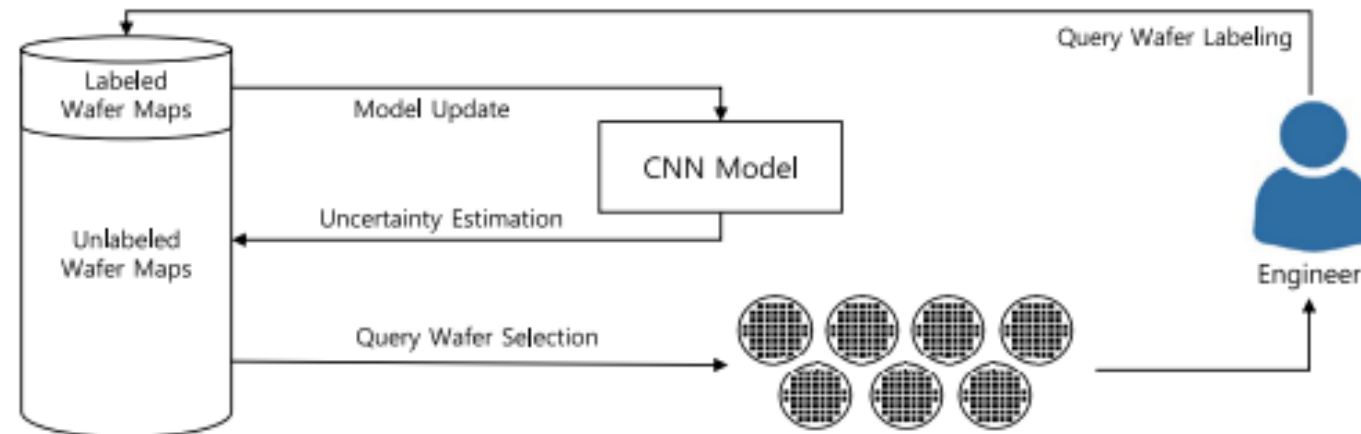


Towards Robust and Reproducible Active Learning using Neural Networks

Munjal, P., Hayat, N., Hayat, M., Sourati, J., & Khan, S. (2022). Towards robust and reproducible active learning using neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 223-232).

Active Learning 이란

- Unlabel data 중 가장 유용한 데이터를 선택하여 학습 데이터셋을 효과적으로 확장하는 방법
- 목적
 - 불필요한 데이터는 학습에서 제외하여 학습 알고리즘의 성능을 높이면서, 전략적 데이터 선택을 통해 라벨링 비용 최소화
 - 지속적인 환경 변화로 인해 발생하는 모델 성능 하락을 보완하기 위해 최소한의 지속적인 라벨링 진행



Overview of the proposed wafer map pattern classification system.

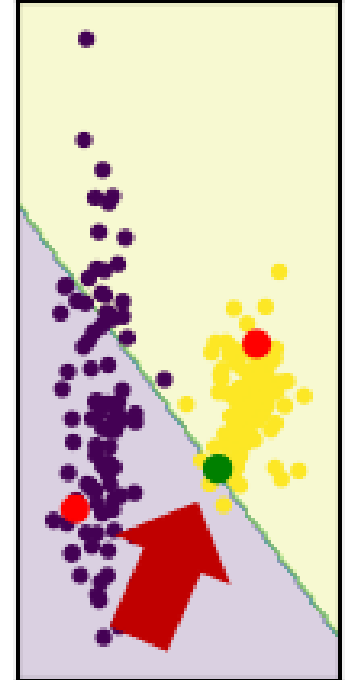
기존 Active Learning 연구에 대한 의문점 및 해당 연구의 기여도

- 논문에서 제기한 의문점
 - 1) 똑같은 데이터셋에서 진행한 랜덤 샘플링 방법이 연구마다 다름
 - 2) 각 논문에서 비교 실험을 진행했을 때 연구마다 방법들의 성능이 각자 다름 ▷ 같은 모델, 기법, 데이터임에도 불구하고 다름
 - 3) 연구를 통해 얻는 결론이 일관적이지 않음 ▷ Uncertainty 기반의 AL이 좋다 vs Diversity 기반의 AL이 좋다
 - 4) 연구에서 사용되는 모델의 regularization에 대한 논의가 충분히 이루어지지 않음
 - 5) 연구들이 weight decay, 하이퍼파라미터 튜닝 과 같이 디테일 한 설정을 하고 있자 않음
- 본 논문은 동일하게 실험을 구성하여 Active Learning의 방법론들을 비교
- 디테일한 튜닝을 통해 랜덤 샘플링이 active learning보다 더 좋은 경우를 제거
- 각 상황, 조건에 따라 가장 좋은 메소드가 다른 다는 것을 확인
- 동일한 조건에서 비교 실험 및 디테일 한 설정 튜닝을 통해 좀 더 일관적이고 reproducible 한 실험을 해야 한다 제언

Pool Based Active Learning Methods

Uncertainty Base

- 목표 : Decision Boundary에 가까운(모델이 헛갈려하는) 샘플을 고르는 것
- 모델 입장에서 데이터의 불확실성(Uncertainty)를 측정하는 방식은 크게 3가지
 - 1) Least Confidence : 최대 확률값이 가장 낮은 데이터 선별
 - 2) Margin Sampling : Top-1 confidence와 Top-2의 confidence의 차이가 낮은 데이터 선별
 - 3) Maximum Entropy : Entropy가 높은 데이터를 선택
- 사전 연구*는 해당 방법을 활용해 모델 예측의 신뢰도가 낮은 것들을 샘플링
- 장점 : 구현하기 간단하며 딥러닝과 연계하여 사용하기 쉽다
- 단점
 - 1) 결정 경계를 만드는 실제 주요한 데이터가 포함되지 않을 수 있음
 - 2) 모델의 아웃풋으로 결정되기에 실제 모델이 어려워 하는 것인지 100% 신뢰할 수 없음(Ex 학습이 잘 안된 모델이 리턴하는 경우)
 - 3) Outlier에 영향을 많이 받음
 - 4) Diversity(데이터 분포) 무시

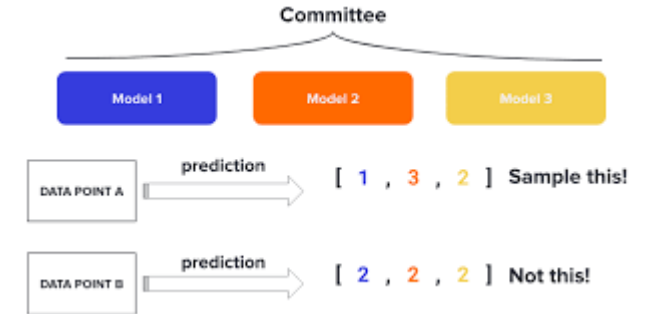


* David D Lewis and William A Gale. A sequential algorithm for training text classifiers. In SIGIR'94, pages 3–12. Springer, 1994. 2

Pool Based Active Learning Methods

Query By Committee Base

- 목표 : 여러 모델을 통해 모델 전체가 어려워하는 데이터를 선별
- 앙상블의 Vote 개념처럼 여러 모델(Committee 위원회)에서 투표
- 장점 : 구현하기 간단하며 앙상블의 효과는 많은 실험을 통해 증명됨
- 단점 : 비용이 큰 Large model 에선 사용하기 힘들
- 사전 연구*는 supervised 모델 여러 개를 사용하여 variance가 큰 데이터 샘플링 진행



$$\operatorname{argmax}_x - \sum_{y \in Y} \frac{Votes(y)}{M} \log\left(\frac{Votes(y)}{M}\right)$$

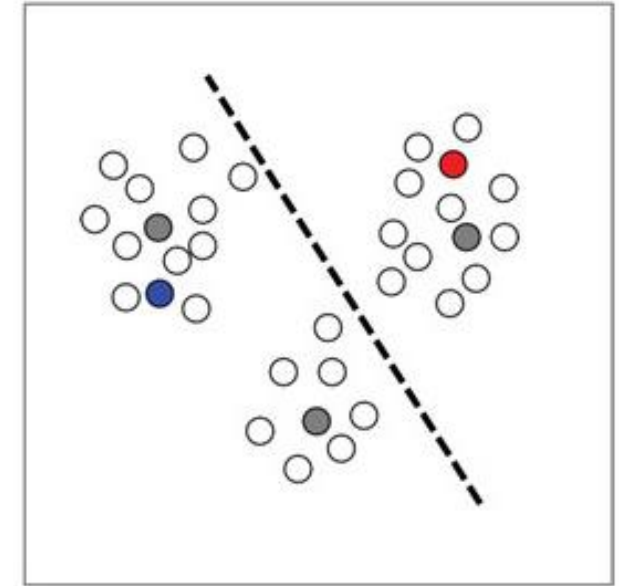
[Vote Entropy 수식]

* William H. Beluch, Tim Genewein, Andreas Nurnberger, " and Jan M. Kohler. The power of ensembles for active learning in image classification. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9368– 9377, 2018. 1, 2, 3, 7

Pool Based Active Learning Methods

Diversity Base

- 목표 : 데이터의 분포를 커버할 수 있는 대표 데이터 선별
- 장점 : 클러스터링과 같은 unsupervised 방법을 이용해 대표되는 데이터를 학습하여 학습의 능률을 향상
- 단점
 - 1) Decision Boundary 근방의 분류에 도움이 되는 정보를 적게 가지고 있음
 - 2) 이상치(Outlier)에 취약
 - 3) 데이터셋의 조건(데이터 밀집도, class imbalance, class간 분류 난이도)에 영향을 크게 받음
 - 밀집도가 높은 Space에서 적은 Sample을 뽑게 된다. (∵ 작은 범위의 set만으로도 데이터 분포가 전부 반영되기에)



(b) Diversity-based selection strategies

Pool Based Active Learning Methods

Coreset*

- Future Expected Loss에 대한 Upper bound를 설정(Generalization Error + Training Error + Core-set Loss) 하여 문제를 정의
- 딥러닝의 특성상 Generalization Error와 Training Error를 0으로 가정 가능하기에 core set loss 만 계산
- 이를 통해 subset에서의 모델 성능과 전체 데이터 셋에서의 성능 차이를 최소화 하는 포인트를 구함

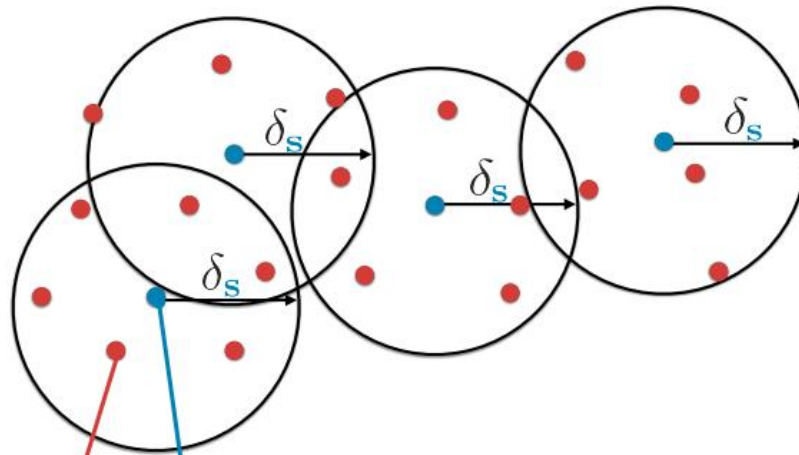
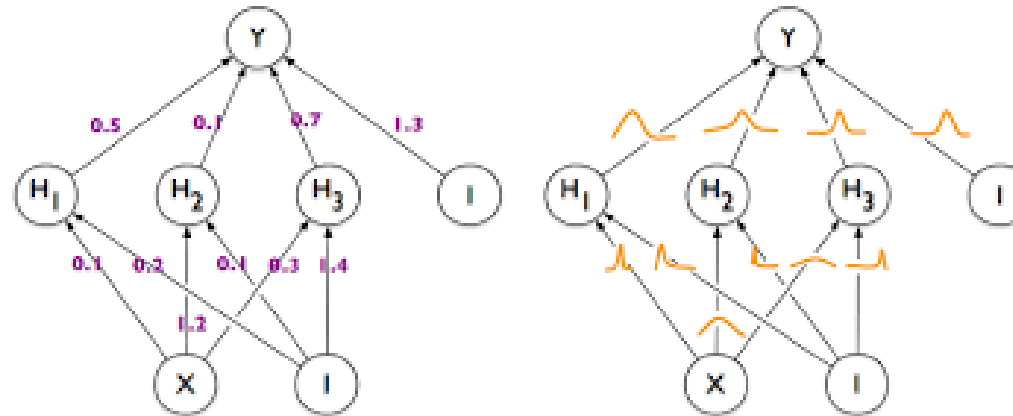


Figure 1: **Visualization of the Theorem 1.** Consider the set of selected points s and the points in the remainder of the dataset $[n] \setminus s$, our results shows that if s is the δ_s cover of the dataset, $\left| \frac{1}{n} \sum_{i \in [n]} l(\mathbf{x}_i, y_i, A_s) - \frac{1}{|s|} \sum_{j \in s} l(\mathbf{x}_j, y_j, A_s) \right| \leq \boxed{\mathcal{O}(\delta_s)} + \mathcal{O}\left(\sqrt{\frac{1}{n}}\right)$

Pool Based Active Learning Methods

Deep Bayesian Active Learning (DBAL)*

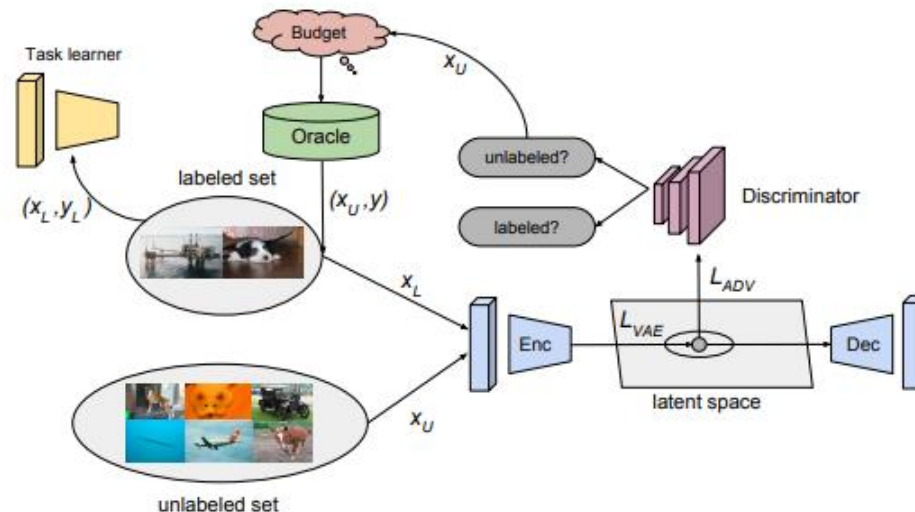
- 테스트 과정에서도 Drop-out 방식을 적용
- 기존의 뉴럴 네트워크는 학습이 끝나면 weight가 고정된 값을 가지게 되지만 베이지안 뉴럴 네트워크는 각 weight 또한 불확실성을 가지는 확률 분포로 표현
- weight가 불확실성을 가지는 확률 분포이므로 최종 출력물인 예측값 또한 불확실성을 가지는 확률 분포
- 예측 값의 엔트로피가 높은 것을 샘플링하여 active learning 진행



Pool Based Active Learning Methods

Variational Adversarial Active Learning*

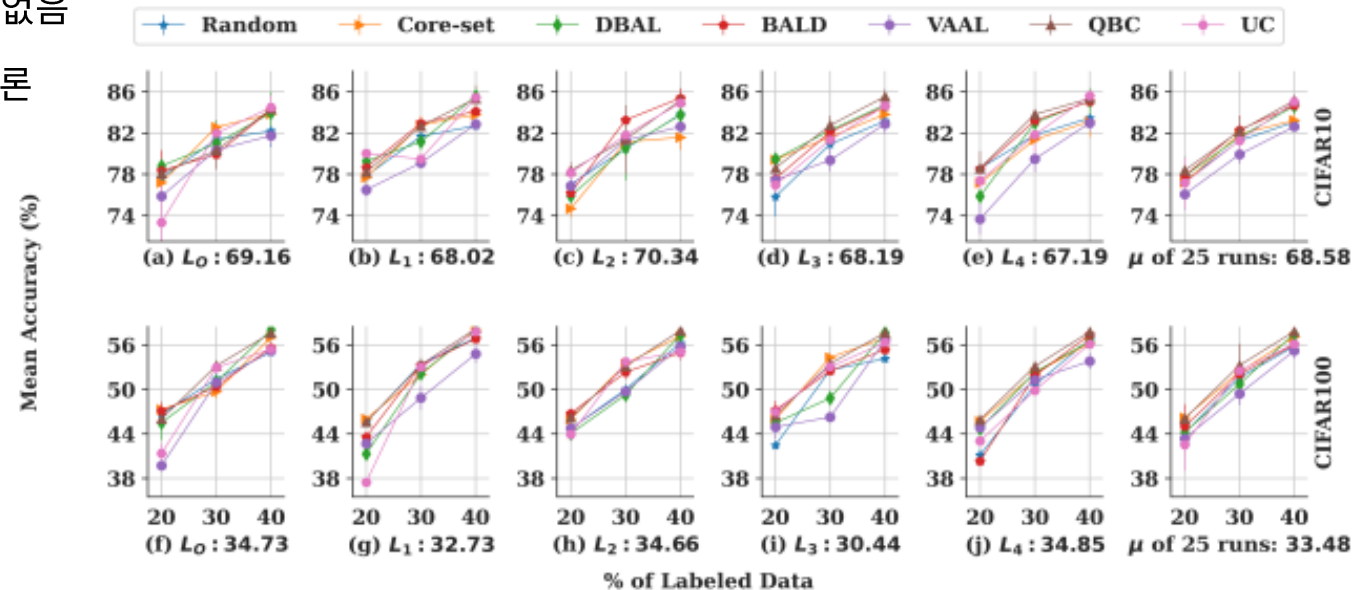
- Generator : labeled와 unlabeled 를 모두 사용해서 VAE를 학습
- Discriminator : label data와 unlabeled data를 구분
- unlabeled dataset pool에서 생성된 discriminator borderline에서 가장 먼 데이터를 샘플링
- data distribution과 uncertainty를 같이 고려



* Sinha, S., Ebrahimi, S., & Darrell, T. (2019). Variational adversarial active learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5972-5981).

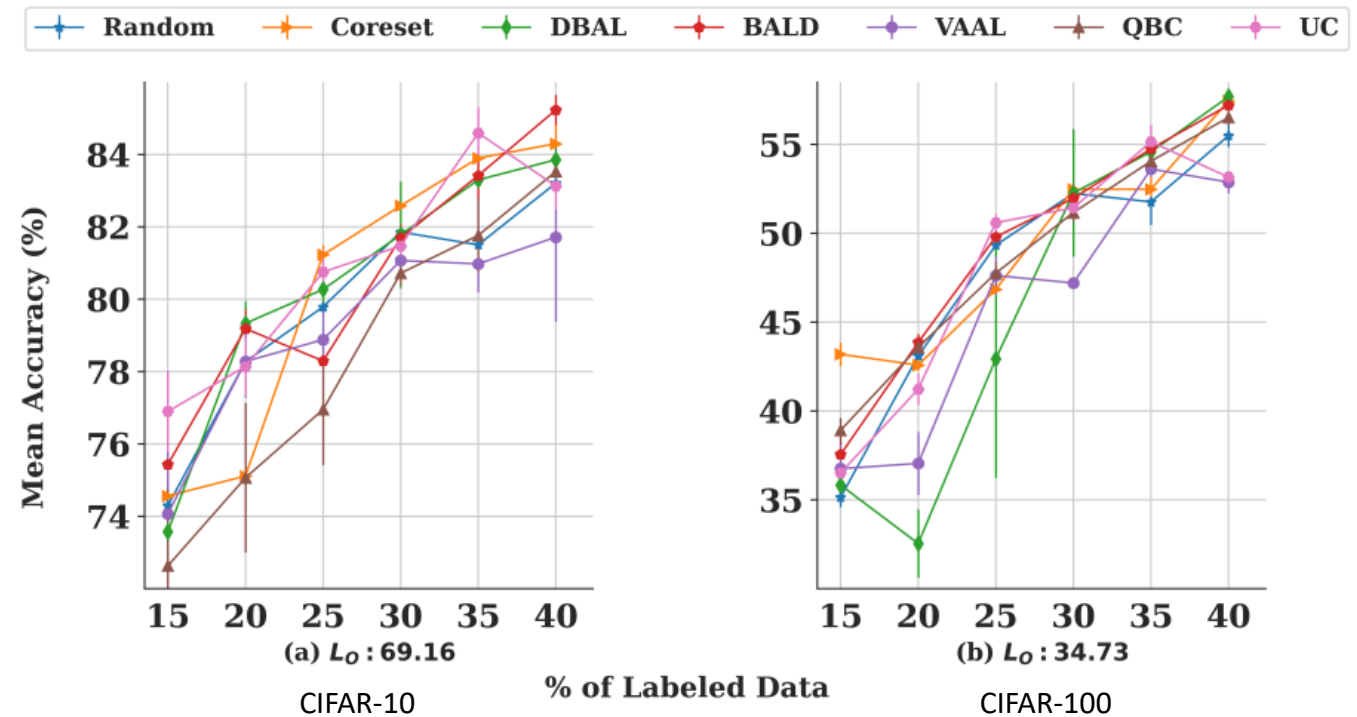
Experiments - Reproducible Setting

- Active Learning은 학습에 사용되는 데이터셋이 매번 변경(계속 증가)되는 상황
- 이러한 상황은 초기 데이터로 정한 하이퍼파라미터는 이후 iteration에 최적이지 아닐 수 있음
- Fig1 : 5번의 random initialize 후, 첫번째 시드 학습 시 AutoML을 사용하여 iteration 마다 50회 random search 하여 best 모델의 파라미터 사용한 실험을 25번 반복한 결과의 평균
- 실험을 통해
 - 모든 실험에서 1등 하는 AL 방법은 존재하지 않음
 - 표준 편차가 1-2% 이기에 각 방법에 대한 성능 차이는 크게 없음
 - 다양한 액티브 러닝 (AL) 방법들과 랜덤 샘플링 기반의 방법론 사이에는 별다른 차이가 존재하지 않음
 - AutoML을 한 이점이 없음(성능 비슷, 복잡성만 증가)



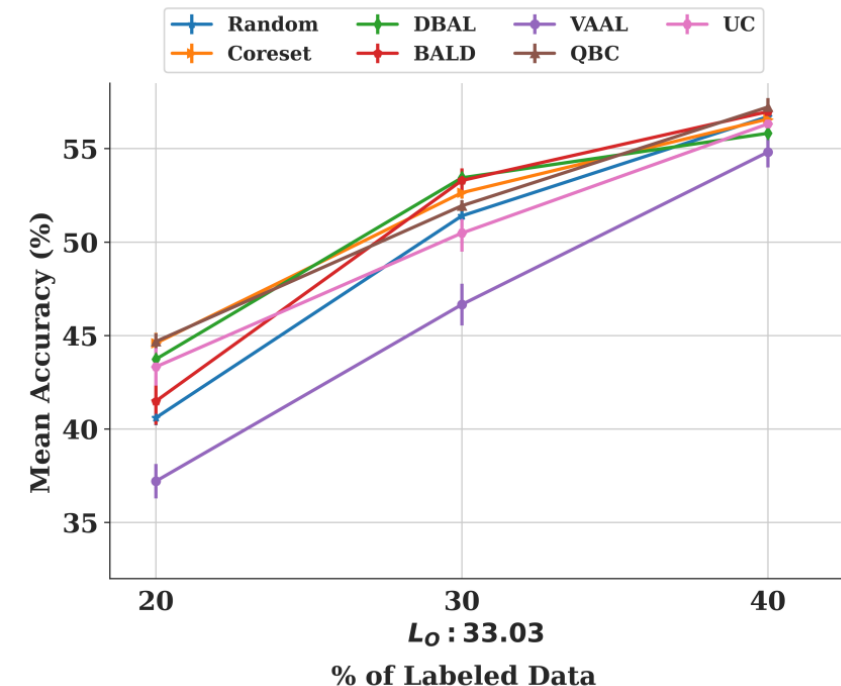
Experiments – Annotation Batch

- 기존의 실험(매 iteration마다 10%씩 추가)와 다르게 5% 씩 추가하는 실험을 진행
- 마찬가지로 각 iteration마다 좋은 방법론이 다르며 랜덤 샘플링보다 안 좋은 기법도 존재
- 추가하는 데이터셋의 크기에 상관없이 기존에 연구된 active learning 방식이 robust하지 않다고 주장



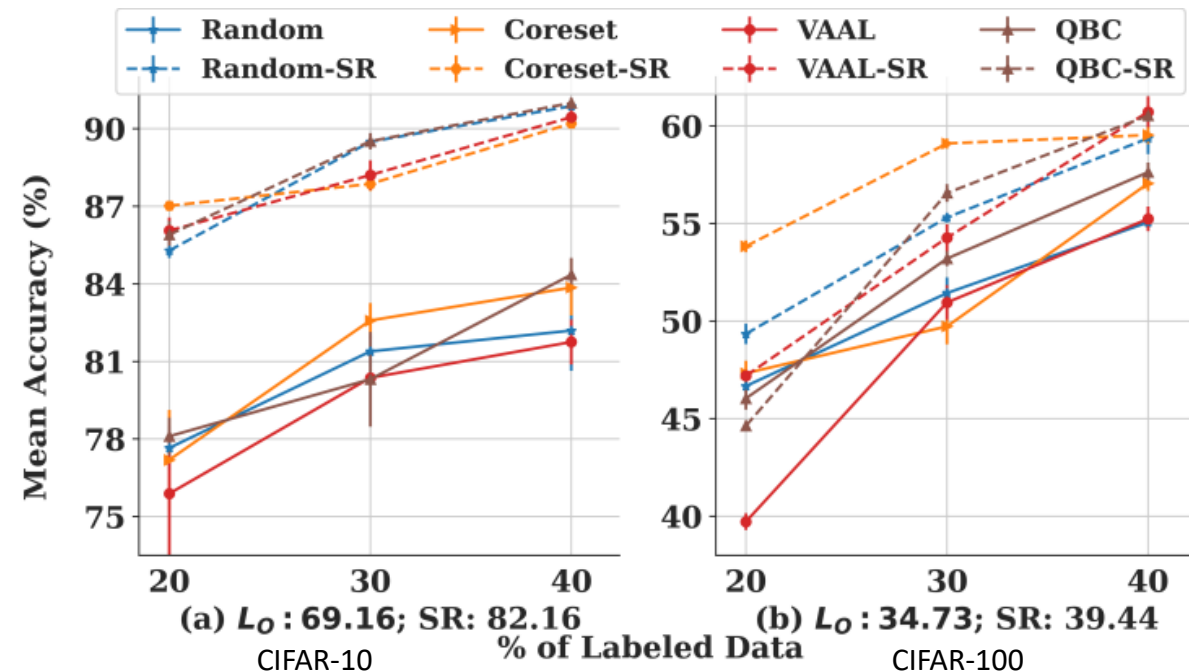
Experiments – Class Imbalance

- long tailed distribution에 대한 active learning을 테스트 하기 위해 CIFAR-100 데이터셋을 변형
- Imbalance 정도 : $\text{samples}[i] = a + b \cdot \exp(\alpha x) / a = 100, x = i + 0.5, \alpha = -0.046, b = 100, i = \text{클래스 순서}$
- 즉, 1번째 클래스는 약 473개, 100번째 클래스는 104개로 4:1의 불균형 비를 가짐
- 실험결과 앞선 실험과 마찬가지로 AL 방법들이 불균형한 클래스 설정에서 일관된 성능을 보이지 않음
- Imbalance 상황일 경우 라벨된 데이터가 많아질수록 오히려 랜덤 샘플링과 active learning 방법 간의 성능 차이가 줄어들음



Experiments – Regularization

- Active learning은 데이터가 부족한 상황을 전제로 하기 때문에 오버 피팅에 취약
- Regularization은 오버 피팅을 막는 역할을 수행해 active learning 성능 향상에 큰 영향을 끼칠 것이라 판단되지만, 기존의 방법론들은 이를 도입하지 않음
- 본 논문에서는 random augmentation과 stochastic weight averaging 기법을 사용하여 사용했을 때의 성능 변화를 관찰 ▷ 실험 결과 정규화 기법은 성능 향상에 크게 기여
- 실험 결과 CIFAR-10에서는 random sampling이 다른 기법들보다 성능이 우수한 것을 확인할 수 있음
- CIFAR-100 또한 random sampling이 좋은 성능을 보이지만 가장 좋은 결과는 아님
- 데이터셋 변화에 따라 방법들의 성능이 바뀌기에 active learning 방법들은 robust 하지 않다 주장



Experiments – Network Architecture

- Active learning back bone network 종류에 따라 성능의 변화에 대한 실험 진행
- CIFAR10에서 ResNet18을 사용할 때 Coreset 방법이 가장 좋지만, ResNet18에 strong regularization을 도입하자 Deep Bayesian 기법이 가장 좋은 성능을 보이고 있음
- 즉, 실험 결과 사용하는 모델 종류에 따라서도 좋은 기법이 다르다는 것을 확인할 수 있음

	Source Architecture						Target Architectures					
Methods	VGG16			WRN-28-2			R18			R18-SR		
	20%	30%	40%	20%	30%	40%	20%	30%	40%	20%	30%	40%
RSB	77.34	80.91	82.05	81.3	81.98	85.68	75.73	79.69	80.69	88.69	90.44	91.8
Coreset	76.18	82.24	85.11	82.32	84.29	86.72	79.27	82.13	85.09	88.48	91.72	92.94
DBAL	78.08	82.42	85.17	80.79	84.54	87.87	75.05	80.69	82.95	89.41	92.23	93.34
VAAL	77.13	79.97	80.55	78.42	83.79	86.58	74.87	79.96	82.24	87.11	90.43	92.19
QBC	78.63	82.66	85.24	81.85	85.06	87.94	76.61	81.77	84.41	88.34	91.37	92.62

Discussion

Under-Reported Baselines

- 여러 active learning 방법들을 재현한 결과 기존에 알려진 결과보다 낮게 나오는 경향을 보이고 있음
- Active learning 방법들은 어떠한 실험 조건 및 환경이 갖춰졌을 때 잘 작동한다는 사실을 알 수 있음

The Role of Regularization

- SGD(확률적 경사 하강법)의 작은 학습률과 함께, Random augmentation과 stochastic weight averaging 가 효과적으로 작용한다는 사실을 알 수 있음

AL Methods Compared To Strong Random Sampling

- 본 연구에서는 실험 환경 및 조건을 잘 세팅한 랜덤 샘플링 방법이 연구된 active learning 방법론들과 거의 유사하거나 조금 떨어지는 수준인 것을 확인

▶ 본 연구를 통해 active learning 구현 시 regularization이 필요하며, 실험 진행 시 민감한 active learning을 동일 선에서 비교하기 위해 일관된 실험 환경 설정 세팅이 필요하다 제언