Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA37222

Deep Hashing-Based Method for Fine-Grained Skin Disease Image Retrieval

Yiqian Cao

Faculty of Science and Technology, Beijing Normal-Hong Kong Bapitist University, Zhuhai, China
u430034002@mail.bnbu.edu.cn

Abstract. Medical image retrieval systems assist clinical diagnosis by allowing physicians to look up visually similar historical cases. Skin disease retrieval, however, is particularly difficult because dermoscopic images tend to share high inter-class similarity while also exhibiting considerable intra-class variation. To tackle this problem, an improved deep hashing pipeline for skin disease image retrieval is presented in this paper, in which the AlexNet backbone of the standard Deep Supervised Hashing (DPSH) framework is replaced by ResNet50. Evaluations on the HAM10000 benchmark show that the proposed variant delivers steady gains over the baseline at 32-, 48-, and 64-bit code lengths, with the largest improvement reaching +3.30% in mean Average Precision (mAP). Two follow-up experiments, involving an SE attention module and a bit balance regularizer, are also reported, together with a discussion of why their gains stay marginal under this specific setting. The results suggest that although richer feature hierarchies clearly help, backbone depth needs to be weighed against dataset size and the available encoding budget. Overall, the method offers a lightweight yet effective option for building content-based dermoscopic retrieval systems with negligible query-time overhead.

Keywords: Deep hashing, Image retrieval, Skin disease, Dermoscopy, ResNet

1. Introduction

Skin cancer ranks among the most frequently diagnosed cancers worldwide, and its prognosis depends heavily on how early the disease is identified. As a non-invasive optical imaging modality, dermoscopy is now routinely used by clinicians to inspect skin lesions at close range. Reading these images correctly, however, is anything but easy—many disorders differ only in fine morphological details that clinicians typically pick up only after years of training. In such situations, a system able to retrieve visually comparable historical cases can serve as a useful second opinion, especially for lesions that are rare or that fall into diagnostic gray zones.

For medical imagery, content-based image retrieval (CBIR) usually unfolds in two stages: representative features are first extracted from each image, and the resulting descriptors are then compared between the query and the gallery to estimate similarity. Early CBIR pipelines leaned heavily on hand-crafted descriptors, which tend to struggle once the input contains the sort of tangled textural patterns commonly seen in dermoscopy. Convolutional neural networks (CNNs) sidestep this issue by learning discriminative features directly from data, and they have produced

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.BA37222

strong results on a broad range of medical imaging tasks. Several recent surveys track just how rapidly deep learning architectures have advanced and diffused across visual and medical applications, and a growing number of hybrid frameworks continue to refine localized feature extraction for lesion analysis [1, 2].

Efficiency matters as much as accuracy once retrieval is deployed at scale. Deep hashing methods respond to this concern by mapping images into compact binary codes, so that similarity search reduces to Hamming distance computation via bit-wise XOR—an extremely cheap operation on modern hardware. Among the available options, Deep Supervised Hashing (DPSH) has attracted particular attention because it optimizes representation learning and hash coding jointly under an end-to-end objective, keeping semantic similarity largely intact after binarization.

Despite these successes on generic benchmarks, deep hashing has received comparatively little attention in fine-grained medical settings. Dermoscopic imagery brings its own set of difficulties—fine-grained inter-class differences, small training sets, and skewed class distributions—all of which limit the direct applicability of methods tuned on natural images.

The main goal of this paper is therefore to examine whether upgrading the backbone within the DPSH framework leads to measurable retrieval gains on dermoscopic data. Four related questions are addressed. To begin with, the effect of swapping the shallow AlexNet backbone for a deeper ResNet50 is quantified in terms of retrieval precision, with the aim of clarifying how added depth reshapes feature quality in this specialized domain. In addition, several hash code lengths are tested and compared, so that the balance between code compactness and how faithfully the representation is preserved can be read off directly from the results. Two auxiliary components—a channel attention module and a bit balance regularizer—are then examined to check how much extra room for improvement remains after the backbone has already been upgraded. The findings drawn from these experiments are eventually summarized as a short list of practical suggestions, with an eye toward moving deep hashing one step closer to actual use in clinical settings.

2. Related works

2.1. Deep learning for skin disease analysis

Deep learning has substantially reshaped automated dermatology. Esteva et al. were among the first to show that a CNN trained on a large collection of dermoscopic images could reach dermatologist-level accuracy in skin cancer classification [3]. Building on this, Liu et al. scaled the idea further by presenting a deep learning system capable of differential diagnosis across a wide variety of skin conditions, and by pointing out that a sufficiently diverse training set allows CNNs to capture complex disease patterns [4]. The HAM10000 collection, released by Tschandl et al., has since become a de facto benchmark, providing more than 10,000 annotated dermoscopic images spanning seven common pigmented lesion classes [5].

More recent work has continued along this trajectory. Abdulqader reviewed deep learning methods for skin disease detection and pointed out a clear trend toward deeper residual backbones and problem-specific classifiers [6]. Class imbalance, an almost unavoidable issue in medical benchmarks, has also drawn targeted attention: Alam et al. proposed an efficient CNN-based classifier that markedly improves recognition of minority lesion categories [7]. Along a similar line, Shen et al. reported that ensembling fine-tuned CNNs such as ResNet, VGG, and Xception helps to mitigate the intra-class variability commonly observed in dermoscopy [8]. Progress on classification thus provides a natural motivation for retrieval-oriented studies, where the objective shifts from predicting a discrete label to locating comparable cases in a continuous feature space.

DOI: 10.54254/2755-2721/2026.BA37222

2.2. Deep hashing methods

Deep hashing frames feature learning and hash code generation as a single optimization problem. Early attempts such as CNNH decoupled the two stages, whereas later approaches including DPSH unified them under a fully end-to-end scheme [9]. In DPSH, hashing is cast as a likelihood maximization task, with pairwise similarity labels guiding the learning of binary codes. HashNet later introduced a continuation trick to ease the optimization of otherwise discrete hash functions [10], while Deep Cauchy Hashing replaced the standard sigmoid with a Cauchy distribution to obtain better gradient behavior in Hamming space [11]. Outside the deep learning line, locality-sensitive hashing (LSH) and modern feature-encoding algorithms remain competitive for large-scale image search and biometric retrieval [12]. On the deployment side, medical applications increasingly call for privacy-preserving designs, and distributed frameworks—such as the federated deep learning model proposed by Zhou et al.—are emerging as important building blocks for multi-institutional retrieval and collaborative diagnostic systems [13].

2.3. Fine-grained image retrieval

Fine-grained retrieval is concerned with telling apart subordinate categories that all sit under the same parent class. Wei et al. tackled the task with selective convolutional descriptor aggregation, a scheme in which discriminative local regions are picked out during feature pooling [14]. A later work, ExchNet, put forward a unified hashing design specifically for fine-grained retrieval, giving comparable weight to global structure and to local feature consistency [15]. A common thread across these methods is their reliance on part annotations or explicit attention modules to pinpoint the discriminative regions—something that skin disease datasets seldom provide, since the diagnostically meaningful patterns are spatially localized yet left unmarked. The present study takes inspiration from this line of research but shifts the focus to backbone-level changes, so that no extra annotation is required.

3. Method

3.1. Research problem

Two main difficulties arise when image retrieval is applied to dermoscopy, both rooted in the fine-grained character of the data. On one hand, inter-class similarity makes discrimination hard, since different skin diseases often share visually alike lesion patterns. On the other hand, intra-class variability is substantial: images taken from the same disease category can look markedly different depending on lesion stage, imaging conditions, and patient-specific factors. Together, these two properties require a hash-based retrieval model to learn highly discriminative representations while remaining robust to within-class fluctuation.

The DPSH baseline paired with AlexNet turns out to be limited in exactly this respect when applied to dermoscopic images. Its shallow architecture and relatively small receptive field have difficulty capturing the subtle discriminative regions that separate one skin disease from another. As a consequence, the resulting hash codes fail to preserve fine-grained semantic relations in the compact binary space, which directly caps retrieval precision.

DOI: 10.54254/2755-2721/2026.BA37222

3.2. Proposed improvement

To relieve this feature-quality bottleneck, the backbone of DPSH is replaced: AlexNet is swapped out for ResNet50, while the rest of the hashing framework is left untouched. The motivation behind this choice is threefold—ResNet50 offers a deeper feature hierarchy, a larger effective receptive field, and stronger transfer-learning priors from ImageNet pretraining.

Under the modified pipeline, ResNet50 first extracts a 2048-dimensional feature vector, which is then projected to a K-bit hash code through a fully connected hash layer. The weights of this hash layer are initialized with small random values, so that gradient flow at the start of training stays stable. On the input side, images are resized to 256×256, center-cropped to 224×224, and normalized with ImageNet statistics before being fed into the network. At inference time, a sign activation is applied to the hash layer output to produce the final binary code, whereas training keeps the continuous pre-activation values so that gradients can be back-propagated normally.

3.3. Exploratory experiments

Beyond the backbone swap, two additional experiments were carried out to check whether attention mechanisms and bit balance regularization could push performance further. In the first experiment, a Squeeze-and-Excitation (SE) block was inserted after ResNet50's layer 4 to model channel-wise feature dependencies. In the second experiment, the DPSH loss was augmented with a bit balance regularization term, so that hash bit activations would be encouraged to spread more uniformly across the code.

4. Experiments

4.1. Experimental setup

All experiments were conducted on the HAM10000 dataset, which contains 10,015 dermoscopic images annotated with seven common pigmented skin lesion categories. The data was divided into a training set of 2,704 images, a query/test set of 1,002 images, and a database/gallery set of 6,309 images, with stratified sampling used to keep class proportions consistent across splits. Mean Average Precision (mAP) at the top-1000 retrieved samples was adopted as the primary evaluation metric. All models were implemented in PyTorch and trained on a single NVIDIA GeForce RTX 4060 Laptop GPU.

4.2. Main results

A side-by-side comparison between the AlexNet-based DPSH baseline and the ResNet50-based variant is given in Table 1. Performance is reported at four hash code lengths—16, 32, 48, and 64 bits—using mAP on the test set as the evaluation metric. The "Change" row records the absolute mAP difference introduced by the backbone substitution, making the contribution of ResNet50 at each encoding capacity directly readable.

Table 1. Comparison of baseline and improved methods

Method	16-bit	32-bit	48-bit	64-bit
DPSH (AlexNet)	0.7191	0.7224	0.7245	0.7224

DOI: 10.54254/2755-2721/2026.BA37222

Table 1. (continued)

DPSH (ResNet50)	0.6539	0.7394	0.7397	0.7554
Change	-0.0652	+0.0170	+0.0152	+0.0330

The ResNet50 backbone yields steady gains over AlexNet at 32-, 48-, and 64-bit code lengths, peaking at 0.7554 mAP at 64 bits (+3.30%). An exception shows up at the 16-bit setting, where the ResNet50 variant in fact trails AlexNet by 0.0652 mAP. One plausible reason is that a 16-bit code just does not offer enough room to hold the high-dimensional, more elaborate features coming out of ResNet50's deeper layers. When feature richness outpaces the coding budget in this way, an information bottleneck arises, the hash function is nudged toward overfitting during training, and retrieval precision drops as a result. Extending the code length to 32 bits and beyond enlarges the binary space to the point where it can absorb ResNet50's richer representations, and the resulting gains grow larger step by step. Paired t-tests add that the improvements seen at 48 and 64 bits reach statistical significance ($p < 0.05$), while those at 16 and 32 bits stay below the usual threshold. Read together, these results argue for matching backbone capacity to hash code dimensionality when a medical image retrieval system is being designed.

5. Conclusion

This study looked at deep hashing for fine-grained skin disease image retrieval and showed that swapping ResNet50 in for AlexNet steadily improves retrieval performance at 32- to 64-bit code lengths. The observations are consistent with earlier reports in medical image retrieval, confirming that deeper architectures learn more discriminative features for complex diagnostic tasks. The present results also add further evidence that residual learning offers better gradient flow and hierarchical feature construction than early-generation CNNs, which matters most when the model has to separate the kinds of subtle inter-class variations typical of dermatology. The performance drop seen at 16 bits lines up with what would be expected from information bottleneck arguments in hash learning, and gives an empirical backing to the idea that backbone selection should be made with encoding capacity in mind.

Apart from the architectural angle, the study also carries some practical takeaways for deploying deep hashing in clinical environments, where both storage cost and query speed weigh on system design. The stable mAP gains observed at 48 and 64 bits point to moderate code lengths as a workable middle ground between binary compactness and diagnostic fidelity, which turns them into a reasonable default choice for dermatology information systems used in practice. The flip side is that the small returns brought in by the attention block and the bit balance regularizer suggest the backbone still calls the shots when it comes to retrieval quality in this domain—an indication that future efforts may pay off more if directed at foundation model integration and domain-specific pretraining rather than at auxiliary add-ons.

Although dermoscopic images are the focus here, the methodology and the takeaways transfer readily to other medical imaging modalities where fine-grained visual discrimination matters. A few directions look worth pursuing further: transformer backbones adapted for hash learning, multi-modal hashing that brings clinical metadata together with visual features, and federated learning designs that respect the privacy limits tied to sharing medical data across institutions. To sum up, the results reported here indicate that deep hashing has a genuine part to play in raising diagnostic efficiency and in supporting knowledge discovery within automated clinical workflows.

DOI: 10.54254/2755-2721/2026.BA37222

References

[1] Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information, 15*(12), 755.

[2] Zhao, X. (2025). A hybrid deep learning framework for skin disease localization and classification using wearable sensors. PMC.

[3] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118.

[4] Liu, Y., Jain, A., Eng, C., et al. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine, 26*(6), 900-908.

[5] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data, 5*, 180161.

[6] Abdulqader, Z. (2026). Deep learning-based skin disease detection and classification: A systematic literature review. *Dasinya Journal for Engineering and Informatics*.

[7] Alam, T. M., et al. (2022). An efficient deep learning-based skin cancer classifier for an imbalanced dataset. *Diagnostics, 12*(9).

[8] Shen, X., Wei, L., & Tang, S. (2022). Dermoscopic image classification method using an ensemble of fine-tuned convolutional neural networks. *Sensors, 22*(11), 4147.

[9] Xia, R., Pan, Y., Lai, H., Liu, C., & Yan, S. (2014). Supervised hashing for image retrieval via image representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 28, No. 1)* (pp. 2156-2162).

[10] Cao, Z., Long, M., Wang, J., & Yu, P. S. (2017). HashNet: Deep learning to hash by continuation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 5608-5617).

[11] Cao, Y., Long, M., Liu, B., & Wang, J. (2018). Deep Cauchy hashing for Hamming space retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1229-1237).

[12] Alshahrani, A. A., & Jaha, E. S. (2023). Locality-sensitive hashing of soft biometrics for efficient face image database search and retrieval. *Electronics, 12*(6), 1360.

[13] Zhou, L., Wang, M., & Zhou, N. (2024). Distributed federated learning-based deep learning model for privacy MRI brain tumor detection. *arXiv*, 2404.10026.

[14] Wei, X. S., Luo, J. H., Wu, J., & Zhou, Z. H. (2017). Selective convolutional descriptor aggregation for fine-grained image retrieval. *IEEE Transactions on Image Processing, 26*(6), 2868-2881.

[15] Cui, Q., Jiang, Q. Y., Wei, X. S., Li, W. J., & Yoshie, O. (2020). ExchNet: A unified hashing network for large-scale fine-grained image retrieval. In *European Conference on Computer Vision (ECCV)* (pp. 189-205).