이미지 검색 애플리케이션을 위한 차원 축소 기능을 갖춘 EfficientNet 모델

가잘 사들루니아 [1], 아흐마드 이자즈 [2], 신석주 [3]
[1,3] 조선대학교, [2] 고려대학교

ghazal.saadloonia@chosun.ac.kr, ijaz@korea.ac.kr, sjshin@chosun.ac.kr(교신저자)

EfficientNet Models with Dimensionality Reduction for Image Retrieval Applications

Ghazal Saadloonia[1], Ijaz Ahmad[2], Seokjoo Shin[3]
[1,3]Chosun University, [2]Korea University

Abstract

Image retrieval is a fundamental task in computer vision that involves searching and retrieving relevant images to a query from a large collection of images. It has applications in different domains, including image search engines, and product recommendation systems. In such systems, deep learning discriminative capabilities can be leveraged; however, the main challenge is their associated computational cost. Therefore, in this work we propose a content-based image retrieval (CBIR) scheme that utilizes pretrained lightweight EfficientNet models to represent images in a high dimensional feature space for retrieval applications. Furthermore, to accelerate the similarity search, we reduce the feature vector size by implementing Principal Component Analysis (PCA) algorithm. The simulation analysis on Corel-1K data shows that our CBIR scheme maintains strong retrieval performance while reducing computational overhead.

Ⅰ. Introduction

Image retrieval is one of the fundamental tasks in computer vision that involves searching and retrieving relevant images to a query from a large collection of images. Over the years, content-based image retrieval (CBIR) algorithms are proposed to address the limitations of traditional text-based indexing methods. In CBIR, each image that is stored in the database has its visual features extracted (such as colors, texture, and shape), followed by feature matching with a query image to retrieve visually similar results. One of the main challenges in CBIR is to extract representative features from the images. Given the recent popularity of artificial intelligence mainly owing it to the success of deep learning (DL), several DL-based CBIR schemes have been proposed [1]. However, DL algorithms are generally characterized as compute intensive tasks; therefore, limiting their applications in CBIR systems.

To address the high complexity associated with DL-based image retrieval schemes, this paper proposes a CBIR scheme that utilizes pretrained EfficientNet models [2] for feature extraction. Also, it integrates principal component analysis (PCA) as a dimensionality reduction method to accelerate the similarity search.

Ⅱ. Proposed Method

Fig. 1. shows our proposed CBIR scheme that mainly consists of two modules: feature extraction and feature matching. For feature extraction, we leverage the powerful discriminative capabilities of DL. A DL model is a parameterized function f_θ that learns to update its parameters θ for mapping an input data $\mathcal{X}$ to an output $\mathcal{Y}$ as closely as possible, i.e., $\{f_\theta: \mathcal{X} \longrightarrow \mathcal{Y} | f_\theta(x_i) \approx y_i\}$, by minimizing a loss function L. The model f usually has a feature extractor module consisting of multiple convolution layers, followed by a classifier module, comprised of fully connected layers. The classifier

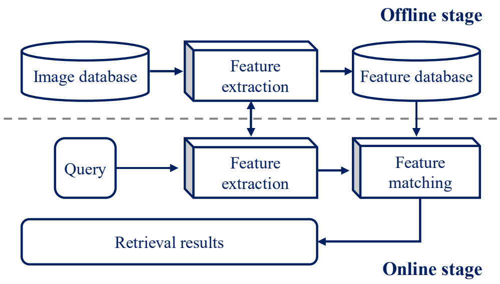

Fig. 1. A high-level illustration of proposed CBIR scheme.

Table. 1. Comparison of the Corel-1K analysis using mAP scores for pre-trained EfficientNet models, including the impact of dimensionality reduction.

Metrics		EfficientNetV1												EfficientNetV2							
		B0		B1		B2		B3		B4		B5		B0		B1		B2		B3	
Param.		4.0		6.6		7.8		10.8		17.7		28.5		5.9		6.9		8.8		13.0	
d		ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d
		1280	397	1280	395	1408	398	1536	406	1792	358	2048	371	1280	367	1280	344	1408	349	1536	360
mAP (%)	L1	0.75	0.74	0.75	0.75	0.77	0.76	0.72	0.72	0.76	0.76	0.78	0.78	0.79	0.79	0.79	0.79	0.76	0.77	0.76	0.75
	L2	0.82	0.82	0.83	0.83	0.81	0.82	0.86	0.86	0.89	0.9	0.87	0.87	0.85	0.86	0.85	0.85	0.83	0.83	0.86	0.86
	L3	0.7	0.7	0.72	0.72	0.63	0.64	0.68	0.68	0.76	0.76	0.76	0.77	0.64	0.64	0.69	0.69	0.69	0.7	0.76	0.76
	L4	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	L5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	L6	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.97	0.96	0.96
	L7	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.96	0.96	0.97	0.97	0.95	0.95	0.96	0.96	0.97	0.97
	L8	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00
	L9	0.88	0.88	0.91	0.91	0.95	0.94	0.95	0.95	0.93	0.93	0.94	0.93	0.92	0.91	0.9	0.9	0.93	0.93	0.94	0.94
	L10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	μ	0.908	0.907	0.912	0.912	0.907	0.908	0.914	0.915	0.927	0.928	0.929	0.929	0.914	0.914	0.915	0.916	0.915	0.916	0.923	0.924

(Param.: number of parameters in million, L1 to L10 are labels in the Corel-1K dataset, μ: mean mAP score of the dataset, d: a feature vector dimensions, ϕ: Original feature vector, φ: reduced feature vector.)

module is task-specific while the feature extractor module can be shared across different tasks. Therefore, we use the classifier of pretrained EfficientNet models proposed in [2], and [3] to represent images in a high-dimensional feature space for retrieval applications e.g., $\phi \longleftarrow f_\theta(\mathcal{X})$. To reduce the dimensionality of the feature space, we applied PCA on constructed feature space i.e., $\varphi \longleftarrow g(\phi)$. In our simulations, we empirically choose 95% of component variance to achieve a balance tradeoff between dimensionality reduction and information loss. Finally, we implement $L1$ distance metric for feature matching as it is computationally simple, yet an efficient metric for the retrieval task.

Ⅲ. Simulation Results

Dataset. To evaluate the performance of proposed CBIR system, we use Corel-1K dataset. The dataset consists of 1,000 images uniformly divided into 10 distinct categories. The dataset contains a wide range of content, including natural scenes, animals and human-made objects. The images are of true color with a consistent size of either 256×384 or 384×256 pixels. The small size of the dataset makes it suitable for demonstrating a more realistic and effective simulation of image retrieval with pretrained models.

Evaluation metric. To quantitatively evaluate our scheme, we use mean average precision (mAP) score. It measures the quality of retrieval by averaging precision scores across multiple queries, with a higher score indicating a better performance. For Q query images, the mAP score is defined as

$$mAP = \frac{1}{Q}\sum_{q=1}^{Q} AP_q\,, \qquad (1)$$

where AP_q is the average precision score that calculates how well relevant images are ranked within the list of retrieved images for a single query.

Performance analysis. Table 1 summarizes the Corel-1K simulation analysis using mAP scores for EfficientNetV1 [2] and EfficientNetV2 [3] models, including the impact of dimensionality reduction. It can be observed that when comparing the same model of V1 and V2 (e.g., V1-B0 versus V2-B0), EfficientNetV2 consistently outperformed EfficientNetV1, both with and without dimensionality reduction. Thus, supporting the general consensus that V2 models are more efficient and accurate. Furthermore, PCA maintained the retrieval accuracy while potentially reducing computational costs. The slight decrease observed in mAP when applying dimensionality reduction indicated a trade-off; however, this difference is insignificant across all evaluated models. It is noteworthy, that the model performance significantly degraded for the labels of Corel-1K dataset that do not correspond to any class of the ImageNet dataset due to the lack of the learned features in the pretrained models.

IV. Conclusion

This study leveraged dimensionality reduction algorithm with pretrained lightweight deep learning models (EfficientNets) for image retrieval application. The simulation analysis showed that EfficientNetV2 is the superior model for image retrieval on the Corel-1K dataset as it offered higher mean average precision. Furthermore, when PCA technique was incorporated for dimensionality reduction, the models generally maintained strong retrieval performance while reducing computational overhead.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

References

[1]. S. R. Dubey, "A Decade Survey of Content Based Image Retrieval Using Deep Learning," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 5, pp. 2687-2704, May 2022.

[2]. M. Tan, and L. Quoc, "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[3]. M. Tan, and L. Quoc, "Efficientnetv2: Smaller models and faster training." International conference on machine learning. PMLR, 2021.