DNA-based Similar Image Retrieval via Triplet Network-driven Encoder

Takefumi Koike
Graduate School of Informatics
Kyoto University Kyoto, Japan
tkoike@easter.kuee.kyoto-u.ac.jp

Hiromitsu Awano
Graduate School of Informatics
Kyoto University Kyoto, Japan
awano@i.kyoto-u.ac.jp

Takashi Sato
Graduate School of Informatics
Kyoto University Kyoto, Japan
takashi@i.kyoto-u.ac.jp

***Abstract*—With the exponential growth of digital data, DNA has emerged as an attractive medium for storage and computing. Design methods for encoding, storing, and searching digital data within DNA storage are thus of utmost importance. This paper introduces image classification as a measurable task for evaluating the performance of DNA encoders in similar image searches. In addition, we propose a triplet network-based DNA encoder to improve accuracy and efficiency. The evaluation using the CIFAR-100 dataset demonstrates that the proposed encoder outperforms existing encoders in retrieving similar images, with an accuracy of 0.77, which is equivalent to 94% of the practical upper limit, and achieves 16 times faster training time.**

I. Introduction

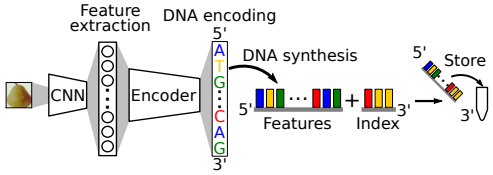
Existing storage methods face limitations regarding lifespan and power consumption against vast amounts of digital data generated daily. Deoxyribonucleic acid (DNA) storage is a promising solution to these challenges. However, accessing this data using computers takes an impractically long time owing to the immense amount of data involved. Therefore, in DNA storage, a method has been proposed for accessing data directly using the structure of DNA without relying on CPUs or GPUs [1].

A DNA-based similar image retrieval is proposed in [2]. This method transforms images into unique single-stranded DNA and evaluates image similarities through the degree of double-stranded DNA hybridization. An encoder facilitates this process by leveraging a hybridization simulator to optimize DNA design. The retrieval is efficient because of the highly parallel nature of hybridization, executing image retrieval scalable.

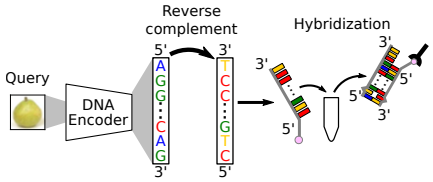
In this paper, we propose a quantitative metric to evaluate the performance of DNA encoding methods in retrieving similar images based on visual similarity. Then, we propose a novel encoder network using deep metric learning [3] for accurate and efficient similar image retrieval.

II. Background and Motivation

Obtaining images based on local features is called the content-based image retrieval (CBIR) task [4]. Literature [2] introduced a method for DNA-based CBIR. Fig. 1(a) illustrates the conversion of feature representations from input images derived from a trained CNN model into DNA sequences. Fig. 1(b) illustrates the DNA-DNA Hybridization (DDH) process, where DNA-converted data hybridizes with a reverse



(a) Feature extraction and encoding



(b) Image retrieval using DDH

Fig. 1: CBIR using DNA. (a) Extract features from images and encode them into DNA using a DNA encoder. (b) Encode the query image and extract similar DNA sequences by hybridization.

complemented query image to form double-stranded DNA, which is retrieved as an image similar to the query.

The existing methods [2] demonstrated similar image retrieval using DNA structure. Here, the retrieval performance evaluation was based on comparing feature vector distances and hybridization yields without considering visual similarity. Therefore, we introduce a comparison process through image classification, focusing on visual similarity to assess the performance of DNA encoders. In addition, we propose a DNA encoder based on a triplet network [3] to improve encoding accuracy. The proposed method offers fast learning based on the Hamming distance of DNA. This is in contrast to the existing method, which uses a time-consuming hybridization simulator during the training of the network.

III. Classification via Triplet Network

A. Image classification using DDH

A representative class-query must be defined for each class to perform the classification task on the DNA sequences in the encoded dataset. These queries should be *mutually exclusive* and *collectively exhaustive* to ensure effective hybridization. We selected the representative query by identifying the most

979-8-3503-4859-0/DATE24/© 2024 EDAA

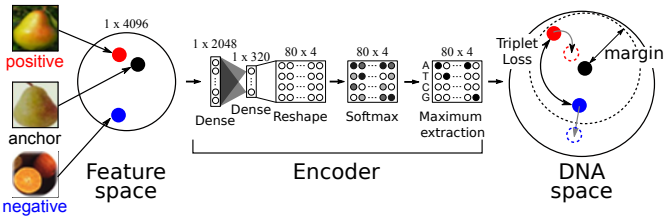


Fig. 2: The proposed network structure. Features extracted by the CNN model are fed into the DNA encoder. Triplets are selected from the encoder output to utilize triplet loss in the training.

frequent base at each position based on the training data since DNA sequences belonging to the same class share common base sequences and are close in terms of Hamming distance. The class-queries are added to a pool of coded images to promote hybridization between the query sequences and the targets. The query with the highest hybridization yield is predicted as the class to which the data belongs.

B. Triplet network-based DNA encoding

The outline of the proposed network is illustrated in Fig. 2. It accepts feature vectors as input and adjusts the distance of the data points in the output embedding space according to the class membership [3]. As its name suggests, the triplet network operates based on the distances among three data points called triplets. The triplets consist of a set of data from the same class ("positive" or $\boldsymbol{x}_+$), a set of data from a different class ("negative" or $\boldsymbol{x}_-$), and a reference data point ("anchor" or $\boldsymbol{x_0}$). The triplet network trains the machine learning model to minimize the distance between the anchor and positive data points while ensuring that the distance between the anchor and negative data points is greater than a predefined minimum distance called the margin. This adjustment enhances the separation of clusters in the embedding space.

In the embedding space, when the data points $\boldsymbol{x_0}, \boldsymbol{x}_+, \boldsymbol{x}_-$ are transformed to $\boldsymbol{x'_0}, \boldsymbol{x'}_+, \boldsymbol{x'}_-$, respectively, the Triplet loss function is expressed using a distance metric, as follows:

$$\mathcal{L}(\boldsymbol{x'_0}, \boldsymbol{x'}_+, \boldsymbol{x'}_-) = \max(0, d(\boldsymbol{x'_0}, \boldsymbol{x'}_+) - d(\boldsymbol{x'_0}, \boldsymbol{x'}_-) + m) \quad (1)$$

where m is the margin, and $d(\cdot)$ is an arbitrary distance function. In this work, the Hamming distance is used. As shown in Fig. 2, during the training, the encoder's outputs are extracted at their maximum values for each column to determine the bases of each column. The Hamming distance, normalized to [0, 1], between the anchor matrix $\boldsymbol{X'_0} = [\boldsymbol{x'_{00}}\, \boldsymbol{x'_{01}} \ldots \boldsymbol{x'_{0L}}]$ and any data matrix $\boldsymbol{X'} = [\boldsymbol{x'_0}\, \boldsymbol{x'_1} \ldots \boldsymbol{x'_L}]$ is expressed by:

$$d(\boldsymbol{X'_0}, \boldsymbol{X'}) = \frac{1}{2L} \sum_{i=0}^{L} ||\boldsymbol{x'_i} - \boldsymbol{x'_{0i}}||_1 \quad (2)$$

where L is the length of a DNA and $||\cdot||_1$ is the L1-norm.

IV. Numerical Evaluation

We compared the proposed method with the state-of-the-art method, primo [2], in terms of classification accuracy and training time. We used the CIFAR-100 dataset for classification tasks. Several CNN models were employed to investigate the impact of different feature spaces on the accuracy of the CBIR.

TABLE I: Classification accuracy (CIFAR-100)

CNN models	Methods	DDH	Software	Ratio
VGG-16	primo [2]	0.437	0.631	0.693
	proposed	**0.517**		**0.819**
EfficientNetV2-S	primo [2]	0.705	0.814	0.866
	proposed	**0.766**		**0.941**

TABLE II: Training time comparison (CIFAR-100)

methods	Training time	
	Average (s/epoch)	Total (min.)
primo [2]	67.8	169
proposed	4.07	10.2

Table I summarizes the evaluation results. The columns labeled DDH and Software represent the accuracy obtained by the proposed method and software CNN models, respectively. The accuracy of the software CNN model serves as the respective upper bound for DDH, and the accuracy difference between DDH and the software explains the performance of the DNA encoders. The proposed method consistently demonstrated better classification accuracy than the existing method, irrespective of the CNN models used. The proposed method is highly effective in capturing visual similarities. The Ratio column presents the normalized accuracy to the software CNN, indicating how close each learning process approached its potential limits. The proposed method achieved a performance higher than 94% while primo with EfficientNet achieved approximately 87%.

In addition, the proposed method requires significantly shorter training time than primo by more than 16 times, as shown in Table II.

V. Conclusion

We proposed a quantitative evaluation method and a novel encoder network for DNA-based similar image retrieval. The proposed method demonstrated a 16x faster training time than primo and achieved a 94% accuracy of the upper limit.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Nos. 23K18462 and 23H03362.

References

[1] L. C. Meiser, B. H. Nguyen, Y.-J. Chen, J. Nivala, K. Strauss, L. Ceze, and R. N. Grass, "Synthetic DNA applications in information technology," *Nature communications*, vol. 13, no. 1, p. 352, 2022.

[2] C. Bee, Y.-J. Chen, M. Queen, D. Ward, X. Liu, L. Organick, G. Seelig, K. Strauss, and L. Ceze, "Molecular-level similarity search brings computing to DNA data storage," *Nature Communications*, vol. 12, p. 4764, 2021.

[3] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Similarity-Based Pattern Recognition: Third International Workshop*, 2015, pp. 84–92.

[4] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. ACM Int. Conf. Multimedia*, 2014, p. 157.