Automatic Tagging Suggestion for Database Enrichment

Emanuele Cuono Amoruso [0009-0008-7302-5862], Sietske Tacoma [0000-0002-9662-8489] and Huib Aldewereld [0000-0001-6339-3241]

HU University of Applied Sciences, Utrecht, The Netherlands
manu.amoruso00@gmail.com, sietske.tacoma@hu.nl, huib.aldewereld@hu.nl

Abstract. The huge number of images shared on the Web makes effective cataloguing methods for efficient storage and retrieval procedures specifically tailored on the end-user needs a very demanding and crucial issue. In this paper, we investigate the applicability of Automatic Image Annotation (AIA) for image tagging with a focus on the needs of database expansion for a news broadcasting company. First, we determine the feasibility of using AIA in such a context with the aim of minimizing an extensive retraining whenever a new tag needs to be incorporated in the tag set population. Then, an image annotation tool integrating a Convolutional Neural Network model (AlexNet) for feature extraction and a K-Nearest-Neighbours classifier for tag assignment to images is introduced and tested. The obtained performances are very promising addressing the proposed approach as valuable to tackle the problem of image tagging in the framework of a broadcasting company, whilst not yet optimal for integration in the business process.

Keywords: Automatic Image Tagging, Image Database, News Broadcasting.

1 Introduction

The increasing use of the Internet has led to an ever-growing number of images shared on the Web. The development of new and effective cataloguing methods is becoming worthy of notice. In this regard, an efficient tagging mechanism, tailored to the specific end-user needs, is crucial. Here we discuss the problem of the management of an image database for a broadcasting company. The specific case concerns "Nederlandse Omroep Stichting" (NOS) and its image database. NOS is one of the broadcasting organizations making up the Dutch Public Broadcasting system. As an organization, NOS is responsible for news, sport, political and events programming on the public service television networks, broadcasting on the main three public television channels. Every day, NOS processes many images in its image database. These need to be stored and catalogued.

This study focuses on a fundamental part of this cataloguing process: image tagging. While tagging plays a crucial role in image retrieval, it is typically entrusted completely to the uploaders. Currently, the uploaders can insert tags in two ways: a free form, where tags are plain text, allowing them to insert any desired text, and a fixed form, where tags must match the predetermined keyword dictionary employed

by the company. These tagging methods allow for a practical implementation of Text-Based Image Retrieval (TBIR), in which the retrieval of an image is based on text queries related to the textual metadata associated with the images themselves [1]. TBIR presents a viable solution for a matter strongly connected with the news realm, namely the constant demand of updates to keep track of current developments. The ever-changing nature of the world requires regular revision of the keyword dictionary. However, TBIR suffers a significant fallacy in the application context, namely the requirement of manual labelling for each image. The manual labelling necessitates an important allocation of human resources, especially in a context where the image acquisition is constant, such as in the realm of a news company. Consequently, this leads to an image tagging process which is not up to date, culminating in a considerable number of untagged images uploaded into the database. Therefore, despite the vast population of images within the database, their usage is impoverished due to the inability to retrieve them effectively. Furthermore, the allowance of unrestricted tagging in the free form case or within a predefined dictionary in the fixed form introduces a potential inconsistency in image tags. For instance, similar images uploaded by different individuals may be associated with a distinct set of tags. While these tags could be an accurate description of the image, the retrieval process diverges substantially based on the tags linked to the images [2].

Alternative methodologies for image retrieval are explored in existing literature, such as Content-Based Image Retrieval (CBIR). CBIR involves the retrieval of images based on their low-level visual features, such as colours, shapes, and space relationships [3]. However, in the aforementioned context, CBIR is not useful for the end-users, usually journalists, who seek appropriate images to complement their work. In practice, journalists want to find a suitable image using keywords directly related to the content of their news item. These keywords should be image tags that reflect the content of the image, thereby expressing their semantic meaning, which entails the analysis and interpretation of the visual content through detection and recognition of objects, image classification and related techniques. To accommodate these requirements, Automatic Image Annotation (AIA) emerges as a potential concept solution. AIA is a technique used to describe images by automatically assigning appropriate semantic tags for images fed to the model [4]. The goal of AIA is to improve efficiency and accuracy in image annotation, which is time-consuming and prone to errors when done manually. Once images are automatically annotated, they can be retrieved using the tags, making image retrieval similar to text document retrieval. Moreover, the application of AIA concepts can address a prominent issue in the news context, namely the need of new tags. While this may not be as critical in other applications, the employment of new tags in the news context is crucial. This is primarily due to the need of staying updated, with the constant emergence of new pieces of information popping up daily.

This study investigates the applicability of AIA in the domain of image tagging, with a specific focus on the tag set expansion within the news context. The primary objective is to determine the feasibility of using AIA in this context to minimize the necessity for extensive retraining whenever a new tag needs to be incorporated in the tag set population. The study is structured as follows. Section 2 provides an overview

of Automatic Image Annotation (AIA) applications, presenting important concepts relevant to the investigated case. Section 3 presents the selected methodologies, introducing the baseline framework employed in this study and formulating the fundamental research questions. The experimental results, aimed at addressing the research questions, are presented in Section 4. Finally, conclusions are drawn in Section 5.

2 Related AIA applications

Applications within the AIA context are diverse and exhibit distinct characteristics. One illustrative example is the Image Annotation tool in which a Wasserstein Generative Adversarial Network (WGAN) is employed as a data augmentation mechanism in the context of an end-to-end image annotation model, as presented by Ke et al. [5]. Further insights into this topic can be found using the framework presented by Cheng et al. [2] (See Figure 1), which enable analysing various kinds of AIA applications, each following a distinct philosophy regarding tag prediction. When considering the applicability of the models shown, it is crucial to understand their suitability within the NOS's context. The most important concepts can be gained from two kinds of models in Figure 1: Nearest neighbour models and Deep learning-based models. Deep learning-based models use deep learning algorithms to derive robust visual features from the images. The deep neural networks allow the handling of high-dimensional feature vectors, enabling the exploration of high-dimensional feature spaces. Through the feature vectors, more complex pattern, such as the presence of a particular object in the scene, can be captured from the images themselves. Instead, Nearest neighbour models retrieve a set of top k similar images from candidate datasets.

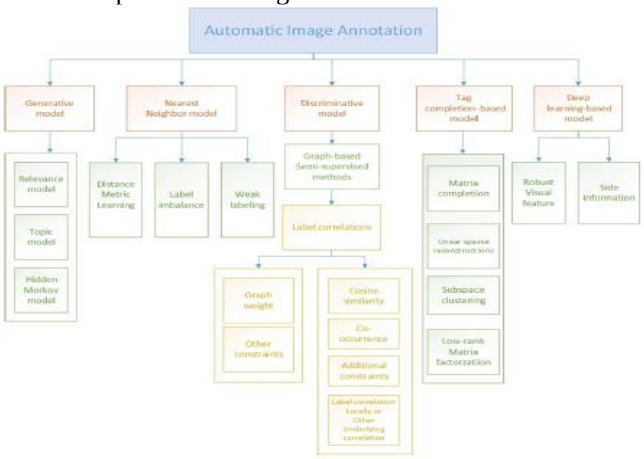

Figure 1 Taxonomy of Automatic Image Annotation techniques from Cheng et al. [2]

The underlying idea of Nearest Neighbour models is that images with similar features are likely to have similar annotations or tags. Thus, by using the existing database of similar images, a new unlabeled image can be appropriately tagged. This concept addresses the need of constant tag update needed in the news context. By employing

the concept of image similarity and using images already present within the database, the process of updating tags can rely on the database population itself. Furthermore, an additional important concept, which could be derived from the models presented [2] is the feature vector extraction done by the Deep Learning-based models. The process of feature extraction could pose a challenge if the features are to be represented manually; however, the employment of Deep Learning techniques to extract relevant patterns from images can aid in improving the representation of the image itself through the feature vector, enhancing the calculation of a more precise distance in the Nearest Neighbour setting.

Therefore, combining Deep learning-based feature extraction and image similarity concepts derived from KNN for tag propagation seems a promising approach. This has been done before in the Siamese Network Architecture, introduced in Koch et al. [6]. This architecture is a powerful approach for a one-shot image recognition system, and it is composed of two identical Convolutional Neural Networks (CNNs) sharing weights, that function as feature extractors for two different images. Subsequently, a distance layer calculates the distance between the two feature vectors generated, enabling the determination of whether the considered images represent the same entity or not. Hence the training focuses on determining an appropriate threshold for classifying the images equality. As a result, the main applications typically revolve around Face Recognition, as proposed in Wu et al [7], or other recognition tasks based on biometric features [8-10], but also in other contests, such as in Liu et al [11], in which this architecture is employed for image classification in a Remote Sensing Scene setting. Due to the application of the fundamental image tagging concepts, the Siamese Network Architecture was contemplated as a potential solution for tackling NOS's task. Nonetheless, despite the promising nature of the concept in theory, the challenges within the experimental setting are non-trivial. A significant challenge is the construction of the reference database needed for conducting image comparisons. This reference database needs to be constructed manually, thereby reintroducing a potential human error into the process. Moreover, within the news context, the setting could greatly vary between images, making it challenging to determine what to compare and when. Therefore, the definition of the problem in the present context makes it unfeasible to develop a tool based on the Siamese Network Architecture.

An example explored for the actual conceptual solution presented in this study is the work of Ma et al. [12], where a comprehensive model composed of a Deep Learning-based feature extractor and a so-called Semantic Extension Model (SEM) is presented. The SEM employs a tag propagation technique inspired by the K-Nearest-Neighbours (KNN) algorithm, as it gathers the feature information pertaining to the images in the database and predicts the tag propagation via a Bayesian-based method. This method functions as an inspiration for the conceptual solution presented in Section 3.

3 Framework Baseline and Methodologies

3.1 Methodologies and Research Questions

The primary objective of this study, as well as of the experiments illustrated in Section 4, is to assess the viability of employing an image tagging tool based on the Deep-Learning feature extractor and the similarity comparison concepts, using KNN comparisons. This tool is compared to a traditional tool based on an image classifier that employs fully connected layers for image classification and tag assignment. Additionally, this study focuses on exploring the flexibility of the tool based on KNN comparison, investigating its behaviour when the tag set expands. This interest is driven by the differences between the tools, particularly the training requirements for the addition of new tags. While a traditional classifier necessitates an extensive retraining each time a new set of tags is introduced, a tool based on a KNN classification methodology may significantly reduce the time and effort to incorporate new tags into the tag set. The study seeks to gain insights into the practical advantages of using a KNN-based approach, which can efficiently adapt to a growing tag database.

The current study presents a solution that aims at addressing specific questions within the domain of the news context. More specifically, the following research questions are tackled:

1. How does the accuracy of the KNN-based automatic image tagging tool compare to that of a more traditional image classifier?

2. Is the KNN-based automatic image tagging tool able to maintain its performance as the number of tags increases?

3. What is the minimum number of images with a new tag required for the KNN-based tool to produce acceptable results, with performances like the ones before the tag insertion?

These research questions are addressed in this study by adopting a straightforward model architecture introduced in next Section 3.2. The primary objective of the model is to extract image features from images in the database and carry-out a single-label annotation via a KNN classification methodology.

The choice for single label is not the only option available. According to Zhang et al. [4], AIA has different kinds of applications. One such application is single label annotation, whereby an image is associated with a single tag, thereby confining the categorization to a single aspect of the whole image. Another is multi-label annotation, which avoids the limitation of a narrow classification, but introduces the complexity of recognizing multiple labels simultaneously. Although multi-label annotation appears to be more specific for the image tagging task under consideration, here we adopt a single label annotation approach due to its direct relevance to the research questions.

The experimental evaluations use two distinct datasets, presented in Section 3.3. The performance scores reported are based on the accuracy of the image tagging tools in the context of image classification. In this scenario, the model is provided with a single image as input and its task is to predict the corresponding tag.

3.2 Model Architecture

Based on the previous considerations, here we introduce a solution combining Nearest Neighbours models and the Deep Learning-based models. Drawing inspiration from the work of Ma et al. [12], as discussed in Section 2. The model architecture employed in this study is based on a Deep-Learning model, AlexNet. The choice of AlexNet is motivated by the work of Ma et al. [12], where the AlexNet is truncated, and the resulting vector from the second fully connected layer (FC2) is considered as feature vector for the following procedures. In this study, we adopt the same approach by removing the last linear layer (see Figure 2) to extract the feature vector. To address the first research question concerning accuracy comparison, we also use AlexNet as a traditional image classifier, without truncating the last fully connected layer. For accuracy comparison, as the first research question requires, the AlexNet is used also as a traditional image classifier, so without the truncation of the last fully connected layer. In both the implementations, AlexNet is retrieved as pre-trained from the torchvision[1] library. For the KNN classification, the KNeighborsClassifier from the scikit-learn library is employed.

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 64, 55, 55]	23,296
ReLU-2	[-1, 64, 55, 55]	0
MaxPool2d-3	[-1, 64, 27, 27]	0
Conv2d-4	[-1, 192, 27, 27]	307,392
ReLU-5	[-1, 192, 27, 27]	0
MaxPool2d-6	[-1, 192, 13, 13]	0
Conv2d-7	[-1, 384, 13, 13]	663,936
ReLU-8	[-1, 384, 13, 13]	0
Conv2d-9	[-1, 256, 13, 13]	884,992
ReLU-10	[-1, 256, 13, 13]	0
Conv2d-11	[-1, 256, 13, 13]	590,080
ReLU-12	[-1, 256, 13, 13]	0
MaxPool2d-13	[-1, 256, 6, 6]	0
AdaptiveAvgPool2d-14	[-1, 256, 6, 6]	0
Dropout-15	[-1, 9216]	0
Linear-16	[-1, 4096]	37,752,832
ReLU-17	[-1, 4096]	0
Dropout-18	[-1, 4096]	0
Linear-19	[-1, 4096]	16,781,312

Figure 2 Feature Extractor structure based on AlexNet

3.3 Benchmark Image Dataset

The benchmark dataset employed for single label annotation is Caltech-256 [13], an object recognition dataset containing a total of 30,607 images, with each category/tag, representing the object recognizable, containing between 31 and 80 unique images, retained from the pytorch library. The dataset is divided into training, validation and test sets, with an 80%-10%-10% distribution. Consequently, the training set consists of 24485 images, while the test and validation sets both containing 3061 images. The split is executed using the torch.utils.data.random_split function. For the enlargement of the tag set population and to test the number of images needed for an accurate tag prediction a non-overlapping subset of Caltech-101 [14] is used.

4 Experiments and results

4.1 KNN vs Traditional Image Classifier

This experiment was conducted to address the first research question, concerning the accuracy of our conceptual solution model compared to the traditional image classifier based on AlexNet. Prior to conducting the experiment, it is anticipated that there might be a drop in performance for the conceptual solution, due to the use of KNN comparison instead of the traditional classification via the fully connected layers. Depending on the magnitude of the performance drop, the conceptual solution could be considered viable because of its flexibility.

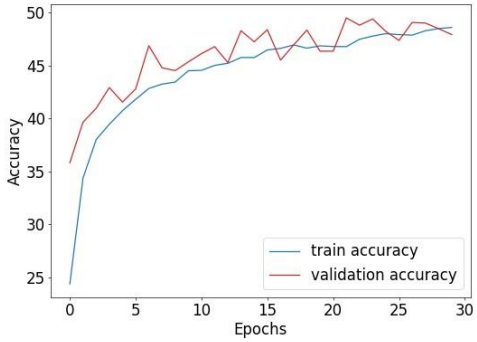

Figure 3 Training and Validation accuracy over the epochs

Regarding the experiment, the baseline AlexNet used as a benchmark is trained over the Caltech-256 dataset. The training process employs an Adam optimizer with a learning rate of 0.01, while the Cross-Entropy Loss function is used as the loss function. Figure 3 illustrates the progression of the accuracy on the training and validation sets over the epochs. Here the training process is stopped after 30 epochs, as its accuracy reaches a plateau, whereas the validation accuracy starts declining, evidencing the occurrence of overfitting. The base AlexNet, without training over the dataset, is used as a feature extractor for the images within the training set. The resulting feature dataset is then used to train the KNN classifier, with a predetermined value of 12 neighbours. The selection of 12 neighbours was determined through experimentation with different numbers of neighbours, and it was found that using 12 neighbours achieved the best performance score over the validation dataset.

Table 1 Accuracy for the two methods

SET/ACCURACY	BENCHMARK	ALEXNET+KNN
Validation	47.91%	43.20%
Test	48.35%	42.56%

Table 1 reports the accuracy results over the test set for both methods. The results indicate a decrease in accuracy for the KNN-based method, achieving a 42.5% accuracy on the test set, whereas a 48.3% accuracy is achieved by the conventional image classifier. As anticipated, a performance drop for the KNN is present, yet its accuracy is acceptable if one considers the advantages offered by this conceptual solution.

4.2 Adding tags

An additional crucial aspect to consider is the performance of the tool when using the KNN approach as the tag set expands. Maintaining a consistent level of performance is a crucial aspect for the adoption of such a tool. This is primarily due to the potential elimination of the necessity for retraining a traditional image classifier with every new tag addition. As a matter of fact, the ability to maintain performance without extensive retraining serves as a compelling factor in favour of employing the proposed tool.

To evaluate the performance of the tool with an expanding tag set, a non-overlapping subset of Caltech-101 is employed. The subset is derived from the larger Caltech-101 dataset by excluding specific images that are already present in the Caltech-256 dataset. The resulting dataset is composed of 3662 new images, with 63 new distinct tags. Both the Caltech-256 dataset and the Caltech-101 subset are partitioned into training and test sets, using a 90%-10% partition ratio. In this case, a separate validation dataset is not extracted as there is no need for hyperparameters fine-tuning, such as the number of neighbours to be employed. Consequently, the sets for Caltech-256 will be composed of 27,552 and 3062 images, respectively for training and test set, and for Caltech-101 subset the division is 3295 training images and 367 test images. Table 2 summarizes the performance analysis of the KNN tool over the two following scenarios: firstly, considering only the Caltech-256 dataset, and secondly, considering the combined dataset of Caltech-256 and Caltech-101 subset.

Table 2 Accuracy with and without expanded tag set

Set/Accuracy	AlexNet+KNN
Caltech-256	40.14%
Caltech-256+Caltech-101 subset	37.94%

A decrease of only 2 percentage points (37.94% vs 40.14%) in the classification accuracy was observed. This appears as marginal in contrast to the significant growth of the tag set population by 63, added to the original 257 tags of Caltech-256. That observation suggests that the 2% decrease can be deemed acceptable and the tag insertion has been successful. Moreover, this performance decrease could be considered acceptable considering the retraining process involved. In this case, the retraining is accomplished by employing the fit function from the scikit-learn library for the KNN classifier. Importantly, this approach does not necessitate the extensive training time required to accommodate additional classes in a traditional classifier.

4.3 Minimum number of images for Tag insertion

The final research question pertains to the minimum number of images required for an effective tag insertion. This specific aspect is crucial for the practical implementation of the proposed solution, as it provides users with guidance regarding the number of tagged images needed to successfully incorporate a new tag into the existing set. To investigate this, the initial training is conducted using the Caltech-256 dataset, while a random subset of stop-sign images from Caltech-101 is used for the tag insertion. It is important to note that the subset selection is entirely random, which may lead to varying accuracy scores for different subsets.

Table 3 Caltech-256 dataset division

Set (from *Caltech-256)*	Ratio -> Number of images (30614 total)
train 1	25% -> 7653
train 2	50% -> 15307
train 3	75% -> 22960
train 4	90% -> 27752

Table 4 Stop-sign subset division

Set (Stop-sign subset)	Ratio -> Number of images (64 total)
train 1 / test 1	25%/75% -> 16/48
train 2 / test 2	50%/50% -> 32/32
train 3 / test 3	75%/25% -> 48/16
train 4 / test 4	90%/10% -> 57/7

The Caltech-256 dataset and the stop-sign subset are partitioned in various degrees, as outlined in Table 3 and Table 4. These divisions are used to evaluate whether a ratio of new images over total images emerges. The experiments involve the computation of the average accuracy of the tag insertion tool for each combination of the partitions shown in Table 3 and Table 4. The results of these experiments are summarized in Table 5, which presents the achieved outcomes. The reported accuracy score is specifically associated with the recognition of the image within the stop-sign subset, thus resulting in higher accuracy scores being obtained because the model is used on a small portion of the image feature space.

Table 5 Average Accuracy for different dataset combinations

Average Accuracy	train 1 (16)	train 2 (32)	train 3 (48)	train 4 (57)
train 1 (7653)	~ 62.25%	~ 81%	~ 78%	~ 88%
train 2 (15307)	~ 51.50%	~ 78.56%	~ 85%	~ 80%
train 3 (22960)	~ 50.6%	~ 78%	~ 80%	~ 80%
train 4 (27752)	~ 43.65%	~ 70%	~ 75%	~ 81%

Table 5 demonstrates that there is no discernible pattern emerging in terms of a specific ratio between the number of new images and the total number of images used, particularly when a larger number of new images is inserted. Instead, the performance of the tool exhibits stability within the range of approximately 40 to 50 images. This

indicates that a threshold for a minimum number of images can be set within this interval. Moreover, the tool appears to be effective even with a small number of images (as depicted in the 1st column of Table 5), as the average accuracy remains comparable to the overall tool accuracy. This observation, in turn, suggests that the tool could be employed even with a limited number of images, with a potential increase in performance as the number of images associated with the specific tag does expand.

5 Conclusions

This study introduced an image annotation tool that integrates a Convolutional Neural Network model, AlexNet, for feature extraction and a K-Nearest-Neighbours classifier for tag assignment to images. Throughout the study, three research questions were investigated and addressed. Firstly, the tool was compared to a baseline image classifier built over AlexNet, showing a comparable performance level. Secondly, the effectiveness of the tool was evaluated as the tag set expanded. Finally, the performance of the tool for a single tag insertion was evaluated, varying the numbers of images used; this investigation aimed at establishing a threshold at which the tool achieves an acceptable performance, defining the minimum number of images required to successfully insert a new tag.

Prospects for further research are discernible. The expansion of the tool capabilities to enable multi-label classification could be explored, thereby expanding the number of tags assigned from the tool to the images. However, this aspect was not explored in the current study due to its focused objective, namely the definition of a general viable solution to the problem presented by NOS. The integration of such a solution should not require extensive research, while the effects of the tag set expansion in the different setting should be inquired.

The presented approach holds the potential to expand NOS's capabilities: integration of similar tools into the operational framework of the organization could empower NOS personnel to embrace new ways of accomplishing their tasks, lifting the burden of manual labour from their shoulders. Consequently, the conceptual solution presented not only addresses the immediate challenge faced by NOS, but also provides a new way of thinking, enabling a new strategy for organizational evolution and enhancement.

Additionally, dedicated efforts could be channeled towards enhancing the performance of the tool to make it more appealing for integration in the company. Notably, the model used as a base is AlexNet, which is a comparatively less aligned with the current advancement and may not be optimally configured to deliver competitive performances with current models. Nonetheless the selection was motivated by its manageability in assessing the problems presented.

Lastly, other performance metrics, such as precision and recall, used in various studies regarding AIA [2], could be considered in order to improve the applicability and interpretation of the results.

References

1. R. Datta, D. Joshi, J. Li, and J.Z. Wang. 2008. “Image retrieval: Ideas, influences, and trends of the new age”. ACM Comput. Surv. 40, 2, Article 5 (April 2008), 60 pages. https://doi.org/10.1145/1348246.1348248
2. Q. Cheng, Q. Zhang, P. Fu, C. Tu & S. Li (2018). “A survey and analysis on automatic image annotation”. Pattern Recognition, 79, 242-259. https://doi.org/10.1016/j.patcog.2018.02.017
3. Latif, Afshan, et al. "Content-based image retrieval and feature extraction: a comprehensive review." Mathematical problems in engineering 2019 (2019). https://doi.org/10.1155/2019/9658350
4. D. Zhang, Md. M. Islam, G. Lu, “A review on automatic image annotation techniques”, Pattern Recognition, Volume 45, Issue 1, 2012, Pages 346-362, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2011.05.013.
5. X. Ke, Z. Jiawei and N. Yuzhen "End-to-end automatic image annotation based on deep CNN and multi-label data augmentation." *IEEE Transactions on Multimedia* 21.8 (2019): 2093-2106. https://doi.org/10.1109/TMM.2019.2895511.
6. G. Koch, R. Zemel, R. Salakhutdinov et al., ”Siamese neural networks for one-shot image recognition”, ICML deep learning workshop, 2015, https://www.cs.utoronto.ca/~gkoch/files/msc-thesis.pdf
7. H. Wu, Z. Xu, J. Zhang, W. Yan and X. Ma, "Face recognition based on convolution siamese networks," 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 2017, pp. 1-5, https://doi.org/10.1109/CISP-BMEI.2017.8302003.
8. C. Lin and A. Kumar, "Multi-Siamese networks to accurately match contactless to contact-based fingerprint images," 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 2017, pp. 277-285, https://doi.org/10.1109/BTAS.2017.8272708.
9. K. Sriskandaraja, V. Sethu, and E. Ambikairajah, "Deep Siamese Architecture Based Replay Detection for Secure Voice Biometric, Interspeech, pp. 671-675, 2018, https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/1819.pdf.
10. M. Fallahi, T. Strufe and P. Arias-Cabarcos, "BrainNet: Improving Brainwave-based Biometric Recognition with Siamese Networks," 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), Atlanta, GA, USA, 2023, pp. 53-60, https://doi.org/10.1109/PERCOM56429.2023.10099367.
11. X. Liu, Y. Zhou, J. Zhao, R. Yao, B. Liu and Y. Zheng, "Siamese Convolutional Neural Networks for Remote Sensing Scene Classification," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 8, pp. 1200-1204, Aug. 2019, https://doi.org/10.1109/LGRS.2019.2894399.
12. Y. Ma, Y. Liu, Q. Xie et al., “CNN-feature based automatic image annotation method”. Multimed Tools Appl 78, 3767–3780 (2019). https://doi.org/10.1007/s11042-018-6038-x
13. G. Griffin, A. Holub, P. Perona, Caltech-256 Object Category Dataset. California Institute of Technology, (2007). https://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001
14. L. Fei-Fei, R. Fergus, P. Perona, “One-Shot Learning of Object Categories”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 4, APRIL 2006, http://vision.stanford.edu/documents/Fei-FeiFergusPerona2006.pdf

Source Code available at: https://github.com/KaminariManu00/Automatic-Tagging-Suggestion-for-Database-Enrichment