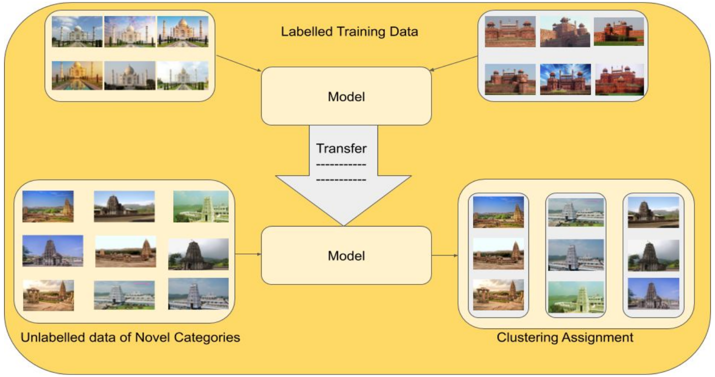
15 - 10 - 2020

Deep Visual Attention Based Transfer Clustering

Akshaykumar Gunari, Shashidhar V Kudari, Sukanya Sanjay Nadagadalli, Keerthi Goudnaik, Ramesh Ashok Tabib, Uma Mudenagudi, Adarsh Jamadandi

KLE Technological University, Hubballi, 580 - 031

akshaygunari@gmail.com, shashidharvk100@gmail.com, sukanyanadagadalli@gmail.com, keerthigoudnaik@gmail.com, ramesh_t@kletech.ac.in, uma@kletech.ac.in, adarsh.jamadandi@kletech.ac.in



CoCoNet '20

ACN'20

Contents:

- ➢ Introduction
- ➢ Motivation
- ➢ Contributions
- ➢ Methodology
- ➢ Discussions
- ➢ Experiment Details Dataset
- ➢ Results
- ➢ Conclusion
- ➢ References

Introduction:

➢ Clustering plays an exceedingly important role in the entire Knowledge Discovery in Databases (KDD) process also as **categorizing data is one of the most rudimentary steps in knowledge discovery.**

➢ The task of **unsupervised image classification** remains an important, and **open challenge in computer vision.**

➢ Clustering is an unsupervised learning task used for exploratory **data analysis to find some unrevealed patterns** which are present in data but cannot be categorized clearly.

➢ Recent developed **deep unsupervised methods allow us to jointly learn representation** and cluster unlabelled data.

➢ Typically, **to learn a better representation, we adopt the auto-encoder** to maximize the mutual information between features.

➢ We propose a methodology **to improvise the technique of Deep Transfer Clustering (DTC)** to the less variant data distribution.

➢ We have discussed the **improvement using attention-based classifiers** rather than regular classifiers as the initial feature extractors in the Deep Transfer Clustering.

Motivation:

- In unsupervised learning, every time a **new set of images belonging to a class arrives**, the autoencoder should be trained again including the new set of arrived images.

- Our method is an addon to the existing Deep Transfer Clustering method [3] which **helps the feature extractor to learn more robust and differentiable features** rather than concentrating on background common features.

- Providing the raw images to the classifier does not imply **what the model should focus to learn**.

- This results in **uncertain behaviour of the model when a small region is modified** in the image as can be observed in the image Fig.1.

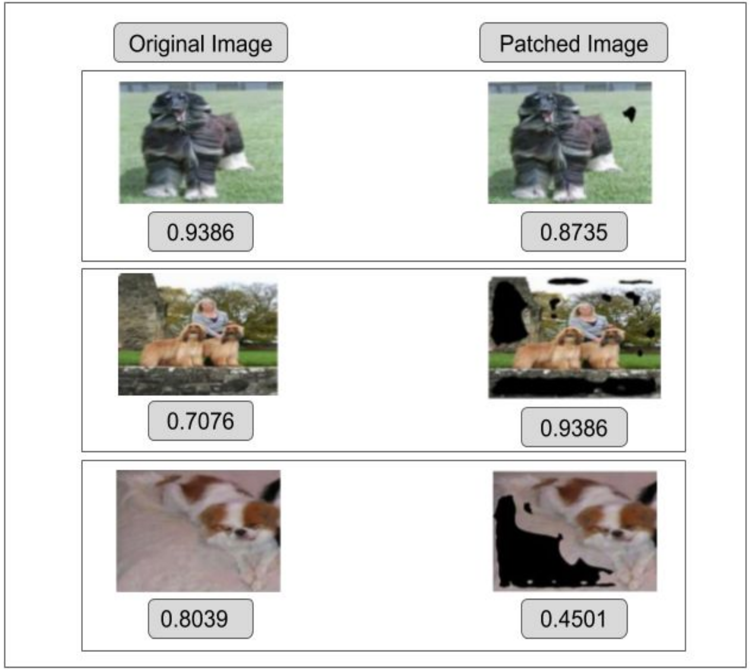


Fig. 1. Uncertain behaviour of the model

Motivation:

- Drastic **decrease in performance** in traditional clustering techniques **where pre-trained classification models are used**, Deep Embedded Clustering for such datasets where the classes have very less variance or high similarity.

- If the model is **not being activated around the intended patterns/objects** in an image, then we can revisit the process of training for required feature extraction.

Parashurameshvara Temple, Bhubaneshwar, mid-7th century

Krimchi temples Udhampur region of Jammu. (9-10 century)

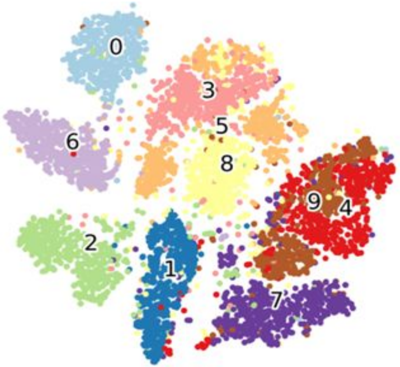


Fig. 2.a. Low Variance Data Distribution

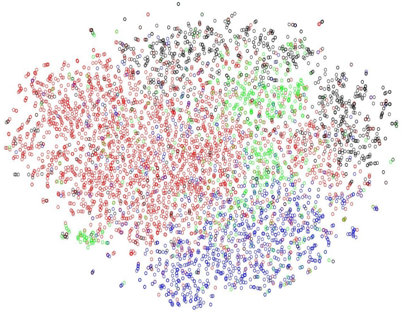
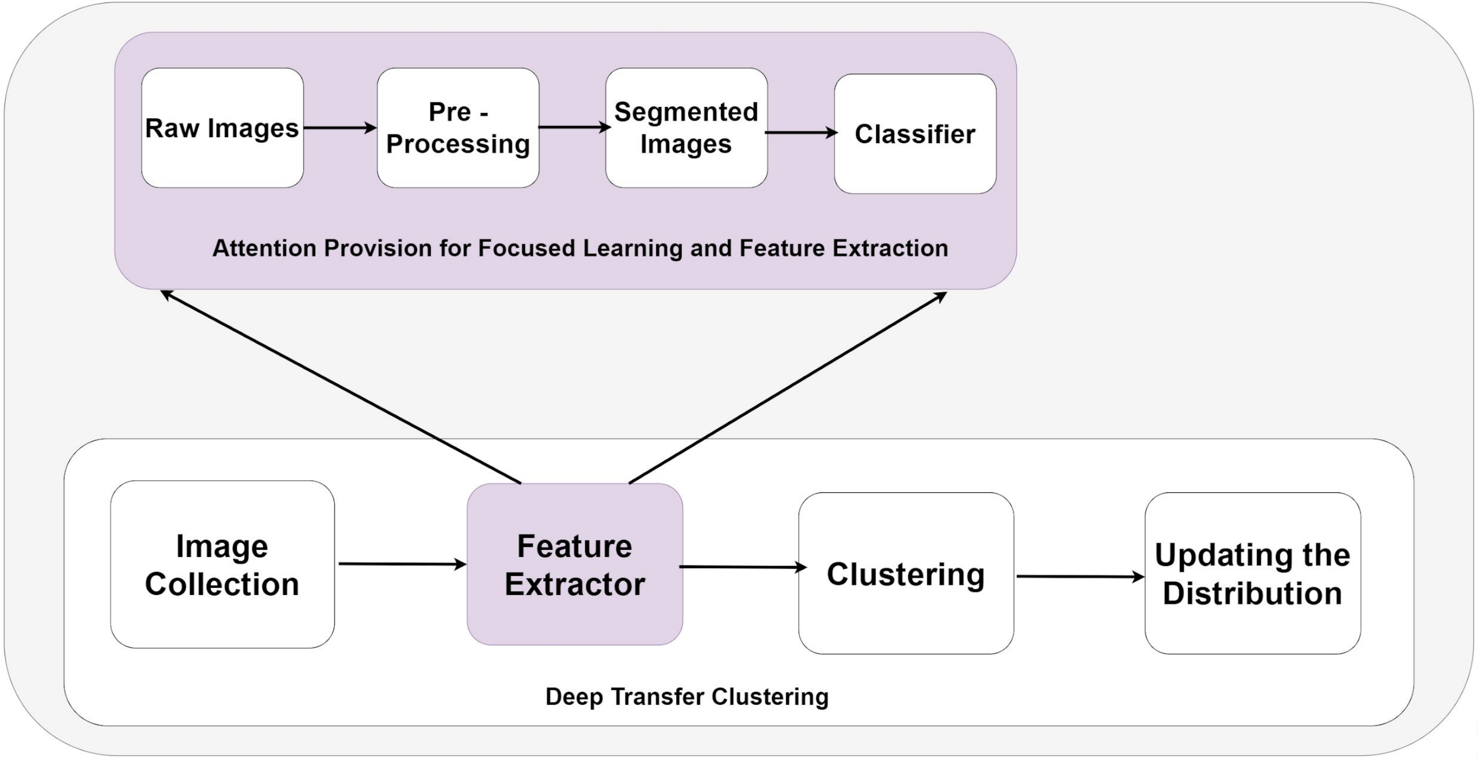


Fig. 2.b. High Variance Data Distribution

Contributions:

- ➢ We **enhance the deep transfer clustering framework** to comprehensively mine various kinds of correlations, and select highly - sensitive data distributions to train the network in a progressive way.
- ➢ **A simple and effective approach for the categorization of unlabelled data is introduced**, by considering it as a deep transfer clustering problem.
- ➢ The main focus was to cluster the **image collection with high similarity** between the two classes.
- ➢ We have achieved this by plugging in a module that makes **the network concentrate on the instructed region of interest rather than a complete image**.
- ➢ The clustering performance based on learned representations from all layers improve, which indicates that **the learned representations can be transferred** across the datasets.

Methodology:



Methodology:

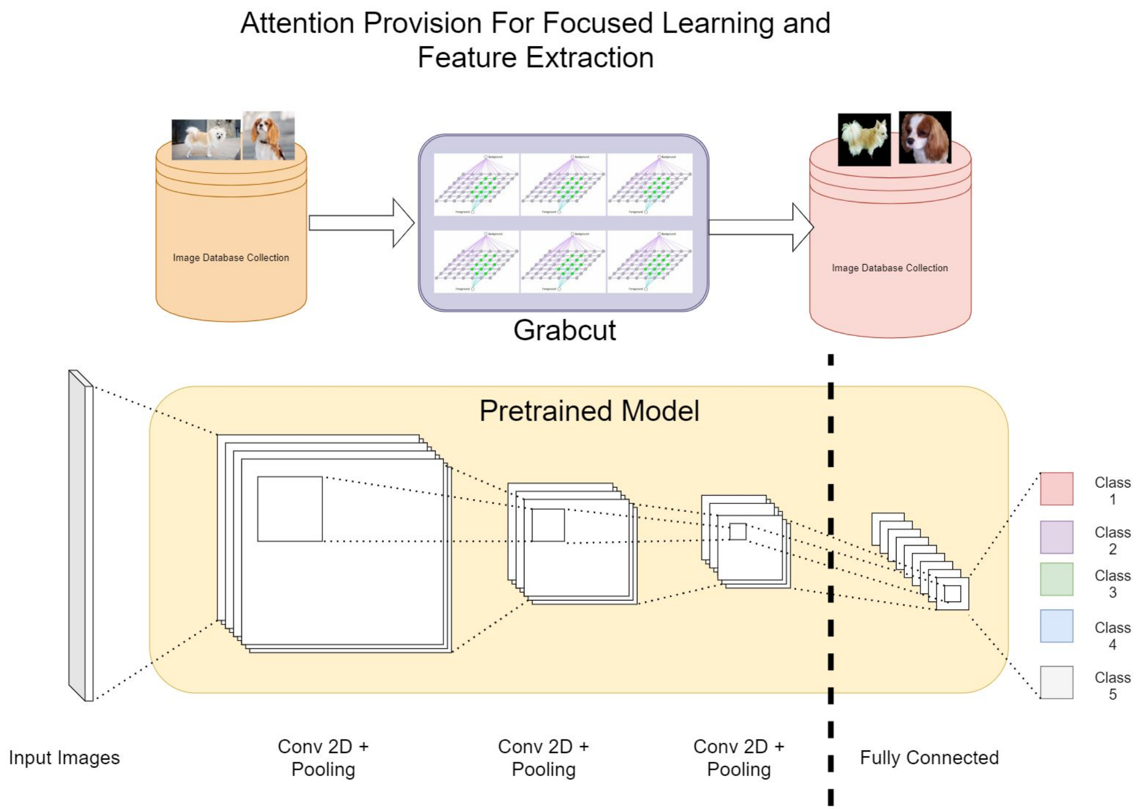


Fig. 2. Attention Provision For Focused Learning

Methodology:

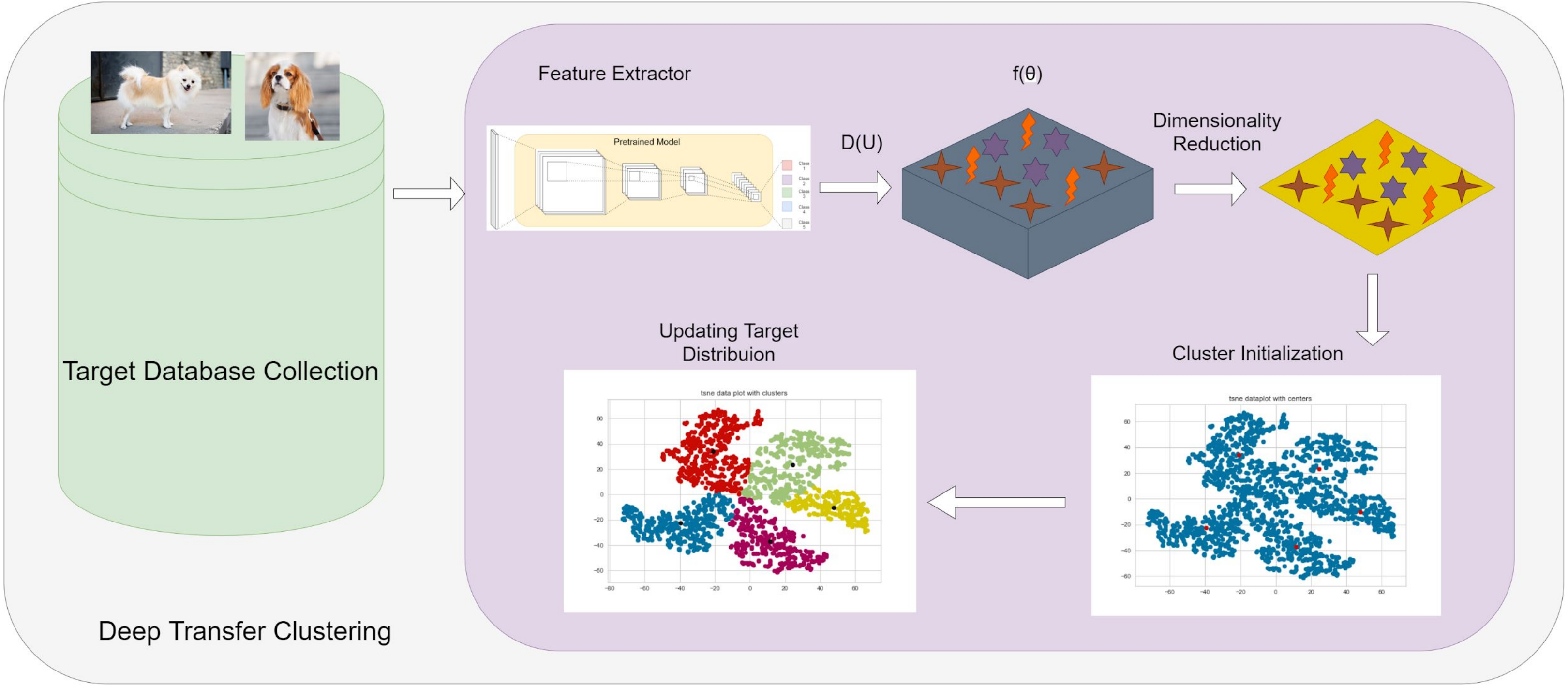


Fig. 2. Deep Transfer Learning

Discussion:

KLE TECH KLE Technological University
Creating Value
Leveraging Knowledge

- Probability of assigning data point

$$p(k|i) \propto \left(1 + \frac{\| z_i + \mu_k \|^2}{\alpha}\right)^{-\frac{\alpha+1}{2}}$$ Eq. (1)

- Objective as KL divergence

$$E(q) = KL(q \parallel p) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} q(k|i) \log \frac{q(k|i)}{p(k|i)}$$ Eq. (2)

- Target distribution q is calculated as

$$q(k|i) \propto p(k|i) \cdot p(i|k)$$ Eq. (3)

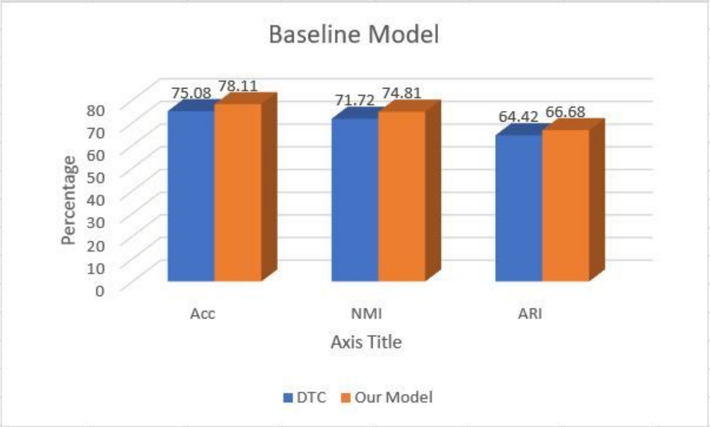
- Equalization effect as the probability of sampling data point

$$q(k|i) \propto \frac{p(k|i)^2}{\sum_{i=1}^{N} p(k|i)}$$ Eq. (4)

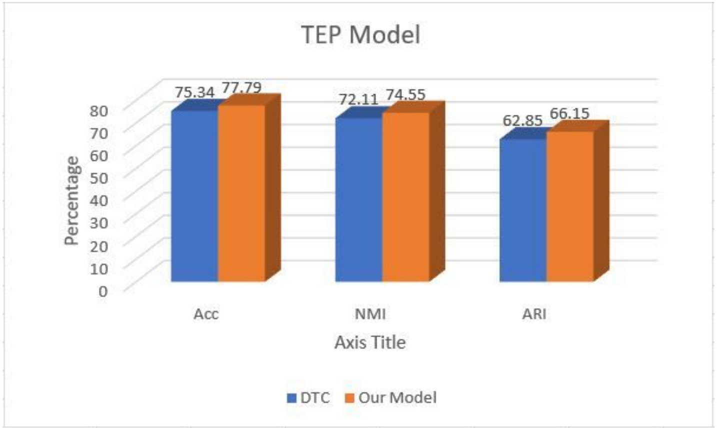
Experiment Details:

- **Dataset**:
  - IDH-10
    - Target Dataset
  - Imagenet [Dog-10]
    - Similar to IDH dataset having high intraclass variance and low inter class variance.
- **Metric**:
  - Algorithm 1 (Attention Provision For Focused Learning):
    - Accuracy
  - Algorithm 2 (Deep Visual Attention Based Transfer Clustering):
    - ACC
    - NMI
    - ARI
- **Inference**:
  - Extracted features could now be fine tuned to update the clusters in target distribution with increased ACC of ~3.08%.

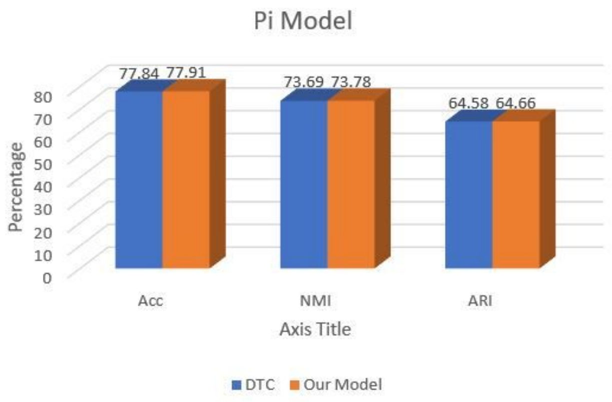
Results:



Baseline Model
Percentage
80
70
60
50
40
30
20
10
0
75.08 78.11
71.72 74.81
64.42 66.68
Acc
NMI
ARI
Axis Title
DTC Our Model



TEP Model
Percentage
80
70
60
50
40
30
20
10
0
75.34 77.79
72.11 74.55
62.85 66.15
Acc
NMI
ARI
Axis Title
DTC Our Model



Pi Model
Percentage
80
70
60
50
40
30
20
10
0
77.84 77.91
73.69 73.78
64.58 64.66
Acc
NMI
ARI
Axis Title
DTC Our Model

Conclusions:

KLE TECH KLE Technological University Creating Value Leveraging Knowledge

- Traditional clustering techniques of using pre-trained models for image clustering will not lead to better results for all the kinds of data, especially for those classes with very high similarity.
- Using pretrained models for feature extraction may not lead us to **extract intended features** capable of helping us meet the objectives of clustering.
- Our proposition consists of **gradual and smooth transforming** a unsupervised objective into a self-supervised one.
- Also, we introduce **additional constraints for selecting an appropriate feature extraction task**, capable of producing meaningful feature representations.
- FOur focus is on completely such datasets where there is less **intra-class variance and low inter-class variance**.

References:

[1] Jianlong Chang, Lingfeng Wang, Gaofeng Meng, Shiming Xiang, and ChunhongPan. Deep adaptive image clustering. In Proceedings of the IEEE international conference on computer vision, pages 5879–5887, 2017.

[2] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasub-ramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar 2018.

[3] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering, 2019.

[4] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. ” grabcut” interactive foreground extraction using iterated graph cuts.ACM transactions on graphics(TOG), 23(3):309–314, 2004.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 618–626, 2017.

[6] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, 128(2):336–359, Oct 2019.

[7] Junyuan Xie, Ross B. Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis.CoRR, abs/1511.06335, 2015.

THANK YOU