Variational Adversarial Active Learning

Sayna Ebrahimi*[1], Samarth Sinha*[2], Trevor Darrell[1]
[1]UC Berkeley, [2]University of Toronto

ICCV 2019 Seoul, Korea

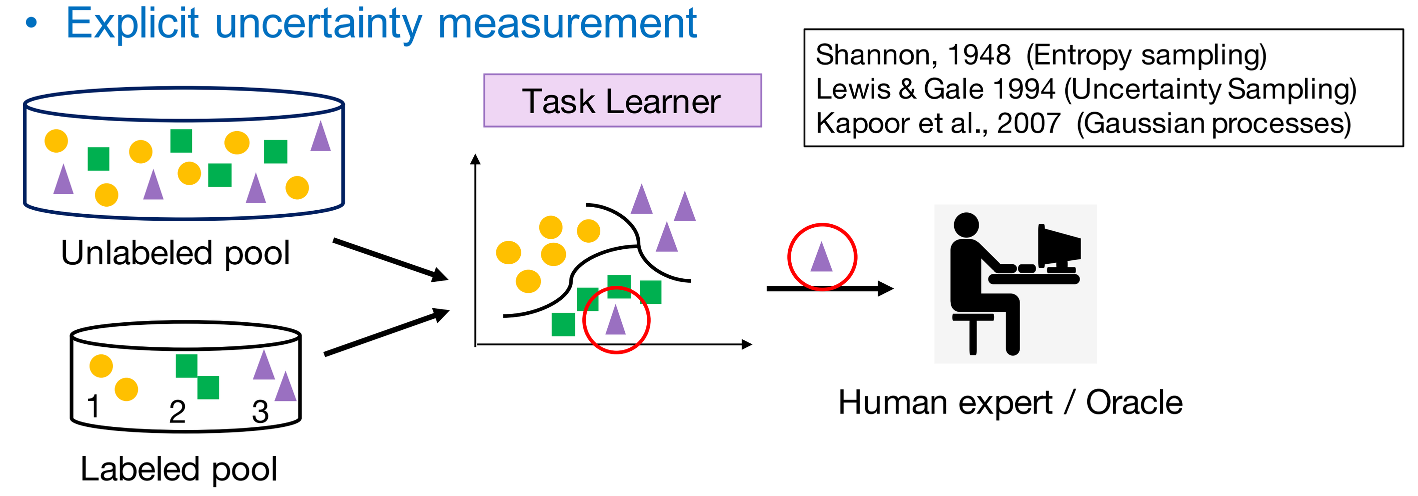
Abstract

- **Motivation:**
  Obtaining labels is expensive and time consuming
- **Problem Statement:**
  Developing a task-agnostic algorithm to querry the most *representative* unlabeled samples for labeling.
- **Strategy:**
  Using adversarial learning to measure *representativesness* of samples without training for the main-stream task.
- **Performance:**
  State-of-the-art on image calssification + semantic segmentation

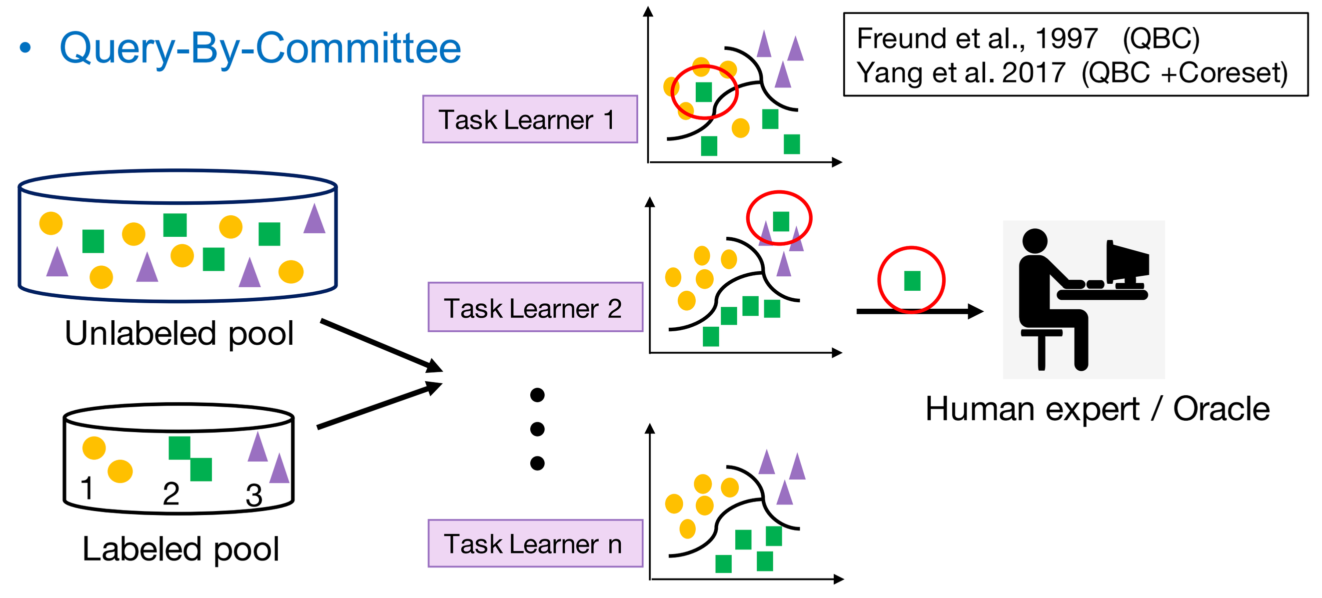
Active Learning

Approaches in active learning have been all *task-dependent.*

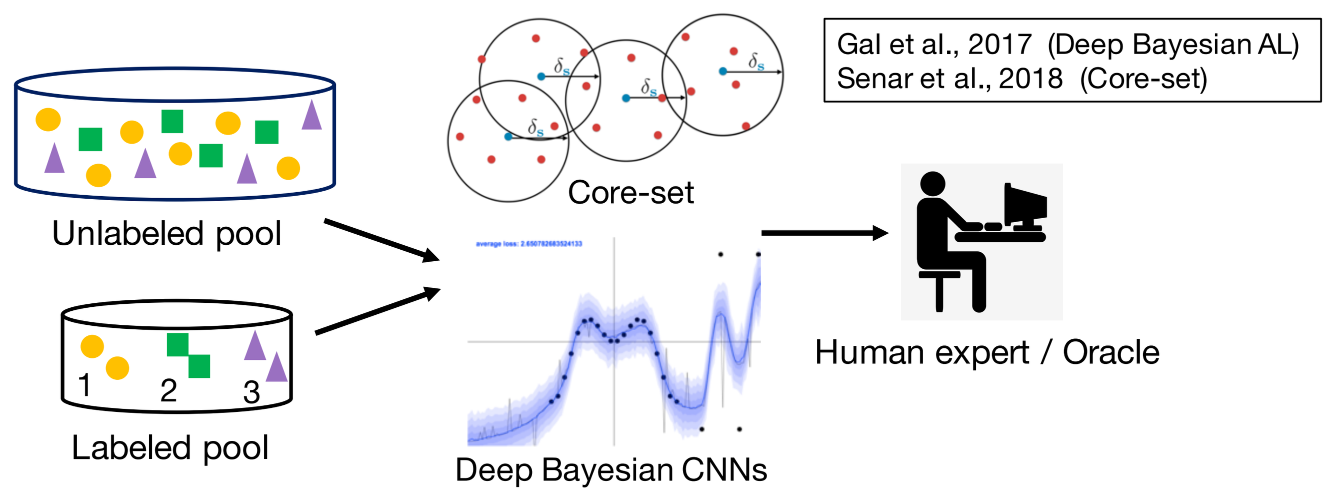
- Explicit uncertainty measurement



- Query-By-Committee



- Representation-based / Bayesian uncertainty



Variational Adversarial Learning (VAAL)

VAE's Objective Function

$$\mathcal{L}_{VAE} = \lambda_1 \mathcal{L}_{\text{VAE}}^{trd} + \lambda_2 \mathcal{L}_{\text{VAE}}^{adv}$$

where:

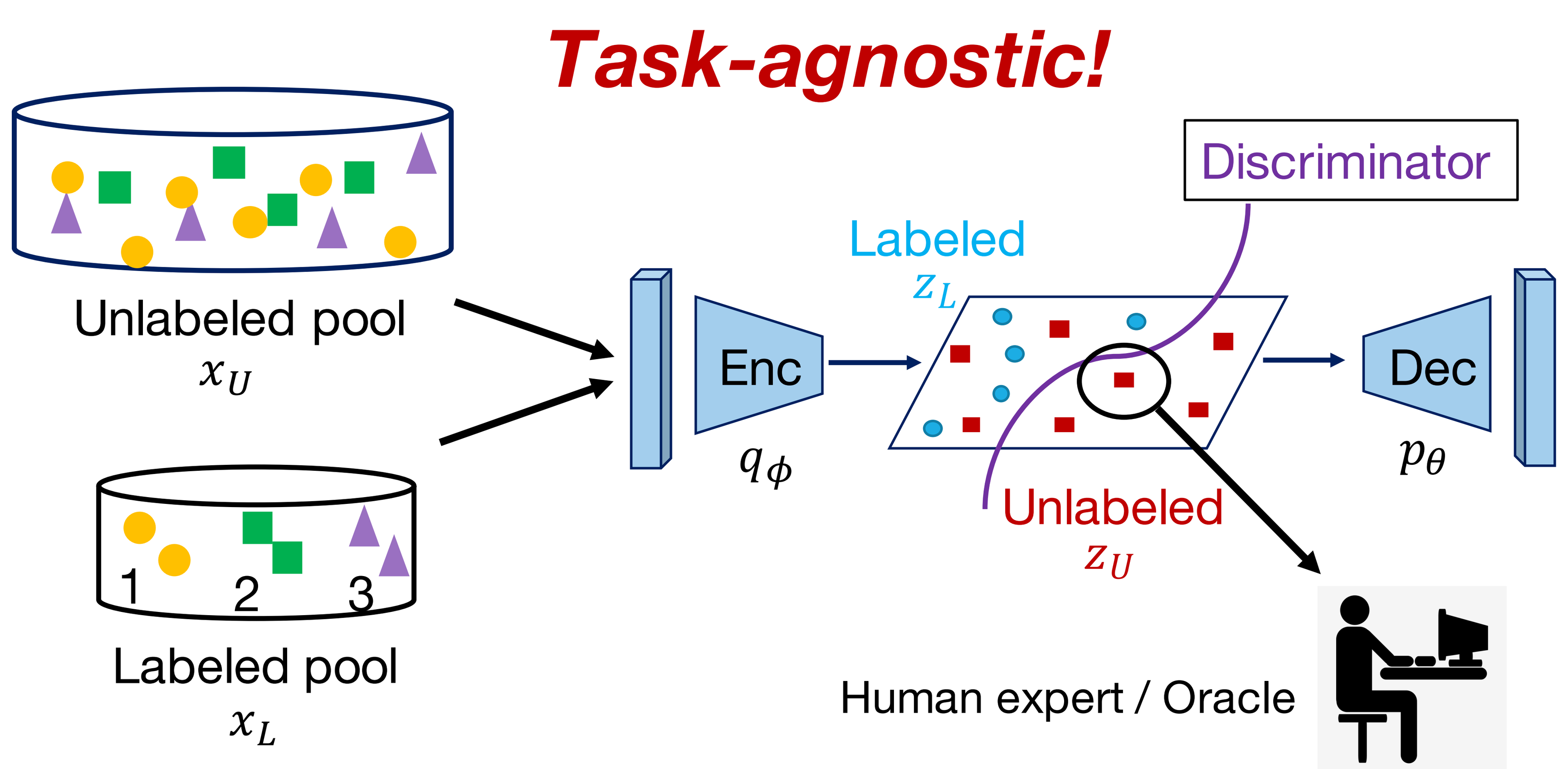
$\mathcal{L}_{\text{VAE}}^{trd}$ → VAE's transductive loss
$\mathcal{L}_{\text{VAE}}^{adv}$ → VAE's adversarial loss
λ_1, λ_2 → Weighting factors

$$\mathcal{L}_{\text{VAE}}^{trd} = \mathbb{E}[\log p_\theta(x_L|z_L)] - \beta\, \mathrm{D}_{\mathrm{KL}}(q_\phi(z_L|x_L)||p(z)) + \mathbb{E}[\log p_\theta(x_U|z_U)] - \beta\, \mathrm{D}_{\mathrm{KL}}(q_\phi(z_U|x_U)||p(z))$$

$$\mathcal{L}_{\text{VAE}}^{adv} = -\mathbb{E}[\log(D(q_\phi(z_L|x_L)))] - \mathbb{E}[\log(D(q_\phi(z_U|x_U)))]$$

Discriminator's Objective Function

$$\mathcal{L}_D = -\mathbb{E}[\log(D(q_\phi(z_L|x_L)))] - \mathbb{E}[\log(1 - D(q_\phi(z_U|x_U)))]$$



Algorithm & Sampling Strategy

Algorithm 1 Variational Adversarial Active Learning

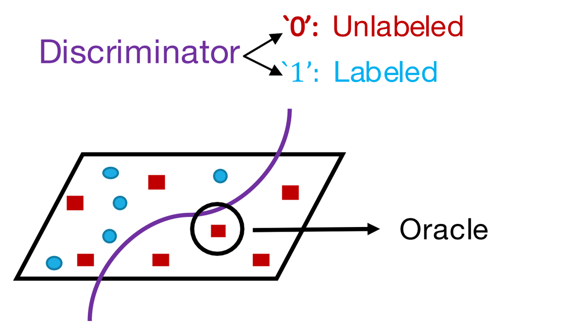
Input: Labeled pool (X_L, Y_L), Unlabeled pool (X_U), Initialized models for θ_T, θ_{VAE}, and θ_D
Input: Hyperparameters: epochs, λ_1, λ_2, α_1, α_2, α_3

1: **for** $e = 1$ to epochs **do**
2: sample $(x_L, y_L) \sim (X_L, Y_L)$
3: sample $x_U \sim X_U$
4: Compute $\mathcal{L}_{\text{VAE}}^{trd}$
5: Compute $\mathcal{L}_{\text{VAE}}^{adv}$
6: $\mathcal{L}_{\text{VAE}} \leftarrow \lambda_1 \mathcal{L}_{\text{VAE}}^{trd} + \lambda_2 \mathcal{L}_{\text{VAE}}^{adv}$
7: Update VAE by descending stochastic gradients:
8: $\theta'_{VAE} \leftarrow \theta_{VAE} - \alpha_1 \nabla \mathcal{L}_{\text{VAE}}$
9: Compute $\mathcal{L}_{\text{D}}$
10: Update D by descending its stochastic gradient:
11: $\theta'_D \leftarrow \theta_D - \alpha_2 \nabla \mathcal{L}_{\text{D}}$
12: Train and update T:
13: $\theta'_T \leftarrow \theta_T - \alpha_3 \nabla \mathcal{L}_{\text{T}}$
14: **end for**
15: **return** Trained $\theta_T, \theta_{VAE}, \theta_D$

Algorithm 2 Sampling Strategy in VAAL

Input: b, X_L, X_U
Output: X_L, X_U

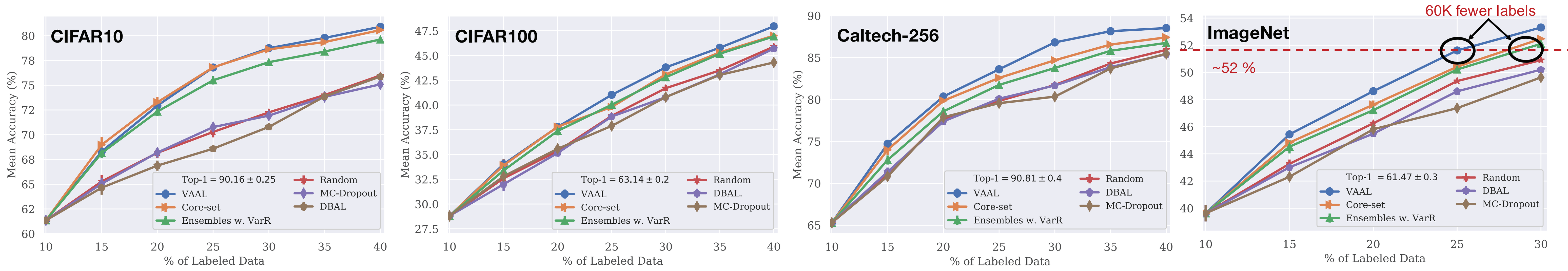
1: Select samples (X_s) with $\min_b\{\theta_D(z_U)\}$
2: $Y_o \leftarrow \mathcal{ORACLE}(X_s)$
3: $(X_L, Y_L) \leftarrow (X_L, Y_L) \cup (X_s, Y_o)$
4: $X_U \leftarrow X_U - X_s$
5: **return** X_L, X_U



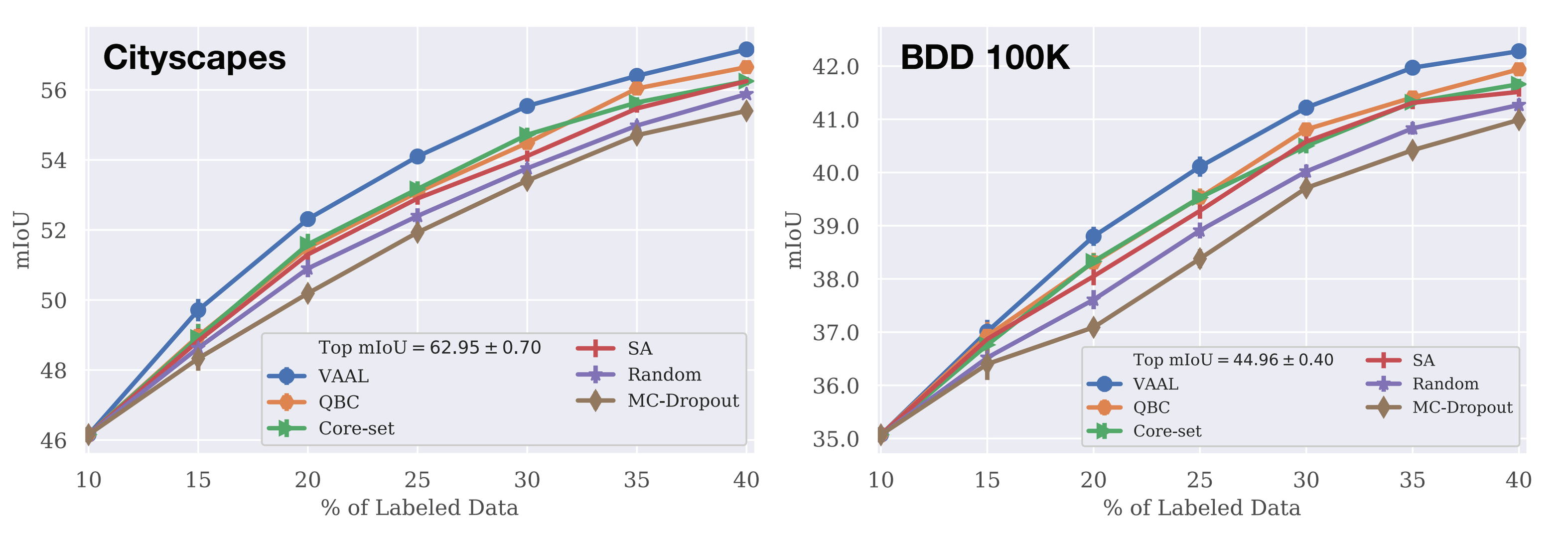
Top k most probable samples predicted as "unlabeled" (with probability closer to 0) will be sent to the oracle

Results on Image Classification and Sematinc Segmentation Benchmarks

- Image Classification Benchmarks



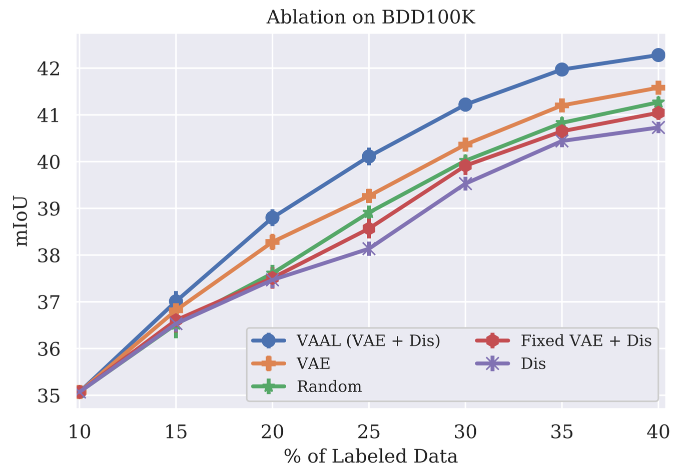
- Semantic Segmentation Benchmarks



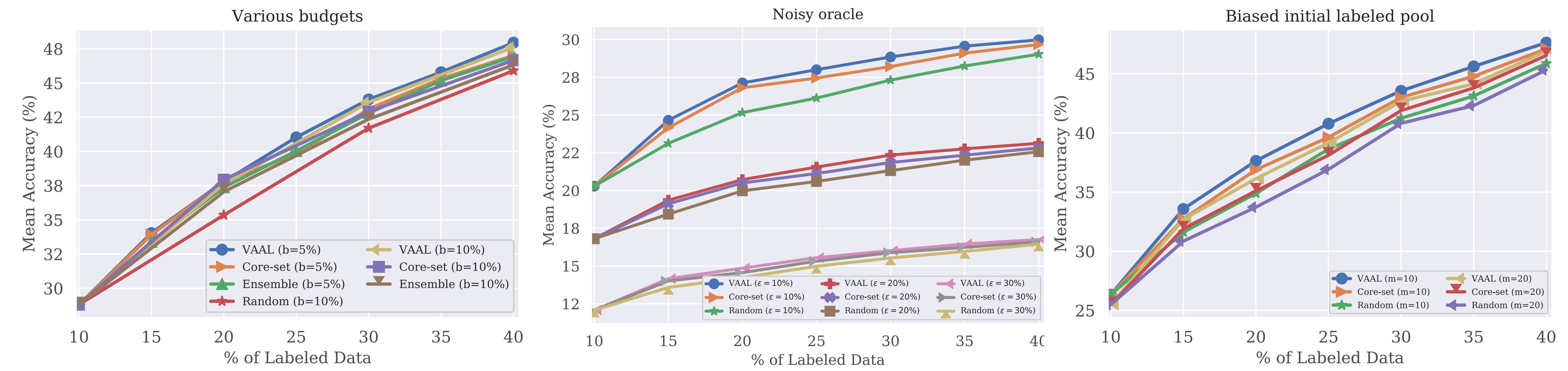
Ablation

Key roles in VAAL in the order of importance:

1. VAE + Discriminator
2. VAE (no Discriminator)
3. Fixed VAE + Discriminator (no adversarial learning) < random
4. Discriminator (no VAE) < random



Robustness & Time Analysis



Method	Wall-clock time (sec)
MC-Dropout (Gal & Ghahramani, '17)	81.05
Core-set (Senar et al. ICLR'18)	75.33
Ensembles using Variation Ratios (Beluch et al. CVPR'18)	20.48
VAAL (ours)	**10.59**

References

[1] Kuo, Weicheng, et al. "Cost-Sensitive Active Learning for Intracranial Hemorrhage Detection." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.
[2] Sener, Ozan, and Silvio Savarese. "Active learning for convolutional neural networks: A core-set approach." ICLR (2017).
[3] Gal, Yarin, Riashat Islam, and Zoubin Ghahramani. "Deep bayesian active learning with image data." ICML (2017)
[4] Beluch, William H., et al. "The power of ensembles for active learning in image classification." CVPR (2018)

Source Code

Code: https://github.com/sinhasam/vaal
Project Page: https://sites.google.com/berkeley.edu/vaal/