**Nearest Neighbour of Different Image Representation**

**Literature review:**

**Submitted by Ankur Chowdhury (1911844042)**

**& Sajid Wasif (1912313642)**

1)

Xianzhong Long et al. [1] described two categories of image classification, including scene recognition and object recognition but they only focus on object recognition problems of image classification. To improve the various image classification performance, they used different models. Among these models, the BoW and SPM model won the greatest popularity and has a wide range of applications in the fields of image retrieval and image classification. They used Spatial Pyramid Matching model (SPM) in Their paper and name their method Nearest Neighbor Basis Vectors Spatial Pyramid Matching (NNBVSPM). They Experiments on different kinds of datasets and demonstrate that classification rate of their soft inner product coding is better than the previous three classical methods.

2)

V´ıt LIST´IK et al. [2] want to prove that although it is possible to reconstruct the image from the semantic feature vector. The task is to generate an image from the semantic feature vector which will be very similar to the original image. The task is the same as for autoencoder with a difference of using pre-trained CNN. They used only the images, not the labels. The original dataset consists of 14M labeled images. They are using random subsets of the dataset. They used pre-trained ResNet for the extraction. Based on their results they concluded that is not possible to reconstruct the private information.

3)

Jia Deng et al. [3] introduce a database named "ImageNet" in this article, which is a sizable ontology of images built on the WordNet structure. The majority of their WordNet’s 80,000 synsets will be filled with an average of 500–1000 crisp, full-resolution images. This will provide them with tens of millions of tagged photographs arranged according to WordNet's semantic hierarchy. They run four different object recognition algorithms in their experiments, which are NN-voting + noisy ImageNet, NN-voting + clean ImageNet, NBNN, and NBNN-100. They exhibit constant improvements as the depth of the embedding function rises, enabling them to attain 60.2% top-1 accuracy on ImageNet classification. They got an average accuracy of 10% of image search results from the Internet, and 99.7% precision is achieved on average. At long last, they will keep looking at more efficient ways to assess the AMT user labels and optimize the number of repetitions required to precisely verify each image in order to speed up the development process even more.

4)

Chengxu Zhuang et al. [4] used a neural network or Local Aggregation (LA) method which nonlinearly embed inputs in a smaller space and they identified close neighbors and background neighbors. They identified two sets of neighbors for an xi and its embedding vi. They used Bi for Nearest-neighbor based identification and Ci for Robustified clustering-based identification, to identify close neighbors, they applied an unsupervised clustering algorithm. They Followed the methods of AlexNet and VGG16 architectures, to add batch normalization (BN) layers in their experiment. They used K-nearest neighbor (KNN) classification results using the embedding features. With all methods, Local Aggregation (LA) performs much better than alternative methods. LA trained ResNet-50 achieves 60.2% top-1 accuracy on ImageNet classification. For the LA method, they consistently see performance gains from both overall deeper structures and from early layers to deeper layers within an architecture.

5)

Mathilde Caron et al. [5] Implemented DeepCluster to train convolutional neural networks unsupervised on massive data like ImageNet and YFCC100M. They train the convnet's parameters by iteratively clustering deep features and using the cluster assignments as pseudo-labels. When they used supervised setting, deeper architectures VGG or ResNet they have a much higher accuracy on ImageNet than AlexNet. They used convnets for mapping raw images to a vector space of fixed dimensionality. On top of the last convolutional layer of a random AlexNet, a multilayer perceptron classifier achieves 12% accuracy on ImageNet whereas the chance is only 0.1%. Finally, they used DeepCluster on image classification, object detection and semantic segmentation on Pascal VOC. DeepCluster methods performs only slightly better than previously published methods. They conclude that when trained on big datasets such as ImageNet or YFCC100M, it outperforms the prior state-of-the-art on every conventional transfer task.

6)

ICSI Zhirong Wu from UC Berkeley (2018) described non-parametric instance discrimination for unsupervised feature learning. The approach obtains a mean average precision of 65.4% with Resnet-50. When m = 1, the accuracy of KNN falls to only 42.5%. The technique performs better than the state-of-the-art on ImageNet and Places for image categorization. On tasks involving semi-supervised learning and object detection, it provides competitive generalization results. With more data and deeper networks, the approach scales effectively and has a small 128-dimensional representation. Some of the results might build upon past research in this area: "Our algorithm beats the state-of-the-art on image classification by a considerable margin, with top-1 accuracy 42.5% on ImageNet 1K and 38.7% for Places 205," said Zhirong Wu of UC Berkeley.

7)

"ImageNet," a sizable ontology of images, is a brand-new image database that Jia Deng et al. (2009) introduce. For visual identification applications like object recognition, image categorization, and object localization, ImageNet might be a useful resource. There has been a huge explosion of data since the advent of the digital age. They want to identify the object class of an image containing an unidentified object by looking up related photos in ImageNet. They plan to have around 50 million clears, varied, full-resolution photos distributed across 50K synsets when ImageNet is finished. In the investigation, there were 12 subtrees, 5247 synsets, and 3. In their conclusions, the authors state that their findings are consistent with past research in the field: "Torralba et al. has shown that, given a large number of images, basic closest neighbor algorithms can obtain reasonable performances despite a high degree of noise. By searching similar photos in ImageNet, we aim to identify the object class of an image containing an unknown object. The group advises that they consider ImageNet to be a new and difficult benchmark dataset for future research because of its high quality, diversity, and size.

8)

The output of a linear SVM trained on ImageNet-CNN features is presented by Bolei Zhou et al. (2014). The researchers show that the internal representations of scene-centric and object-centric neural networks are different. On all of the existing scene benchmarks, the team offers cutting-edge performance employing deep features. Places is a scene-centric image dataset that the authors introduce. It is 60 times bigger than the SUN database. The researchers demonstrate that a CNN network trained using the Places database outperforms scene-centered benchmarks significantly. Three datasets were used in the analysis. The findings of the authors could complement earlier research on this subject: "In Section 4, we demonstrate new scene classification performance when training deep features from millions of labeled scene photos. We display the replies of the units at various CNN layers," Zhou explained. According to Zhou and colleagues, these two networks perform considerably differently on several recognition standards. Future research will compare and contrast the RF at various layers of the object-centric network with the well-known scene- and object-centered neural cortical circuits seen in the human brain.

9)

In "The Pascal Visual Object Classes Challenge," Mark Everingham et al. (2009) noticed that whereas the VOC dataset was cited in 15 publications in CVPR07, it was cited in 27 papers in CVPR08—nearly tripling in popularity from year to year. Everingham and colleagues examine some of the challenge's complaints, examine how the task is changing as a result of the taster competitions, and offer suggestions for ways to enhance and broaden the dataset and challenge. The PASCAL1 Visual Object Classes (VOC) Challenge consists of two parts: an annual competition and workshop and a publicly accessible collection of photos and annotation. Pixel-level segmentations of the visible region of all included items were added to a subset of images from each of the primary datasets. The analysis comprised 439 annotated individuals. Some of their conclusions are said to support earlier studies on this topic: "INRIA entered two approaches employing the same channels, but they were integrated differently. According to Everingham, a few of the 2006 entries approached the categorization task as detection: "There is a car here, so the image contains a car."

10)

Target detection, according to Ningwei Wang et al. (2021), is a crucial step in deep learning. They document experiments on the LISA Neighbor Feature Pyramid Network. The experiment makes use of NFPN, ResRoIE, and RFP in contrast to some widely used networks that are built on two stages. Pyramid Network (FPN), which is top-down based, recognizes each feature layer separately. By combining feature data from many levels, the Recursive Feature Pyramid (RFP) fusion approach lessens imbalance. The researchers' findings in some ways seem to support earlier study on the topic: "symmetry is utilized to merge knowledge acquired from two components. To choose the most desirable characteristic and anchor box, parallel anchor-based and anchor-free branches run in symmetry, Wang recommended.

**References:**

1. <https://link.springer.com/article/10.1007/s11042-012-1289-4>
2. <http://poseidon2.feld.cvut.cz/conf/poster/poster2018/proceedings/Poster_2018/Section_IC/IC_057_Listik.pdf>
3. [J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li,... - Google Scholar](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=3.%09J.+Deng%2C+W.+Dong%2C+R.+Socher%2C+L.-J.+Li%2C+K.+Li%2C+and+L.+Fei-Fei.+Imagenet%3A+A+large-scale+hierarchical+image+database.+2009.+1%2C+13&btnG=)
4. [Local Aggregation for Unsupervised Learning of Visual... - Google Scholar](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Local+Aggregation+for+Unsupervised+Learning+of+Visual+Embeddings&btnG=)
5. [Unsupervised Learning | Neural Computation | MIT Press](https://direct.mit.edu/neco/article-abstract/1/3/295/5492/Unsupervised-Learning)