Tutorial on Image Clustering

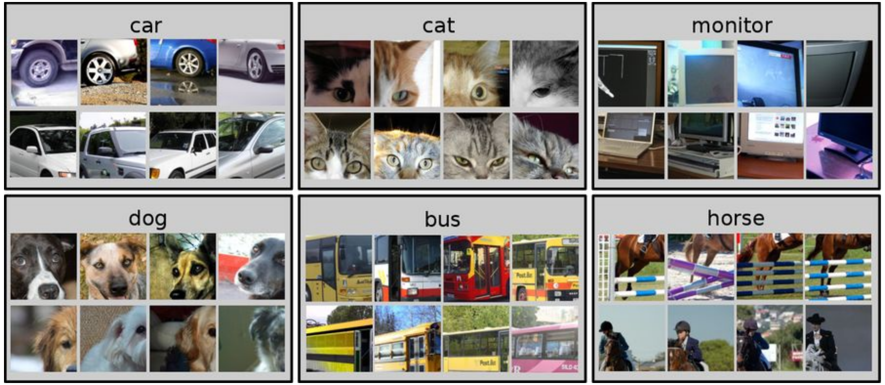
by Joris Guérin

Which is the best CNN feature extractor?

Experiment description - datasets used

Natural object classification

VOC 2007



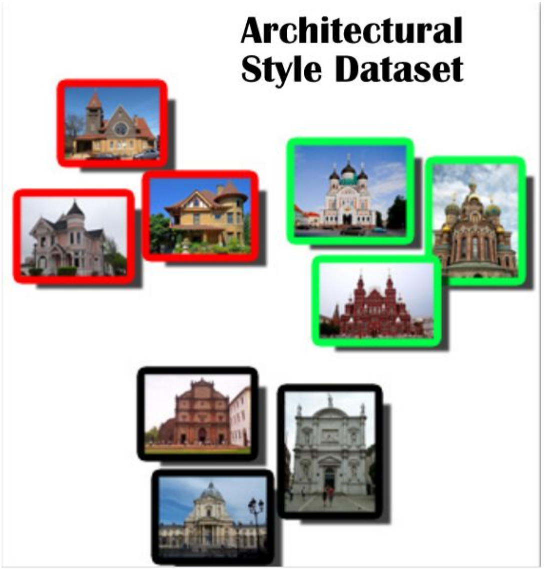
Coil 100

Experiment description - datasets used

Scene recognition

MIT 67

Architectural style (JTU, Shanghai)



Experiment description - datasets used

Fine grained

Caltech UCSD Birds 200

Flower dataset (Oxford)

Experiment description - datasets used

Face recognition

UMist (Sheffield)

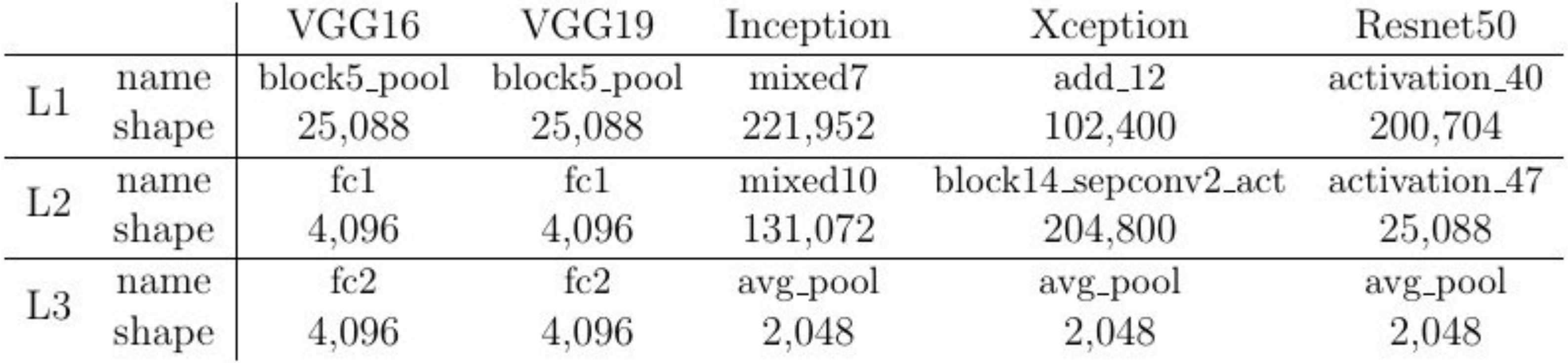
FEI (São Bernardo do Campo)

Experiment description - Experiments setup

Architectures used:

VGG16, VGG19, ResNet50, Inception, Xception

Layers used:



		VGG16	VGG19	Inception	Xception	Resnet50
L1	name	block5_pool	block5_pool	mixed7	add_12	activation_40
	shape	25,088	25,088	221,952	102,400	200,704
L2	name	fc1	fc1	mixed10	block14_sepconv2_act	activation_47
	shape	4,096	4,096	131,072	204,800	25,088
L3	name	fc2	fc2	avg_pool	avg_pool	avg_pool
	shape	4,096	4,096	2,048	2,048	2,048

Experiment description - Experiments setup

Clustering methods used:

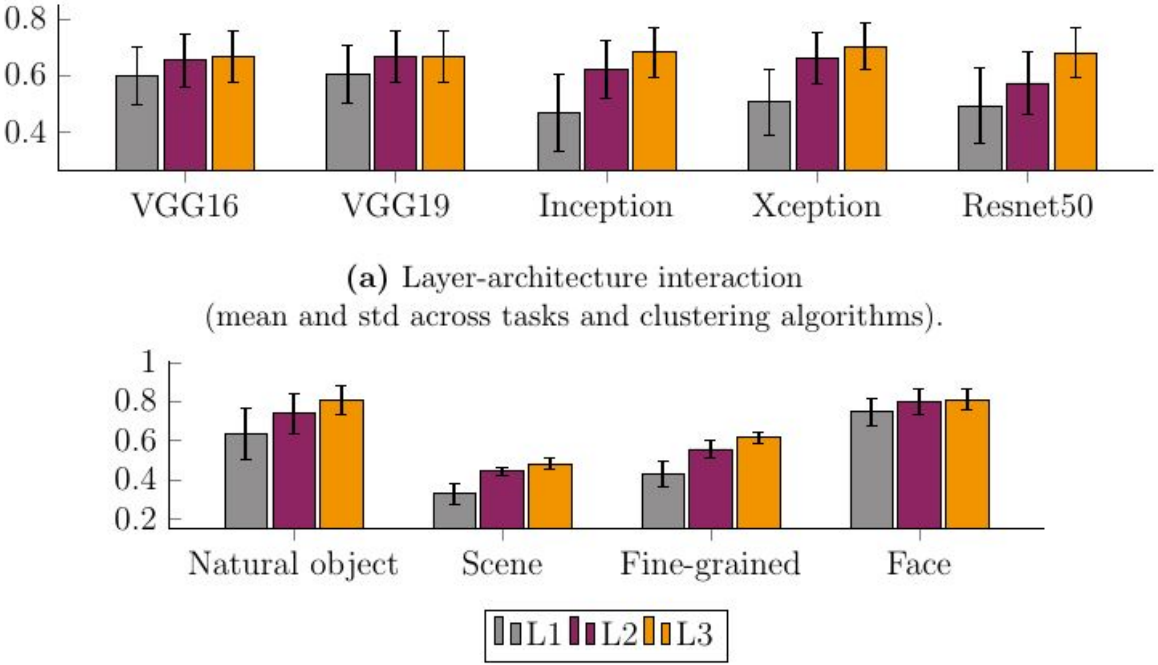
KMeans and **Agglomerative clustering**

Metrics used:

NMI and **Purity**

Results summary

Layer choice:

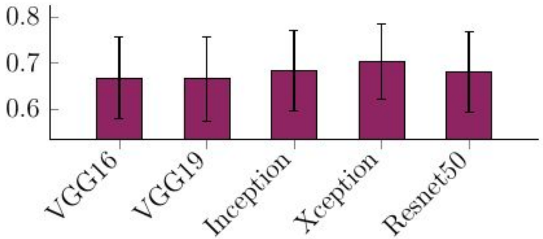


(a) Layer-architecture interaction
(mean and std across tasks and clustering algorithms).

(b) Layer-task interaction
(mean and std across architectures, datasets and clustering algorithms).

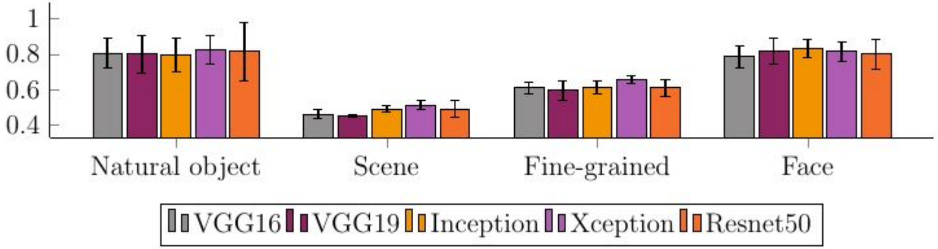
Conclusion: **Use last layer**

Results summary

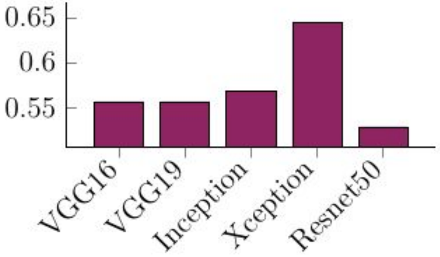


(a) mean and std across tasks and clustering algorithms

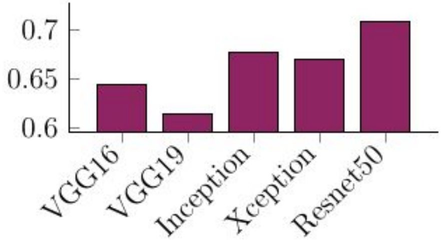
Architecture choice:



(b) Architecture-task interaction (mean and std across datasets and clustering algorithms)



(a) Birds - Agglomerative clustering



(b) Flowers - Agglomerative clustering

Conclusion: **We don’t know?!**

Possible strategies for new unsupervised dataset?

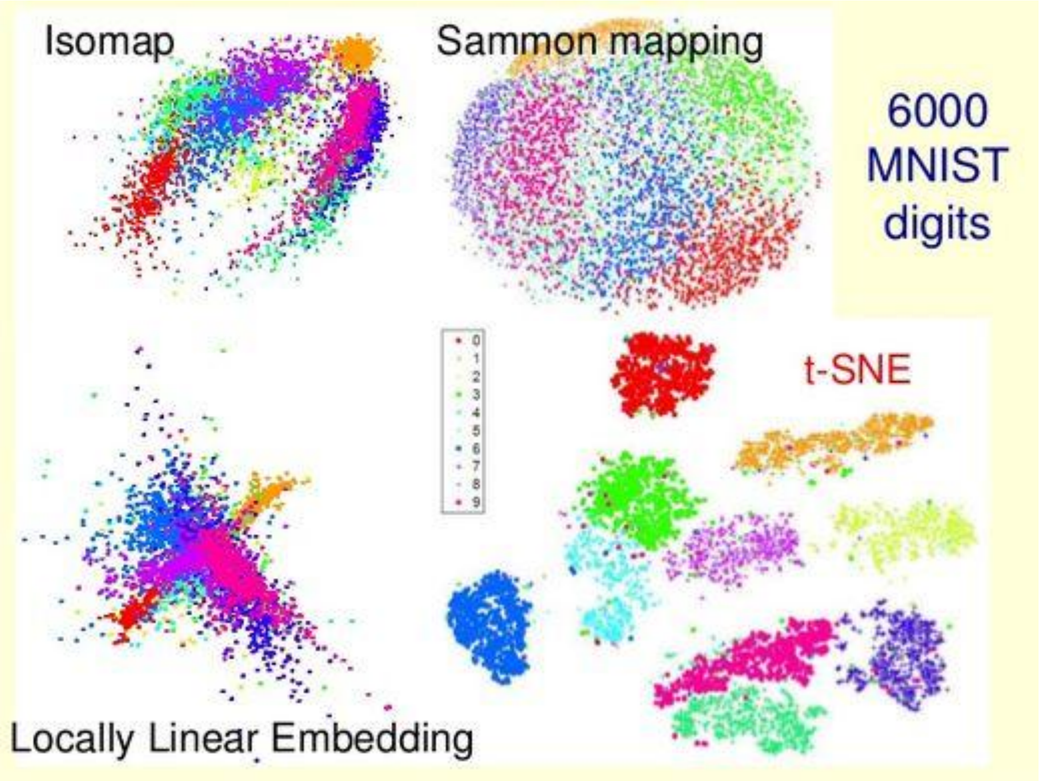
Selecting hyperparameters:

Supervised case:
Cross validation

Unsupervised case:
Follow the leader
(online learning)

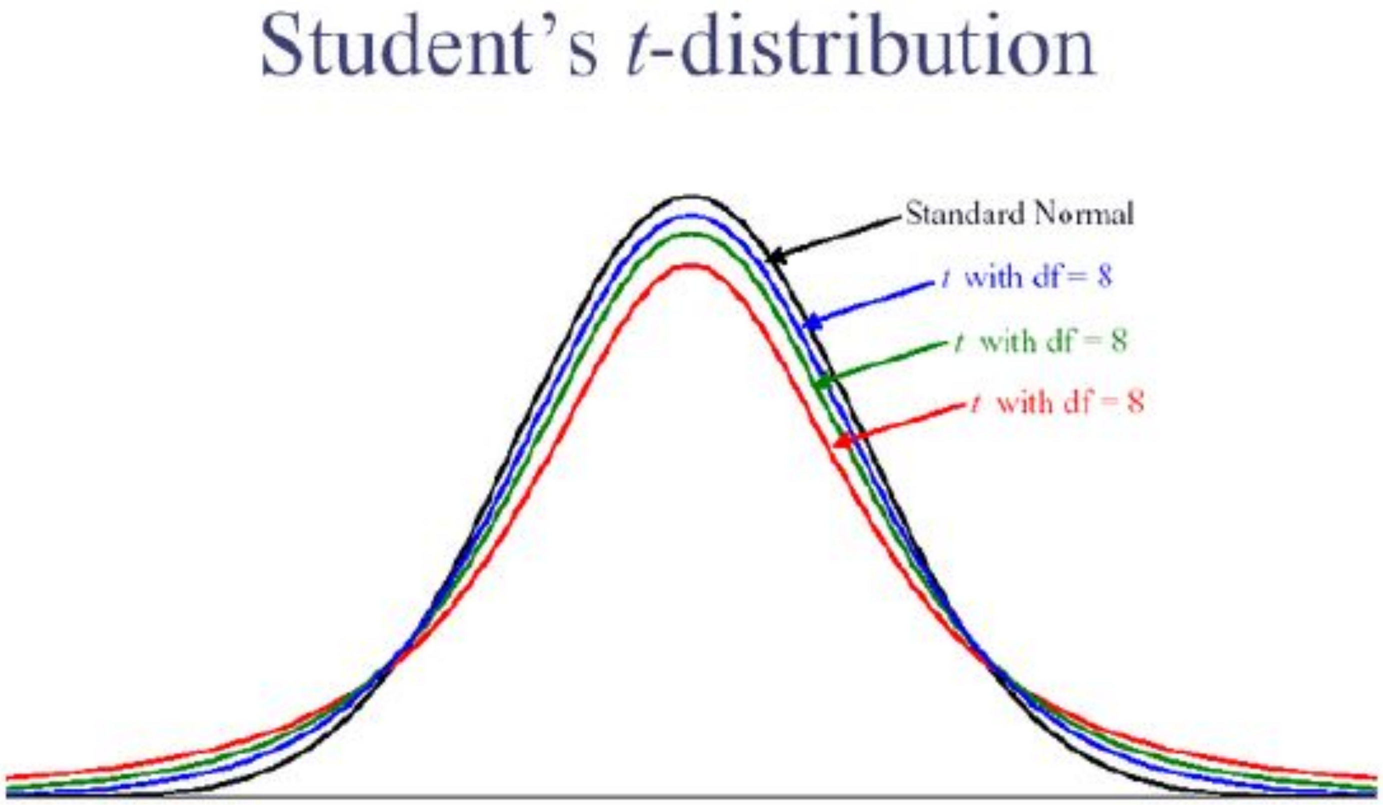
Dimensionality reduction

Objectives and overview



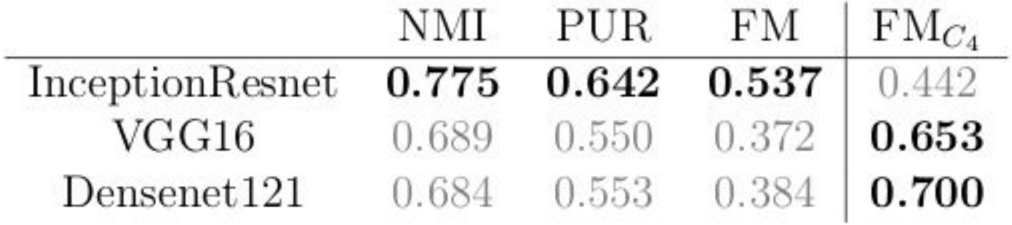
- Methods to visualize high dimensional data.
- Transform data into a 2D or 3D space

t-distributed Stochastic Neighbor Embedding (t-SNE)

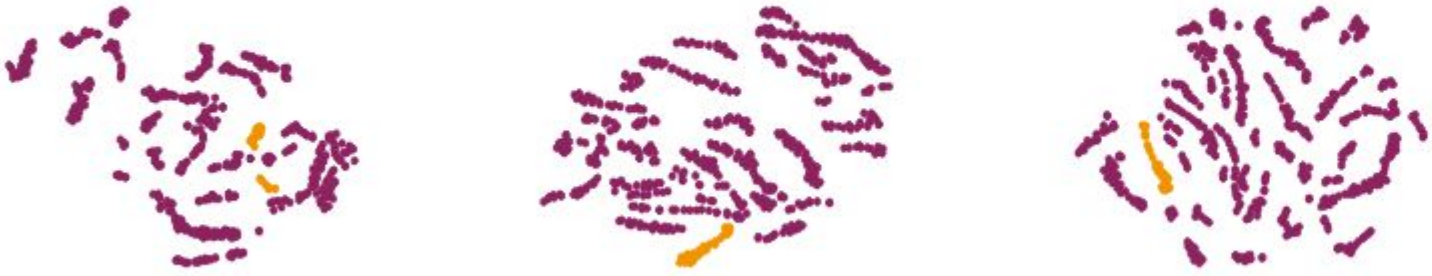


Ensemble of feature extractors

Why does it makes sense?



	NMI	PUR	FM	FM_{C_4}
InceptionResnet	**0.775**	**0.642**	**0.537**	0.442
VGG16	0.689	0.550	0.372	**0.653**
Densenet121	0.684	0.553	0.384	**0.700**

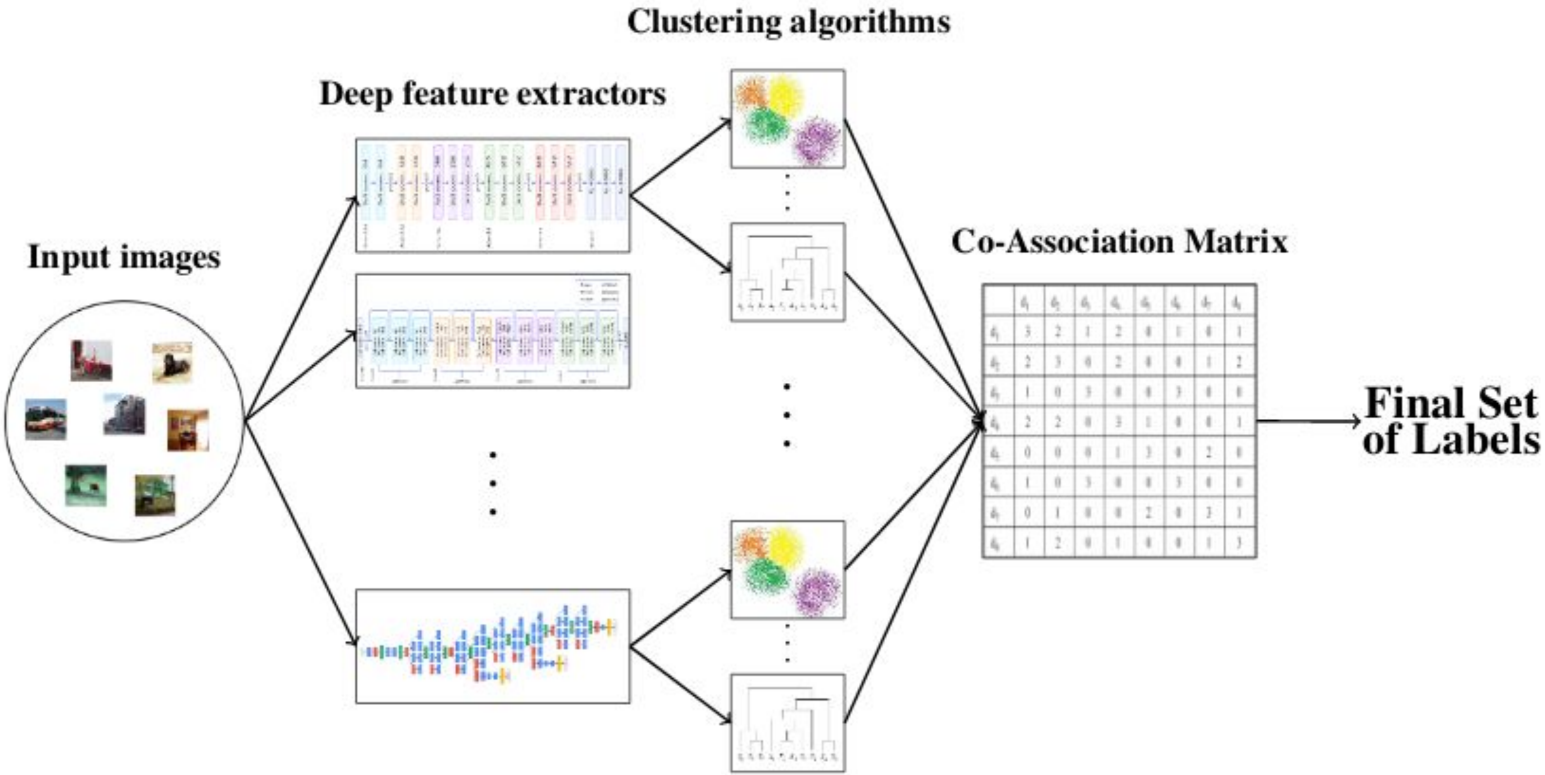


(a) InceptionResnet

(b) VGG16

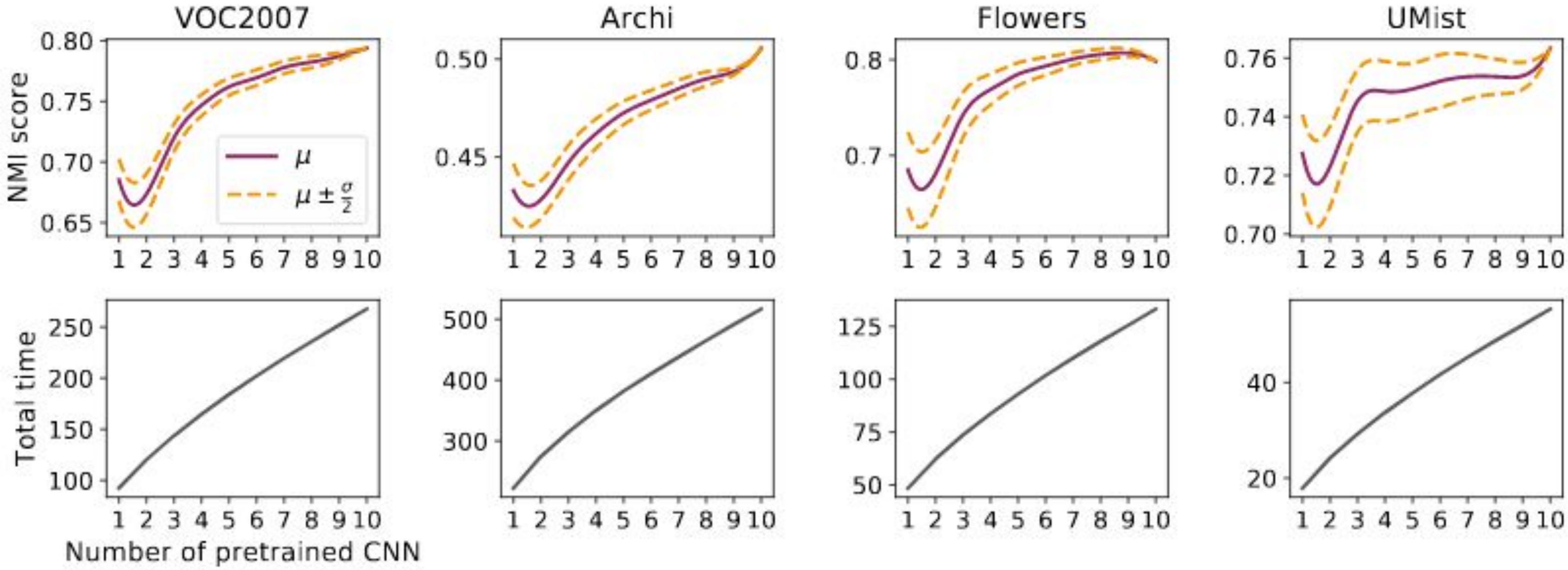
(c) Densenet121

Methodology



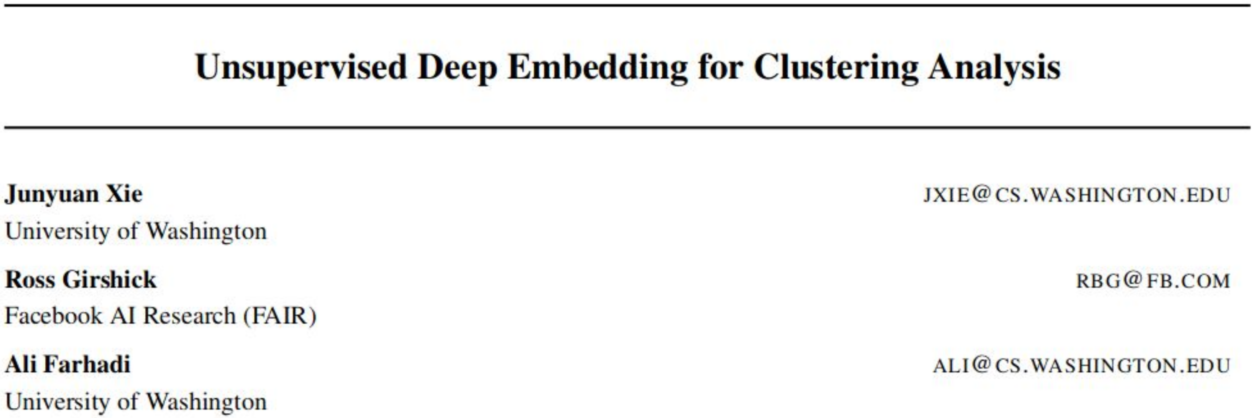
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
d_1	3	2	1	2	0	1	0	1
d_2	2	3	0	2	0	0	1	2
d_3	1	0	3	0	0	3	0	0
d_4	2	2	0	3	1	0	0	1
d_5	0	0	0	1	3	0	2	0
d_6	1	0	3	0	0	3	0	0
d_7	0	1	0	0	2	0	3	1
d_8	1	2	0	1	0	0	1	3

Results


VOC2007
Archi
Flowers
UMist
NMI score
μ
$\mu \pm \frac{\sigma}{2}$
0.80
0.75
0.70
0.65
0.50
0.45
0.8
0.7
0.76
0.74
0.72
0.70
Total time
250
200
150
100
500
400
300
125
100
75
50
40
20
1 2 3 4 5 6 7 8 9 10
Number of pretrained CNN

Deep end to end clustering

Deep Embedded Clustering (DEC)



Unsupervised Deep Embedding for Clustering Analysis

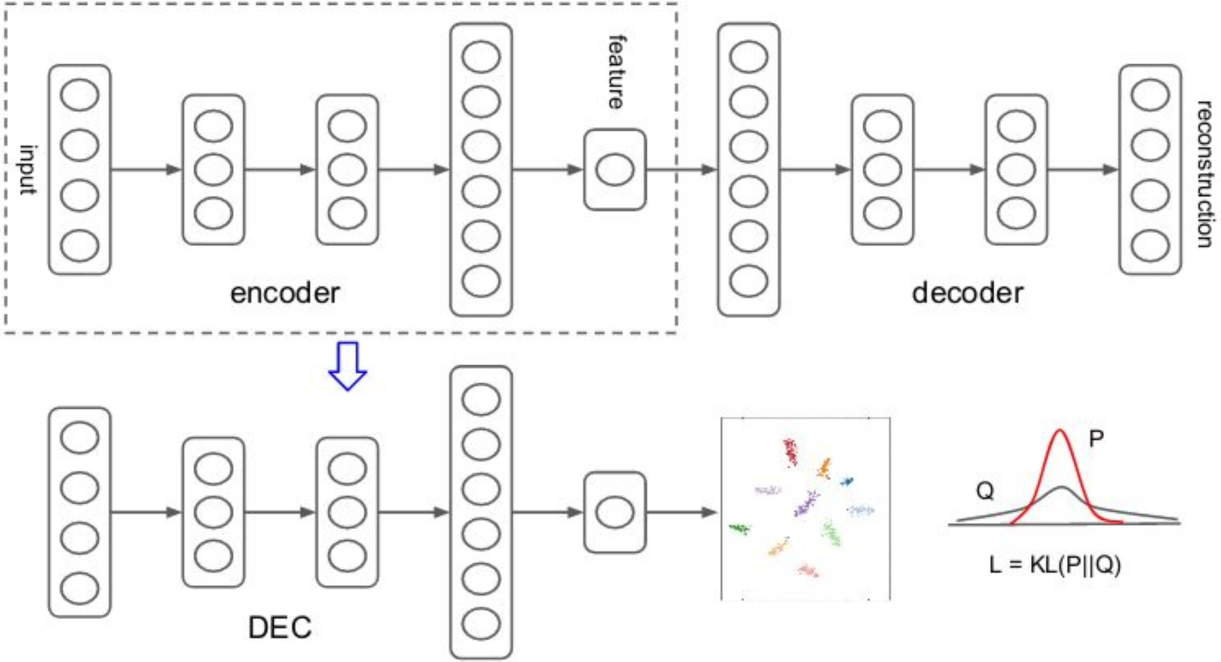
Junyuan Xie JXIE@CS.WASHINGTON.EDU
University of Washington

Ross Girshick RBG@FB.COM
Facebook AI Research (FAIR)

Ali Farhadi ALI@CS.WASHINGTON.EDU
University of Washington

Objective: Learn jointly cluster assignment and new data representation

Autoencoder initialization



1- Denoising AE

2- End-to-end reconstruction AE

Joint optimization

- Soft assignment: Student t-distribtuion (similarity between embedded point z_i and centroid μ_j)

$$q_{ij} = \frac{(1+\left\|z_i-\mu_j\right\|^2/\alpha)^{-(\frac{\alpha+1}{2})}}{\sum_{j'}(1+\left\|z_i-\mu_{j'}\right\|^2/\alpha)^{-(\frac{\alpha+1}{2})}}$$

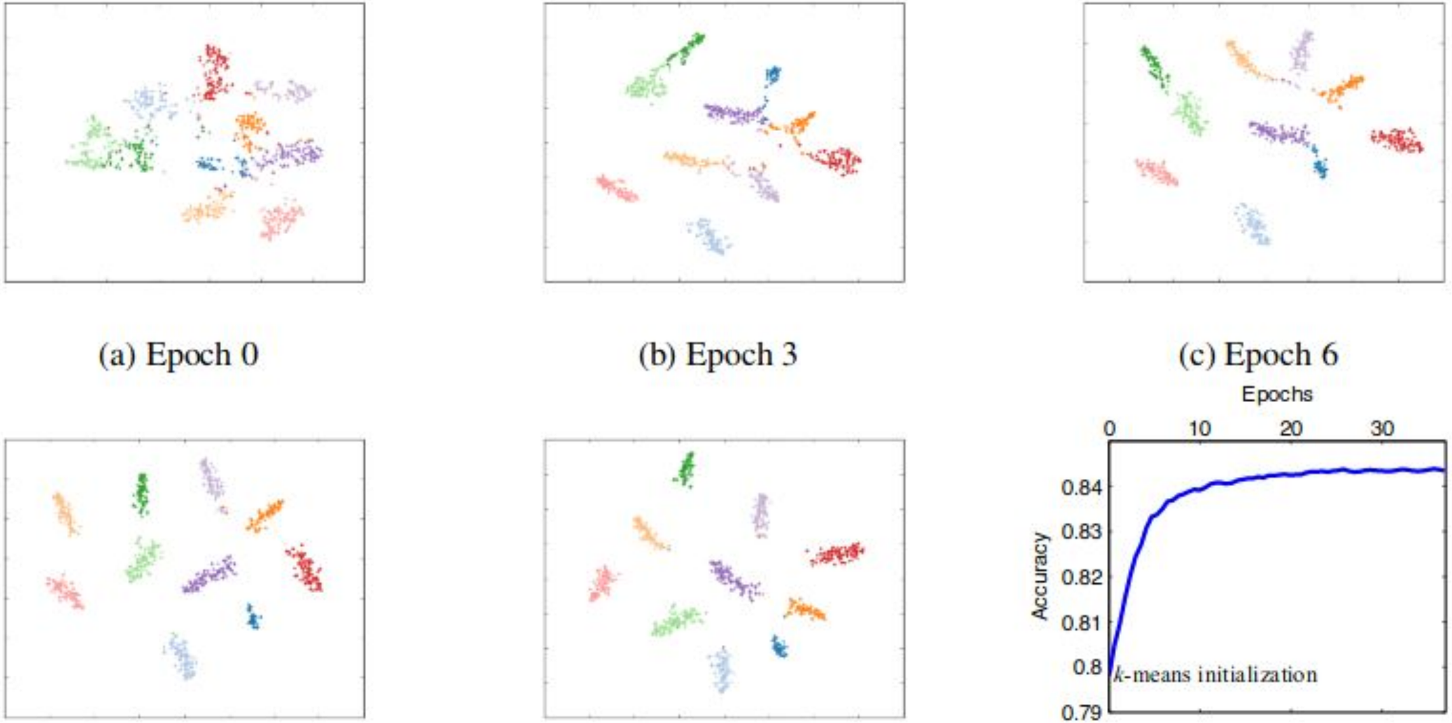
- Target distribution: Strengthen high confidence prediction, normalize loss contribution of each cluster

$$p_{ij} = \frac{q_{ij}^2/f_j}{\sum_{j'} q_{ij'}^2/f_{j'}};\ f_j = \sum_j q_{ij}$$

- optimization: Gradient descent to update θ and μ_j to minimize

$$\mathcal{L} = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Results MNIST



(a) Epoch 0

(b) Epoch 3

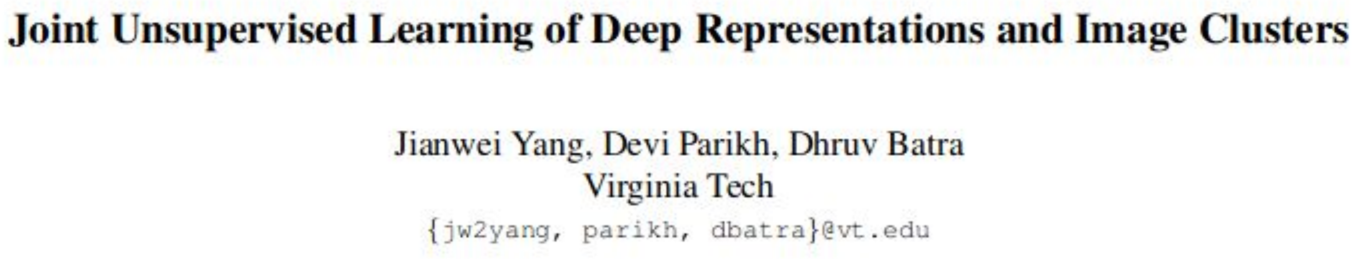
(c) Epoch 6

(d) Epoch 9

(e) Epoch 12

(f) Accuracy vs. epochs

JULE for networks gathering



Joint Unsupervised Learning of Deep Representations and Image Clusters

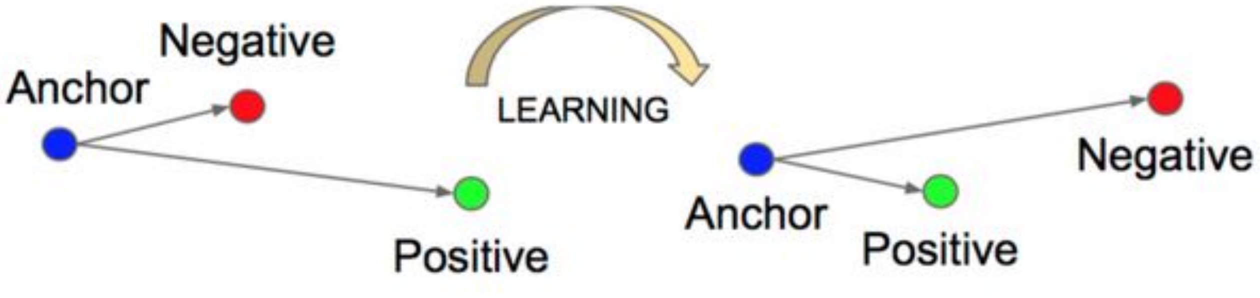
Jianwei Yang, Devi Parikh, Dhruv Batra
Virginia Tech
{jw2yang, parikh, dbatra}@vt.edu

Objective: Learn jointly cluster assignment and new data representation

Triplet loss

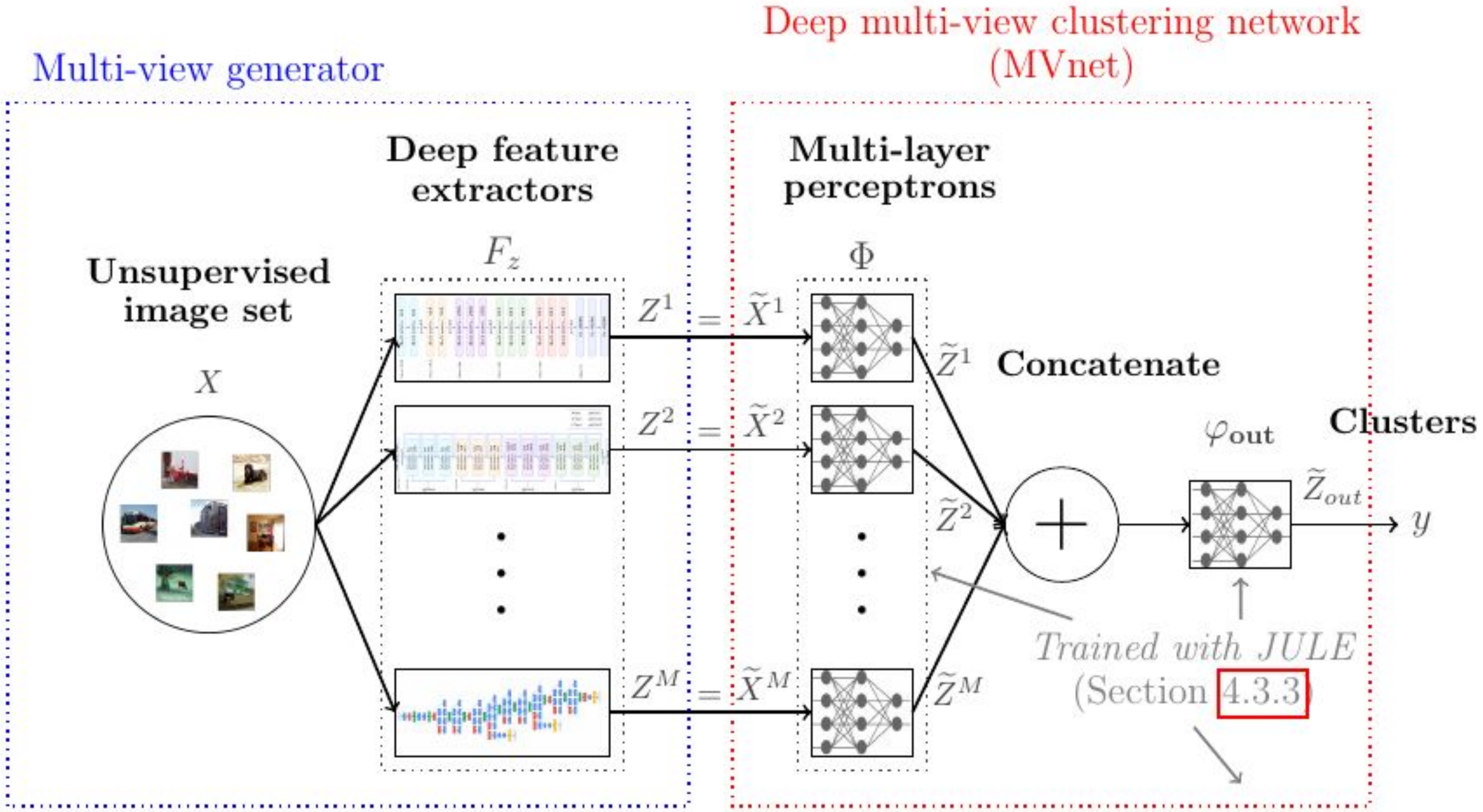
$$Loss = \sum_{i=1}^{N} \left[\| f_i^a - f_i^p \|_2^2 - \| f_i^a - f_i^n \|_2^2 + \alpha \right]_+$$

Schroff et al.

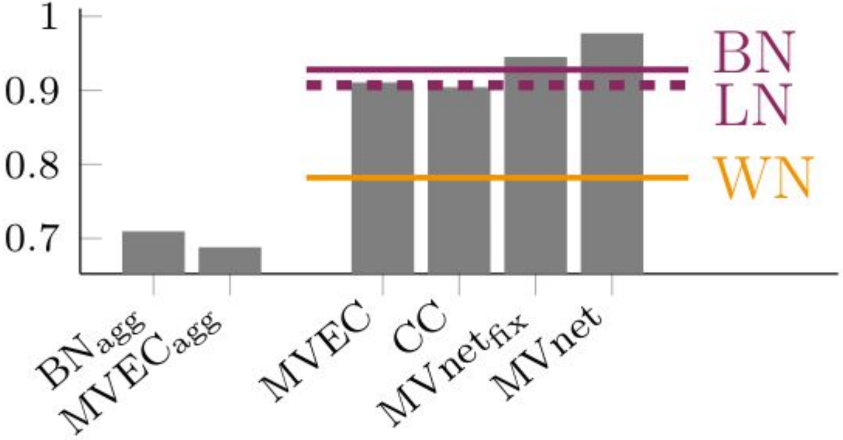


Schroff et al.

Deep Multi-view clustering



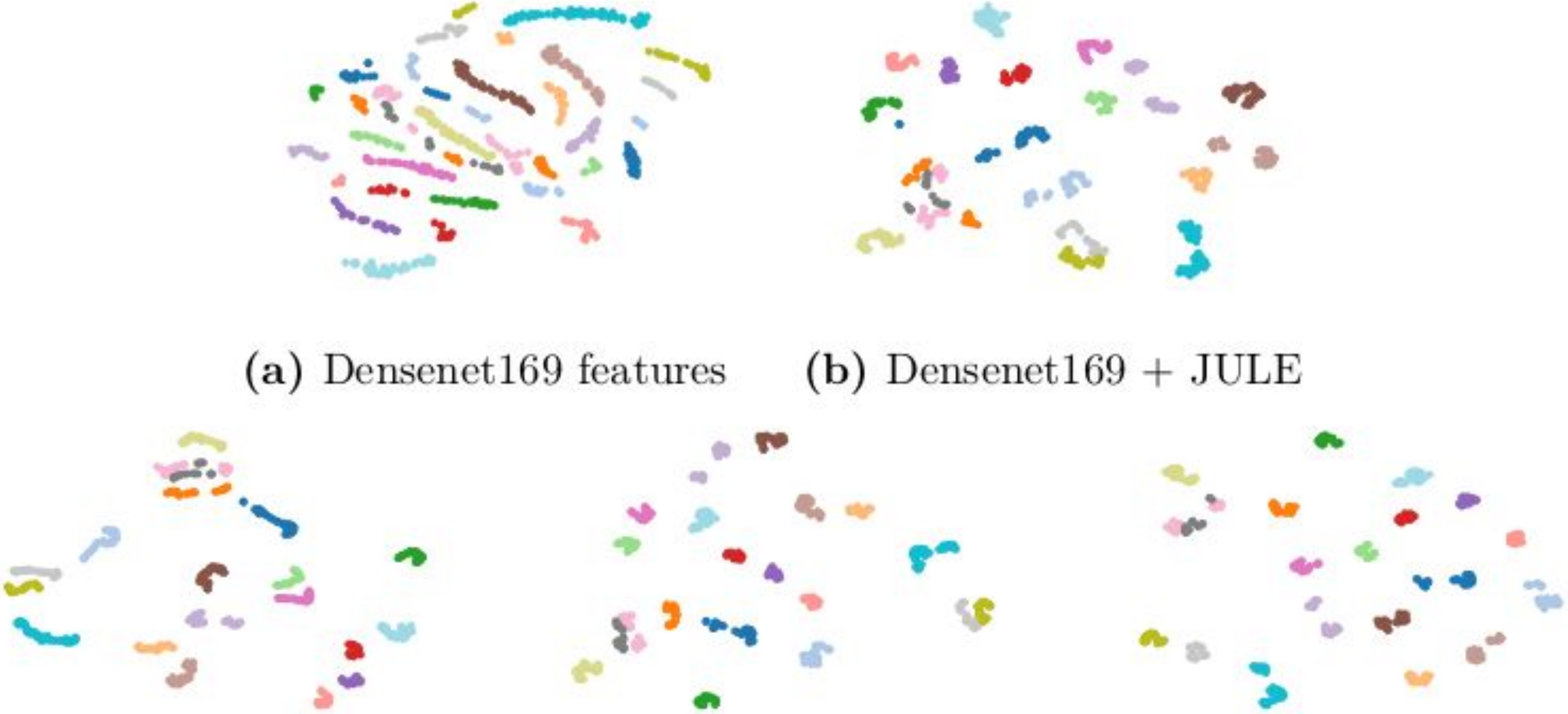
Results on UMist



(g) UMist

Final scores:

Purity : 0.967
NMI : 0.984



(a) Densenet169 features **(b)** Densenet169 + JULE

(c) Concat **(d)** MVnet$_{\text{fix}}$ **(e)** MVnet

Conclusions

Multiple deep CNN feature extractors
+
Deep Clustering (JULE)

Initial scores:

Purity : 0.503
NMI : 0.663

Final scores:

Purity : 0.967
NMI : 0.984