Image Corpus Representative Summarization

Anurag Singh, Lakshay Virmani & A.V. Subramanyam*

Netaji Subhas University of Technology, Delhi, India & Indraprastha Institute of Information Technology*, Delhi, India

anurags.it@nsit.net.in, lakshayv.it@nsit.net.in, subramanyam@iiitd.ac.in

1. Introduction

The selection of representatives from a corpus of images in an essential requirement for efficient representation, navigation and exploration.

Applications

- Web image collection for e-commerce, tourism and travel exploration, story-telling from personal album collections, online image recommendation systems.
- Summarization of a dataset can help train models without trading-off much on accuracy as the diversity of data is maintained while saving huge computational resources.

2. Problem Overview and Challenges

- Given a collection of images, we aim to find a subset summary of these images.
- The problem is challenging because:
  - A good summary must cover various aspects of an image set such as relevance and diversity.
  - Redundancy is very hard to find and learn in an image corpus. Unlike in videos, there is no temporal relationship among images in a set.

3. Proposed Algorithm

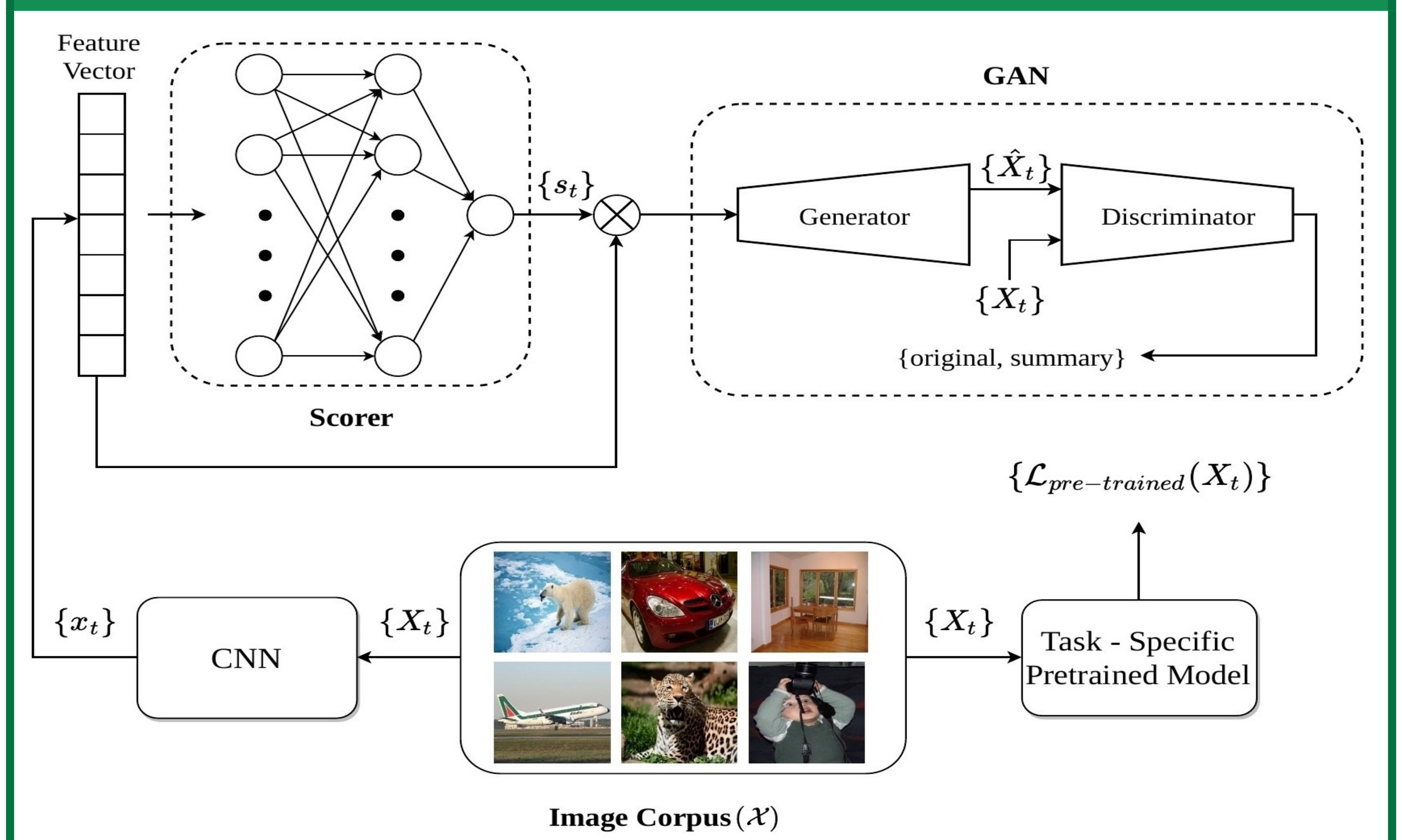

Figure 1: Architecture of the Proposed Model

- The scorer assigns a relative importance score to each image such that higher the score, more the likelihood of the image being present in the summary.
- The pre-trained model is used when a task specific summary needs to be generated.

4. Training The Model

- Losses used: Reconstruction Loss $\mathcal{L}_{reconstruct}$, Loss of GAN $\mathcal{L}_{GAN}$, Regularization Loss $\mathcal{L}_{sparsity}$ and Task-Specific Loss $\mathcal{L}_{task-specific}$.
- The SUM_{gen} variant of our model uses $\mathcal{L}_{reconstruct} + \mathcal{L}_{GAN} + \mathcal{L}_{LR}$, while SUM_{gen}^{DPP} uses $\mathcal{L}_{reconstruct} + \mathcal{L}_{LR} + \mathcal{L}_{DPP} + \mathcal{L}_{GAN}$ and SUM_{task} employs $\mathcal{L}_{reconstruct} + \mathcal{L}_{LR} + \mathcal{L}_{GAN} + \mathcal{L}_{task-specific}$ for training.

Regularization Loss

This loss regularizes the number of images that form the summary.

$$\mathcal{L}_{sparsity} = \mathcal{L}_{LR} + \delta \mathcal{L}_{DPP} \tag{1}$$

where,

$$\mathcal{L}_{LR} = \left\| \frac{1}{n} \sum_{t=1}^{n} s_t - \sigma \right\|_2$$

where $\delta \in \{0, 1\}$ and $\mathcal{L}_{DPP}$ is Determinantal Point Process (DPP) loss [1].

Task-Specific Loss

A task specific summary would be used to perform certain task. In the following, we assume a task specific summary where the task is classification.

$$\mathcal{L}_{task-specific} = \frac{(1-s)\mathcal{L}_{pre-trained}(X)}{\beta} \tag{2}$$

where $\mathcal{L}_{pre-trained}(X)$ is the loss obtained from the task specific pre-trained model and β is a hyper-parameter.

5. Results

Variant	SUM_{gen}	SUM_{gen}^{DPP}	SUM_{task}
CIFAR100	0.307	0.292	**0.274**
VOC	0.565	0.546	**0.542**

Table 1: Reconstruction Error for different variants of our model at $\sigma = 0.1$

Method	Original	KMeans	SSDS	HyperSphere	Ours
CIFAR10	***89.12***	78.24	79.34	79.13	**80.61**
AwA2	***92.50***	87.35	88.61	88.90	**89.50**

Table 2: Comparison results of Classification Accuracy at $\sigma = 0.1$

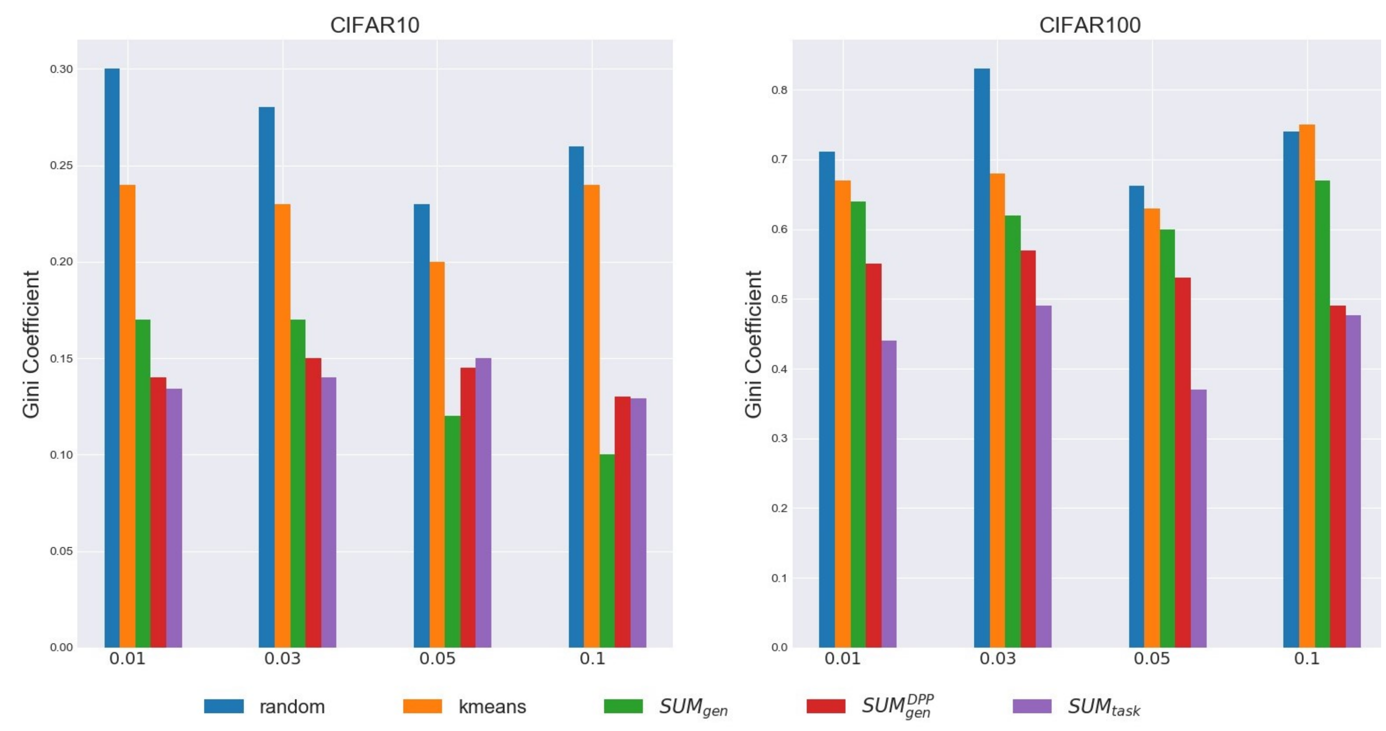

Figure 2: Gini index for different datasets and σ values. Proposed model gives best (lowest) Gini index compared to K-means and random methods.

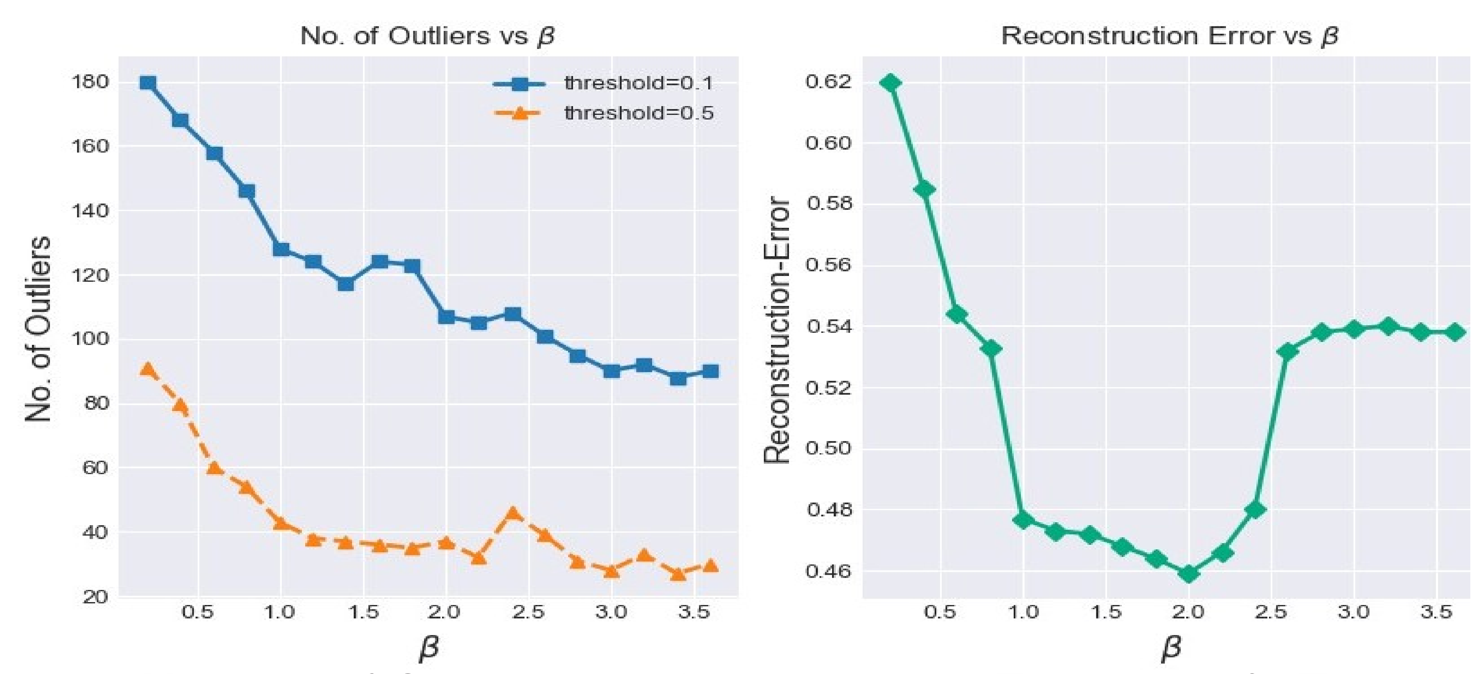

Figure 3: Variance of Outliers and Reconstruction Error with β. The thresholds used here are 0.1 and 0.5, all images with cross-entropy loss ($\mathcal{L}_{pre-trained}$) greater than the threshold are considered to be outliers.

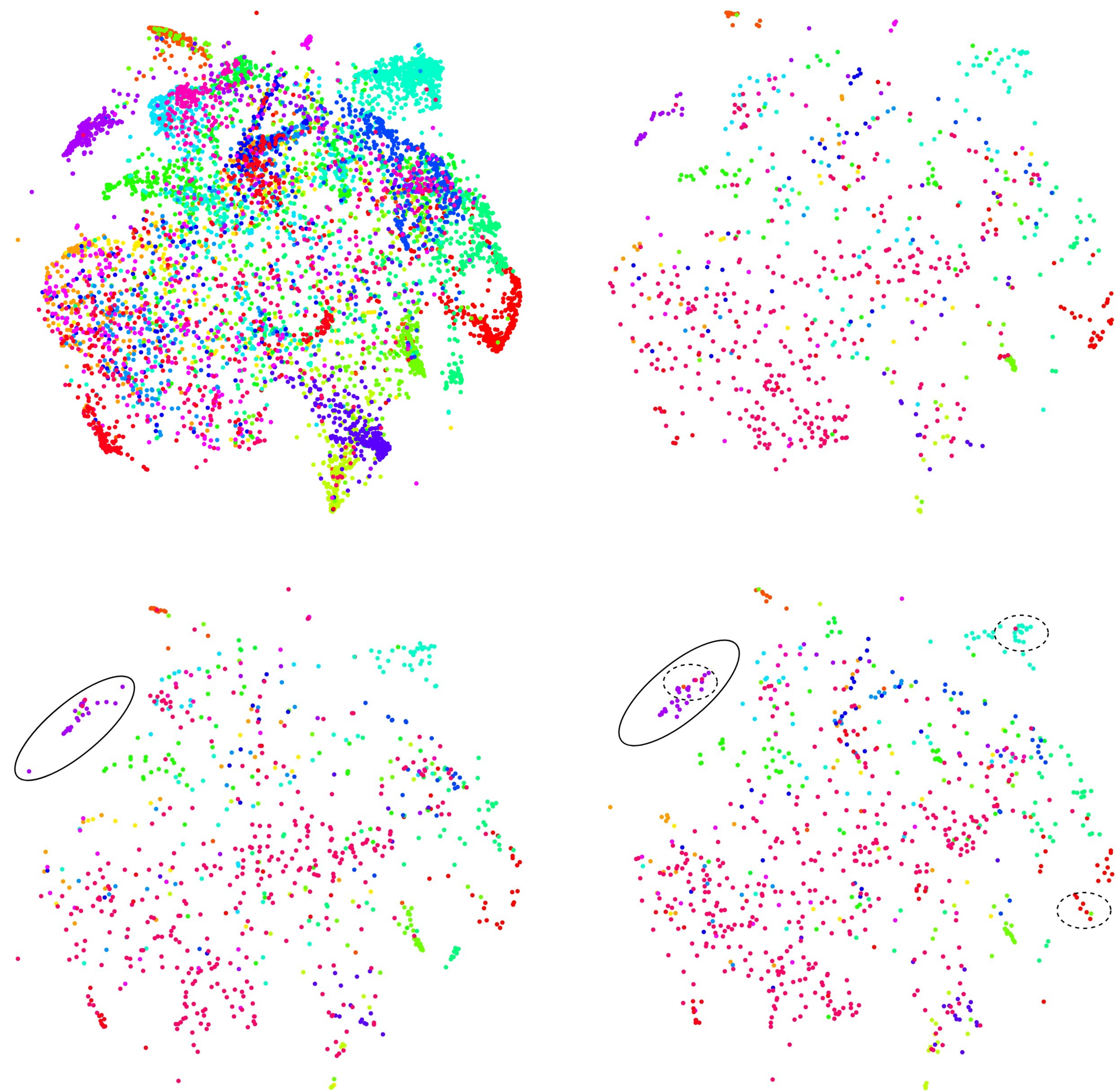

Figure 4: t-SNE plot for VOC2012. Full dataset, summary at 5% with $SUMgen$, SUM_{gen}^{DPP} and SUM_{task} variants. Different colors represent different classes.

6. References

[1] Alex Kulesza, Ben Taskar, et al. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.