Seminar Paper: Group 13

Deep Clustering for Unsupervised Learning of Visual Features[1]

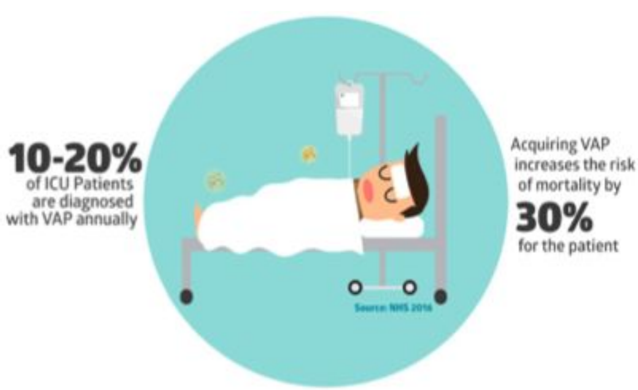
Members:
Suraj Shah
Mentors:
Dr. Mathias Unberath

Clinical Collaborators:
Dr. Jim Fackler
Dr. Jules Bergmann
Dr. Ferdinand Hui

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Project Overview

Topic: Assessing Ventilator-Associated Pneumonia (VAP) in the PICU



Chest X-Rays: empirical standard in pneumonia diagnosis
Mechanical Ventilation: critical life sustaining ICU therapy, however, with host of problems that leads to VAP
Objective: prepare automatic classifier to be used in the PICU to monitor onset of disease progression and detect VAP

Paper Summary and Background

- Methods performed by the **Facebook AI Research Group ,** submitted to EECV (European Conference on Computer Vision) 2018

- Proposed **DeepCluster**, a novel clustering method

- Utilizes pre-trained Convolutional Neural Networks (CNNs) with standard clustering methods on top (k-means)

- **Outperforms state-of-the-art** image classification networks

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Selection Motivation

- Aims to provide **solution to internet-scale image classification** problems *(highly relevant to my project’s datasets)*

- Supervised learning is **highly dependent on labeling and annotation** which is sparse in real-life problems

- Takes novel approach to network training and is **considered** one of the **most successful approaches** thus far

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Paper's Approach

1. Overview of Technical Methods

2. Implementation Details

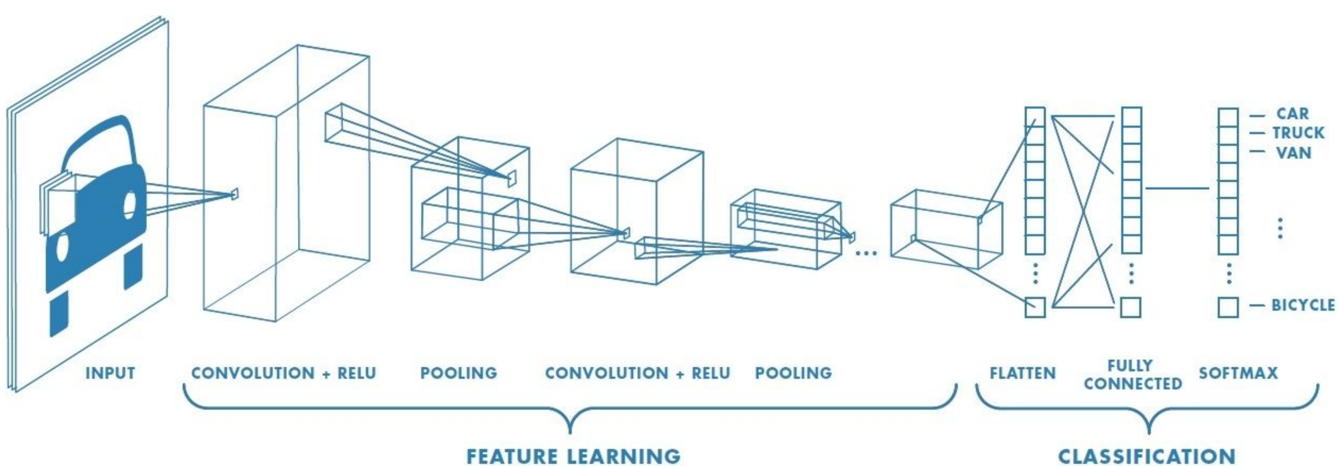
3. Experiments

4. Results and Comparison to Prior Literature

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Paper's Approach: Overview of Technical Methods

Overview of CNN Architecture

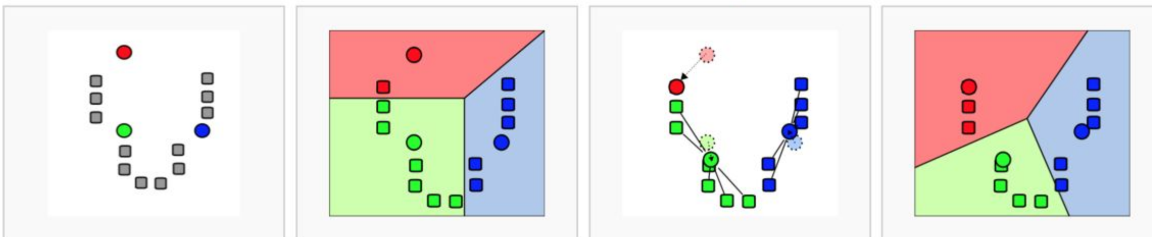


[2] Hatami, N., Gavet, Y. and Debayle, J. (2019). Classification of Time-Series Images Using Deep Convolutional Neural Networks. [online] arXiv.org. Available at: https://arxiv.org/abs/1710.00886. Accessed 20 Feb. 2019

Paper’s Approach: Overview of Technical Methods

K-Means

Demonstration of the standard algorithm



1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3. The centroid of each of the *k* clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.

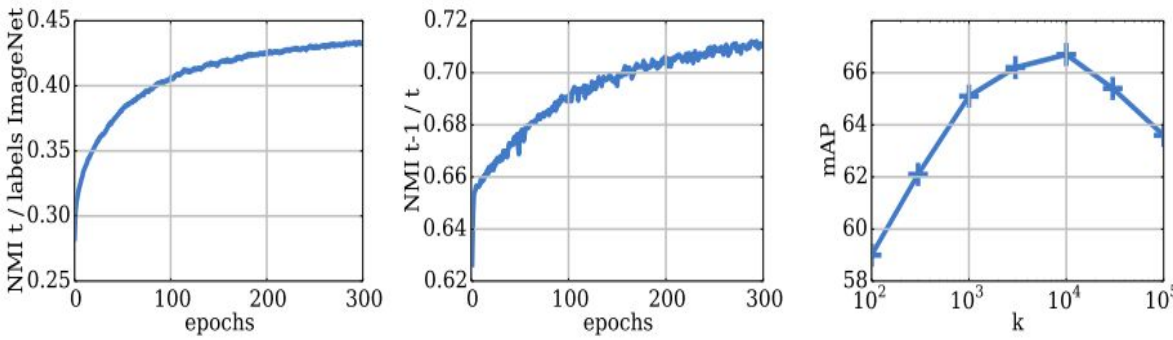
[3] Landman, Nathan, Hannah Pang, and Christopher Williams. "K-Means Clustering." Brilliant Math & Science Wiki. 2018. 20 Apr. 2019 <https://brilliant.org/wiki/k-means-clustering/>.

Paper’s Approach: Implementation Details

- Network: utilize the pre-trained AlexNet and VGG-16 models (compare both because of trade-off in computational complexity)
- Data Acquisition: train on ImageNet (1.3M images with 1000 classes)
- CNN Parameters: dropout, L2 penalization, momentum = 0.9
- Feature regularization: PCA-reduction to 256 and whitened
- Training: over 500 epochs, update clusters after each epoch

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Paper's Approach: Experiments



(a) Clustering quality (b) Cluster reassignment (c) Influence of k

NMI (Normalized Mutual Information) is the parameter that is measured above. NMI measures the information shared between two different assignments, i.e. a measure of independence

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Paper's Approach: Experiments

	ImageNet					Places				
Method	conv1	conv2	conv3	conv4	conv5	conv1	conv2	conv3	conv4	conv5
Places labels	–	–	–	–	–	22.1	35.1	40.2	43.3	44.6
ImageNet labels	19.3	36.3	44.2	48.3	50.5	22.7	34.8	38.4	39.4	38.7
Random	11.6	17.1	16.9	16.3	14.1	15.7	20.3	19.8	19.1	17.5
Pathak *et al.* [38]	14.1	20.7	21.0	19.8	15.5	18.2	23.2	23.4	21.9	18.4
Doersch *et al.* [25]	16.2	23.3	30.2	31.7	29.6	19.7	26.7	31.9	32.7	30.9
Zhang *et al.* [28]	12.5	24.5	30.4	31.5	30.3	16.0	25.7	29.6	30.3	29.7
Donahue *et al.* [20]	17.7	24.5	31.0	29.9	28.0	21.4	26.2	27.1	26.1	24.0
Noroozi and Favaro [26]	**18.2**	28.8	34.0	33.9	27.1	23.0	32.1	35.5	34.8	31.3
Noroozi *et al.* [45]	18.0	**30.6**	34.3	32.5	25.7	**23.3**	**33.9**	36.3	34.7	29.6
Zhang *et al.* [43]	17.7	29.3	35.4	35.2	32.8	21.3	30.7	34.0	34.1	32.5
DeepCluster	12.9	29.2	**38.2**	**39.8**	**36.1**	18.6	30.8	**37.0**	**37.5**	**33.1**

Comparison to other unsupervised learning models within the layers

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Paper's Approach: Results

Method	Training set	Classification		Detection		Segmentation	
		FC6-8	ALL	FC6-8	ALL	FC6-8	ALL
Best competitor	ImageNet	63.0	67.7	43.4†	53.2	35.8†	37.7
DeepCluster	ImageNet	72.0	73.7	51.4	55.4	43.2	45.1
DeepCluster	YFCC100M	67.3	69.3	45.6	53.0	39.2	42.2

Comparison across performance yields that DeepCluster outperforms competitors on all tasks

Method	AlexNet	VGG-16
ImageNet labels	56.8	67.3
Random	47.8	39.7
Doersch *et al.* [25]	51.1	61.5
Wang and Gupta [29]	47.2	60.2
Wang *et al.* [46]	–	63.2
DeepCluster	**55.4**	**65.9**

Comparison of VGG-16 and AlexNet shows the performance increase in using a more complex model

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Personal Critique: Areas of Improvement

- **Lack of literature review into image clustering**
  - Paper continually references other competitors and papers in discussion of results, but does not give significance to *why DeepCluster is different*
  - Would have been beneficial in understanding *why DeepCluster outperforms*
- **Lack of visual schematics**
  - Mentions the network implementation but does not give a visual representation of *how weights are updated* in the CNN
  - Had to search for understanding on my own to supplement review material

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Personal Critique: Areas of Improvement

- **Sole use of hard assignments**
  - Would have been interesting to compare if a probabilistic (soft) clustering approach would have been more useful in updating the weights
- **Resetting clusters after each epoch**
  - The method they chose here is very computationally intensive and perhaps inefficient given the number of epochs they train for
  - Perhaps they could have computed the sample overlap between old and new clusters and assign the cluster IDs translationally *(full reset would not be necessary)*

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Personal Critique: Positives

- **Approach**: novel and fundamental to understanding of how we can combine learning methods to perform classification
- **Training and Testing**: utilized both AlexNet and VGG-16, as well as validated on image sets that were quite different in composure
- **Linear Activation of Specific Layers**: ability to discern where in the convolutional layers that performance starts to outperform

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep Clustering for Unsupervised Learning of Visual Features." arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Personal Takeaways

- Learned how clustering can be applied on top of a CNN’s outputs to produce a more robust model for classification
- An improved understanding of the challenges associated with different image datasets and how to improve one’s own model based on these parameters
- Limitations with different architectures of CNNs and how to best optimize the architecture for the task encountered

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

Future Steps

- Replace hard assignments with weighted K-means or GMM-EM (Gaussian Mixture Model with Expectation Maximization) to allow more distribution in weight updating
- Test on purely unannotated and un-labeled data to see if similar performance can be achieved (pure unsupervised learning)
- Establish similar performance levels with refined networks such as ResNet and DenseNet for both improvements in efficiency and performance

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv:1807.05520 [cs.CV]. Proc. ECCV (2018).

601.456 CIS II Spring 2019

Engineering Research Center for Computer Integrated Surgical Systems and Technology