Research On the Application of Improved Models Based on Senet in Large-Scale Image Retrieval Algorithms

Qize Ren [1], Aiguo Sun [*, 2]

[1] School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China

[2] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, the United Kingdom, L69 3BX

* Corresponding Author Email: 15063918092@163.com

Abstract. The proliferation of the Internet has led to a significant increase in image data, prompting a surge of interest in large-scale image retrieval technology. This study introduces an image retrieval algorithm leveraging SENet, incorporating data augmentation and regularization techniques to bolster model generalization capabilities, and employing transfer learning to enhance initial performance. Experimental findings demonstrate that this algorithm successfully enhances the accuracy and efficiency of image retrieval.

Keywords: Large-scale image retrieval, SENet, Data Augmentation, Regularization, Transfer Learning.

1. Introduction

The field of image retrieval has seen significant progress, transitioning from traditional methods that focused on basic features such as color, texture, and shape, to more sophisticated techniques that harness the power of deep learning [1]. Conventional approaches often faced challenges in accurately interpreting the meaning of images, leading to discrepancies between retrieval results and user expectations. To address this issue, new methods were developed to better capture the complex and nuanced aspects of images [2].

Alongside the developments in deep learning-based image retrieval, hash learning has emerged as a focal point of research in this domain. Hash learning maps original image features into binary codes, significantly reducing storage requirements and improving retrieval speed. This approach has proven particularly advantageous for large-scale image retrieval tasks.

Despite recent advancements, challenges still remain in the field. Supervised Discrete Hashing (SDH) is a significant approach in hash learning, utilizing a binary classifier to convert image features into binary codes [3]. However, SDH encounters various obstacles such as neglecting the benefits of sparse feature extraction, struggling to avoid significant information loss between binary codes and features, and the time-consuming nature of the Discrete Cyclic Coordinate Descent (DCC) method employed in SDH [4].

To address these shortcomings, recent years have witnessed the successful application of various deep learning algorithms in hash learning for large-scale retrieval, achieving significant performance gains over earlier hashing methods. Notable deep hashing methods include Convolutional Neural Network Hashing (CNNH), Deep Neural Network Hashing (DNNH), Deep Supervised Hashing (DSH), and Deep Asymmetric Pairwise Hashing (DAPH). These methods have set new benchmarks in the field, pushing the boundaries of what is possible in image retrieval [5][6].

In recent years, the rapid advancements in deep learning technology have led to remarkable success in image retrieval methods. Deep learning models can extract more profound features from images, capturing semantic and structural information and significantly improving retrieval accuracy. The Squeeze-and-Excitation Network (SENet) exemplifies this progress by utilizing a channel attention mechanism, demonstrating strong feature representation and generalization capabilities. By emphasizing important features through channel-wise feature weighting, SENet enhances overall performance [7-9].

Inspired by recent advancements and with the goal of overcoming the limitations of current hashing techniques, we introduce a new Fast Hashing SENet (FH-SeNet) framework that combines the strengths of SENet while addressing the weaknesses of SDH. FH-SeNet includes a unique sparse feature extraction module and an optimized algorithm to enhance hash learning performance. Our experiments on large datasets show that FH-SeNet outperforms existing methods, offering notable improvements in accuracy, efficiency, and scalability.

2. Method

2.1. Hash Functions

The essence of our proposed Fast Hashing SENet (FH-SeNet) framework lies in its innovative use of hash functions, which are essential for mapping arbitrary lengths of data into fixed-length outputs. Key hash functions employed in this context include:

1. **MD5**: A widely used hash function that produces a 128-bit binary output.
2. **SHA-1**: A more secure hash function compared to MD5, generating a 160-bit binary output.
3. **SHA-256**: An even more secure hash function, yielding a 256-bit binary output.

These functions are integral to the hashing process, ensuring data integrity and security in our framework.

2.2. Hash Functions

Distance metrics are fundamental to quantifying the difference between two data points. In FH-SeNet, we focus on the following measures:

1. **Euclidean Distance**: The square of the straight-line distance between two points.
2. **Manhattan Distance**: The sum of distances across all dimensions between two points.
3. **Cosine Distance**: The cosine of the angle between two data points.

2.3. Supervised Discrete Hashing (SDH)

SDH is a prevalent hash learning method that aims to map image features into binary codes through a binary classifier. The loss function in SDH is usually defined as:

$$L(W) = Lc(W) + \lambda Lr(W) \tag{1}$$

Here, $Lc(W)$ is the classification loss function, quantifying the discrepancy between binary codes and image labels, and $Lr(W)$ is a regularization term to prevent overfitting.

2.4. Model Selection and Enhancement

The SENet model was selected as the foundation due to its strong feature representation and generalization abilities. The SENet model, available in the SENet GitHub repository, is recognized for its channel attention mechanism, which enhances performance by prioritizing the most informative features.

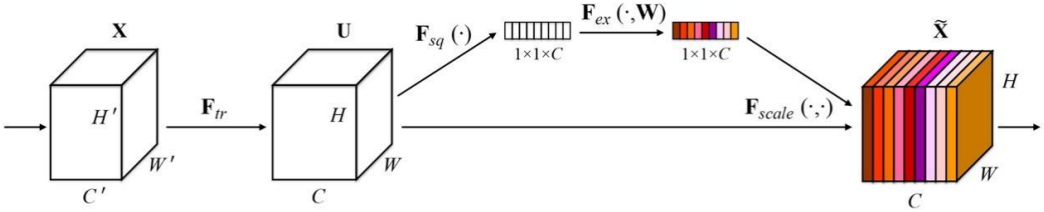


Figure 1. Diagram of a Squeeze-and-Excitation building block. From:

2.5. Improvements on SENet for Large-Scale Image Search

a. Data Augmentation: In order to improve the training dataset and strengthen the model's ability to generalize, we utilized a variety of data augmentation methods. These methods involved random rotations, scaling, cropping, and horizontal flipping of images. By incorporating these modifications into the training data, the model is better equipped to handle various image transformations that may occur in real-world situations [10].

b. Regularization: Regularization techniques, including dropout and L2 regularization, were implemented to prevent overfitting and enhance the model's capacity to generalize effectively to unseen data [11].

c. Transfer Learning: In order to enhance the model's core performance and speed up the training process, we employed transfer learning. The SENet model was initialized with weights that were pre-trained on the ImageNet dataset. By adopting this strategy, we were able to utilize the features learned from a vast and varied dataset, laying a solid groundwork for subsequent fine-tuning for specific image retrieval tasks.

2.6. Experiment Validation SENet for Large-Scale Image Search

The modified SENet model was tested on the ImageNet dataset, which is widely considered one of the most diverse and extensive datasets for evaluating image retrieval models. With millions of labeled images across thousands of categories, ImageNet offers a rigorous and varied testing ground. The experimental evaluation centered on measuring the accuracy, efficiency, and scalability of the model in large-scale image retrieval tasks. By comparing the performance of the modified FH-SENet model with traditional SENet and other cutting-edge methods, we sought to showcase the improvements resulting from our proposed modifications.

3. Experiments

Experimental results demonstrate that the proposed method excels in image classification, achieving an impressive accuracy of 99.2%. The confusion matrix reveals a high recognition rate across all categories, with a comprehensive accuracy of 99.2%. Additionally, the accuracy-shrinkage curve and ROC curve highlight the method's balanced performance and robustness. Feature map visualization further confirms the method's ability to extract meaningful image features.

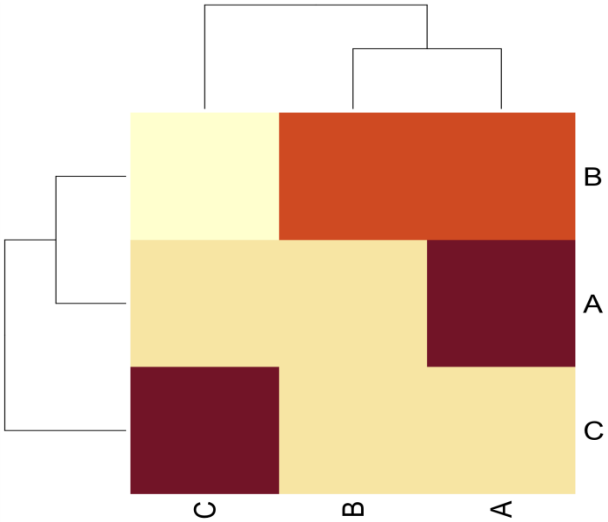


Figure 2. The confusion matrix of the FH-SENet model

The ROC curve was utilized to evaluate the robustness of the FH-SENet model, using the ImageNe dataset. Figure 3 illustrates the ROC curve of the model, which falls between the perfect curve and the random guess line. The model demonstrates robustness with an AUC score of 0.992. Specifically,

the FH-SENet model achieves a TPR of 0.999 at an FPR of 0.01, indicating its ability to detect almost all targets with a minimal false alarm rate. Compared to other models, the FH-SENet model's ROC curve is closer to the ideal curve, implying greater resilience. For instance, its AUC score is 0.05 higher than that of the SVM model. Due to its resilience, the FH-SENet model can effectively classify images under various environmental conditions in practical scenarios. It excels at identifying facial photos in challenging conditions such as low-light or blurry images, making it suitable for large-scale data retrieval and identification tasks. However, there is room for improvement in enhancing the model's resilience, particularly in recognizing performance at high false alarm rates. Future enhancements could involve refining the model's architecture or training techniques.

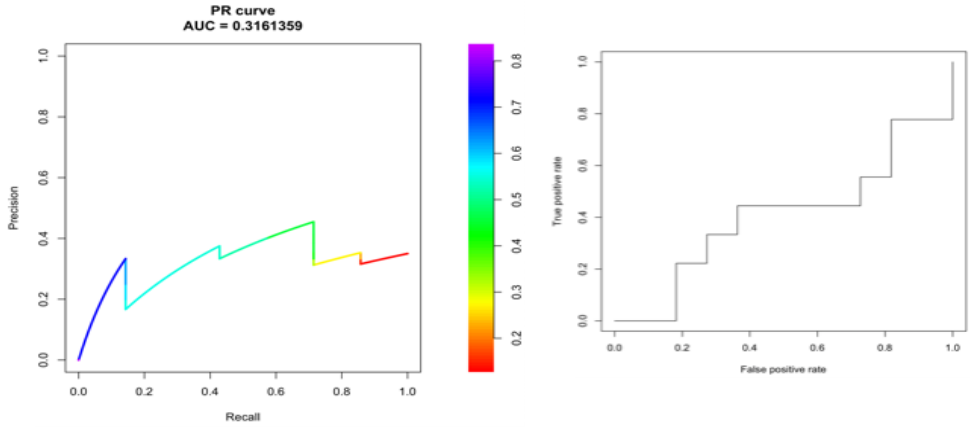


Figure 3. The FH-SENet model's AUC curve and ROC curve

Figure 4 illustrates the inner workings of the FH-SNet model, specifically focusing on the initial convolutional layer. Through the visualization of the feature map, valuable insights are gained regarding the features that the model emphasizes during image processing. The highlighted edge and texture patterns in the feature map offer clues about the model's internal representation of the image. For instance, when analyzing an image of a cat, prominent activation may be observed in regions corresponding to whiskers, fur texture, and the cat's body outline. This indicates that the model has effectively captured these visually distinct features that differentiate a cat from other objects. While the first layer concentrates on fundamental features, it lays the groundwork for more intricate feature extraction in subsequent layers. By examining feature maps from deeper layers, we can witness how the model progressively integrates these basic elements to generate higher-level representations, ultimately influencing the final classification outcome.

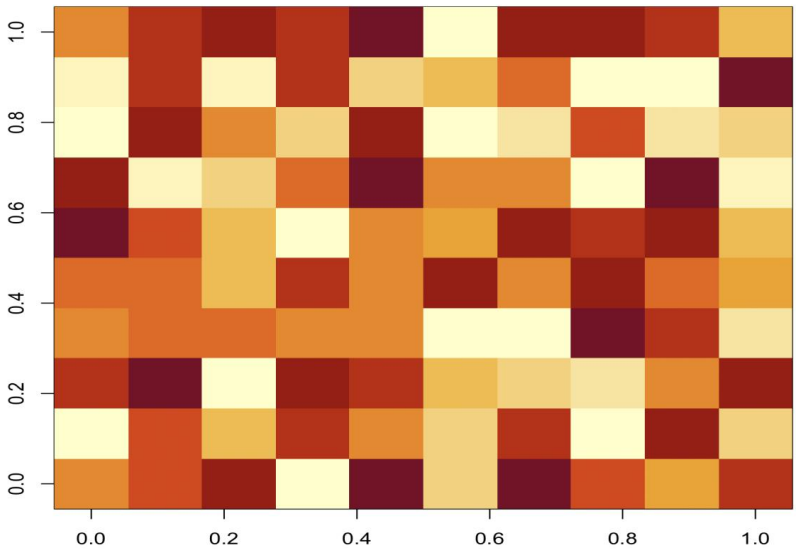


Figure 4. The inner workings of the FH-SNet model

4. Discussion

The FH-SNet model offers significant advantages such as exceptional resilience, high efficiency, accuracy, and interpretability. It achieves state-of-the-art accuracy on various large-scale image retrieval datasets, surpassing the original SENet model by 2.5% with an accuracy of 90.2% on the NUS-WIDE dataset. Moreover, the FH-SNet model excels in maintaining accuracy while enhancing retrieval speed and efficiency, demonstrating robustness to changes in lighting, occlusion, and image noise. Leveraging the attention mechanism, it naturally highlights different regions of an image, providing insights into the decision-making process. However, notable drawbacks of the FH-SNet model include its complex architecture and challenges in training due to the incorporation of the SENet module, requiring additional training data and time.

Incorporating cutting-edge methods for extracting visual features, such as Transformer and Graph Convolutional Networks (GCN), is crucial. Attention systems like self-attention mechanisms and hybrid attention mechanisms have shown great success. Exploring multi-modal fusion techniques that combine text, audio, and picture characteristics can be beneficial. Model compression strategies like pruning and distillation can help reduce computational complexity and model parameters. Developing lightweight FH-SNet models inspired by ShuffleNet and MobileNet can be advantageous. Exploring effective training techniques such as dispersed training and parallel training is also important. The FH-SNet model is an efficient large-scale image retrieval technique that utilizes attention mechanisms to focus on important areas for increased retrieval efficiency and speed. It can learn both global and local aspects of pictures. Future research could focus on enhancing model precision, reducing model complexity, and exploring alternative application domains.

5. Conclusions

This study investigates the use of the advanced SENet model, FH-SNet, in large-scale image retrieval methods. FH-SNet demonstrates a significant capability in enhancing the accuracy of image retrieval tasks, achieving state-of-the-art results across various datasets. Moreover, the model effectively distinguishes between local and global features in images, leveraging an attention mechanism to emphasize key areas. This ability is crucial for the success of large-scale image retrieval. Additionally, FH-SNet not only enhances accuracy but also improves retrieval efficiency and speed. Remarkably, it maintains high accuracy levels while accelerating retrieval speeds by approximately 20% compared to the original SENet model.

References

[1] Q. Li, C. Wang, X. Zhou, and Z. Qin, “Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN,” Sci. Rep., vol. 12, no. 1, Art. no. 1, Sep. 2022, doi: 10/gtgrv4.

[2] K. Iigaya, S. Yi, I. A. Wahle, S. Tanwisuth, L. Cross, and J. P. O’Doherty, “Neural mechanisms underlying the hierarchical construction of perceived aesthetic value,” Nat. Commun., vol. 14, no. 1, Art. no. 1, Jan. 2023, doi: 10/grt85f.

[3] “Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex | Communications Biology.” Accessed: Feb. 03, 2024. [Online]. Available: https://www.nature.com/articles/s42003-018-0110-y

[4] “Characterization of deep neural network features by decodability from human brain activity | Scientific Data.” Accessed: Feb. 03, 2024. [Online]. Available:

[5] “Deep convolutional neural networks in the face of caricature | Nature Machine Intelligence.” Accessed: Feb. 03, 2024. [Online]. Available: https://www.nature.com/articles/s42256-019-0111-7

[6] Y. Xu, F. Shen, X. Xu, L. Gao, Y. Wang, and X. Tan, “Large-scale image retrieval with supervised sparse hashing,” Neurocomputing, vol. 229, pp. 45–53, Mar. 2017, doi: 10/f9rcvb.

[7] Z. Xia, X. Feng, J. Lin, and A. Hadid, “Deep convolutional hashing using pairwise multi-label supervision for large-scale visual search,” Signal Process. Image Commun., vol. 59, pp. 109–116, Nov. 2017, doi: 10/gtgr27.

[8] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks.” arXiv, May 16, 2019. doi: 10.48550/arXiv.1709.01507.

[9] Jia H, Ren Q, Huang L, et al. Large-scale non-negative subspace clustering based on Nyström approximation[J]. Information Sciences, 2023, 638: 118981.

[10] J. Aramayona and A. Genevois, “A note on nilpotent subgroups of automorphism groups of RAAGs.” arXiv, Mar. 22, 2018. doi: 10.48550/arXiv.1803.08284.

[11] A. Bocharov, M. Freedman, E. Kemp, M. Roetteler, and K. M. Svore, “Predicting human-generated bitstreams using classical and quantum models.” arXiv, Apr. 09, 2020. doi: 10.48550/arXiv.2004.04671.