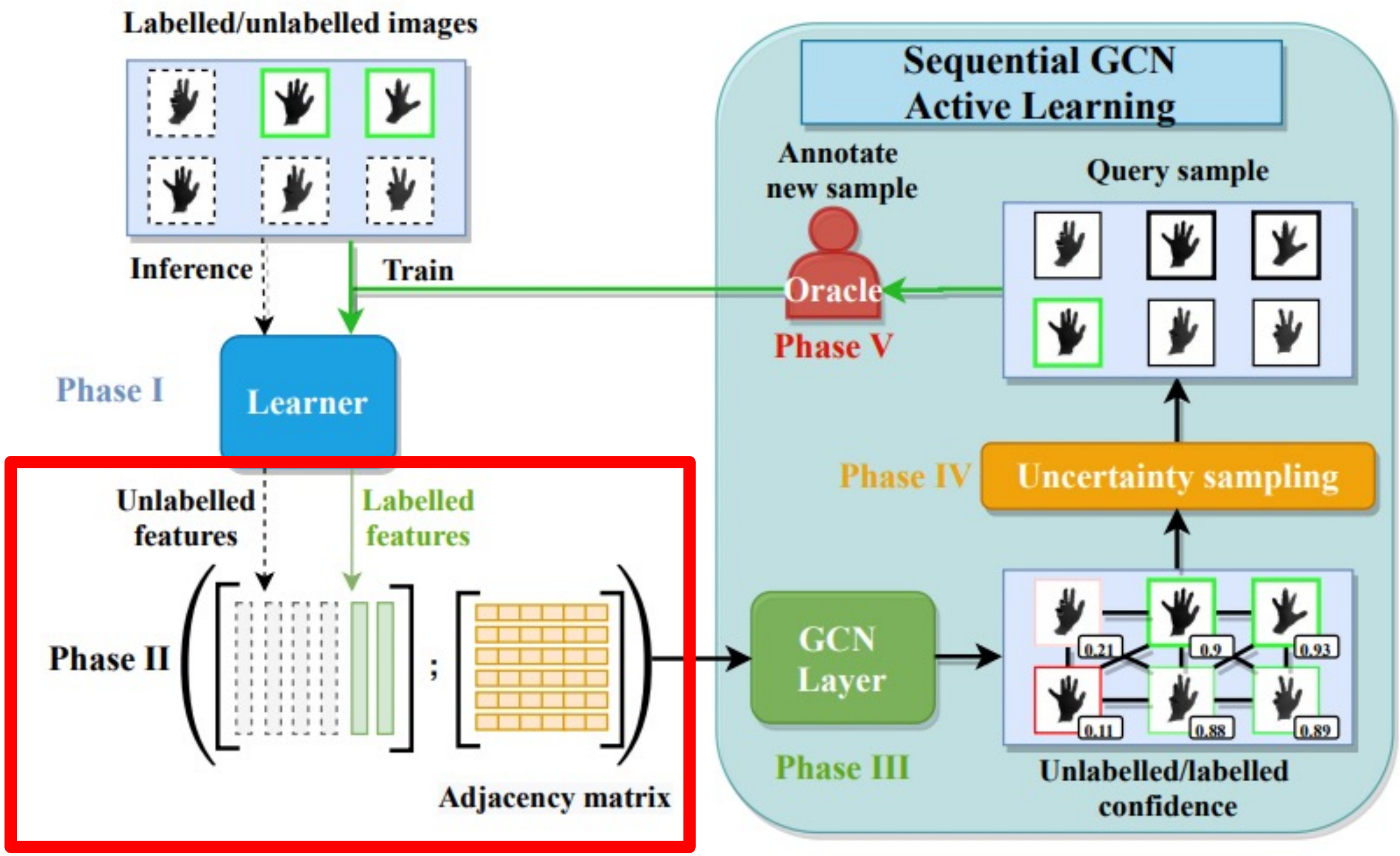
Sampler

- Graph Convolutional Network



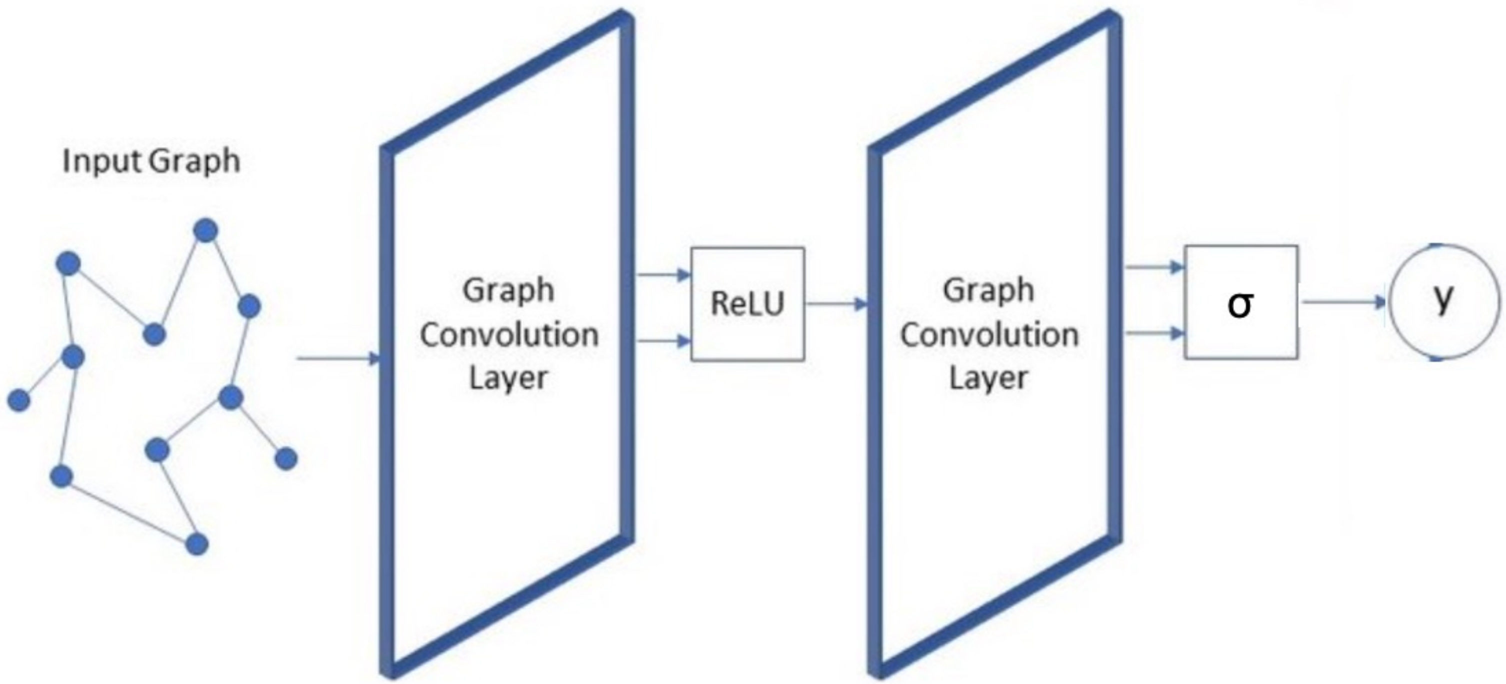
Sampler

- Graph Convolutional Network



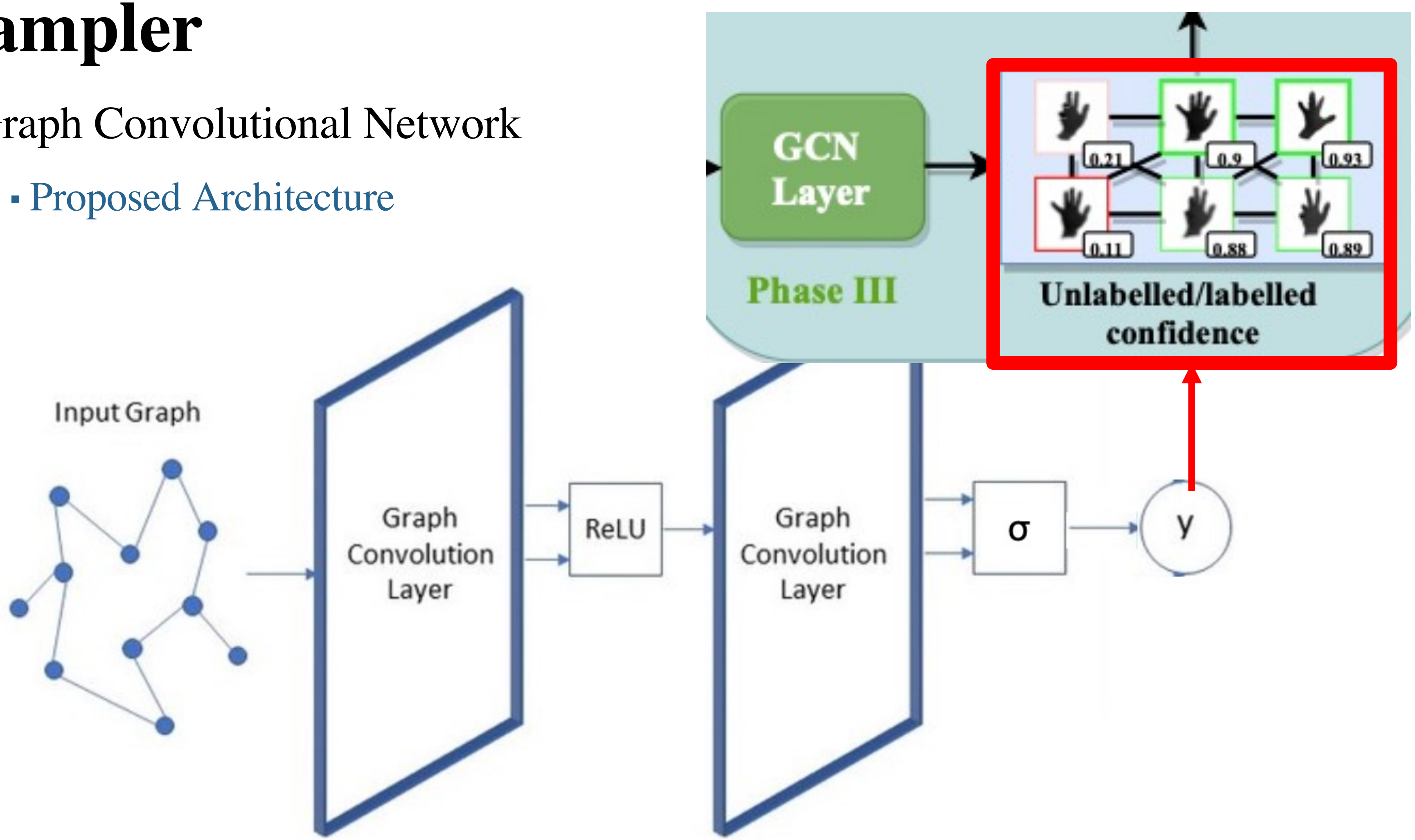
Sampler

- Graph Convolutional Network
  - Proposed Architecture



Sampler

- Graph Convolutional Network
  - Proposed Architecture



Sampler

- UncertainGCN: Uncertainty sampling on GCN

$$\mathbf{D}_L = \mathbf{D}_L \cup \underset{i=1\cdots b}{\arg\max} |s_{margin} - f_{\mathcal{G}}(\mathbf{v}_i; \mathbf{D}_U)|.$$

- CoreGCN: CoreSet sampling on GCN

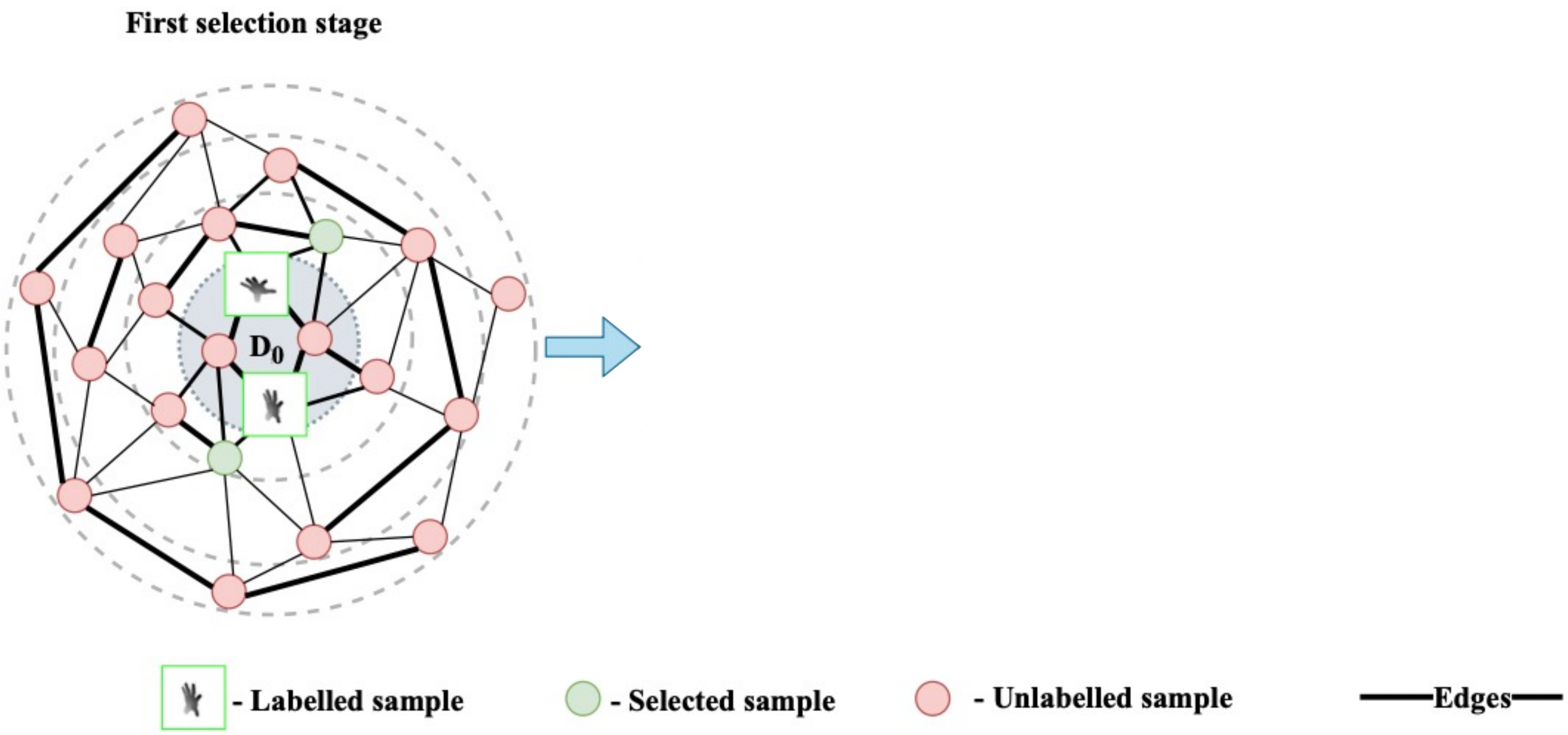
$$\mathbf{D}_L = \mathbf{D}_L \cup \underset{i \in \mathbf{D}_U}{\arg\max} \min_{j \in \mathbf{D}_L} \delta(f^1_{\mathcal{G}}(A, \mathbf{v}_i; \Theta_1), f^1_{\mathcal{G}}(A, \mathbf{v}_j; \Theta_1))$$

UncertainGCN: Uncertainty sampling on GCN

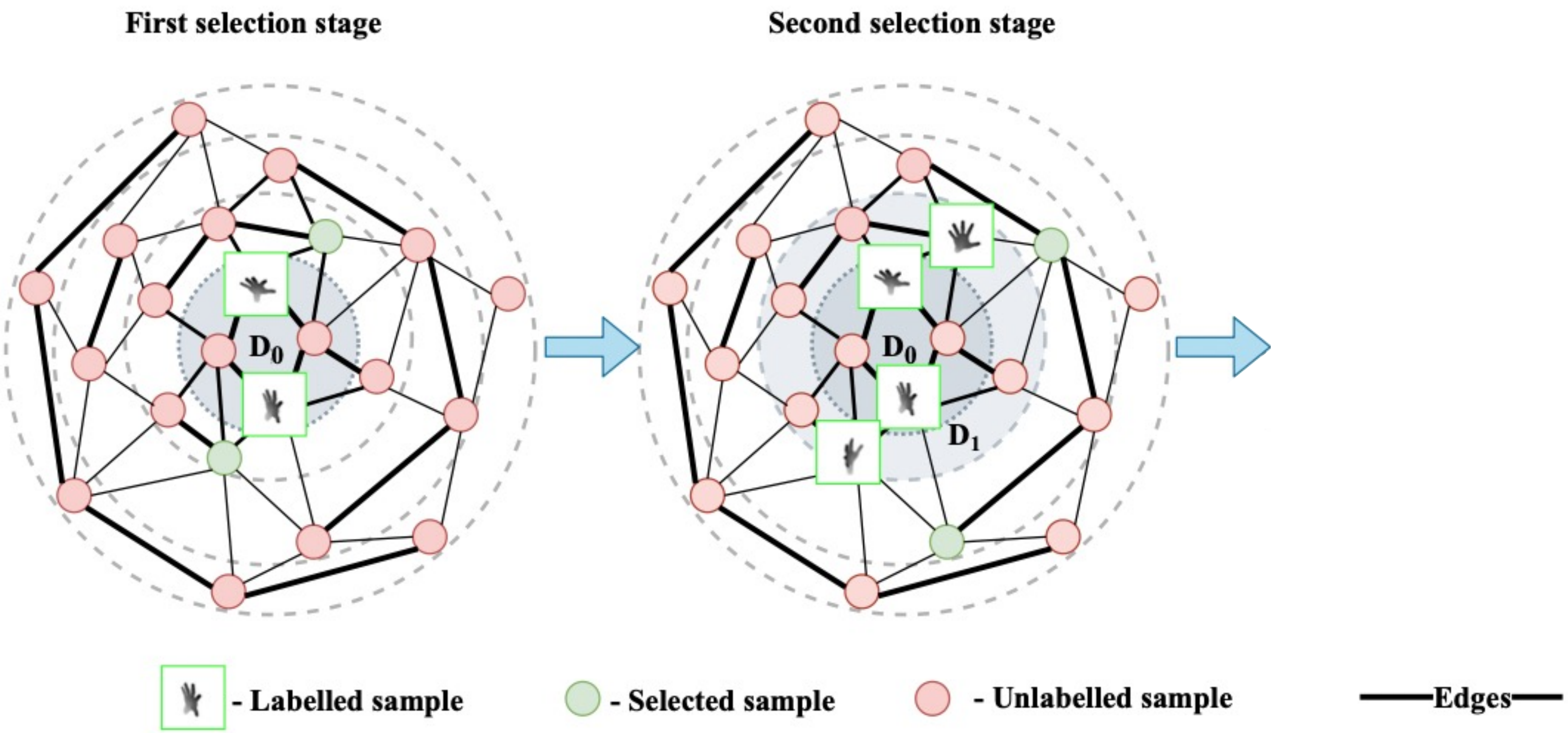
- After training on GCN, move to selection.
- From D_U, draw confidence scores $f_g(v_i; D_U)$ as output of GCN.
- Using this score, select the unlabeled images with UncertainGCN.
- Apply the following equation:

$$\mathbf{D}_L = \mathbf{D}_L \cup \underset{i=1\cdots b}{\arg\max} |s_{margin} - f_{\mathcal{G}}(\mathbf{v}_i; \mathbf{D}_U)|.$$

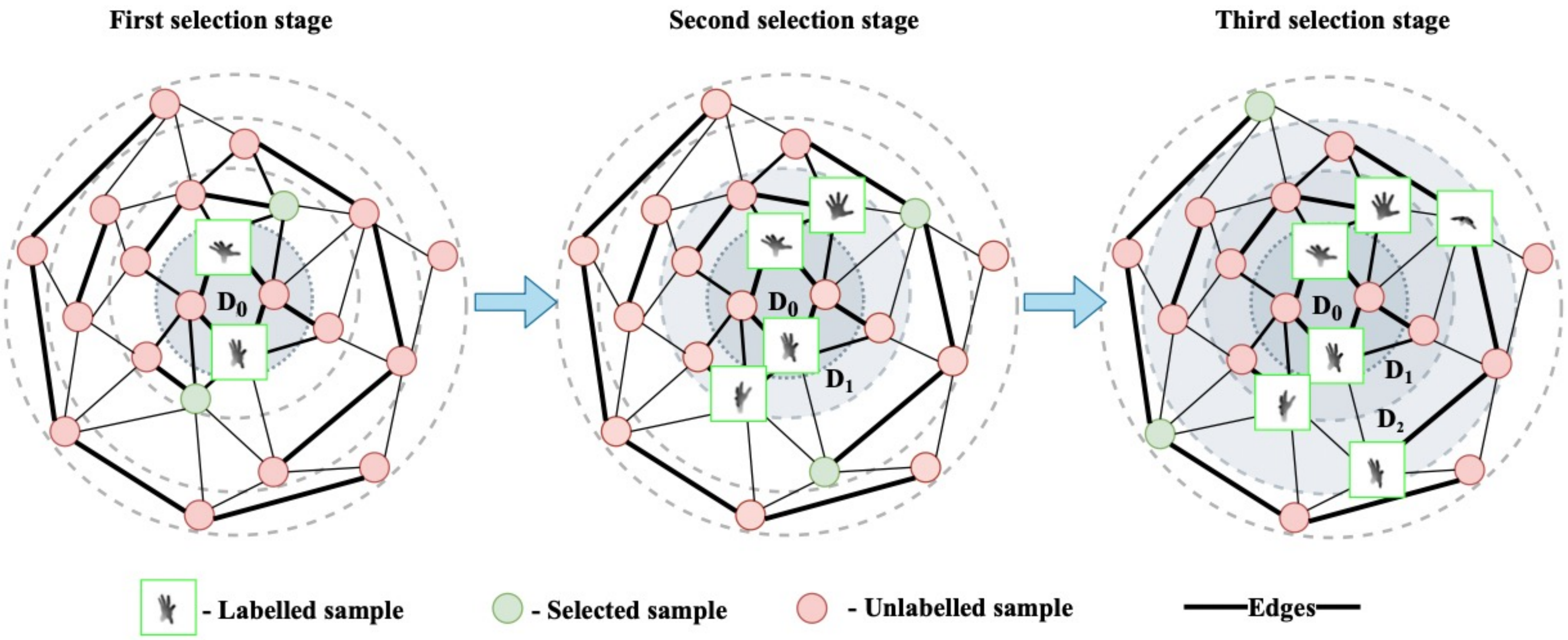
UncertainGCN: Uncertainty sampling on GCN



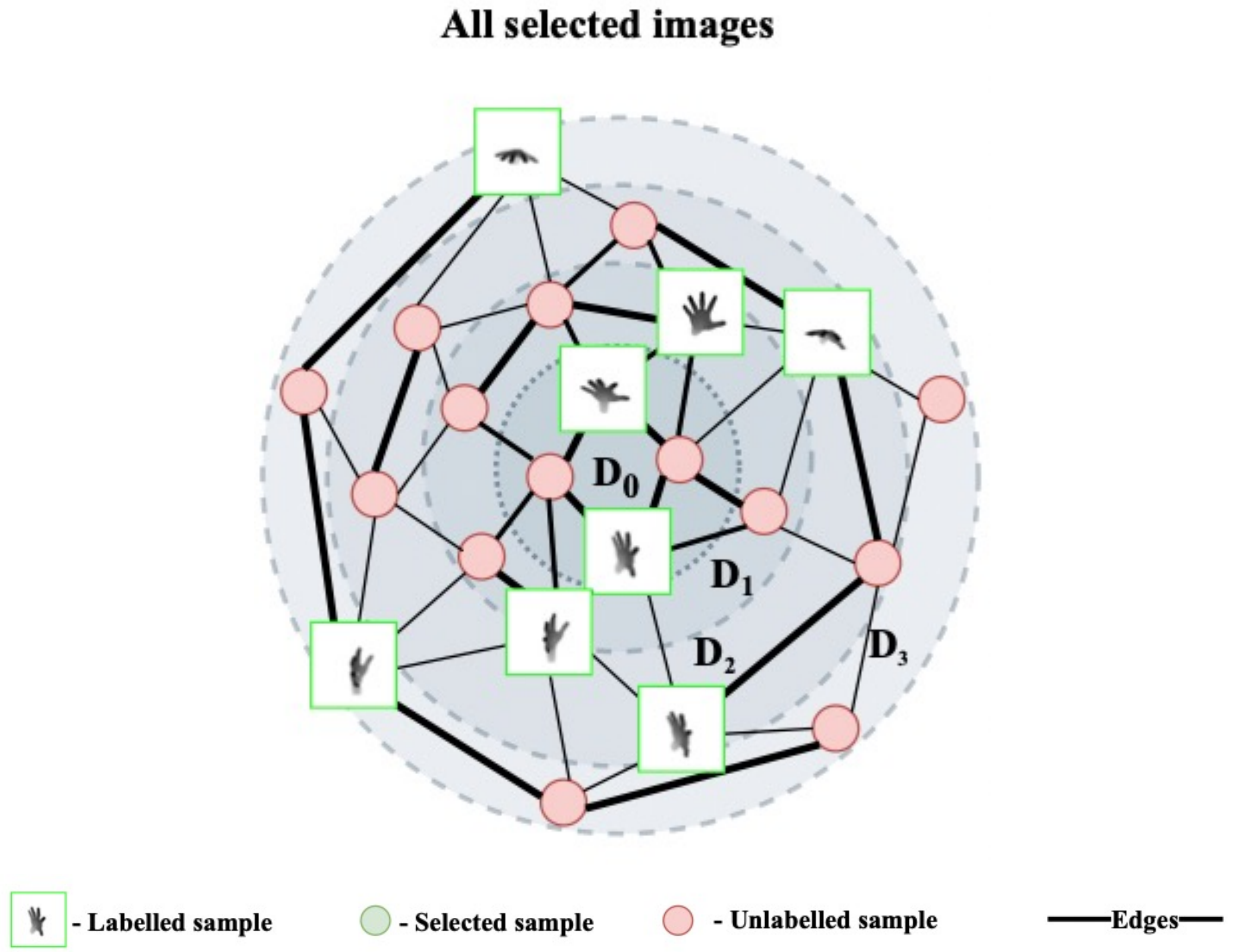
UncertainGCN: Uncertainty sampling on GCN



UncertainGCN: Uncertainty sampling on GCN



UncertainGCN: Uncertainty sampling on GCN



CoreGCN: CoreSet sampling on GCN

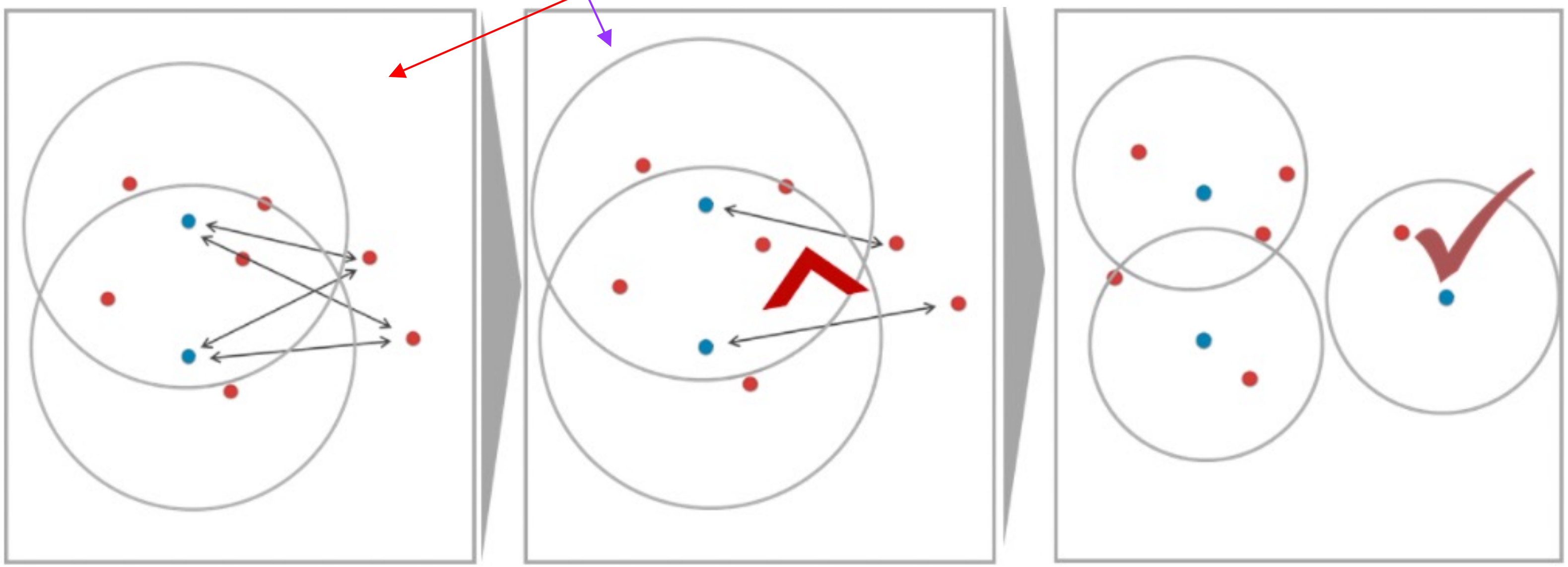
- Approach CoreSet technique
  - To integrate geometric information between the labeled and unlabeled graph representation.
  - Apply the following equation:

$$\mathbf{D}_L = \mathbf{D}_L \cup \underset{i \in \mathbf{D}_U}{\arg\max} \min_{j \in \mathbf{D}_L} \delta(f_{\mathcal{G}}^1(A, \mathbf{v}_i; \Theta_1), f_{\mathcal{G}}^1(A, \mathbf{v}_j; \Theta_1))$$

CoreGCN: CoreSet sampling on GCN

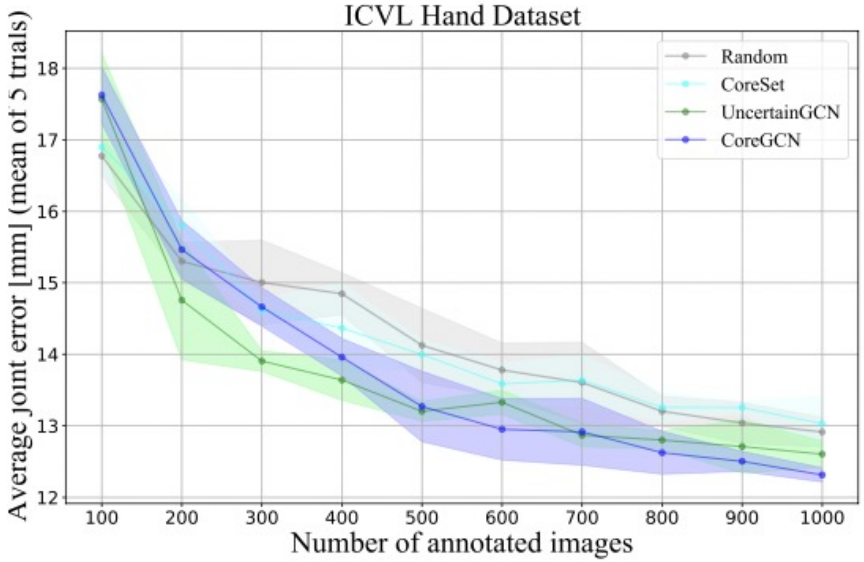
- Approach CoreSet technique
  - To integrate geometric information between the labeled and unlabeled graph representation.
  - Apply the following equation:

$$\mathbf{D}_L = \mathbf{D}_L \cup \boxed{\arg\max_{i \in \mathbf{D}_U} \boxed{\min_{j \in \mathbf{D}_L} \delta(f^1_{\mathcal{G}}(A, \mathbf{v}_i; \Theta_1), f^1_{\mathcal{G}}(A, \mathbf{v}_j; \Theta_1))}}$$



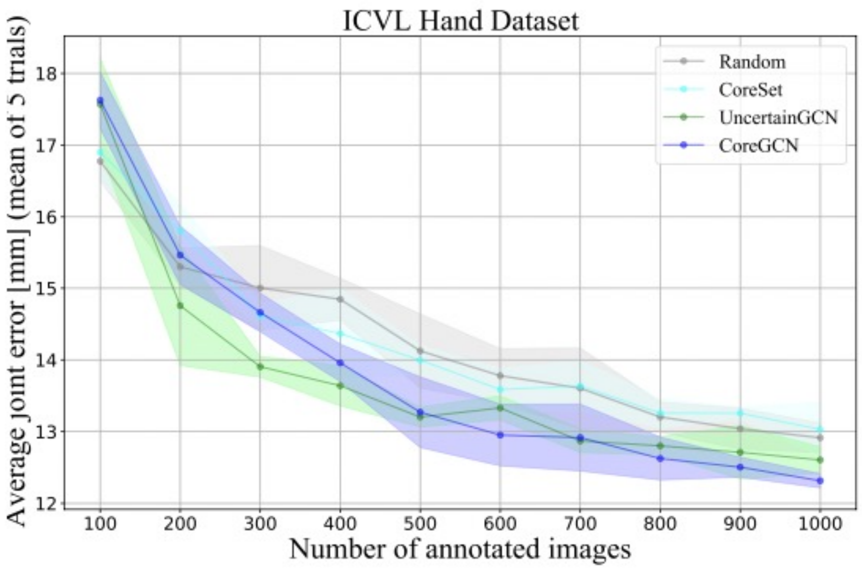
Experiment

- 3D Hand Pose Estimation



Experiment

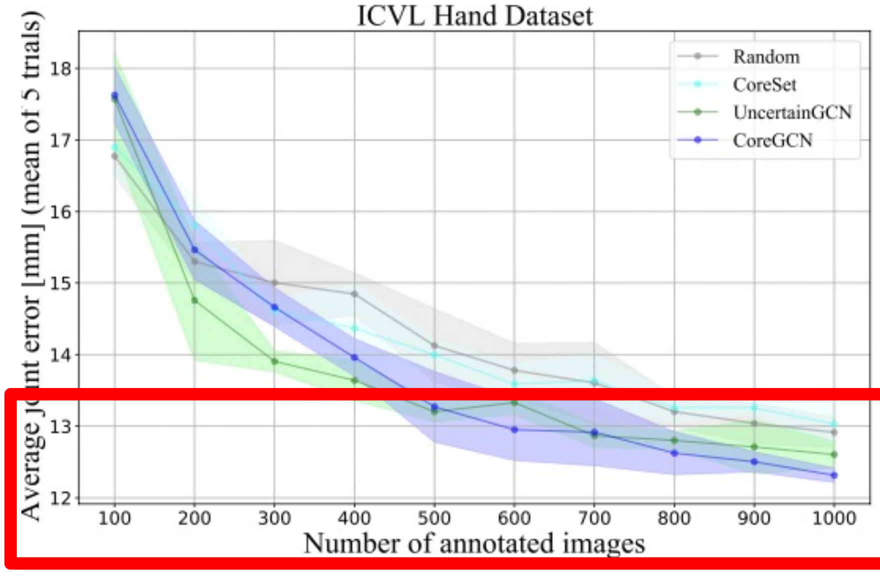
- 3D Hand Pose Estimation



Methods	Mean error (mm)			Input	Type
	ICVL	MSRA	NYU		
DeepModel [46]	11.56	-	17.04	2D	R
DeepPrior [27]	10.4	-	19.73	2D	R
Ren-4x6x6 [14]	7.63	-	13.39	2D	R
Ren-9x6x6 [42]	7.31	9.7	12.69	2D	R
DeepPrior++ [26]	8.1	9.5	12.24	2D	R
Pose-Ren [3]	6.79	8.65	11.81	2D	R
DenseReg [42]	7.3	**7.2**	10.2	2D	D
CrossInfoNet [6]	6.73	7.86	10.08	2D	R
JGR-P2O [8]	6.02	7.55	**8.29**	2D	D
3DCNN [11]	-	9.6	14.1	3D	R
SHPR-Net [4]	7.22	7.76	10.78	3D	R
HandPointNet [9]	6.94	8.5	10.54	3D	R
Point-to-Point [12]	6.3	7.7	9.10	3D	D
V2V [24]	6.28	7.59	8.42	3D	D

Experiment

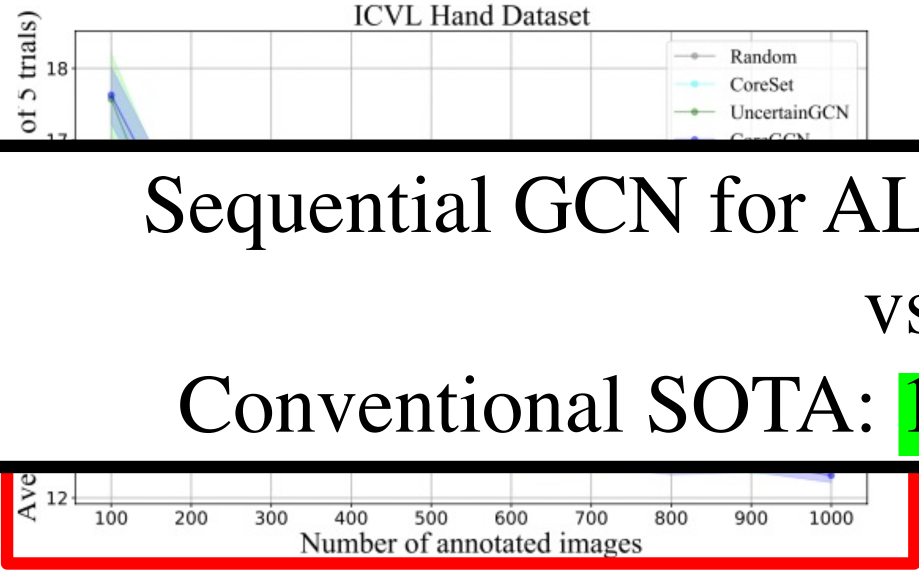
- 3D Hand Pose Estimation



Methods	Mean error (mm) ICVL	MSRA	NYU	Input	Type
DeepModel [46]	11.56	-	17.04	2D	R
DeepPrior [27]	10.4	-	19.73	2D	R
Ren-4x6x6 [14]	7.63	-	13.39	2D	R
Ren-9x6x6 [42]	7.31	9.7	12.69	2D	R
DeepPrior++ [26]	8.1	9.5	12.24	2D	R
Pose-Ren [3]	6.79	8.65	11.81	2D	R
DenseReg [42]	7.3	**7.2**	10.2	2D	D
CrossInfoNet [6]	6.73	7.86	10.08	2D	R
JGR-P2O [8]	6.02	7.55	**8.29**	2D	D
3DCNN [11]	-	9.6	14.1	3D	R
SHPR-Net [4]	7.22	7.76	10.78	3D	R
HandPointNet [9]	6.94	8.5	10.54	3D	R
Point-to-Point [12]	6.3	7.7	9.10	3D	D
V2V [24]	6.28	7.59	8.42	3D	D

Experiment

- 3D Hand Pose Estimation



Methods	Mean error (mm) ICVL	MSRA	NYU	Input	Type
DeepModel [46]	11.56	-	17.04	2D	R
DeepPrior [27]	10.4	-	19.73	2D	R
Ren-4x6x6 [14]	7.63	-	13.39	2D	R
Ren-9x6x6 [42]	7.31	9.7	12.69	2D	R
HandPointNet [9]	6.94	8.5	10.54	3D	R
Point-to-Point [12]	6.3	7.7	9.10	3D	D
V2V [24]	6.28	7.59	8.42	3D	D

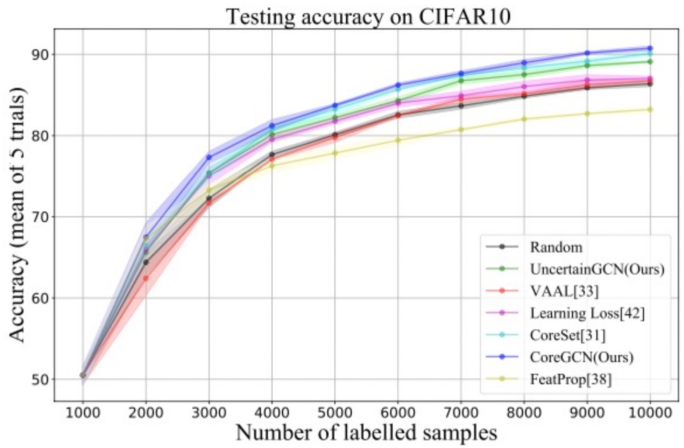
Sequential GCN for AL: 1,000 labeled images

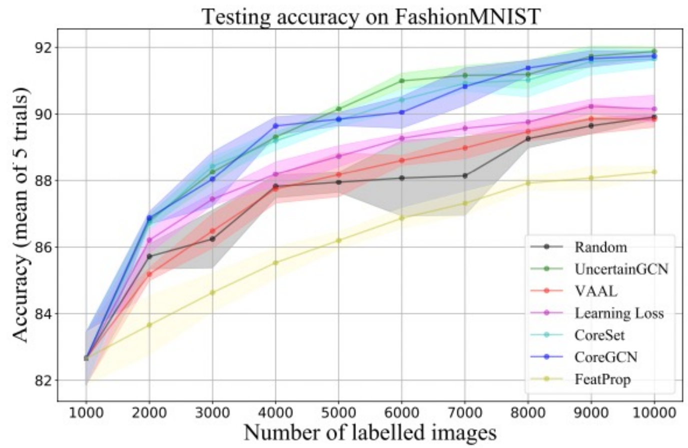
vs.

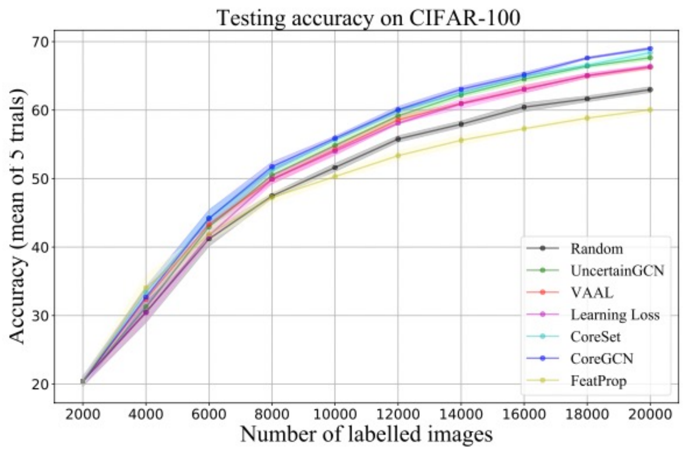
Conventional SOTA: 16,004 labeled images

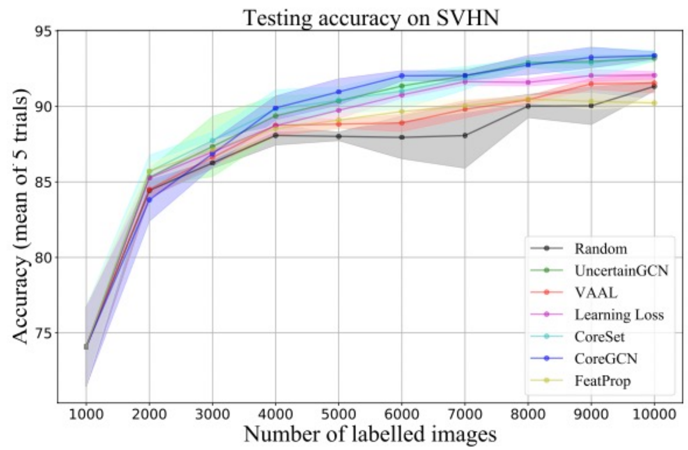
Experiment

- Image Classification









Conclusion

- Benefit of Active Learning
  - The more important data is picked, the lower the cost and the higher the performance will be.
- Compared with Semi-supervised learning
  - Active Learning with Pseudo-Labels for Multi-View 3D Pose Estimation
    - Meta Reality Labs
- Limitations
  - Need more elaborate method.
- Future work
  - Apply to Interacting-hand Pose Estimation.

References

1. Razvan Caramalau, Binod Bhattarai, and Tae-Kyun Kim. Sequential Graph Convolutional Network for Active Learning. In CVPR, 2021.
2. Sener and Silvio Savarese. Active Learning for Convolutional Neural Networks: A Core-set approach. In ICLR, 2018.
3. Donggeun Yoo and In So Kweon. Learning Loss for Active Learning. In CVPR, 2019.