Viabiltiy of Invariant Information Clustering of Human Actions in Images

Jørgen Hoel, David Lung
Computer Science and Media Technology
Malmö Universitet
Malmö, Sweden
Email: j.hoel@hotmail.com, lung.david90@gmail.com

Abstract

This report investigates a novel method designed to overcome the challenge of clustering unlabeled image data in the context of HAR (human activity recognition). Aiming at leveraging the information based clustering technique IIC (invariant information clustering) within a neural network framework [1], we attempt to apply this unsupervised technique on the challenging domain of categorizing diverse human actions without the need for manual labeling. This has the potential of not only streamlining the data preprocessing pipeline, but could also ensure that the model is adaptable to real-world applications where labeled data is often either scarce or costly to obtain. Despite the extensive experimentation and evaluation, we faced difficulties in achieving a robustness and performance for the target HAR dataset, unlike what has been shown in other domains. The findings shed light on the complexities inherent to HAR, and highlight challenges needed to be overcome for the successful application of IIC.

I. Introduction

The most simple deep learning methods are supervised, meaning that they need manually labeled data for training, limiting their usefulness in many areas. This is especially true for large-scale image classification and for segmentation where labeling is very costly, leading to a need for alternative methods [2]. Unsupervised clustering, on the other hand, attempts to group data points into classes without any prior knowledge of class-membership. Previous works have tried to combine clustering algorithms with deep learning methods, for instance as using k-means objectives to start training a network. However, simply combining clustering and representation learning methods often results in degenerate solutions, where one cluster dominates [3].

Invariant Information Clustering (IIC) is a method that can address these issues, as it is a clustering algorithm that trains a function, such as a neural network, to assign inputs into a group without needing labels [1]. It utilizes information theory and an objective function based on the mutual information between paired image samples [1], [4]. Because of the high-level nature of this method, the input data as well as transforming function can be of many types, and in the context of a discrete clustering space, the mutual information objective used for training can be calculated exactly.

In this report, experiments and results from adaptation of the IIC method to a dataset containing images of 15 classes of human activities are discussed. This involves the steps taken to increase the likelihood that the model is able to learn the features of the images, and methods for improving robustness such as the use of an overclustering technique. Details regarding the preprocessing techniques to provide the model with paired and diverse samples are explained, as well as the models tested. In addition to clustering, results from a supervised run with a similar model architecture are presented.

In spite of the efforts to adapt IIC to the chosen HAR dataset, none of the IIC models showed promising performance. The supervised models were far more successful, suggesting that the difficulties have to do with the inherent challenge of unsupervised methods, and a lack of discriminiatory power in the IIC objective. Elaboration of the results and interpretations are presented towards the end of the report.

II. Related Work

A. Clustering overview

Clustering is one of the fundamental techniques in data analysis and in machine learning, and is used for grouping similar samples together. It can fulfil various purposes, for example data exploration, pattern recognition, and data compression. By organizing unlabeled data into different groups, it becomes easier to understand the underlying structure and relationships within the dataset, which can allow further insights and more informed decision-making [5].

While IIC aims to learn a suitable representation space, the most well-known clustering techniques, such as K-means or DBSCAN, use metrics directly from the input space to compute clusters.

K-means is a simple and efficient algorithm that partitions data into k clusters by iteratively assigning data points to the nearest cluster center and updating cluster centroids. The k-means algorithm is widely recognized and utilized as a primary clustering method, and there exists numerous extensions proposed in literature. [6]

DBSCAN, which stands for Density-Based Spatial Clustering of Applications with Noise, is also commonly used in machine learning and data mining. It identifies clusters based on the density of data points in the feature space, making it effective in identifying clusters of arbitrary shape and is

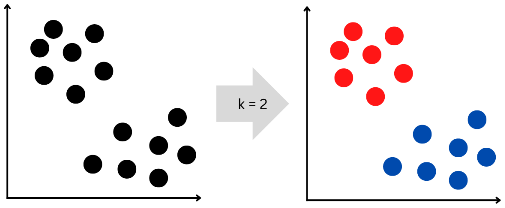


Fig. 1. Adapted from [6].K-means clustering of data in two clusters

robust to noise and outliers. However, it tends to struggle with datasets of varying densities or with high-dimensional data due to the curse of dimensionality [7].

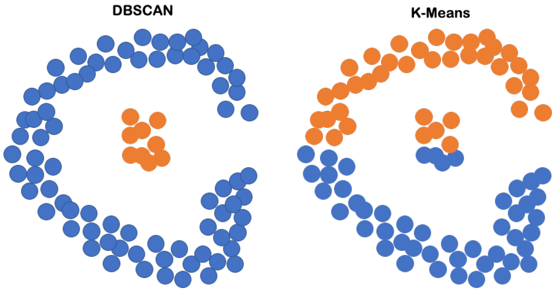


Fig. 2. Adapted from [8]. How clusters are being assigned in DBSCAN clustering vs. how clusters are being assigned in K-Means Clustering

Unlike K-means, which assume convex cluster shapes, and DBSCAN, which relies on density, IIC is not limited by assumptions about cluster shapes. This flexibility allows IIC to directly be applied to data with complex and non-linear structures, such as image data.

B. Invariant representations in images

Invariant representations refers to the goal of extracting essential information from images while disregarding irrelevant variations. This leads to more robust and reliable analysis. These representations enable algorithms to recognize objects or patterns despite changes in appearance, thereby enhancing the generalization capability of machine learning models [9]. In the implementation of IIC, this is accomplished by generating paired samples, where the model learns to recognize the relevant part of the image.

The importance of learning invariant representations in unsupervised feature learning and deep learning has previously been demonstrated [10]. There was shown a necessity for robust feature representations that are capable of working with diverse variations in input data. This included changes in viewpoint, scale, rotation, illumination, and occlusions. The invariant representations are crucial for the generalization of machine learning models across different conditions and environments.

C. Previous work on IIC in images

Our experiments inherit from the training regime formulated in [1], see figure 3. The authors were able to prove the usefulness of IIC on images from various domains, and set benchmark records at the time for unsupervised clustering of several dataset. The robustness was also verified by using semi-supervised settings. The advantages of the supporting overclustering was demonstrated through ablation studies, showing that it greatly improved the robustness of the model. The code was made available [11], which was useful and much was adapted when building models for the new experiments presented here.

Another report found that while image-based IIC can be effective, the algorithm is sensitive to certain types of non-uniform input data, and requires careful attention towards the transformations applied in image generation [12]. This highlights that while IIC may be effective, it can require extensive testing and may fail with certain datasets or insufficient models.

In addition to scientific publications, a practical implementation of IIC has been made available through the code sharing platform Github [13]. This work builds directly upon the repository in [11]. A comprehensive implementation of IIC for images, specifically the MNIST dataset [14], is detailed. This resource provides an accessible and detailed tutorial of the IIC algorithm, complete with code examples and applications to different image datasets. The repository aids both in understanding the algorithm but also serves as a valuable tool for practitioners looking to apply IIC in their own projects.

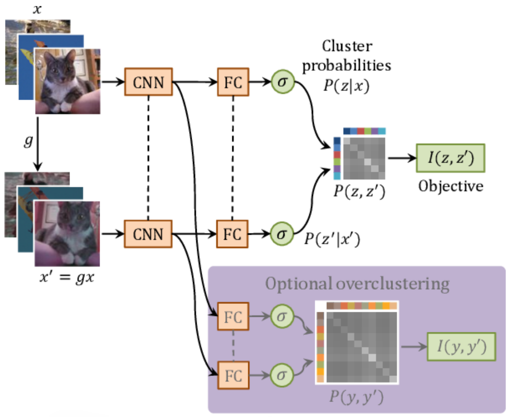


Fig. 3. Adapted from [1]. The workflow used for training the model in IIC. The original image and its augmented version are passed through the same function, in this case a neural network. The softmax function assigns class-probabilities to both images, resulting in the probability-matrix P. This matrix can be used as input for the invariant information objective. There is also a second output head used for overclustering, which has a regularizing effect on the model.

D. Image-based HAR

The overwhelming majority of research in image-based HAR has been focused on supervised techniques [15]. This approach has been widely adopted due to its simplicity and effectiveness at enabling precise activity classification. There is also a lower number of studies on still image data, with most studies focusing on video based data. Human actions generally have a temporal element associated, and can be very difficult to classify without any added context. Video-based techniques such as 3D ConvNets can make efficient use of the time-dependency, and has achieved high performance in areas such as anomaly detection or crime classification in surveillance videos [16].

Other examples of HAR research include both handcrafted methods, such as SURF (Speeded Up Robust Features), and deep learning [17]. Additionally, transfer learning, human pose-based techniques, and neural network-based pose estimation have been explored to enhance accuracy and reduce resource requirements. Ongoing research continues to advance HAR methodologies [18].

III. Method

IIC offers a method of clustering mutual information in paired data [4]. In the case of unsupervised learning with images, it aims to group data points such that each cluster is invariant to certain transformations applied to the data. The following outlines the main components and working principles of IIC, and then the details of the models used.

A. IIC Objective

The core idea of IIC is to maximize the mutual information between the cluster assignments of transformed versions of the same input data. Using this objective encourages the model to learn clusters that are invariant to transformations, i.e. it will consistently assign an image and its disturbed version to the same cluster. This is the intuition behind how the model learns to categorize new samples, by assigning it to the most similar cluster based on the information criteria.

The mutual information between two variables U and V [4] is defined as:

$$I(U,V) = \sum_{u \in U} \sum_{v \in V} P(u,v) \log \frac{P(u,v)}{P(u)P(v)}. \tag{1}$$

In the context of IIC, U and V represent the cluster assignments of the original and transformed versions of the input data. In the cases were there can be set a fixed number of classes, these take the shape of n-dimensional probability vectors, where n is the presumed number of classes. Setting the objective to maximize $I(U,V)$ can thus establish a clustering model that is invariant to transformations.

IIC is said to be naturally robust against clustering degeneracy, where one cluster dominates or some clusters disappear [1], [12]. As the objective of maximizing information is parallell to maximizing entropy, it should naturally prevent all images from being assigned to the same class.

B. Models and architecture

The method outlined above generalizes to any function that transforms the input into class probabilities. Because the target dataset is image data, there is a lot of precedence for using neural networks based on convolutional layers, as these functions are achieving state of the art results in HAR from image-based data such as videos

Although it potentially is worthwhile to experiment on constructing a new architecture for the base-model, it is often more time-efficient to test for one or more pre-existing models. The extra resources can then be better spent on fine-tuning model parameters and data preprocessing [19].

All experiments were conducted with the Pytorch library, [20], using the ResNet18 model [21] as a backbone. This is the shallowest of the ResNet models, which makes it the fastest to train. This family of models are utilizing skip connections, see figure 4. This allows for the training of deep networks without suffering from the vanishing gradient problem [21]. The architecture also leverages identity mappings which lets parts of the network learn to use the identity function when appropriate, improving training efficiency.

When using existing models, one has the choice of using pre-trained parameters in the architecture, or tuning new ones from scratch. The network has been trained on ImageNet [22], a large visual database designed for use in object recognition research. Thus the convolution filters are adept at picking extracting general characteristic features from real objects, and it can be sufficient to keep these weights ”frozen”, and only ad some final fully connected layers at the end for processing of the extracted features and final classification. This can therefore reduce the training time, as it means that only a fraction of the networks total parameters have to be updated. It was decided that this approach was sufficient for the supervised experiment due it’s aim of proving the ability of the relatively shallow ResNet18 architecture. For the unsupervised experiments, as they were the focal points of investigation, it was decided to expend the extra resources to train all model parameters.

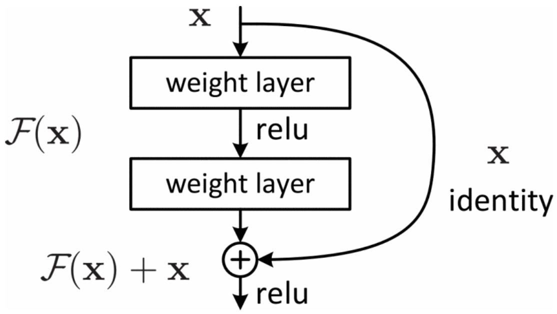


Fig. 4. Adapted from [21]. Illustration of the residual connections utilized in ResNet18.

C. Overclustering

Similar to [11], [13], the overclustering technique was used. This refers to occasionally training the model with a higher number of output clusters than the presumed number of ground truth clusters. It essentially helps the model when working on noisy data, where it may appear to have more classes than in the default output layer. Visually, the process is described in figure 3, and practically, it involves having an extra output layer in the model, with considerably more output nodes (classes). The model will then be instructed to train with this auxiliary output head a given amount of the time.

D. Data

The target dataset [23] consists of 12600 still images, each labelled into 1 out of 15 classes. The classes describe human activities such as using a laptop or drinking. The activities take place in various scenes, both indoor scenes, such as in a home, concert hall, and outdoor scenes. In several images, there are several of the class-member activities take place simultaneously, e.g. sitting while clapping, making classification challenging even to the human eye.

The fact that the dataset is labelled makes it suitable for early testing of IIC in HAR. This allows the model to be evaluated by how well the clusters capture the ground-truth labels. All models could thus be evaluated by measuring accuracy and inspection of confusion matrix.

IV. Results

A. IIC

1) Experimental Setup: The experiments involved multiple runs of the IIC setup with and without overclustering and varying numbers of training epochs. Clusters were assigned based on the frequency of classes within the clusters. The models were trained and evaluated on different subsets of features, to address if the number of classes lead to a different performance.

In the experiments, the models were trained on three different subsets of the dataset, each designed to evaluate the clustering performance under varying levels of class complexity and sample size.

- Full Dataset: This subset contains all 15 original classes. It represents the full range of the dataset and labels.
- Merged Classes: In this subset, similar classes were merged into higher-level features. For instance, 'biking' and 'cycling' were combined into a single class 'transportation'. This resulted in a total of 9 classes, with all samples from the original dataset included. This setup tests the model's ability to generalize across higher-level class definitions, while merging some of the similar and potentially most confusing classes.
- Reduced Dataset: This subset contains only three categories, each consisting of 2 original classes. This means the model was trained on only 2/5 of the total data. This setup focuses on evaluating the model's performance when dealing with fewer classes and a subset of semantically very distinct classes; 'eating or drinking', 'being stationary' and 'transportation'.

For all three subsets, the original images were preprocessed and scaled to normal values using the PyTorch transformation module. The transformations applied to the paired images include a variety of augmentations and randomness to ensure robustness of the trained models, and include:

- Random Horizontal Flip: Flips the image horizontally with a certain probability.
- Random Vertical Flip: Flips the image vertically with a certain probability.
- Random Rotation: Rotates the image by a random degree within a specified range.
- Random Crop: Crops the image to a random size and aspect ratio.
- Color Jitter: Randomly changes the brightness, contrast, saturation, and hue of the image.

These transformations were chosen as they have been noted to help a model learn invariant features and increase its generalizing abilities [1]. The paired data, i.e., transformations of the original images, were systematically altered to include these variations.

Feature Subset	Total Accuracy
A (15 classes)	12.2 %
B (9 classes)	9.0 %
C (3 classes)	39.0 %

TABLE I
Clustering accuracies for different subsets of available data.

2) IIC: Despite multiple runs and variations in experimental parameters, the resulting accuracy was generally low, with clustering accuracies being only slightly better than random guessing (i.e. zero training). The results for the best performing models, tuned by overclustering regime and epoch-number, are displayed in table I. Notably, the most promising results were obtained when using all 15 classes in the dataset. In this scenario, the clustering accuracy reached 12%, which was twice as good as random guessing (6.3%). However, as can be seen in the confusion matrix in figure 5, there is a significant degeneracy in the clustering. One cluster, primarily representing the "fighting" category, dominates the labelling, although a few other categories such as 'cycling' and 'using_laptop' received assignments. Experimentation with a smaller neural network architecture, based solely the basic convolutional building blocks and pooling layers in Pytorch, yielded faster degeneration, and into only class, indicating the sensitivity of the clustering algorithm to model choice.

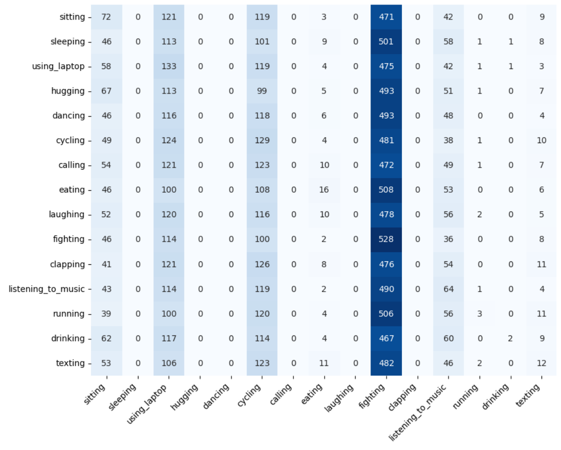


Fig. 5. Illustration of the samples predicted labels (x-axis) and true labels (y-axis), for the highest performing IIC-model on the full dataset.

3) Interpretation: The observed degeneracy in clustering suggests challenges in effectively capturing the underlying structure of human actions in images using IIC. This highlights potential biases or limitations in the dataset or clustering approach. Despite efforts to optimize experimental parameters and explore different subsets of classes, the results exemplify the difficulty in achieving meaningful clustering of HAR based on image data with the unsupervised methodology.

In order to further gain insight into the decision-making of the model, some samples from the same clusters are shown in figures 6 and 7, taken from the relatively best model trained on 15 classes. Despite its limitations, its possible that some visual patterns or features are shared among images within the same cluster. On the whole, while there in some instances appear to be logical patterns among the misclassification, such as the appearance of a phone leading to a sample being wrongly classified as ’texting’, there are just as many unexplainable classifications. This leads to the conclusion that the model in general was not able to learn the invariant features among the same class samples.

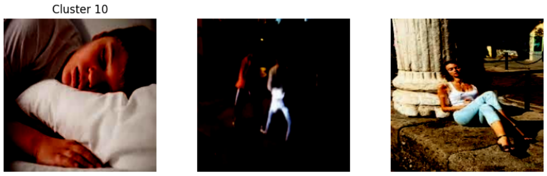


Fig. 6. 3 randomly selected pictures that were grouped in the same cluster that picture people sleeping in 2 out of those 3, while the third one is not discernible.

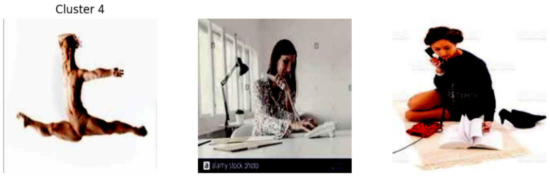


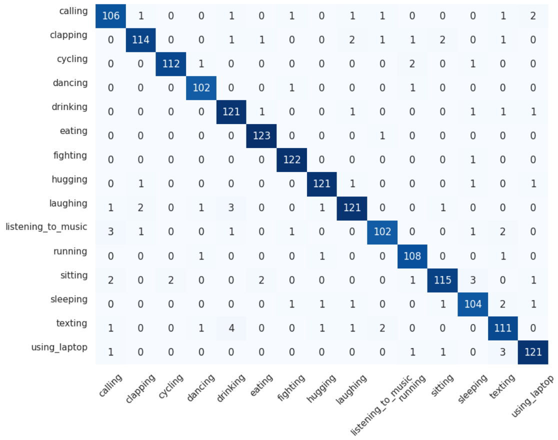
Fig. 7. 3 randomly selected pictures that were grouped in the same cluster that could group pictures of people talking on the phone. It seems that the model pays more attention into the position of the body than to the phone object.

B. Supervised Classification

1) Comparison: In contrast to the unsupervised clustering experiments, the results from the supervised classification model were positive. The confusion matrix is shown in figure 8.

The supervised ResNet18 network achieved a high accuracy of 95% after training for 25 epochs on the same dataset. Notably, both the unsupervised clustering and supervised classification tasks utilized the ResNet18 architecture, suggesting that the underlying model capacity was not the primary factor contributing to the large gap in performance.

2) Interpretation and Implications: The big contrast in performance between the unsupervised clustering and supervised classification tasks raises some important considerations. The high accuracy attained by the supervised model shows the discriminative power of the ResNet18 architecture on the target dataset and the feasibility of learning meaningful representations from a HAR perspective. The discrepancy in performance between the two tasks suggests that the objective of the unsupervised clustering task may be a lot more more challenging or not even sufficient for a model of this scale to learn representative features effectively. This shows the complex nature of unsupervised learning tasks, especially in a domain such as image-based HAR, where the underlying structure is intricate.



	calling	clapping	cycling	dancing	drinking	eating	fighting	hugging	laughing	listening_to_music	running	sitting	sleeping	texting	using_laptop
calling	106	1	0	0	1	0	1	0	1	1	0	0	0	1	2
clapping	0	114	0	0	1	1	0	0	2	1	1	2	0	1	0
cycling	0	0	112	1	0	0	0	0	0	0	2	0	1	0	0
dancing	0	0	0	102	0	0	1	0	0	0	1	0	0	0	0
drinking	0	0	0	0	121	1	0	0	1	0	0	0	1	1	1
eating	0	0	0	0	0	123	0	0	0	1	0	0	0	0	0
fighting	0	0	0	0	0	0	122	0	0	0	0	0	1	0	0
hugging	0	1	0	0	0	0	0	121	1	0	0	0	1	0	1
laughing	1	2	0	1	3	0	0	1	121	0	0	1	0	0	0
listening_to_music	3	1	0	0	1	0	1	0	0	102	0	0	1	2	0
running	0	0	0	1	0	0	0	1	0	0	108	0	0	1	0
sitting	2	0	2	0	0	2	0	0	0	0	1	115	3	0	1
sleeping	0	0	0	0	0	0	1	1	1	0	0	1	104	2	1
texting	1	0	0	1	4	0	0	1	1	2	0	0	0	111	0
using_laptop	1	0	0	0	0	0	0	0	0	0	1	1	0	3	121

Fig. 8. Confusion matrix from the best supervised model.

V. Conclusion

The investigation into the application of IIC into the domain HAR was not able to show promising results. It has been noted that IIC tends to be sensitive to the type of input, and further requires very careful attention during training to perform well [12]. The good results and benchmarks achieved previously [1] appear to not be readily transferable to this task. It is possible that with meticulous experimentation with model parameters, different selection of models, and/or different data preprocessing, that a better result was feasible.

The well-known difficulty [15] of HAR in images likely add to the challenge. Human actions can be highly variable and involve subtle differences that are difficult for an unsupervised model to capture without labeled data. IIC relies on the assumption that different transformations of the same image should remain in the same cluster. However, this might not provide enough discriminative power.

Although it is also possible that the chosen neural network model was not complex enough, and therefore not able to learn the nuanced features in the IIC setting, the almost exact same architecture was able to achieve high performance in the supervised setting.

Further explorations of IIC in HAR could try to test models that are fed with different types of transformations. If the quality and appropriateness of the image transformations is not sufficient, the model may not be able to learn the relevant features. The sobel transformation, although found effective in the aforementioned successful runs [1], was for instance not implemented here. Alternatively, a different choice of HAR dataset could be necessary for good performance, as there was encountered a number of samples that potentially could belong to several classes.

While HAR is a challenging task, it is also of much importance, and as the number of real world datasets is insufficient, the application of unsupervised techniques in this domain holds potential. The area of HAR in images remains underexplored, and this investigation into the use of IIC indicates that it may not be the most promising approach. The results suggest that IIC struggles with the complexity of human actions in images, and the need for experimenting with different unsupervised techniques. A future application of this method in HAR will require innovative approaches and substantial breakthroughs.

References

[1] X. Ji, J. F. Henriques, and A. Vedaldi, "Invariant information distillation for unsupervised image segmentation and clustering," *CoRR*, vol. abs/1807.06653, 2018. [Online]. Available: http://arxiv.org/abs/1807.06653

[2] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320303642

[3] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," 2019.

[4] E. G. Learned-Miller, "Entropy and mutual information," Department of Computer Science, University of Massachusetts, Amherst, Amherst, MA 01003, Sep. 2009, abstract.

[5] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217311815

[6] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80 716–80 727, 2020.

[7] D. Deng, "Dbscan clustering algorithm based on density," in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, 2020, pp. 949–953.

[8] Yufeng, "Understanding dbscan and implementation with python," *Towards Data Science*, 2024. [Online]. Available: https://towardsdatascience.com/understanding-dbscan-and-implementation-with-python-5de75a786f9f

[9] W. Du, H. Chen, and H. Yang, "Learning invariant representation for unsupervised image restoration," 2020.

[10] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives," *ArXiv*, vol. abs/1206.5538, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:4493778

[11] X. Ji, J. F. Henriques, and A. Vedaldi, "Invariant information clustering," 2019, gitHub repository. [Online]. Available: https://github.com/xu-ji/IIC

[12] Y. Dimitrov, "Feasibility study of "invariant information clustering for unsupervised image segmentation"," master thesis, Delft University of Technology, 10 2021, to reference this document use: http://resolver.tudelft.nl/uuid:21248d5f-bfde-4805-811f-e0db76289d67.

[13] A. Naumov, "Iic tutorial," 2021, gitHub repository. [Online]. Available: https://github.com/vandedok/IIC_tutorial

[14] Y. LeCun, C. Cortes, and C. J. Burges, "Mnist handwritten digit database," 2010, accessed: 2024-05-25.

[15] F. Scheidegger, R. Istrate, G. Mariani, L. Benini, C. Bekas, and C. Malossi, "Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy," *arXiv*, 2018, submitted on 26 Mar 2018.

[16] A. Goyal, M. Mandal, V. Hassija, M. Aloqaily, and V. Chamola, "Captionomaly: A deep learning toolbox for anomaly captioning in social surveillance systems," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 207–215, 2024.

[17] S. Megrhi, W. Mseddi, and A. Beghdadi, "Spatio-temporal surf for human action recognition," 12 2013.

[18] A. Siyal, Z. Bhutto, S. Muhammad, M. S. S. Syed, A. Iqbal, F. Mehmood, A. Hussain, and S. Ahmed, "Still image-based human activity recognition with deep representations and residual learning," *International Journal of Advanced Computer Science and Applications*, vol. 11, pp. 471–477, 06 2020.

[19] I. Oztel, G. Yolcu, and C. Oz, "Performance comparison of transfer learning and training from scratch approaches for deep facial expression recognition," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, 2019, pp. 1–6.

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[23] M. Nagadia, "Human action recognition (har) dataset," 2022, accessed: 2024-04-25. [Online]. Available: https://www.kaggle.com/datasets/meetnagadia/human-action-recognition-har-dataset/