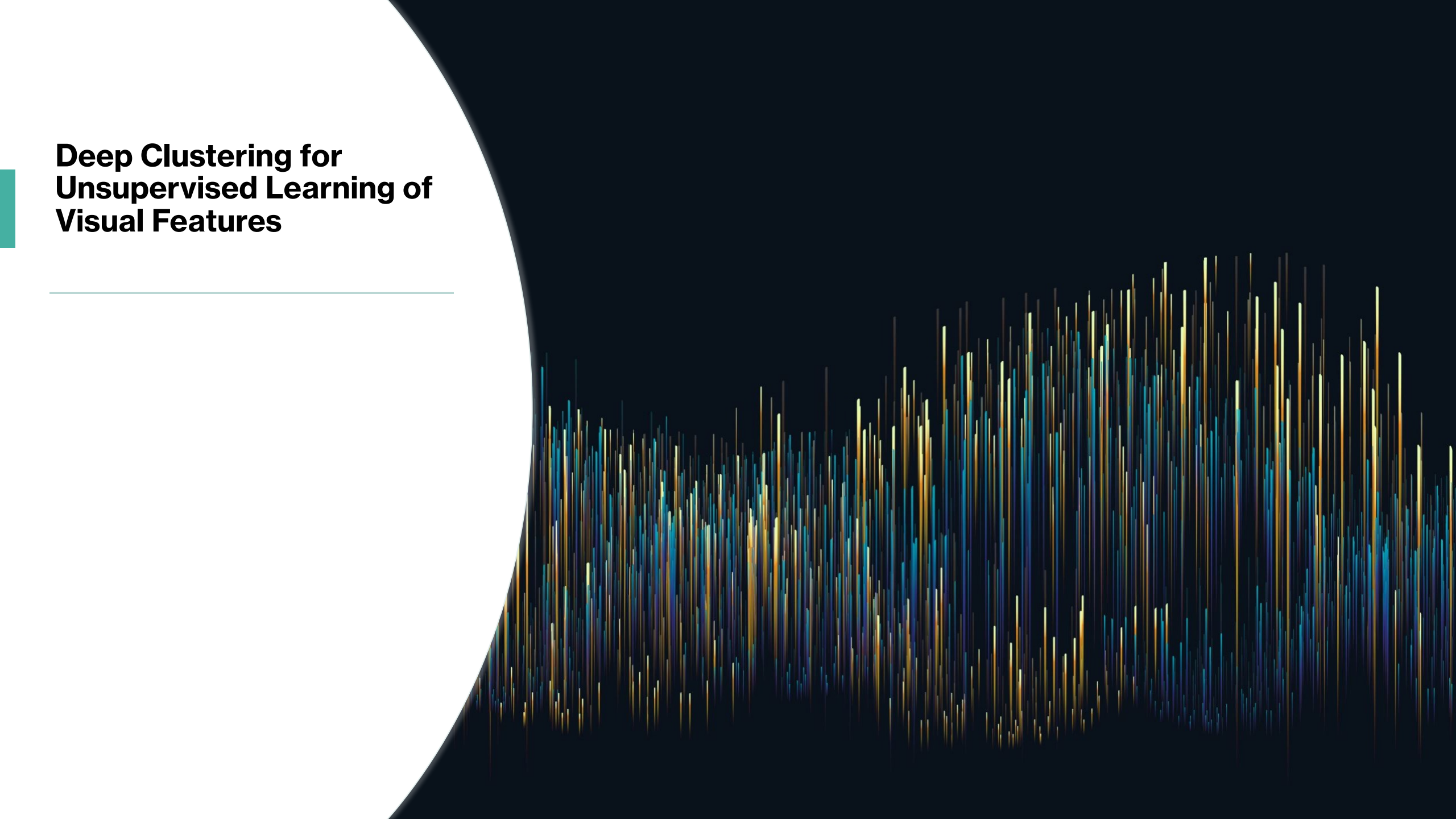
Deep Clustering for Unsupervised Learning of Visual Features



Motivation

01

Traditional supervised learning methods need large annotated datasets.

02

Annotations are done manually.

03

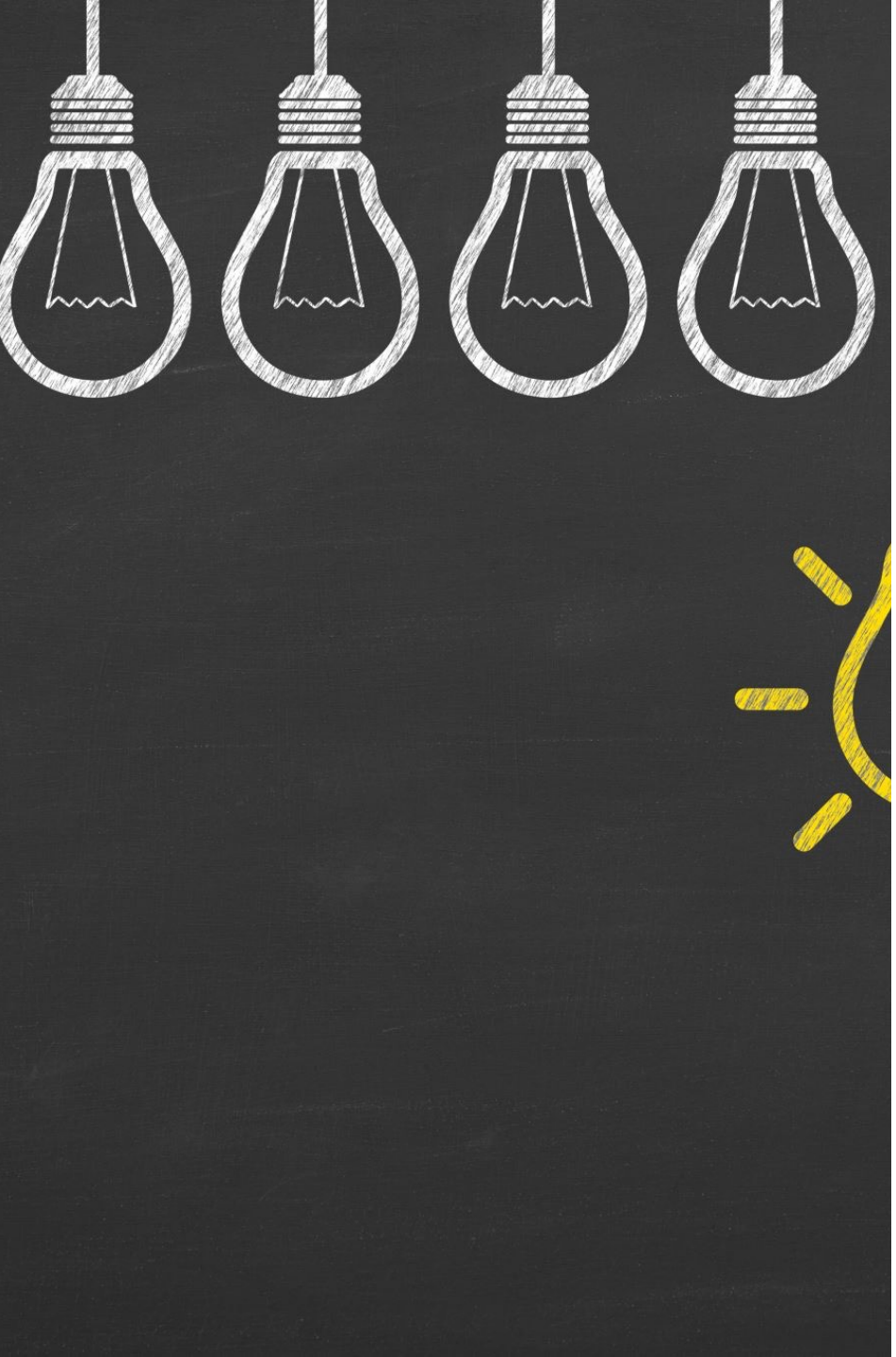
Missing annotations hinders the learning

Problem Statement

- Most deep learning models require a large amount of labelled data to train, which are tough and time consuming to obtain.
- Unsupervised learning can be an alternative approach that leverages unlabelled data to learn useful representations of visual features.
- Problem is to develop a deep clustering method that can effectively learn visual features without the need for annotations.

Solution

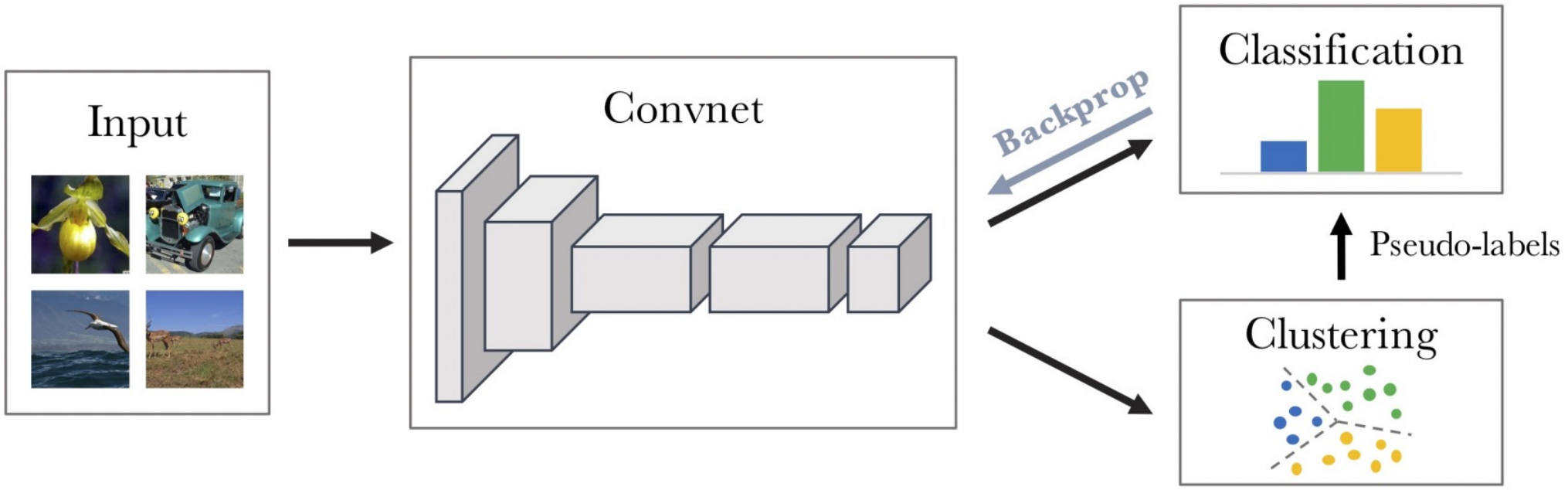
- Use clustering algorithm such as Kmeans, to cluster image representations.
- Use these cluster assignments as pseudo-labels for the data and train the model end-to-end.



Scope

- Adapt unsupervised learning to the end-to-end training of visual features.
- Develop DeepCluster, a clustering method that jointly learns the parameters of the network and the cluster assignments of the resulting features.
- DeepCluster iteratively groups the features with a standard clustering algorithm and uses these cluster assignments as labels.

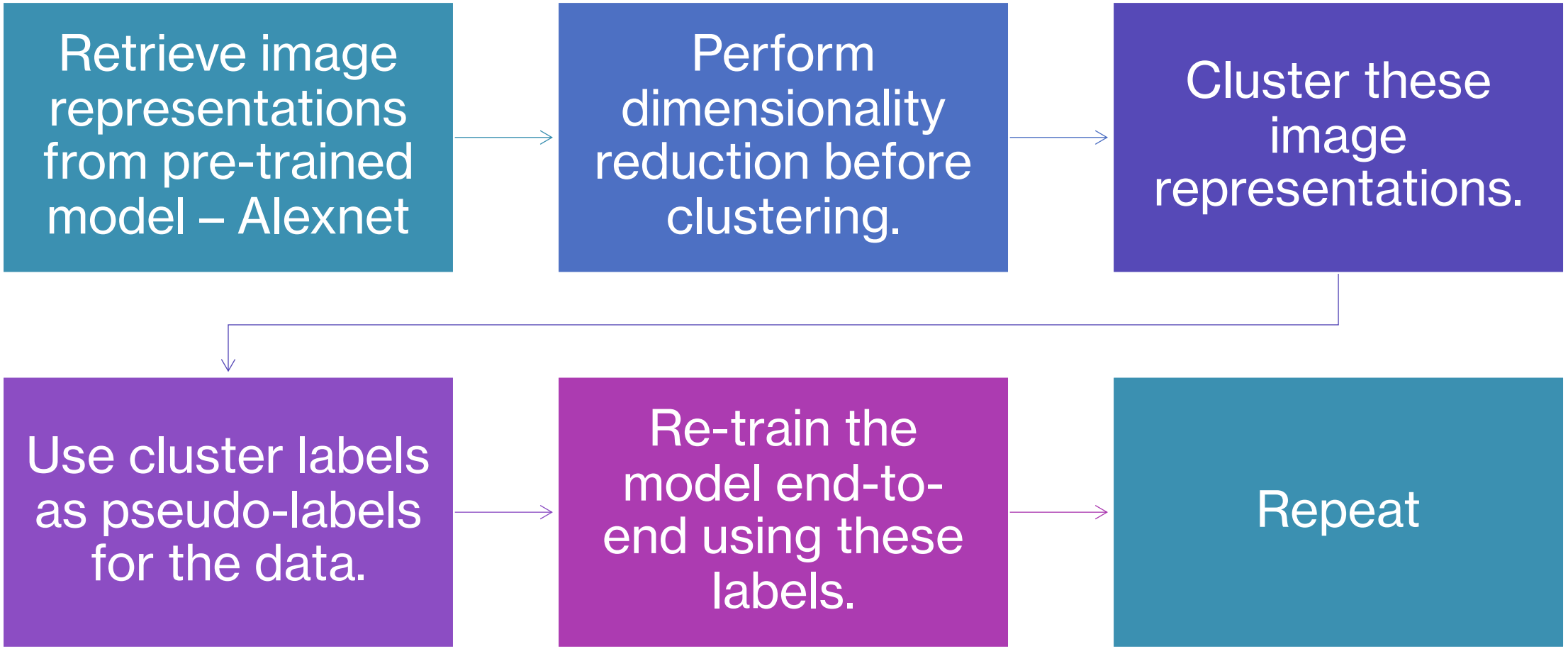
Deep Clustering for Unsupervised Learning of Visual Features



Components

- Dataset – CIFAR10
- Convnet – Alexnet
- Clustering Algorithm – Kmeans
- Dimensionality Reduction - PCA

Implementation



Experiments

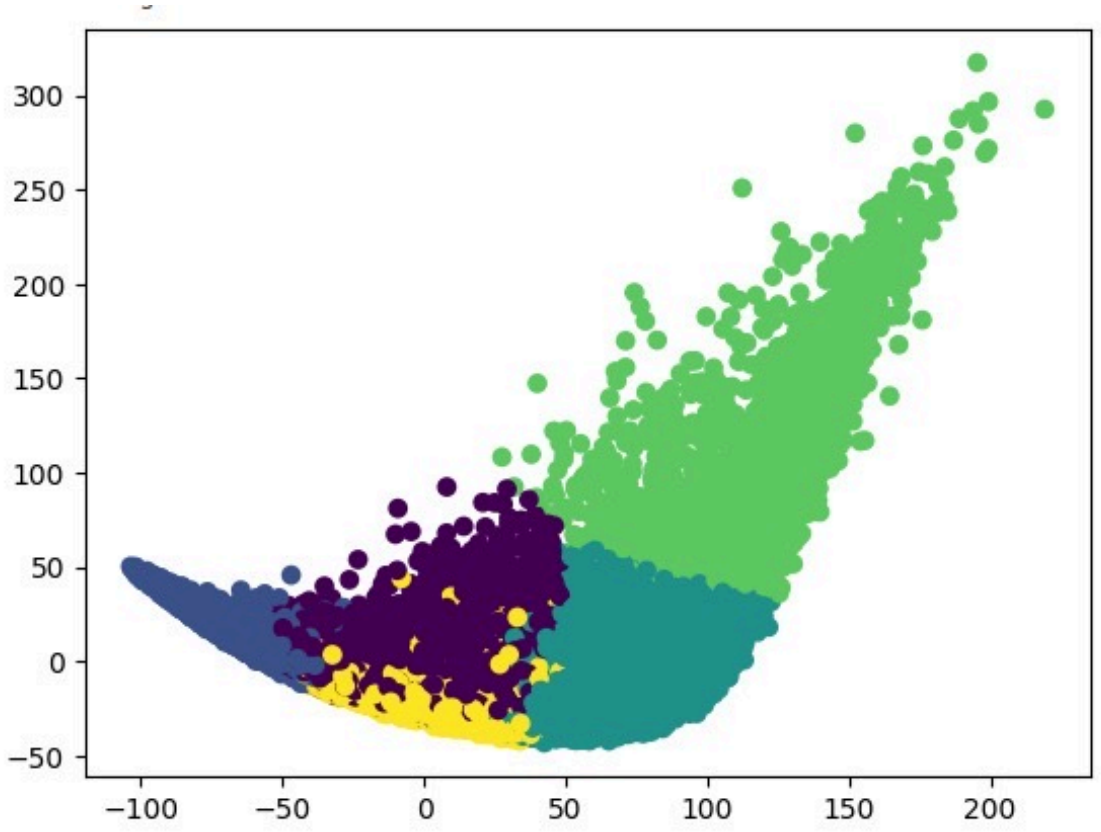
- Number of epochs
- Number of reassignments between epochs
- Choosing the number of clusters
- Relation between clusters and labels (NMI)

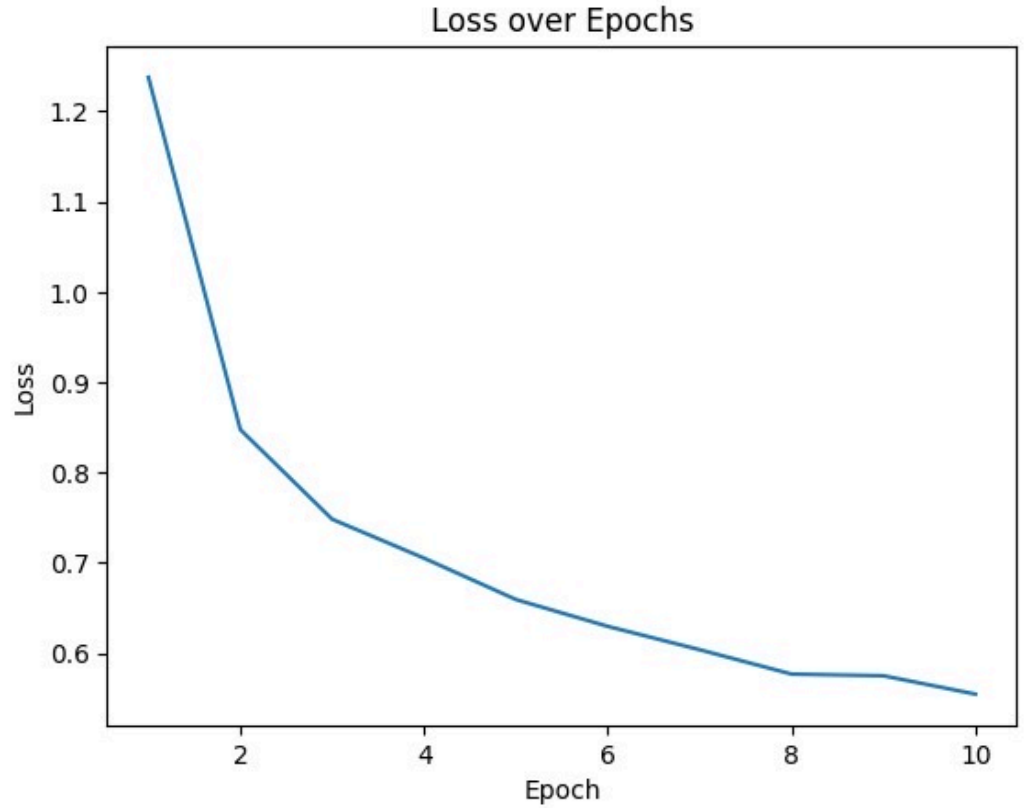
Results / Analysis

Number of clusters

Clusters	Epochs	Test Loss	Accuracy (%)
5	10	0.002	91
7	3	0.003	84
8	10	0.003	86
10	10	0.005	79

Best Result (k = 5)






```
Test set: Average loss: 0.002, Accuracy: 9128/10000 (91%)
-------------- ITERATION 5 for K = 5 done! ------------------
```

Results / Analysis

Effect of epochs (k = 5 , 8 , 10)

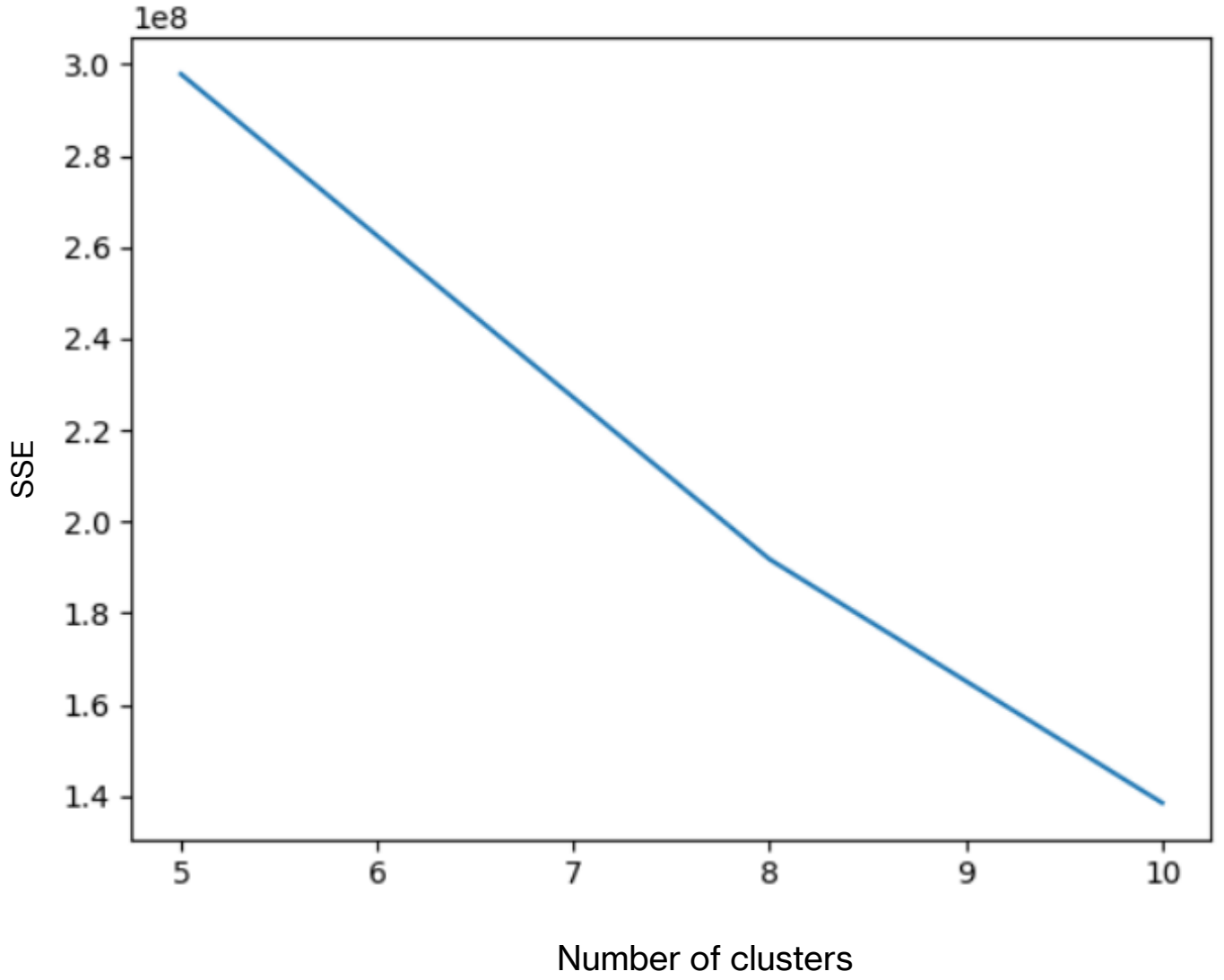
Epoch	Test Loss	Accuracy
3	0.004	79
5	0.003	86
10	0.003	86

Epoch	Test Loss	Accuracy
3	0.002	88
5	0.002	89
10	0.002	91

Epoch	Test Loss	Accuracy
3	0.009	67
5	0.005	79
10	0.01	57

Results / Analysis

k vs SSE (epoch 5)



Analysis

- Initial accuracy of pre-trained AlexNet was pretty low, because AlexNet is trained on ImageNet which has 1000 classes, we had to modify and fine-tune the last layer for CIFAR10 with 10 class output.
- We can see general upward trend in NMI graphs, suggesting fewer and fewer image representation vectors are switching clusters, i.e. clusters are getting stabilized every iteration, suggesting network is learning better representation.
- We see better results with lower k values, because CIFAR10 is well separated dataset (images in same class are very similar to each other)

REFERENCES:

- [1] Devin A. et al. Deep Clustering for Unsupervised Learning of Visual Features. In: Proceedings of the 31st International Conference on Machine Learning (ICML 2014), JMLR Workshop and Conference Proceedings vol 32, 2014.

THANK YOU!