Set Based Stochastic Subsampling

Bruno Andreis[1], Seanie Lee, A. Tuan Nguyen[2], Juho Lee[1,3], Eunho Yang[1,3], Sung Ju Hwang[1,3]

[1]KAIST, South Korea

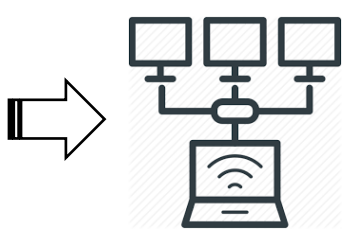
[2]University of Oxford

[3]AITRICS, South Korea

Motivation

Many machine learning algorithms must deal with **huge volumes of data** both in terms of **dimensionality** and the **number of instances**.

Difficult Scene needs analyzing



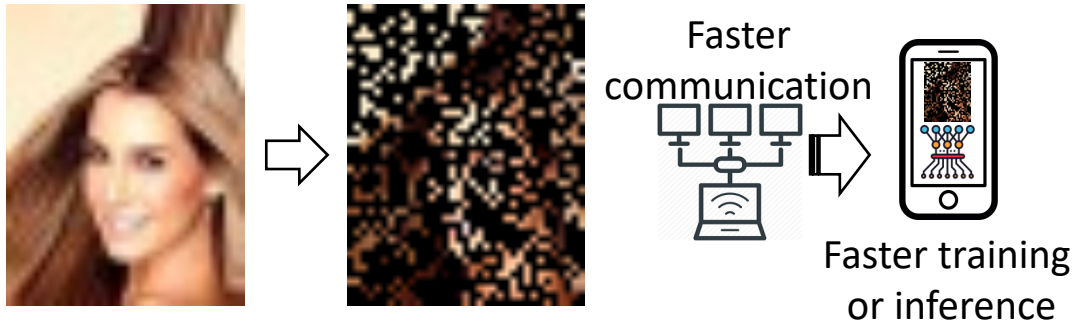
Sending data to Server

Data reduction both at the instance and feature level

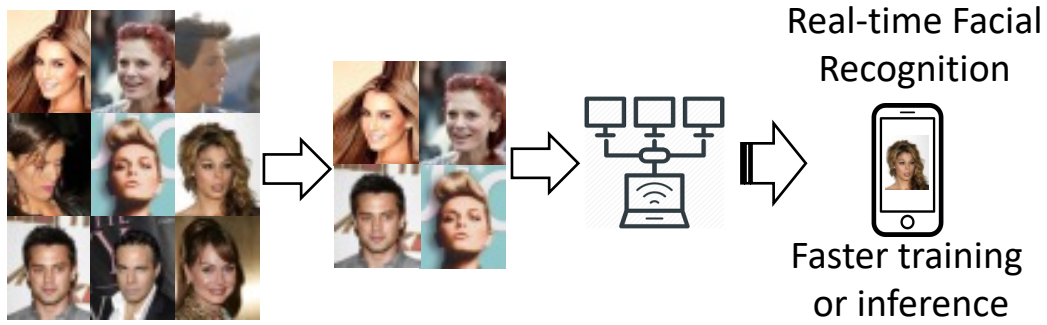
This can serve as a **memory, computational and communication** burden on resource constrained devices as well as real-time applications.

Problem Statement

Given a set $D = \{d_1, d_2, \dots, d_n\}$ we want to select a **core subset** $D_s = \{s_1, s_2, \dots, s_k\}$ that is most representative of the original set.



Feature Selection



Instance Selection

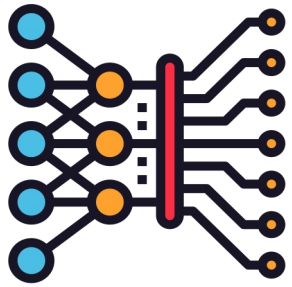
Given that $\ell(\cdot, \mathrm{D})$ is the loss for a task when performing with the whole set D. We aim to select the core subset such that $\ell(\cdot, \mathrm{D_s})$ **approximates** $\ell(\cdot, \mathrm{D})$.

Task Constrained Objective Function

Suppose we have a distribution of sets $p(D)$. For each $D \sim p(D)$, we aim to learn to select a subset $D_s \sim p(D_s|D)$ such that $\ell(\cdot, \mathbf{D_s})$ is **minimized**.

D

D_s



$\ell(\cdot, D_s)$

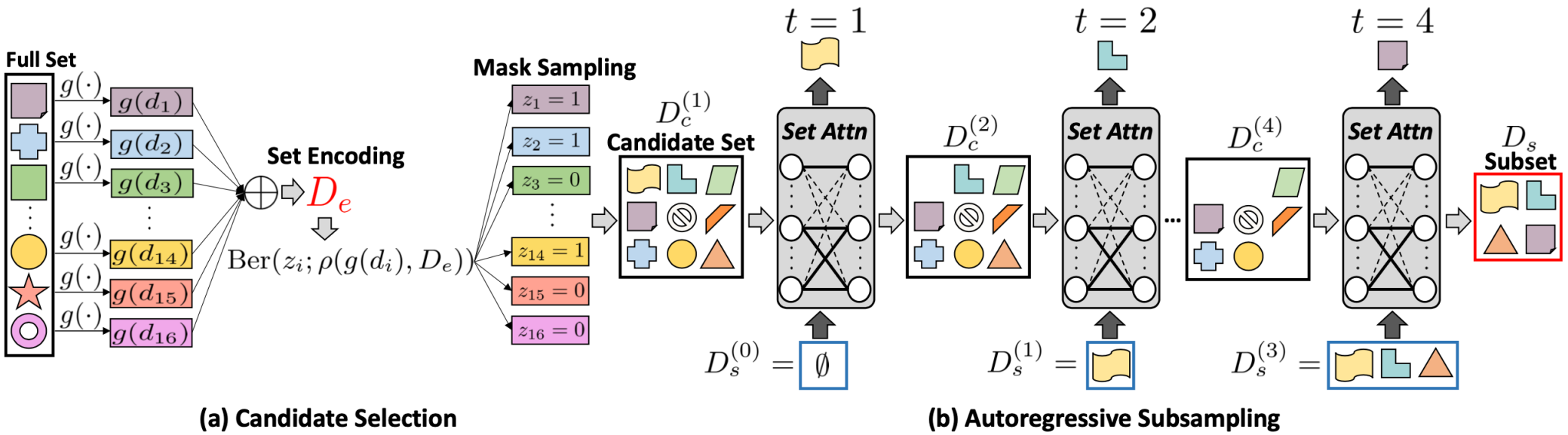
Classification Network

$$\mathbb{E}_{\boldsymbol{p(D)}}\left[\mathbb{E}_{\boldsymbol{p(D_s|D)}}[\ell(\cdot, \boldsymbol{D_s})\ ]\right]$$

This objective function is used as a proxy for the various tasks such as classification, reconstruction, object detection, semantic segmentation etc.

Set Based Two Stage Subset Selection Scheme

Independent selection of subsets results in **redundancy** such as repetition of already selected elements. Hence we introduce **a two stage selection approach**.

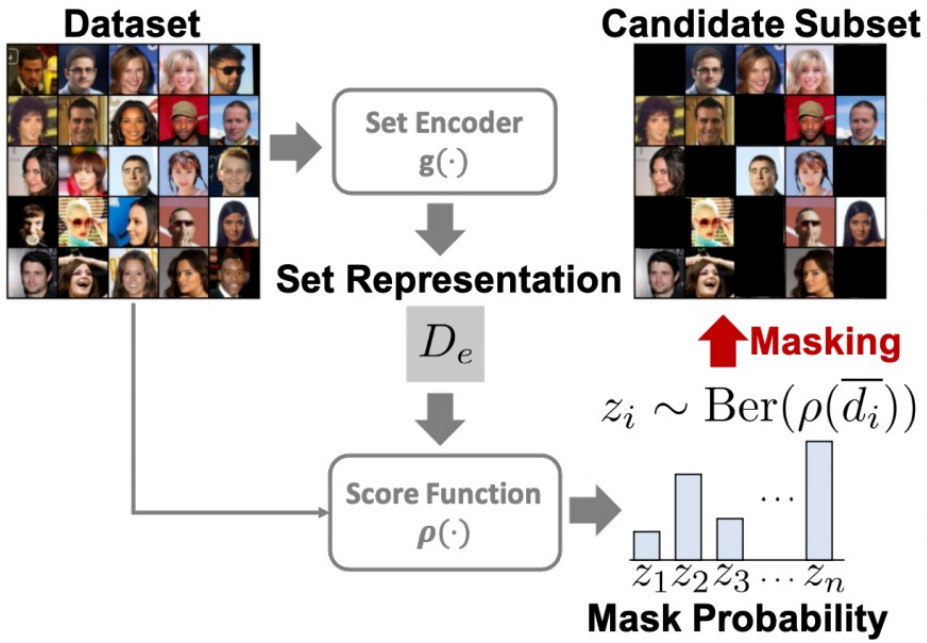


Candidate Selection of potential elements in the given set (Cond. Independent).

Autoregressive Subset Selection selection of the core subset (Cond. Dependent).

Candidate Selection

For a given set $D = \{d_1, d_2, \dots, d_n\}$ a mask $Z = \{z_1, z_2, \dots, z_n\}$ is computed that indicates which elements belong in the candidate set.

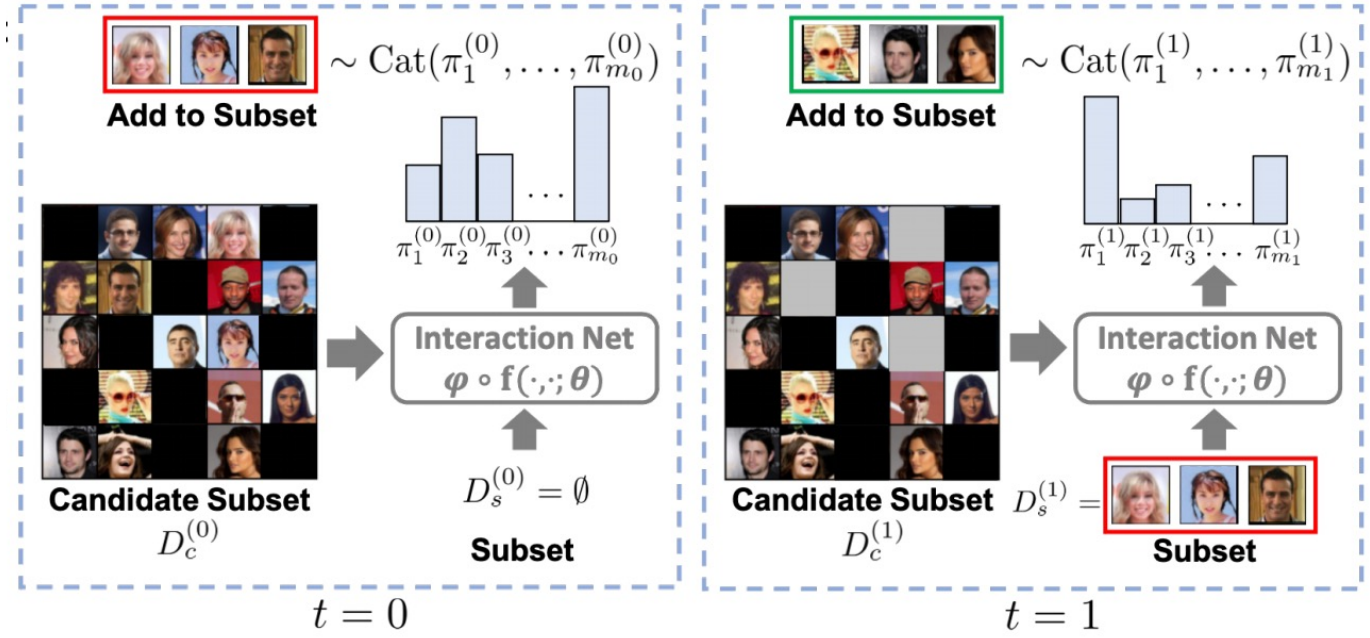


$$p(z_i|d_i, D) = \mathrm{Ber}\Big(z_i; \rho\big(d_i, r(D)\big)\Big)$$

The latent variables for the mask $\{z_1, z_2, \dots, z_n\}$ are **conditionally independent** given the original set D.

Autoregressive Subset Selection

For a given candidate set, we iteratively select the elements of the core subset from $D_c \setminus D_s^{(i-1)} = \{w_1, w_2, \ldots, w_m\}$.



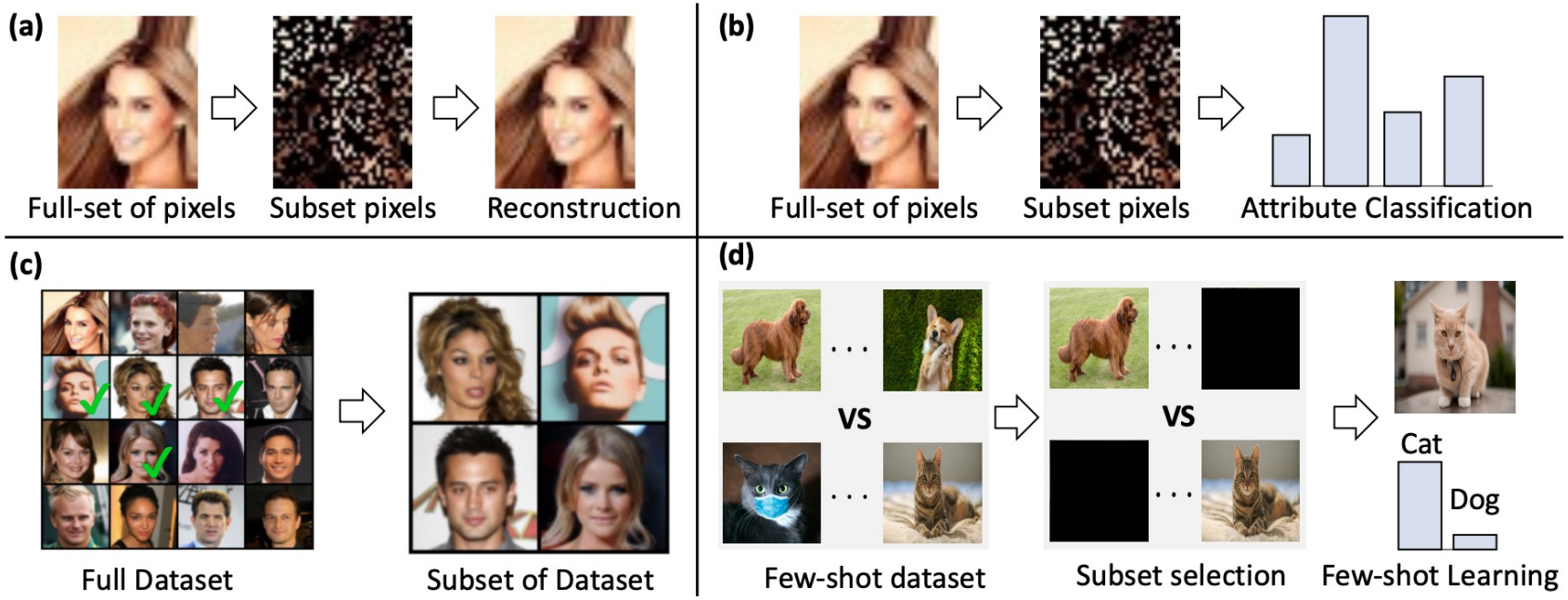
$$p\left(s_i = w_j | D_c, D_s^{(i-1)}\right) = p_j$$

$$p_j = \frac{f\left(w_j, D_c, D_s^{(i-1)}\right)}{\sum_{j'=1}^{m} f\left(w_{j'}, D_c, D_s^{(i-1)}\right)}$$

This stage models the interactions between the elements of the set and performs selection that avoids redundancy.

Stochastic Subset Selection: Applications

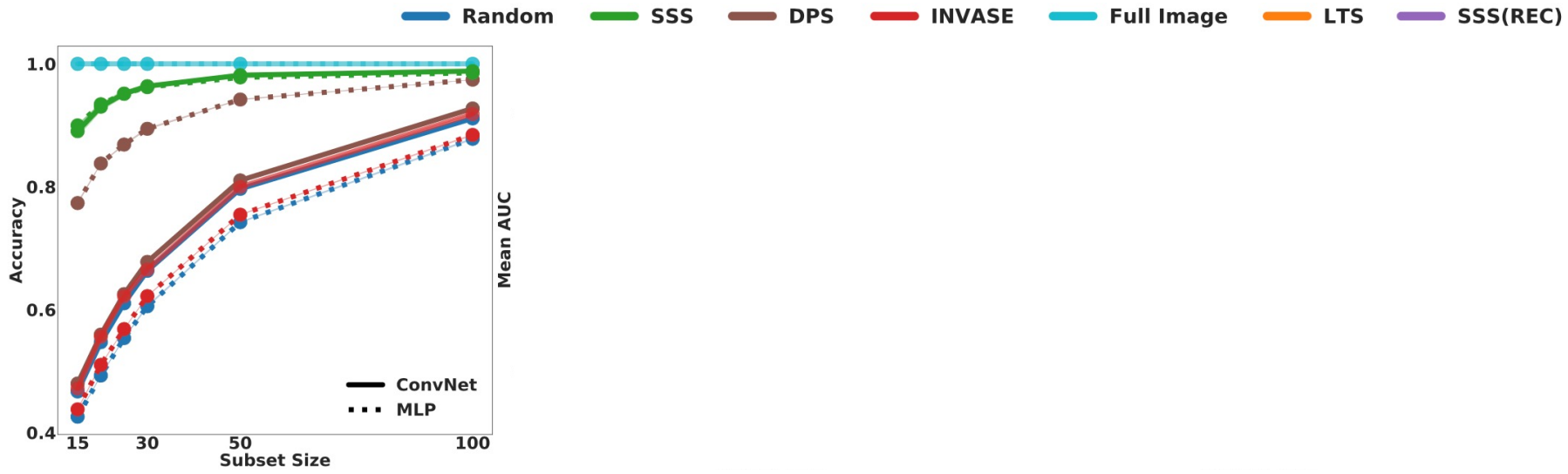
We apply stochastic subset selection to various applications such as set reconstruction, classification and dataset distillation(DD) for both feature and instance selection.



In all the applications, we use the selected subset as a proxy for the full set and constrain it with the task objective.

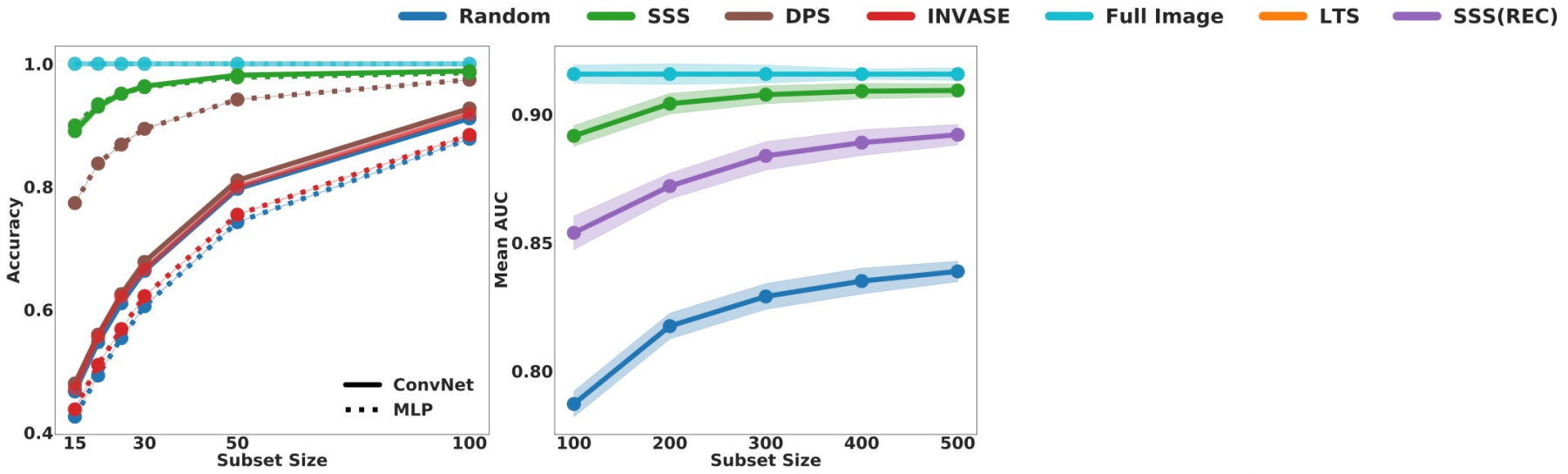
Feature Selection

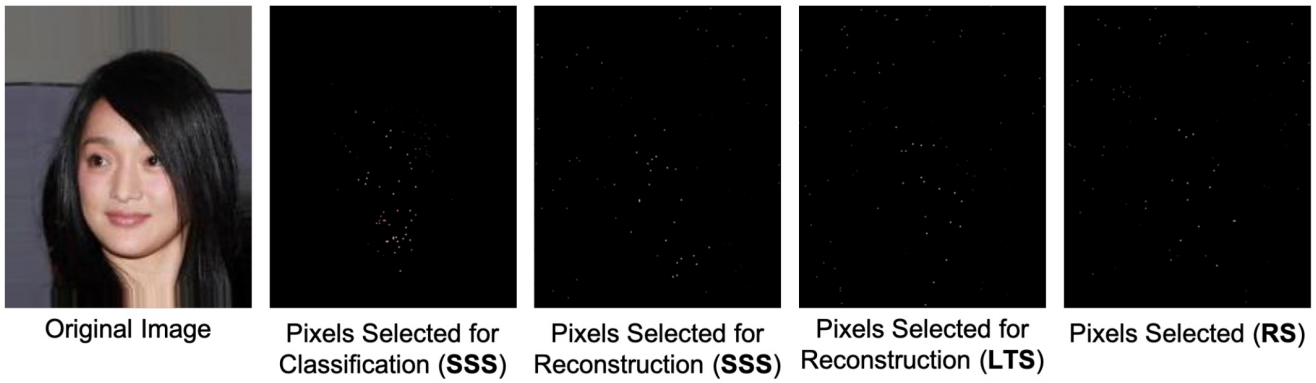
We compare our model with random sampling, learning to sample, DPS, INVASE as well using the full input set on reconstruction and classification tasks.



Feature Selection

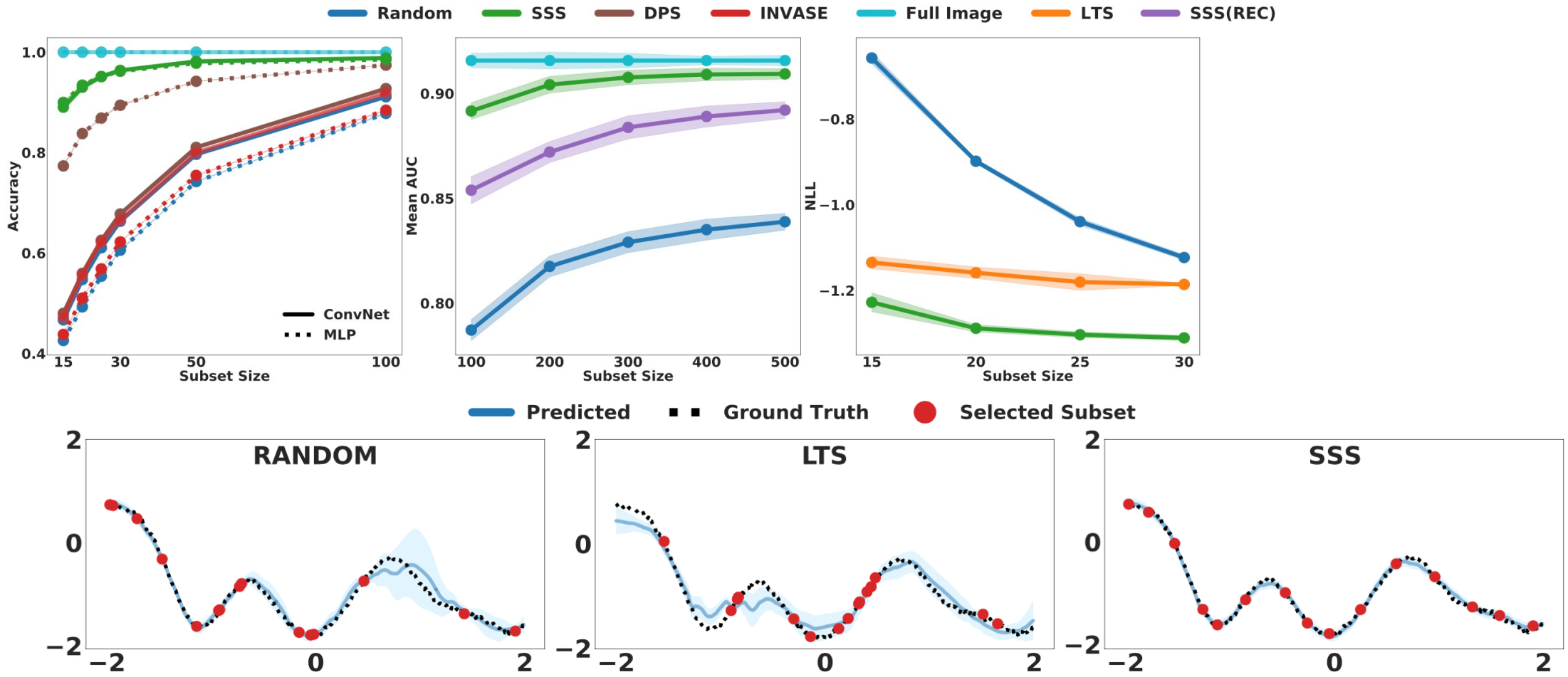
We compare our model with random sampling, learning to sample, DPS, INVASE as well using the full input set on reconstruction and classification tasks.





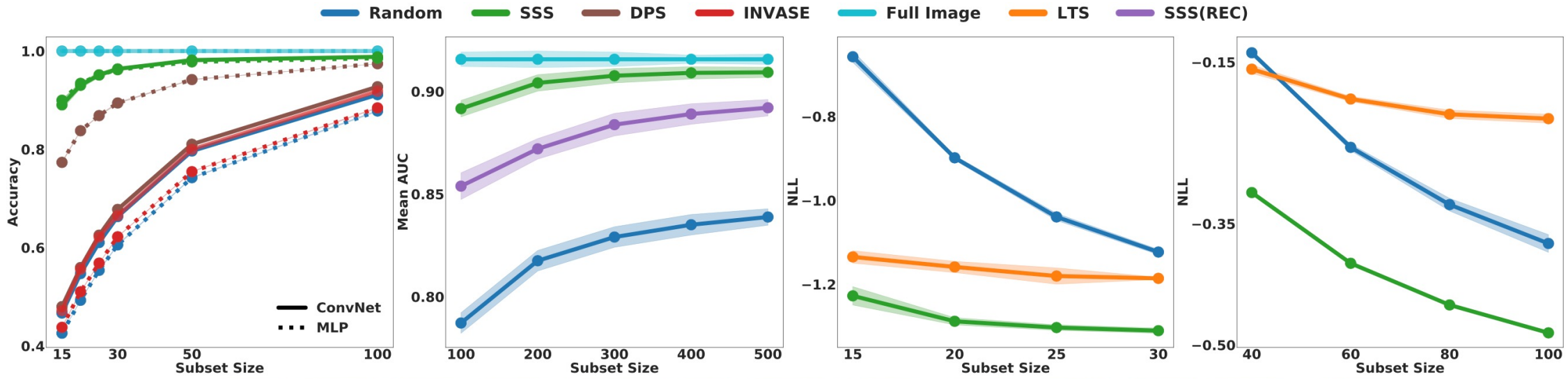
Feature Selection

We compare our model with random sampling, learning to sample, DPS, INVASE as well using the full input set on reconstruction and classification tasks.



Feature Selection

We compare our model with random sampling, learning to sample, DPS, INVASE as well using the full input set on reconstruction and classification tasks.



Instance Selection

We perform experiments on both classification and reconstruction based on a few selected instances and compare with FPS and Random Sampling.

#Instances	2	5	10	15	20	30
FPS	6.50	4.51	3.07	2.75	2.71	2.29
Random	3.73	1.16	0.90	0.38	0.39	0.20
SSS	**2.53**	**1.02**	**0.59**	**0.33**	**0.24**	**0.17**

FID score on Mini-ImageNet (Generation)

#Instances	1	2	5	10
FPS	0.432	0.501	0.598	0.636
Random	0.444	0.525	0.618	0.663
SSS	**0.475**	**0.545**	**0.625**	**0.664**

Accuracy on Mini-ImageNet (Classification)

Selected Instances from a randomly constructed set of size 200