

PENet: A phenotype encoding network for automatic extraction and representation of morphological discriminative features

Zhengyu Zhao^{1,2,3}  | Yuanyuan Lu^{1,2}  | Yijie Tong^{1,2} | Xin Chen⁴ | Ming Bai^{1,2,3,5}

¹Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management (Chinese Academy of Sciences), Institute of Zoology, Chinese Academy of Sciences, Beijing, China

²Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

³University of Chinese Academy of Sciences, Beijing, China

⁴Cangzhou Normal University, Cangzhou, Hebei Province, China

⁵Northeast Asia Biodiversity Research Centre, Northeast Forestry University, Harbin, China

Correspondence

Ming Bai

Email: baim@ioz.ac.cn

Funding information

Northeast Asia Biodiversity Research Center, Grant/Award Number: NABRI202203; China Postdoctoral Science Foundation, Grant/Award Number: 2022M713134; National Key R&D Program of China, Grant/Award Number: 2022YFC2601200; National Natural Science Foundation of China, Grant/Award Number: 32270468 and 32200354; National Science & Technology Fundamental Resources Investigation Program of China, Grant/Award Number: 2019FY100400, 2019FY101800, 2022FY100500 and 2022FY202100

Handling Editor: Daniele Silvestro

Abstract

1. Digitalized natural history collections serve as vital ecological and evolutionary research resources. Specimen retrieval based on morphological features allows for the rapid acquisition of similar specimens from these collections, aiding in maximizing the utilization of their resources and catering to the requirements of related research. However, achieving this objective requires effective feature extraction and representation techniques.
2. We developed a phenotype encoding network (PENet), a deep learning-based model that combines hashing methods to automatically extract and encode discriminative features into hash codes.
3. We evaluated the performance of PENet on six data sets, including a newly constructed beetle data set (6566 images), which covers over 60% of the genera within the six subfamilies of Scarabaeidae. Phenotype encoding network showed high performance in feature extraction and image retrieval, allowing users to input an image of a specimen and efficiently retrieve all specimens with similar morphology. Two visualization methods, t-SNE and Grad-CAM, were used to evaluate the representation ability of the hash codes. Additionally, by using the hash codes generated from PENet, a phenetic distance tree was constructed based on the beetle data set. The result indicated that the hash codes could reveal the phenetic distances and relationships among categories to a certain extent.
4. PENet provides an automatic and efficient method to extract and represent morphological discriminative features. The generated hash code can be used as a low-dimensional carrier of these features, enabling efficient specimen retrieval. Moreover, the distance information carried by these hash codes suggests their potential in systematics, deserving further exploration.

KEY WORDS

deep learning, discriminative features, encoding, feature-based retrieval, hash

Zhengyu Zhao and Yuanyuan Lu are co-first authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
 © 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

1 | INTRODUCTION

The continuous advancement of specimen digitization has significantly facilitated the utilization of natural history collections, providing an invaluable resource for related research (Allan et al., 2019; Beaman & Cellinese, 2012; Hedrick et al., 2020; Nelson & Ellis, 2019; Suarez & Tsutsui, 2004). Once desired specimens are obtained from these collections, they can be analysed by combining relevant metadata (e.g. dates, locations and climate), thereby aiding in the investigation of complex ecological and evolutionary issues (Holmes et al., 2016; Lister, 2011; Schindel & Cook, 2018; Winker, 2004). In certain situations, researchers may seek to rapidly retrieve a series of specimens from natural history collections that exhibit similar morphologies to their study subjects, such as studies related to phenotypic convergence (Lamichhaney et al., 2019) or evolutionary changes (Ozgo & Schilthuizen, 2012). However, existing retrieval methods are inefficient in retrieving large numbers of specimens with similar morphological features (Chen et al., 2005; Hall, 1974; Mallik et al., 2010). This hinders the rapid acquisition of desired materials and the maximization of the vast value held within natural history collections. Morphological feature-based retrieval methods have the potential to overcome the limitations of current approaches (Latif et al., 2019). The effective implementation of this process requires consideration from two aspects: the extraction of discriminative phenotypic features and the comparison and matching of these extracted morphological features (Jan et al., 2020; Meharban & Priya, 2016).

Traditional methods for feature extraction heavily rely on expert knowledge, which can be time-consuming and restrict the broader application of these extracted features (Hawkins, 2000). Additionally, the morphological features of specimens can be highly diverse, presenting a challenge in rapidly identifying similar specimens within large-scale specimen repositories. In recent years, with the progress of deep learning techniques (Jordan & Mitchell, 2015), many effective algorithms for extracting biological features have emerged and found practical applications in natural history collections (Walker et al., 2022). These algorithms have achieved significant successes in tasks such as fast specimen identification (Younis et al., 2018), extraction of label information from collections (Owen et al., 2020) and measurement of functional traits (Weeks et al., 2023), thereby optimizing the utilization of specimens. This indicates that deep learning methods are efficient in extracting morphological features. Moreover, since the features extracted by deep learning methods are digitized, they offer greater scalability, making them suitable for retrieving specimens with similar attributes. However, the feature vectors extracted from digitized collections may have a high number of dimensions, which poses challenges in directly utilizing these features. Comparing and matching high-dimensional features require substantial computational resources and time. Therefore, it is necessary to explore methods for obtaining low-dimensional representations of these features to address these challenges.

In computer science, hashing methods are commonly employed to handle complex, high-dimensional data and vectors by reducing

their dimensionality to hash codes while preserving important information (Chi & Zhu, 2018; Knott, 1975). The hashing methods make processing and analysis more efficient, especially in tasks such as sample retrieval (Tang et al., 2015). The hash code is a sequence of 0/1 digits, for example, '11010011101011', where '1' can be regarded as representing a certain characteristic present in the image and '0' represents its absence. As a result, combining deep learning models as feature extractors with hash codes as feature representations has the potential for faster retrieval of sample images (Luo et al., 2020).

In this study, we propose an end-to-end phenotype encoding network with the backbone of the latest deep learning architecture Swin transformer (Liu et al., 2021), which can automatically extract high-dimensional features from input images and convert them into hash codes. We have applied six data sets (Beetle, Fungi, Butterfly, Flower, Bird, and Merged data sets) to explore the application of hash codes in two aspects. First, we verified the ability of hash codes to retrieve specimens at a large scale in scenarios such as the natural history collections in six data sets and demonstrated the application cases of using hash codes to retrieve specimens in the simulated database. In this way, even without prior knowledge about the query image, the hash code can match individuals in the database that have similar morphological features to it. Next, to further explore the representation ability of hash codes, we employ two distinct visualization methods. One showcases the collective representation ability of the hash codes, while the other highlights the performance of each individual bit within the code, thus indicating that hash codes serve as effective carriers of features. Furthermore, we conduct an exploration of the potential of hash codes in generating a subfamily-level phenetic distance tree, using a beetle data set that encompasses six major subfamilies of Scarabaeidae as an illustrative example.

2 | MATERIALS AND METHODS

2.1 | Data sets and data preprocessing

In this study, we used a total of six image data sets (Table 1), including the beetle data set (Zhao et al., 2023), fungi data set (Picek et al., 2022), butterfly data set (Gerald, 2022b), flower data set (Nilsback & Zisserman, 2008), bird data set (Gerald, 2022a) and a merged data set composed of the previous five data sets. The beetle data set was newly constructed, while the others were obtained from publicly available data sets. To verify the feature extraction capabilities of different deep learning models, we conducted tests on beetle and bird data sets, which exhibit significant differences in the number of categories. The quantity of these categories can be determined based on diverse classification criteria, which may extend beyond taxonomy, and the outcomes of the model are closely linked to these criteria. To evaluate the retrieval performance of PENet, we tested it on all data sets. Furthermore, to explore the potential

TABLE 1 Names, numbers of samples, numbers of categories and the usage of the data sets.

Data set	Quantity	Number of categories	Usage
Beetle data set	6566	7	Validation of feature extraction capability Evaluation of PENet's retrieval capability Exploration of hash code applications
Fungi data set	7685	20	Evaluation of PENet's retrieval capability
Butterfly data set	10,035	75	Evaluation of PENet's retrieval capability
Flowers data set	7370	102	Evaluation of PENet's retrieval capability
Bird data set	62,388	400	Validation of feature extraction capability Evaluation of PENet's retrieval capability
Merged data set	94,044	604	Evaluation of PENet's retrieval capability

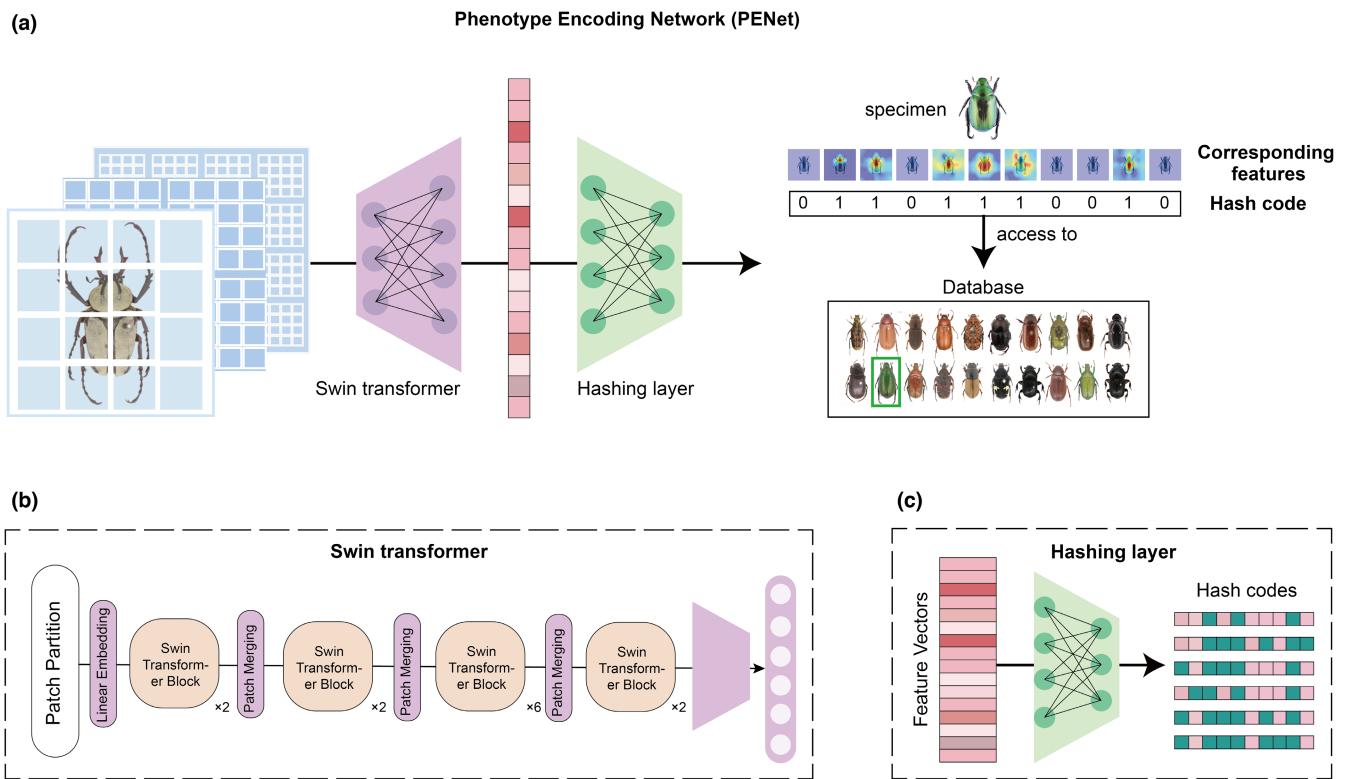


FIGURE 1 Illustrations of the PENet. (a) The PENet pipeline. (b) The architecture of the Swin transformer. (c) Hashing layer mapping feature vectors to hash codes.

applications of hash codes, we used the beetle data set as an example to investigate the value of distance information carried by hash codes. Detailed descriptions of the data sets are provided in the Supporting Information.

To ensure the suitability of images from various sources for model training, we conducted preprocessing steps. First, we resized the images to ensure uniform length and width, employing a solid colour filling strategy to prevent image distortion. Next, the data sets were divided in a ratio of 7:2:1, allocating 70% of the data for training (training set), 20% for validation (validation set) and 10% for testing (test set). This division ensures that the model is trained on a sufficiently large amount of data while also having enough data for validation and testing to assess its performance.

2.2 | Validation of the extraction capability of the Swin transformer

The quality of features extracted by deep learning models plays a crucial role in specimen retrieval performance. Here, our work is mainly based on the Swin transformer (Liu et al., 2021), a state-of-the-art vision transformer model with a hierarchical design. This distinctive architecture enables it to show high performance in image feature extraction (Han et al., 2023). As depicted in Figure 1b, the Swin transformer mainly consists of four stages, each encompassing Patch Merging operations (except the first stage, which utilizes a linear layer) and multiple Swin transformer blocks (Figure S1b). During the processing flow, the Swin transformer segments the input image

into smaller patches to capture local details. These patches are then mapped into fixed-length vectors by Linear Embedding and input to the first stage of the Swin transformer. After the four stages of processing, these feature vectors are mapped to a high-dimensional representation. Among these stages, Patch Merging aims to reduce the resolution of the feature vectors before they are fed into the subsequent three stages. Once Patch Merging is done, Swin transformer blocks are used to extract features. More details of the swin transformer architecture can be found in the [Supporting Information](#).

To evaluate the feature extraction capabilities of the Swin transformer, we utilized two data sets of significantly different sizes: the beetle data set (six categories) and the bird data set (400 categories). Moreover, we conducted a comparative analysis between the Swin transformer and two well-established convolutional neural network models, namely AlexNet (Krizhevsky et al., 2012) and ResNet (He et al., 2016). AlexNet, a classic model in computer vision's early days, and ResNet, an innovative model incorporating residual connections, have both been widely employed for feature extraction tasks. By comparing the Swin transformer with these models, we can effectively assess its feature extraction performance.

To show more intuitively the feature extraction ability of different models, we used the classification accuracy performance of each model to demonstrate their respective feature extraction effects. Accuracy values were obtained on the test set that was not involved in the training process. Additionally, the confusion matrix is a commonly used tool for evaluating the performance of classification models (Sokolova & Lapalme, 2009), as it provides insight into the model's classification performance across different categories. Therefore, we also performed a confusion matrix analysis on the test set to further clarify the ability of different models to differentiate each category in the data set. To further improve the speed of model convergence, we loaded weights that were pretrained on ImageNet and trained for 50 epochs with a batch size of 64, and we used data augmentation strategies including random rotation, random flipping and random centre cropping in the training process. During the training process, the parameters were selected and saved for all models based on achieving the highest accuracy.

2.3 | PENet and model training

Based on the Swin transformer, we propose a new network combined with the hash method, called the PENet, which can represent the extracted features by a series of binary numbers ([Figure 1a](#)). In PENet, the Swin transformer is used to perform feature extraction on the input images ([Figure 1b](#)). The hashing method downscale the extracted high-dimensional vectors to binary codes as output, with each image corresponding to one hash code ([Figure 1c](#)).

Specifically, we adjusted the input dimension of the PENet to $224 \times 224 \times 3$, where 224×224 is the length and width of the input image, and three is the number of channels. Then, each input image was divided into 56×56 patches, where each patch is $4 \times 4 \times 3$, ensuring that there is no intersection between the patches. At this point,

the dimension of the input features was changed to $56 \times 56 \times 48$. After that, these features were processed into $56 \times 56 \times 96$ -dimensional features by the Linear Embedding layer and fed into Swin transformer blocks for feature extraction. After four stages of processing, the extracted features were passed through the global average pooling layer to generate a 768×1 -dimensional vector containing the high-dimensional features of the input image. Finally, we added a hash layer after the global average pooling layer to downscale the final obtained high-dimensional vectors to hash codes. The length of the hash code can be changed by setting the number of neurons in the hash layer. In summary, when training PENet, labelled images are required. Once the training is completed, PENet serves an end-to-end model that takes images as input and produces binary hash codes as output.

In the training process of PENet, we utilized the loss function proposed by Liu et al., which takes image pairs as input (Liu et al., 2016). This loss function quantifies the distance between image pairs based on contrastive loss and is in the form of distance-similarity product minimization. Let the original image pairs be denoted as I_1, I_2 with their corresponding output as binary hash codes $\mathbf{b}_1, \mathbf{b}_2$. The loss for the original image pairs can be defined as follows.

$$L(\mathbf{b}_1, \mathbf{b}_2, y) = \frac{1}{2}(1 - y)D(\mathbf{b}_1, \mathbf{b}_2) + \frac{1}{2}y \max(m - D(\mathbf{b}_1, \mathbf{b}_2), 0), \\ \text{s.t. } \mathbf{b}_j \in \{+1, -1\}^k, j \in \{1, 2\}, \quad (1)$$

where $y=0$ is defined when the original image pairs are similar (i.e. have the same label) and $y=1$ otherwise. $D(\cdot, \cdot)$ denotes the Hamming distance between $\mathbf{b}_1, \mathbf{b}_2$, m is a margin threshold parameter, and k is the length of the binary code. Here, the Hamming distance indicated the number of different characters in the corresponding positions of two equal strings, that is, the number of different bits in the two hash codes (Bookstein et al., 2002). The Hamming distance between two hash codes, x and y , is denoted as:

$$D(x, y) = \sum x_i \oplus y_i. \quad (2)$$

In this formula, $i=0, 1, \dots, n-1$, x and y are hash codes, n is the length of the hash code, and \oplus denotes the exclusive or (XOR) operation.

When similar images are mapped to different hash codes or dissimilar images are mapped to similar hash codes, they are punished by this loss function. It should be noted that when dissimilar images are mapped onto similar hash codes, they receive the punishment of this loss function only when the Hamming distance between their hash codes is below the threshold m . Based on this loss function, the model is trained by randomly selecting pairs of images as input. If two images have similar features, their hash codes keep close to each other; otherwise, they are pushed far away. This approach allows the model to convert the degree of similarity or dissimilarity of features between image pairs into distance measurements, and accurately map high-dimensional features to the hash codes (see [Supporting Information](#) for more details of the loss function).

2.4 | Fast retrieval of specimen images

To demonstrate the ability of hash codes to retrieve specimens, we tested six data sets (Beetle, Fungi, Butterfly, Flower, Bird, and Merged data set) using hash codes generated by PENet. The training set was used to adjust the model weight parameters, while the test set and the validation set were used to evaluate the performance of the model. We did this by computing the Hamming distance between the test set hash codes and the validation set hash codes. To further evaluate the model's retrieval capability, we used the mean Average Precision (mAP) as a metric, which measures the quality of retrieval results and the accuracy of ranking (Luo et al., 2020). The mAP is calculated as the average of the average precision of each query over all queries, which is calculated as:

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q}, \quad (3)$$

where Q is the number of queries, and the Average Precision is calculated as:

$$AP = \frac{\sum_{k=1}^n P(k) \cdot rel(k)}{N}, \quad (4)$$

where: n is the total number of retrieved items, k is the rank in the sequence of retrieved items, P(k) is the precision at rank k, which is the ratio of the number of relevant items among the top k retrieved items to k, rel(k) is an indicator function that is 1 if the item at rank k is relevant, and 0 otherwise, N is the total number of relevant items in the data set.

Specifically, for each data set, every hash code generated from the test set served as a query. During the retrieval, the query hash code was compared with every hash code in the validation set, and the Hamming distance was calculated for each comparison. The images in the validation set were then sorted according to their distance from the query, with the most similar images appearing at the top of the ranking. The mAP was then calculated based on the sorted results. The size of the test set is equivalent to the number of queries. To achieve optimal performance with different hash code lengths, we conducted separate training sessions for PENet each time the hash code length was modified. By varying the length of the hash code, we were able to assess PENet's performance in terms of information compression and discrimination. A shorter hash code length reduces storage requirements and computational complexity, but it may lead to a loss of discriminative power. Conversely, a longer hash code preserves more information but can introduce redundancy. This allows us to determine the ideal hash code length that strikes the right balance for a specific application or task. We trained the model for 45 epochs and saved the parameter settings that yielded the best performance for each hash code length (Table S3). Furthermore, for a more comprehensive demonstration of PENet's retrieval performance, we carried out additional comparisons involving the retrieval capabilities of PENet, AlexNet, and ResNet on the bird data set (refer to Supporting Information 5.1 for details).

To further demonstrate the effectiveness of hash codes in retrieval and storage, we first selected 1000 images from the Merged data set that were not used in the training process to serve as a simulated database and converted them into hash codes via the PENet. Next, we selected an untrained image from each of the five data sets (Beetle, Fungi, Butterfly, Flower, and Bird data sets) and converted them into hash codes. Finally, the hash code generated in the previous step was used as a query to retrieve the simulated database and return the five most similar images among them.

2.5 | Verification and visualization of hash codes representation capability

To demonstrate the overall representation capability of the hash codes, we applied the t-distribution stochastic neighbour embedding algorithm (t-SNE) to visualize the generated hash codes (Laurens & Hinton, 2008). This is currently a widely used method for dimensionality reduction and visualization, applied not only in machine learning research but also in the field of biology, including areas such as single-cell sequencing data (Do & Canzar, 2021; Kobak & Berens, 2019) and proteomics data (Rahmatbakhsh et al., 2021). The t-SNE algorithm is a non-linear dimensionality reduction technique, capable of mapping high-dimensional data points to a two-dimensional plane. The algorithm preserves the original structure of the data, at least at a local scale, by clustering items that are similar. Here, we have chosen to use 64-bit hash codes for t-SNE visualization, with each hash code representing a 64-dimensional feature vector.

The feature extraction process in deep learning has long been considered to have relatively low interpretability. As a result, some methods have emerged to attempt to visualize the extracted features, among which the gradient-weighted class activation mapping (Grad-CAM) algorithm is commonly used (Selvaraju et al., 2020). Grad-CAM generates heatmaps by calculating feature map gradients and weighting them with average pooling, it can help to understand which features are utilized by the network. Here, a total of 30 species of beetles in six subfamilies were selected from the beetle data set as test data for illustration. We used PENet to convert these beetle images into 64-bit hash codes and generated heatmaps using the Grad-CAM algorithm to display each bit of the hash code features.

2.6 | Constructing phenetic distance trees based on hash codes

As the hash codes contain the discriminative features extracted by the model and the distance of morphological differences among different categories, the generated hash codes from different taxa could form a morphological distance matrix. Therefore, we explored the potential of hash codes in representing phenetic distances. Here, we randomly selected two species from each subfamily of the beetle data set to generate 64-bit hash codes using PENet (Table S4). Based on these hash codes, we constructed a phenetic distance tree, with one species of

Staphylinidae selected as the outgroup. The hash code of the outgroup was also generated using PENet. Moreover, we employed the Grad-CAM method to highlight the positions of features corresponding to Staphylinidae. These selected species did not participate in the PENet training process to avoid overfitting. We used Nona (Goloboff, 1993) run via Winclada (Nixon, 1999) to perform heuristic searches to find the most parsimonious trees (MPT). MPTs were found with a heuristic search using the commands 'hold1000000', 'mult*1000', 'hold/10', 'mult*max'. Furthermore, to further demonstrate the potential of hash codes in phenetic distance studies, we additionally selected images of several other species and constructed phenetic distance trees based on these data (Supporting Information 5.2).

3 | RESULTS

3.1 | Feature extraction capability of different deep learning models

After 50 training epochs, all three models, namely AlexNet, ResNet, and Swin transformer, gradually converged, with respective training times of 101, 123, and 126 min (Figure S2). In the validation set

of the beetle data set, their highest accuracy was 90.58%, 94.98%, and 97.8%, respectively (Figure 2a). The confusion matrix analysis revealed that the Swin transformer outperformed the other two models in the performance of each subfamily, with accuracy above 95% for every subfamily (Figure 2b-d). However, the other two models mainly had prediction errors concentrated in the subfamilies Dynastinae, Melolonthinae, and Rutelinae. In particular, their prediction accuracy for the subfamily Dynastinae was relatively poor, with AlexNet at 78% and ResNet at 90%, compared to the accuracy of other subfamilies.

For bird data sets with a larger number of categories, AlexNet, ResNet, and Swin transformer also converged, with training times of 218, 487, and 478 min, respectively. The relatively simple network structure of AlexNet led to a shorter training time but a lower accuracy of only 88.5%, compared to ResNet's 96.05% and Swin transformer's 98.6% (Figure 2a). In confusion matrix analysis, the Swin transformer still performs the best, with only 25 mispredicted samples out of 2000 (Table S6).

For these two data sets, the Swin transformer outperforms the other two models in both accuracy and confusion matrix analysis, indicating that the Swin transformer has stronger feature extraction capability.

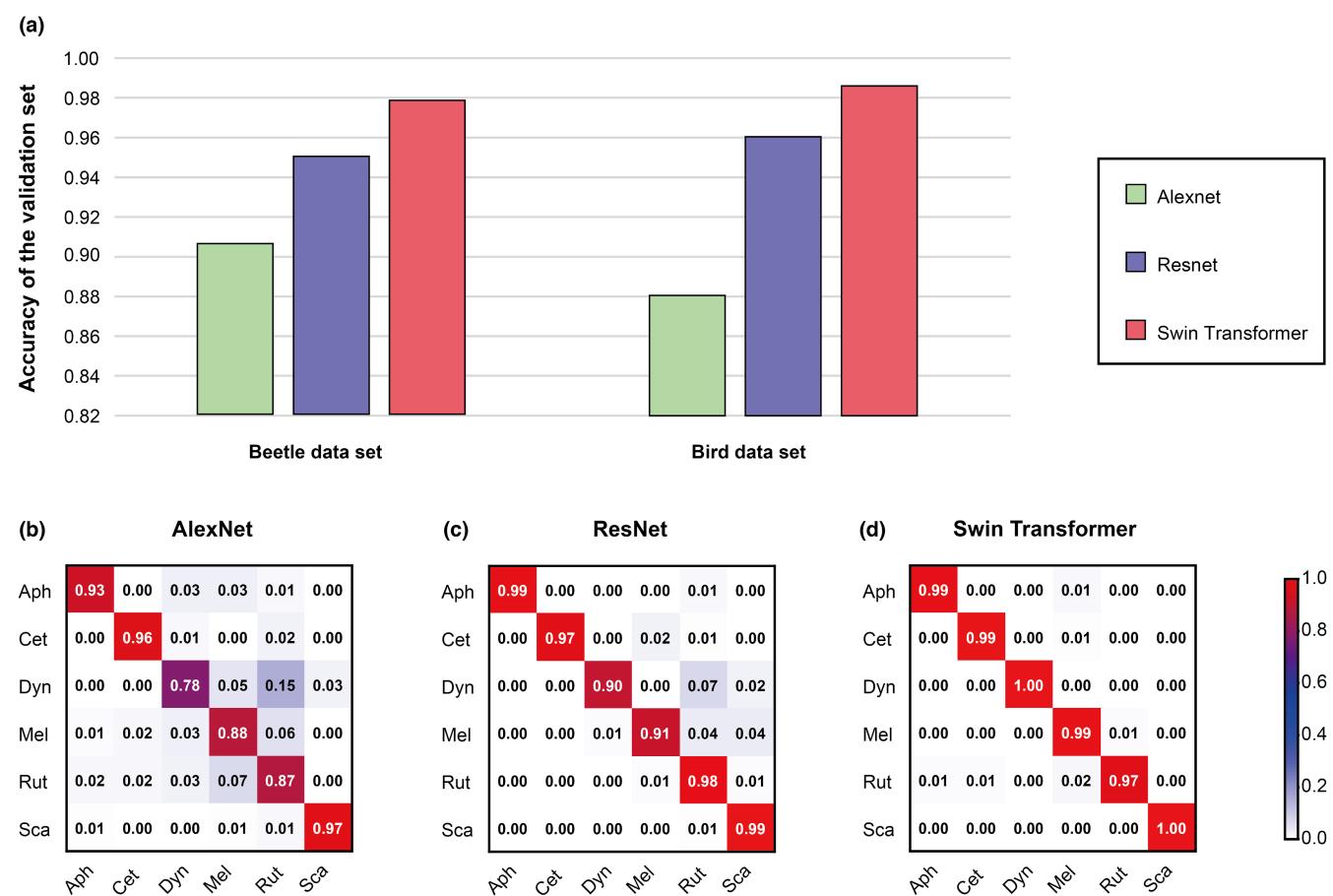


FIGURE 2 Comparison of the classification performance of the AlexNet, ResNet, and Swin transformer. (a) The highest accuracy of the three models in the validation set. (b-d) Confusion matrix analysis of the three models in the beetle data set.

3.2 | Image retrieval performance of PENet

For different data sets, the mAP values of hash codes retrieval are somewhat different. The highest mAP value on the beetle data set occurs with a hash code length of 48, reaching 98.2% (Figure 3a). For the Fungi, Butterfly, Flower, Bird, and Merged data sets, the highest mAP values are achieved with hash code lengths of 128, reaching 88.9%, 99%, 95.4%, 97.6%, and 94.8%, respectively (Figure 3b-f). Among them, we observed that the highest mAP value for the Fungi data set with the complex backgrounds was 88.9%, suggesting that the background of the data set can impact the retrieval accuracy of hash codes to some extent. Although the mAP value of the Fungi data set is relatively low, its impact on the mAP value of the merged data set is weakened because mAP takes into account the average precision of all categories.

Across all data sets, the mAP values exhibit a plateau pattern, where they gradually increase as the length of hash codes increases, and then reach a stable level. This suggests that the length of the required hash code does not increase infinitely as the number of categories increases. While our study did not explore longer hash code lengths, increasing the hash code length moderately can enhance performance, especially as the number of categories grows. However, it is important to avoid excessively extending the hash code length. In the six data sets analysed in this study, it was found that once the hash code length exceeded 64 bits, the changes in mAP values became negligible (ranging from 0.1% to 0.9%). Therefore, in the simulated database scenario, we utilized 64-bit hash codes as a demonstration, as they possess sufficient discriminatory power while avoiding excessive information redundancy. After retrieving data from a simulated database using hash codes, the top five results all correspond to the same category as the hash code used for the retrieval. Furthermore, among

the retrieval results of the same category based on the hash codes, the images that rank higher are more similar to the query image in terms of their features (Figure S3).

3.3 | Visualization of hash codes

After reducing the dimensionality of the hash codes, which contain 64 feature values, using t-SNE, we can observe that the groups of categories with similar features are still effectively clustered together (Figures S4 and S5). This indicates that the hash codes carry sufficient discriminative features and provide a good representation of them. Furthermore, the visualization of discriminative features corresponding to each bit of the hash code also supports this perspective. For the 30 selected species, 64 heatmaps were generated for each, one of these species was displayed in Figure 4, and the rest of the results were detailed in Supporting Information. Here, the areas with the highest intensities in the heatmaps of status 1 in each bit of hash code cover the features extracted from the PENet, Status 0 means that these samples do not contain the features extracted in these hash code bits. After studying these heatmaps, it is further indicated that some bits of hash codes pointing to the real discriminative features which have been used in traditional ways, for example, bit 1: legs, bit 16: the shape of the prothorax, bit 39: the centre of the body, and bit 44: the end of the abdomen.

3.4 | Phenetic distance tree based on hash codes

We utilized PENet to convert 13 beetle images that were not part of the training set into hash codes (Table S4). The hash codes of images belonging to the same category exhibited a similar

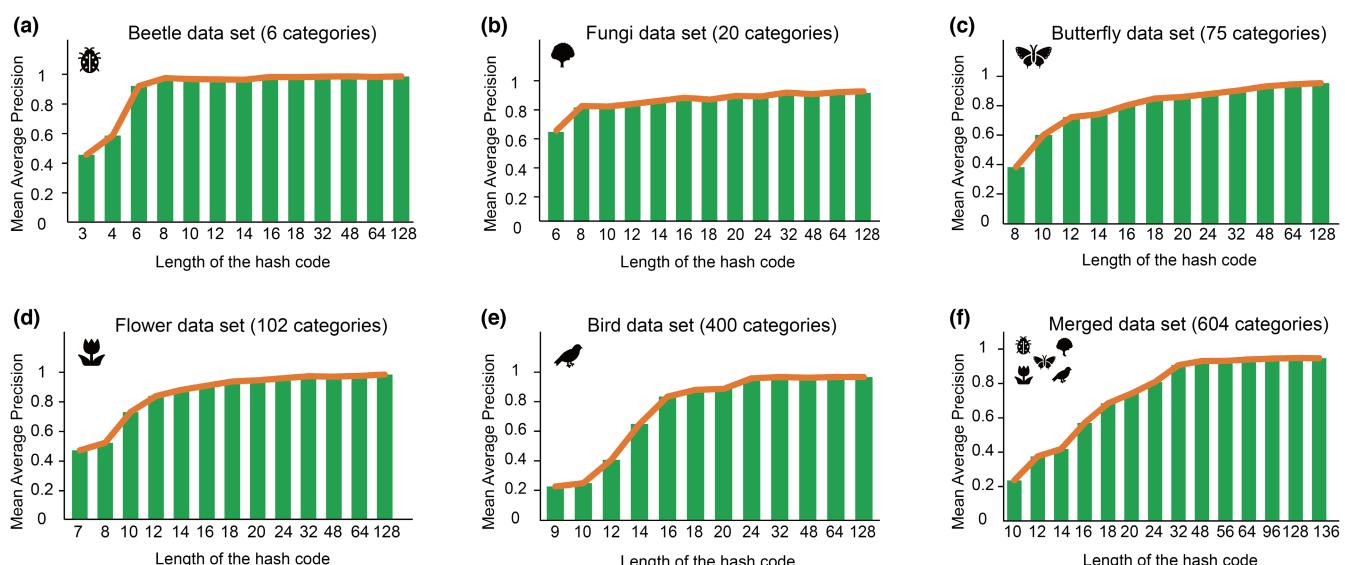


FIGURE 3 Histograms obtained under different hash code bits with their corresponding mAPs. (a–e) correspond to the beetle, fungi, butterfly, flower, and bird data sets, respectively, and (f) is merged from these five data sets. The number of categories in each data set is in parentheses after each heading.

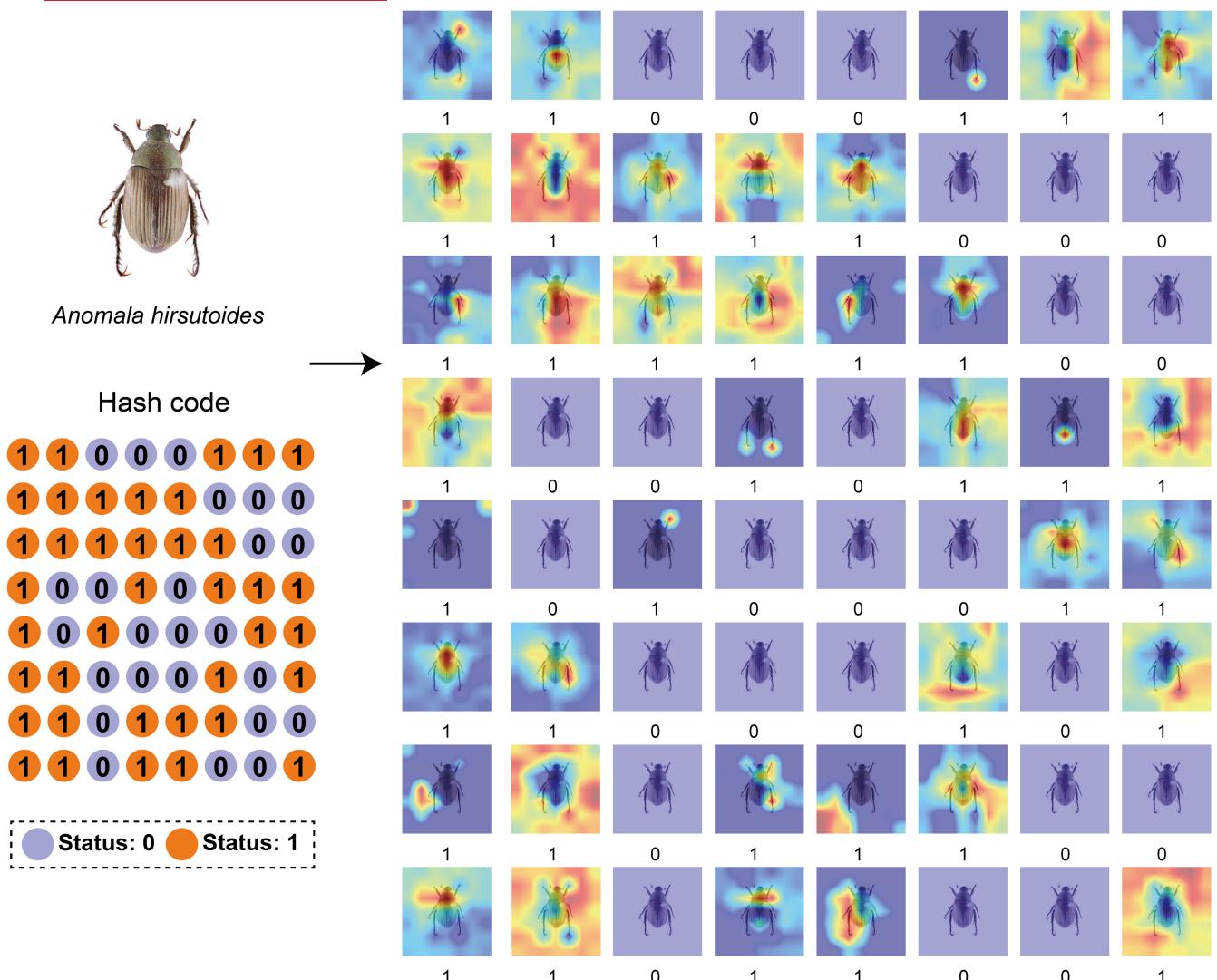


FIGURE 4 A 64-bit hash code is generated from a specimen by the PENet, where each bit corresponds to some features. The arrangement order of each bit in the hash code is from left to right, then from top to bottom.

pattern, whereas those of different categories showed some distinctions. For image that falls outside of the known categories (Staphylinidae), which can be considered out-of-distribution, its hash code differs from those of the known categories. To further investigate the value of hash codes in systematics, we constructed a phenetic distance tree. Maximum parsimony analysis of the 64 hash codes yielded two most parsimonious trees (tree length = 100 steps, CI = 0.55, RI = 0.72). Morphological characters (Hash codes) were optimized with parsimony on the first of the two most parsimonious trees (Figure 5), showing only unambiguous changes. Black circles indicate nonhomoplasious changes, and white circles indicate changes in homoplasious characters. The number above the branch represents hash code numbers, below the branch is hash code status (0 or 1).

From the topological structure of the phenetic tree constructed by hash codes, it is shown that all pairs of taxa belonging to the same subfamily were clustered together. Within the family Scarabaeidae, two primary lineages were clustered: the coprophagous lineage

(Scarabaeinae + Aphodiinae) and the phytophagous lineage (Cetoniinae + Melolonthinae + Rutelinae + Dynastinae). Within the phytophagous lineage, Melolonthinae, Rutelinae, and Dynastinae were clustered together, with Dynastinae and Rutelinae forming sister groups. The topological structure of the phenetic tree constructed by hash codes is similar to the phylogenetic trees revealed from molecular and morphological data, indicating that the hash codes contain additional distance information (Figure 5).

4 | DISCUSSION

The implementation of automated and efficient discriminative feature extraction, along with mathematical encoding of extracted features, will further accelerate the development of research that relies on the dependence of morphological features. Here, we developed a novel deep learning model called PENet, which can rapidly extract discriminative features and represent them efficiently

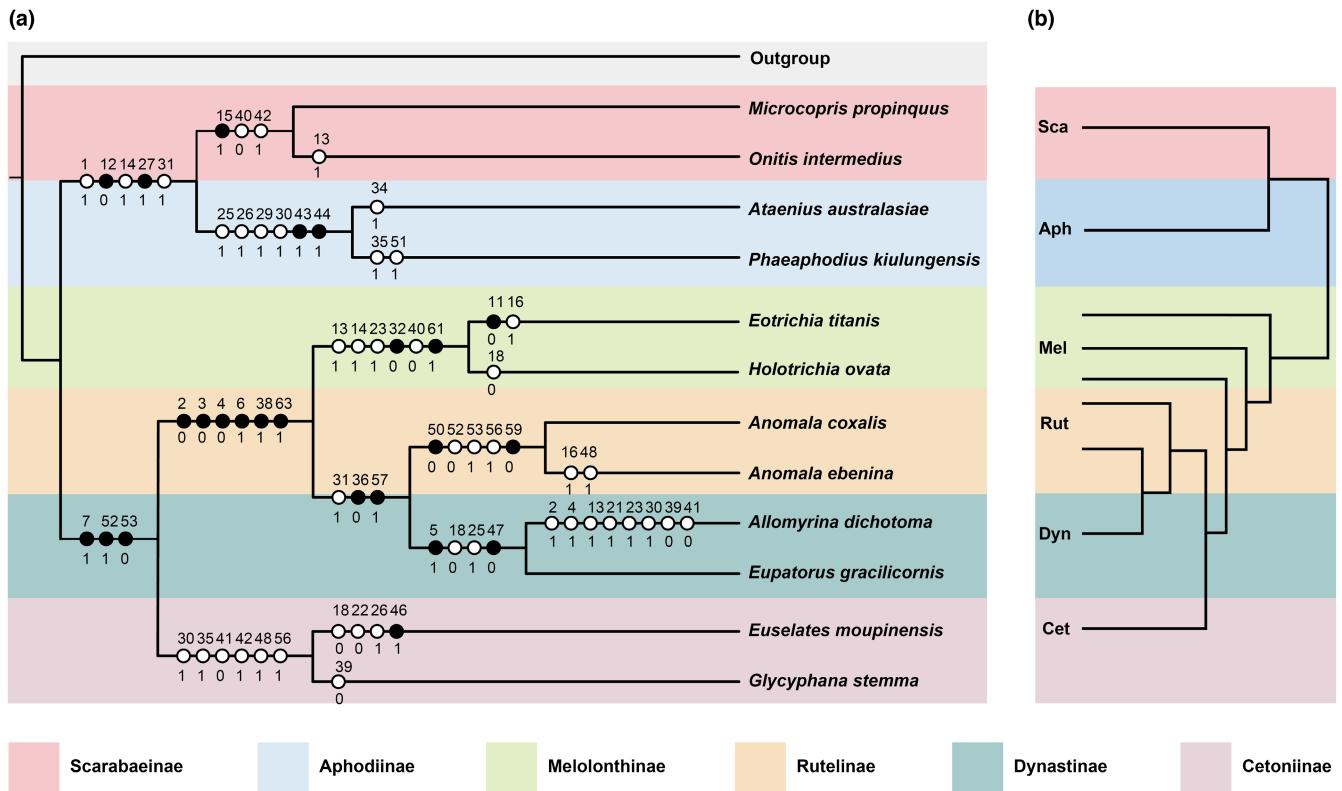


FIGURE 5 (a) Phenetic distance tree based on hash codes. Two species were selected for each subfamily with a hash code length of 64. The family Staphylinidae was selected as outgroup. (b) The simplified relationships among six subfamilies revealed by the existing phylogenetic tree (Ahrens et al., 2014).

using hash codes. Our study has shown that the hash codes generated by PENet are an efficient carrier of the extracted discriminative features, as they encode these features as sequences. This encoding enables fast retrieval performance and facilitates comparisons and matches with minimal computational resources. In addition, we utilized hash codes to construct a simple phenetic distance tree and discovered their potential for further application in systematics.

As the feature extractor of PENet, the Swin transformer outperforms conventional convolutional neural networks by excelling at capturing comprehensive and informative features from input data. This can be seen in the confusion matrix analysis of the beetle data set, where the Swin transformer performs well on each subfamily (Figure 2d). In contrast, AlexNet and ResNet show imbalanced performance on each subfamily, with the main prediction errors occurring in Dynastinae and Melolonthinae. For Dynastinae, which had the smallest number of images, the imbalanced data distribution may have been a contributing factor to the poor performance of the AlexNet and ResNet models. For Melolonthinae, as the largest subfamily with numerous species (~630 genera), there may be more extensive variations in the morphological features of its species, which could have impacted the performance of the models. Unlike AlexNet and ResNet, the Swin transformer excels in handling data imbalance and complex feature data. This may be attributed to its unique architecture, which segments images into

small tokens to extract features and incorporates self-attention mechanisms to capture the interdependencies between different tokens, allowing for more effective feature extraction and weighting (Liu et al., 2021; Vaswani et al., 2017). Furthermore, some studies have shown that, unlike convolutional neural networks which focus on extracting texture information from images, transformer-based models place greater emphasis on extracting shape information from the global image (Baker et al., 2018; Tuli et al., 2021). Therefore, on the whole, Swin transformer has greater potential for extracting discriminative features, and it also endows PENet with improved performance (Table S5).

Natural history collections play a crucial role as essential resources for studying the morphological traits of species (Heberling, 2022; Holmes et al., 2016; Lister, 2011), and recent advancements in their digitization have significantly facilitated related research (Page et al., 2015; Soltis, 2017). These advancements have also paved the way for rapid specimen retrieval. Our newly proposed method, PENet, offers a fast and efficient approach to retrieving specimens based on their morphological features. By employing PENet, we extract morphological features from digital specimen images and convert them into hash codes. These hash codes effectively represent the extracted features and demonstrate strong representational power. Through retrieval tests conducted on six data sets, we have demonstrated the effectiveness of these hash codes in capturing discriminative information and enabling the retrieval of similar

categories. Even in cases where we lack information about the query image, the hash code can match individuals in the database with similar morphological features to the query image. Consequently, once we generate a library of corresponding hash codes from existing digitized specimens, we can swiftly retrieve specimens based on their morphological features. This retrieval method significantly improves efficiency and reduces storage space requirements. For instance, considering the simulated database we constructed, compressing all images to approximately 220×220 pixels still requires approximately 8000 times the storage space needed for the hash codes alone.

In addition to enabling fast specimen retrieval, hash codes also have the potential to be further applied in systematics. Unlike most existing species classification models, PENet takes into account the distance relationship between different categories during the training process of generating hash codes. Therefore, the hash codes not only represent the extracted discriminative features but also carry distance information between different categories. Tests conducted on six subfamilies of the Scarabaeidae demonstrated that hash codes can be used to generate a phenetic distance tree. When compared with the existing phylogenetic tree, the phenetic distance tree showed some similarities in the division of certain major branches: two basic lineages, sister group relationship between Rutelinae and Dynastinae (Ahrens et al., 2014; Lu et al., 2023; Mckenna et al., 2015). And the position of Cetoniinae is similar to some morphological-based phylogenetic results, which further indicate that the hash codes could reveal the phenetic distances and relationships among categories to a certain extent (Howden, 1982; Meinecke, 1975). However, we emphasize that our experiments are preliminary, and the analysis is based on partial information, which is insufficient to encompass all the features of the test species. Future research could expand on this approach by augmenting the data sets, like multi-angle photography and microscopic imaging, to acquire comprehensive morphological information. This will enable us to further substantiate the efficacy of hash codes in extracting phylogenetic information.

There are still some issues to consider in the implementation of PENet in practice. When utilizing PENet for specimen retrieval, a critical aspect to consider is that the hash codes must originate from the same trained model to ensure comparability across different collections. Moreover, the retrieval performance is significantly influenced by the length of the hash code and the specificity of the training data. Therefore, it is essential to determine the optimal hash code length based on the specific requirements of diverse fields. Additionally, attention should be given to the selection of training data, aiming to minimize the impact of hash code length on PENet's performance. When addressing the inclusion of new data entries, the necessity of retraining the model depends on the characteristics of the data. Generally, if the existing model's training data sufficiently encompasses the distribution of the new data, fine-tuning the model might be sufficient for adaptation, obviating the need for a full retraining. However, it is important to note that updating the model could alter the hash

codes generated, potentially affecting the efficacy of the retrieval approach. This involves striking a balance between the imperative for model updates and the preservation of stable and comparable hash codes. Additionally, while the highlighted regions emphasized by the Grad-CAM method are similar to expert-perceived features, additional confirmation should be required to ascertain whether the extracted features match the phenotypic information to be studied.

5 | CONCLUSIONS

Overall, our newly developed end-to-end PENet model demonstrates high performance in feature extraction and fast retrieval, with the potential for broader applications in systematics. Hash codes carry both discriminative features and phenetic distance information while maintaining a low-dimensional representation, allowing efficient morphological information retrieval with a minimal storage cost. PENet provides an effective solution for fast retrieval of natural history collections. Furthermore, it can be further applied to explore morphological features, supporting research on macro-morphological evolution and mimicry. In future research, we will continue to explore the extended applications of hash codes and consider unsupervised training methods that rely solely on morphological distance information to train PENet, further increasing its scope of applications.

AUTHOR CONTRIBUTIONS

Zhengyu Zhao, Yuanyuan Lu and Ming Bai designed the study and wrote the paper. Zhengyu Zhao performed the training of the model, and Yuanyuan Lu interpreted the results. Zhengyu Zhao, Yijie Tong and Xin Chen completed the data collection. Xin Chen managed the equipment for model training. All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

We would like to thank Jing Li and Ning Liu from the Institute of Zoology, Chinese Academy of Sciences for technical support in computers. And we would like to thank Chuanbu Gao for providing some beetle images for this study. Shiqing Qiao, Yiyao Zhang and Zi Jin from Shenyang Normal University, and Jiateng Zhao from Hebei University for data collecting. This research was funded by the National Key R&D Program of China (No. 2022YFC2601200); the National Natural Science Foundation of China (Nos. 32270468, 32200354); the National Science & Technology Fundamental Resources Investigation Program of China (Nos. 2022FY202100, 2022FY100500, 2019FY100400, 2019FY101800); the project of the Northeast Asia Biodiversity Research Centre (NABRI202203); and China Postdoctoral Science Foundation (No. 2022M713134).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14235>.

DATA AVAILABILITY STATEMENT

All PENet code and associated data are deposited at <https://doi.org/10.57760/sciedb.11460> (Zhao, 2023).

ORCID

Zhengyu Zhao  <https://orcid.org/0000-0003-2335-447X>

Yuanyuan Lu  <https://orcid.org/0000-0003-0648-5531>

REFERENCES

- Ahrens, D., Schwarzer, J., & Vogler, A. P. (2014). The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B: Biological Sciences*, 281(1791), 20141470. <https://doi.org/10.1098/rspb.2014.1470>
- Allan, E. L., Dupont, S., Hardy, H., Livermore, L., & Smith, V. (2019). High-throughput digitisation of natural history specimens. *Biodiversity Information Science and Standards*, 3, e37337. <https://doi.org/10.3897/biss.3.37337>
- Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS Computational Biology*, 14(12), e1006613. <https://doi.org/10.1371/journal.pcbi.1006613>
- Beaman, R., & Cellinese, N. (2012). Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, 209, 7–17. <https://doi.org/10.3897/zookeys.209.3313>
- Bookstein, A., Kulyukin, V. A., & Raita, T. (2002). Generalized hamming distance. *Information Retrieval*, 5(4), 353–375. <https://doi.org/10.1023/A:1020499411651>
- Chen, Y., Bart, H. L., & Teng, F. (2005). A content-based image retrieval system for fish taxonomy. *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 237–244. <https://doi.org/10.1145/1101826.1101864>
- Chi, L., & Zhu, X. (2018). Hashing techniques: A survey and taxonomy. *ACM Computing Surveys*, 50(1), 1–36. <https://doi.org/10.1145/3047307>
- Do, V. H., & Canzar, S. (2021). A generalization of t-SNE and UMAP to single-cell multimodal omics. *Genome Biology*, 22(1), 130. <https://doi.org/10.1186/s13059-021-02356-5>
- Gerald, P. (2022a). Birds 400 SPECIES-image classification. <https://www.kaggle.com/datasets/gpiosenka/100-bird-species>
- Gerald, P. (2022b). Butterfly & moths image classification 100 species. <https://www.kaggle.com/datasets/gpiosenka/butterfly-image-s40-species>
- Goloboff, P. (1993). NONA, version 2.0. A tree searching program. Distributed by the author.
- Hall, A. V. (1974). Museum specimen record data storage and retrieval. *Taxon*, 23(1), 23–28. <https://doi.org/10.2307/1218085>
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>
- Hawkins, J. A. (2000). A survey of primary homology assessment: Different botanists perceive and define characters in different ways. In R. Scotland & T. Pennington (Eds.), *Homology and systematics: Coding characters for phylogenetic analysis* (pp. 22–53). Taylor and Francis.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Heberling, J. M. (2022). Herbaria as big data sources of plant traits. *International Journal of Plant Sciences*, 183(2), 87–118. <https://doi.org/10.1086/717623>
- Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., Kennedy, J., Clarke, J. A., Cook, J. A., Blackburn, D. C., Edwards, S. V., & Davis, C. C. (2020). Digitization and the future of natural history collections. *BioScience*, 70(3), 243–251. <https://doi.org/10.1093/biosci/biz163>
- Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., LaBarbera, K., Wommack, E. A., Martins, F. M., Crawford, J. C., Mack, K. L., Bloch, L. M., & Nachman, M. W. (2016). Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25(4), 864–881. <https://doi.org/10.1111/mec.13529>
- Howden, H. F. (1982). Larval and adult characters of *Frickius Germain*, its relationship to the Geotrupini, and a phylogeny of some major taxa in the Scarabaeoidea (Insecta: Coleoptera). *Canadian Journal of Zoology*, 60(11), 2713–2724. <https://doi.org/10.1139/z82-347>
- Jan, M., Zainal, N., & Jamaludin, S. (2020). Region of interest-based image retrieval techniques: A review. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 9(3), 520. <https://doi.org/10.11591/ijai.v9.i3.pp520-528>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Knott, G. D. (1975). Hashing functions. *The Computer Journal*, 18(3), 265–278. <https://doi.org/10.1093/comjnl/18.3.265>
- Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, 10(1), 5416. <https://doi.org/10.1038/s41467-019-13056-x>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Lamichhaney, S., Card, D. C., Grayson, P., Tonini, J. F. R., Bravo, G. A., Näpflin, K., Termignoni-Garcia, F., Torres, C., Burbrink, F., Clarke, J. A., Sackton, T. B., & Edwards, S. V. (2019). Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 374(1777), 20180248. <https://doi.org/10.1098/rstb.2018.0248>
- Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., Zafar, B., Dar, S. H., Sajid, M., & Khalil, T. (2019). Content-based image retrieval and feature extraction: A comprehensive review. *Mathematical Problems in Engineering*, 2019, 1–21. <https://doi.org/10.1155/2019/9658350>
- Laurens, V. D. M., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2605), 2579–2605.
- Lister, A. M. (2011). Natural history collections as sources of long-term datasets. *Trends in Ecology & Evolution*, 26(4), 153–154. <https://doi.org/10.1016/j.tree.2010.12.009>
- Liu, H., Wang, R., Shan, S., & Chen, X. (2016). Deep supervised hashing for fast image retrieval. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2064–2072. <https://doi.org/10.1109/CVPR.2016.227>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Lu, Y., Ahrens, D., Shih, C., Shaw, J. J., Yang, X., Ren, D., & Bai, M. (2023). A cretaceous chafer beetle (Coleoptera: Scarabaeidae) with exaggerated hind legs—Insight from comparative functional morphology into a possible spring movement. *Biology*, 12(2), 237. <https://doi.org/10.3390/biology12020237>

- Luo, X., Wu, D., Chen, C., Deng, M., Huang, J., & Hua, X. (2020). A survey on deep hashing methods. *ACM Transactions on Knowledge Discovery from Data*, 17(1), 1–50. <https://doi.org/10.1145/3532624>
- Mallik, J., Samal, A., & Gardner, S. L. (2010). A content based image retrieval system for a biological specimen collection. *Computer Vision and Image Understanding*, 114(7), 745–757. <https://doi.org/10.1016/j.cviu.2010.01.006>
- Mckenna, D. D., Farrell, B. D., Caterino, M. S., Farnum, C. W., Hawks, D. C., Maddison, D. R., Seago, A. E., Short, A. E. Z., Newton, A. F., & Thayer, M. K. (2015). Phylogeny and evolution of Staphyliniformia and Scarabaeiformia: Forest litter as a stepping stone for diversification of nonphytophagous beetles: Evolution of Staphyliniformia and Scarabaeiformia. *Systematic Entomology*, 40(1), 35–60. <https://doi.org/10.1111/syen.12093>
- Meharban, M. S., & Priya, S. (2016). A review on image retrieval techniques. *Bonfring International Journal of Advances in Image Processing*, 6(2), 7–10. <https://doi.org/10.9756/BIAIP.8136>
- Meinecke, C.-C. (1975). Olfactory sensilla and systematics of the Lamellicornia (Insecta, Coleoptera). *Zoologische Verhandlungen*, 82(1), 1–42. <https://doi.org/10.1007/BF00995905>
- Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 374(1763), 20170391. <https://doi.org/10.1098/rstb.2017.0391>
- Nilsback, M.-E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–729. <https://doi.org/10.1109/ICVGIP.2008.47>
- Nixon, K. (1999). *WinClada ver. 1.0000*. Published by the Author, Ithaca, NY, USA, 2002.
- Owen, D., Groom, Q., Hardisty, A., Leegwater, T., Livermore, L., van Walsum, M., Wijkamp, N., & Spasić, I. (2020). Towards a scientific workflow featuring natural language processing for the digitisation of natural history collections. *Research Ideas and Outcomes*, 6, e58030. <https://doi.org/10.3897/rio.6.e58030>
- Ożgo, M., & Schilthuizen, M. (2012). Evolutionary change in Cepaea nemoralis shell colour over 43 years. *Global Change Biology*, 18(1), 74–81. <https://doi.org/10.1111/j.1365-2486.2011.02514.x>
- Page, L. M., MacFadden, B. J., Fortes, J. A., Soltis, P. S., & Riccardi, G. (2015). Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience*, 65(9), 841–842. <https://doi.org/10.1093/biosci/biv104>
- Picek, L., Šulc, M., Matas, J., Jeppesen, T. S., Heilmann-Clausen, J., Læssøe, T., & Frøsliev, T. (2022). Danish fungi 2020-not just another image recognition dataset. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1525–1535. <https://doi.org/10.1109/WACV51458.2022.00334>
- Rahmatbakhsh, M., Gagarinova, A., & Babu, M. (2021). Bioinformatic analysis of temporal and spatial proteome alternations during infections. *Frontiers in Genetics*, 12, 667936. <https://doi.org/10.3389/fgene.2021.667936>
- Schindel, D. E., & Cook, J. A. (2018). The next generation of natural history collections. *PLoS Biology*, 16(7), e2006125. <https://doi.org/10.1371/journal.pbio.2006125>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Soltis, P. S. (2017). Digitization of herbaria enables novel research. *American Journal of Botany*, 104(9), 1281–1284. <https://doi.org/10.3732/ajb.1700281>
- Suarez, A. V., & Tsutsui, N. D. (2004). The value of museum collections for research and society. *BioScience*, 54(1), 66–74. [https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
- Tang, J., Li, Z., Wang, M., & Zhao, R. (2015). Neighborhood discriminant hashing for large-scale image retrieval. *IEEE Transactions on Image Processing*, 24(9), 2827–2840. <https://doi.org/10.1109/TIP.2015.2421443>
- Tuli, S., Dasgupta, I., Grant, E., & Griffiths, T. L. (2021). Are convolutional neural networks or transformers more like human vision? *arXiv*, arXiv:2105.07197. <http://arxiv.org/abs/2105.07197>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv*, arXiv:1706.03762. <http://arxiv.org/abs/1706.03762>
- Walker, B. E., Tucker, A., & Nicolson, N. (2022). Harnessing large-scale herbarium image datasets through representation learning. *Frontiers in Plant Science*, 12, 806407. <https://doi.org/10.3389/fpls.2021.806407>
- Weeks, B. C., Zhou, Z., O'Brien, B. K., Darling, R., Dean, M., Dias, T., Hassena, G., Zhang, M., & Fouhey, D. F. (2023). A deep neural network for high-throughput measurement of functional traits on museum skeletal specimens. *Methods in Ecology and Evolution*, 14(2), 347–359. <https://doi.org/10.1111/2041-210X.13864>
- Winker, K. (2004). Natural history museums in a postbiodiversity era. *BioScience*, 54(5), 455. [https://doi.org/10.1641/0006-3568\(2004\)054\[0455:NHMIAP\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0455:NHMIAP]2.0.CO;2)
- Younis, S., Weiland, C., Hoehndorf, R., Dressler, S., Hickler, T., Seeger, B., & Schmidt, M. (2018). Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Botany Letters*, 165(3–4), 377–383. <https://doi.org/10.1080/23818107.2018.1446357>
- Zhao, Z. (2023). Codes for PENet: A model for rapid retrieval of natural history collections (version V1). Science Data Bank. <https://doi.org/10.5776/scencedb.11460>
- Zhao, Z., Tong, Y., Lu, Y., & Bai, M. (2023). Beetle dataset (version V1). Science Data Bank. <https://doi.org/10.5776/scencedb.07795>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Supplementary text, tables, figures and references.

How to cite this article: Zhao, Z., Lu, Y., Tong, Y., Chen, X., & Bai, M. (2023). PENet: A phenotype encoding network for automatic extraction and representation of morphological discriminative features. *Methods in Ecology and Evolution*, 00, 1–12. <https://doi.org/10.1111/2041-210X.14235>