Deep-learning technology provides insights into the morphological evolution of birds

Jiao Sun[123*]
[1] CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China;
[2] Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan 430074, China;
[3] University of Chinese Academy of Sciences, Beijing 100049, China.
* Correspondence: sunjiao19@mails.ucas.ac.cn
ORCID: https://orcid.org/0000-0002-5028-8132

Abstract

The evolution of biological morphology is critical for understanding the diversity of the natural world, yet traditional analyses often involve subjective biases in the selection and coding of morphological traits. This study employs deep learning techniques, utilizing a pretrained ResNet34 model capable of recognizing over 10,000 bird species, to explore avian morphological evolution. We extracted weights from the model's final fully connected (fc) layer to create vector representations of avian species and assessed their similarities using cosine similarity metrics. The results demonstrated multiple clustering patterns with or without biological meaning. Some clustering results are consistent with traditional classifications based on morphology, some are consistent with modern cladistic classifications, and some show behavioural and ecological similarities. We utilized the variance of vectors based on Euclidean distance to assess the morphological disparity among various taxa and evaluated the association between morphological disparity and species richness. The result showed Despite these insights, some clusters indicated the influence of non-biological image features on clustering outcomes. This study underscores the potential and limitations of using deep learning approaches in morphological evolution studies.

Key words: bird, biodiversity, clustering, deep learning, morphological evolution.

Introduction

The evolution of biological morphology plays a crucial role in shaping the diverse natural world we observe today. It provides insight into the adaptation and survival of species over time, influencing various ecological interactions and the functioning of ecosystems. Traditionally, analyses of morphological evolution have involved subjective elements, as even quantitative analyses based on morphological traits require human intervention in the selection and coding of these traits (Clark et al., 2023; Crouch & Ricklefs, 2019). This subjectivity can introduce biases, affecting the accuracy and reliability of evolutionary interpretations.

To address these limitations, our research employs advanced deep learning technologies, specifically Convolutional Neural Networks (CNNs). CNNs, popularized by LeCun et al. (1989), are designed to automatically learn hierarchical features from image data, making them exceptionally suited for visual recognition tasks. By utilizing CNN image classification models, we can leverage the learned weights as indicators of morphological traits of various species, providing a more objective basis for understanding biological evolution.

In our study, we utilized the model trained by Sun (2025) and calculated the cosine similarity between species using the weights extracted from the last fully connected layer (fc). This methodology enabled us to perform hierarchical clustering based on the cosine similarities, yielding insights into the morphological evolution of avian species. Furthermore, we will utilized Euclidean distance-based vector variance to assess the morphological disparity among families and orders. Spearman's rank correlation coefficient (ρ) was then employed to evaluate the association between morphological disparity and species richness.

While this approach shows promise, it is important to note that it has some limitations, some results may lack biological meaning. However, as the development of deep learning technologies and more photos from citizen science, there is substantial potential for these methods to enrich our understanding of morphological evolution in the future.

Materials and Methods

Materials

In our study, we utilized the model by Sun (2025), which is based on the ResNet34 architecture and is capable of recognizing over 10,000 bird species with an accuracy of approximately 90%. The original weight data for this model is available on Hugging Face, a popular platform for sharing machine learning models and datasets. The model was trained on a dataset based on IOC (International Ornithological Congress) World Bird List 10.1 (Gill et al., 2021), while we reassigned the orders and families of all species to align with the IOC 15.1 (Gill et al., 2025).

We began by extracting the weights from the final fully connected layer (fc) of the ResNet34 model. Each species's weights were treated as a vector to analyze the relationships between different avian species based on these representations.

Similarity analysis

To assess the similarity between various species, we employed cosine similarity. Initially, all vectors were subjected to L2 normalization to ensure that they each had a unit length. Following this normalization, we performed dot product calculations on the normalized vectors. This method is equivalent to computing the cosine similarity of the original vectors, providing a meaningful metric for evaluating the relationships among the species. For this implementation, we utilized built-in functions from the PyTorch library, which facilitated efficient computation (Ansel et al., 2024).

Next, we conducted agglomerative hierarchical clustering using the average linkage method to merge clusters. This was executed with the hierarchical clustering functionalities implemented in the SciPy library (Gommers et al., 2025). The hierarchical structure of the clusters was output in Newick format, a widely used format in computational biology for tree structures. Finally, we utilized ETE3 to export the clustering dendrogram in SVG format (Huerta-Cepas et al., 2016).

To analyzing the clustering result, we applied a recursive top-down analysis to the tree, evaluating each internal node for taxonomic “purity.” For a given node, we defined taxonomic purity as the proportion of the majority taxon. A node with a taxonomic purity of more than 85% was considered taxonomically consistent. For taxonomically consistent nodes, all of their child nodes were excluded from further checks. For such nodes, we further examined all species to identify and annotate taxonomical outliers, which belong to taxa that different from the majority taxon of this branch. Finally, we conducted manual review to confirm whether outliers had biological similarity with the majority taxa of their branches and what kinds of similarity do they have. The above pipeline was carried out in both order-level and family-level.

Disparity analysis

Before proceeding with this analysis, we removed 122 species whose weight vectors were identified as lacking biological significance during the manual review process. The specific reasons are detailed in the discussion section.

To assess the morphological disparity among taxa, we employed Euclidean distance-based vector variance. Additionally, we calculated the Pearson correlation coefficient and the Spearman’s rank correlation coefficient to examine the relationship between diversity and morphological disparity. We computed the Spearman's coefficient and linear, log-linear, and power-law relationships of the Pearson coefficient both including and excluding monospecific taxa. Based on the correlation results, we fitted the relationship function using the model that exhibited the highest correlation. This analysis was conducted at both order and family levels.

Result

Our clustering process was successfully conducted, resulting in a comprehensive hierarchical clustering output. The agglomerative clustering technique applied to the cosine similarity measures of the weight vectors yielded a dendrogram that illustrates the relationships between the different avian species based on their morphological features learned by the ResNet34 model.

In the taxonomic consistence analysis, we identified a total of 391 branches with high taxonomic consistency at the family level and 94 branches at the order level. Additionally, we found 474 family-level outlier species and 533 order-level outlier species. The clustering result with collapsed high-purity branches are illustrated in Figure 1.

The orders and families with the greatest and least disparity are as follows (excluding monospecific taxa): The three orders with the greatest disparity are Anseriformes, Charadriiformes, and Pelecaniformes. Conversely, Struthioniformes, Aegotheliformes, and Apterygiformes had the least disparity.

At the family level, the three families with highest disparity included Laridae, Anatidae, and Ciconiidae, while the families with the least disparity were Atrichornithidae, Salpornithidae, and Melampittidae.

Regarding the relationship between order-level disparity and diversity, the Spearman's coefficient was calculated at 0.60, with a p-value of 1.40×10^{-5}. For family-level analysis, the Spearman's coefficient was 0.70, with a p-value of 1.05×10^{-39}. After removing all monospecific taxa, the Spearman's coefficient decreased to 0.51, with significant p-values of 3.4×10^{-16} in family-level and 6.4×10^{-4} in order-level. The Pearson coefficients for power-law relationships after the removal of monospecific taxa were found to be closest to the Spearman coefficients. All coefficients are listed in Table 1. Therefore, we derived the following power-law fitted relationship functions (Figure 2):

Fitted model in family level: $Y = 2.3 \times 10^{-4} \bullet X^{0.2038}$
Fitted model in order level: $Y = 3 \times 10^{-4} \bullet X^{0.1343}$

Discussion

Similarity analysis

The results of our clustering analysis highlight some critical insights as well as important limitations of deep learning techniques in studying morphology. One significant concern is the issue of interpretability in deep learning models. These models often seek local optima rather than global solutions, leading to clustering outcomes that may not possess real biological meanings.

Currently, the discussion section covers only a limited number of taxa, and we aim to report potential patterns that may carry biological relevance. Notably, the four genera referred to as "fulvetta," which were traditionally considered similar and related (Pasquet et al., 2006), did not fall into a single cluster. Instead, we found that *Fulvetta* and *Lioparus* clustered with most parrotbills (Paradoxornithidae), aligning with their modern taxonomic classification. On the contrary, the genera *Schoeniparus* (Pellorneidae) and *Alcippe* (Alcippeidae) fell into two separate but adjacent clusters, which suggests that deep neural networks can tell the morphological differences among the "fulvettas."

Another intriguing finding involves the *Pseudopodoces humilis*, a species morphologically similar to the genus *Podoces*, while is classified within the family *Paridae* based on molecular phylogeny. However, this species clustered with the "snowfinches" (*Onychostruthus*, *Pyrgilauda*, *Montifringilla*), indicating possible behavioural and ecological similarities. Both are secondary cavity-nesting birds, often found cohabiting with members of the family Ochotonidae in the Tibetan Plateau (Lu et al., 2011).

We observed that Galliformes and Tinamiformes were nested into a single cluster, pointing to similar morphology traits according to the model. Moreover, the clustering of most species of Strigiformes and Caprimulgiformes *s. l.* (including Steatornithiformes, Nyctibiiformes, Podargiformes and Aegotheliformes, but not Apodiformes) fell into adjacent clusters, potentially suggesting morphological convergence due to adaptations for nocturnality. Alternatively, this could merely reflect that most images were captured at night, leading the deep learning model to consider them similar.

Nonetheless, several groupings identified in our analysis evidently lack biological significance. For instance, *Nymphicus hollandicus* was shown to be most similar with *Melopsittacus undulatus*, while not aligned with other members of Cacatuidae. This is likely due to their wide captivity, leading the model to learn human presence or man-made objects in the images. Additionally, many extinct species clustered together, possibly due to their representation via skeletal images, artistic reconstructions, or other non-biological patterns. Furthermore, some of the recently described species or newly separated cryptic species grouped together, which might reflect insufficient image data, leading to underfitting of the model.

This highlights the duality of using Convolutional Neural Networks (CNNs) in morphological analysis: they can capture morphological differences that may not be discernible to the human eye, while they are also influenced by data noise. This underscores the necessity for human researchers to use their biological knowledge and practical experience to distinguish meaningful patterns from the noise when utilizing deep learning for morphological evolution studies. Thus, researchers can maximize the strengths of these technologies while avoiding their limitations.

Looking ahead, we plan to expand our analysis by developing code to assess the morphological disparity among different orders. We believe that this methodology may also contribute to the study of avian vocalizations.

Morphological Disparity

In our analysis of the correlation between morphological disparity and species richness, we removed monospecific taxa from consideration. The variance of a vector group containing only a single vector is inherently zero, representing a statistical technicality rather than indicating a lack of morphological evolution in their evolutionary history. Such zero values result in a series of repeated minimum ranks within Spearman's rank correlation analysis, leading to an overall correlation coefficient that deviates from the norm. After excluding monospecific taxa, Spearman's rank correlation coefficients decreased from 0.60 at the order level and 0.70 at the family level to approximately 0.51. This suggests that the original correlation was inflated by a substantial number of zero variances. While the correlation values declined after removal, the significance remained extremely robust, with the p-value of 3.4e-16 at the family level and of 6.4e-4 at the order level. The Spearman's rank correlation coefficients for both morphological disparity and species richness remained at 0.51, indicating a potential stability of this value across different taxonomic levels. It suggested a moderate positive correlation between morphological disparity and species richness.

The Pearson correlation coefficient was higher under the power-law relationship (log-log model), approximately 0.51, closely aligning with the Spearman rank correlation coefficient and exceeding the value of around 0.45 seen in the log-linear relationship ($Y \sim \log(X)$ model). This indicates that the relationship between the data showed higher correlation when examined on the power-law relationship.

The exponents of 0.2038 and 0.1343 in the fitted models being less than 1 imply a marginally decreasing trend in morphological diversity as species richness increases. This phenomenon may be linked to limitations in ecological niches and limited resource availability. It suggests that morphological disparity rapidly expanded in the earlier phases of adaptive radiation, while newer species in later stages tended to exhibit greater similarity in morphological traits.

Code accessibility

Code: https://github.com/sun-jiao/osea_morpho_evo
Model: https://huggingface.co/sunjiao/osea

Conflict of interests

The author has no conflict of interests.

Author contribution statements

JS designed the project, programmed the python script and drafted the manuscript.

Supplementary files

- class_similarity.csv: similarity matrix of all “classes” (species)
- morphology_clustering.tre: the clustering result
- tree_output.svg: the image version of the clustering result
- tree_output_analysed_family.svg: clustering with family and outlier annotations
- tree_output_analysed_order.svg: clustering with order and outlier annotations
- disparity_family.csv: morphological disparity (vector variances) of families
- disparity_order.csv: morphological disparity (vector variances) of orders

References

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., … Chintala, S. (2024). PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, 929–947. https://doi.org/10.1145/3620665.3640366

Clark, J. W., Hetherington, A. J., Morris, J. L., Pressel, S., Duckett, J. G., Puttick, M. N., Schneider, H., Kenrick, P., Wellman, C. H., & Donoghue, P. C. J. (2023). Evolution of phenotypic disparity in the plant kingdom. *Nature Plants*, *9*(10), 1618–1626. https://doi.org/10.1038/s41477-023-01513-x

Crouch, N. M. A., & Ricklefs, R. E. (2019). Speciation Rate Is Independent of the Rate of Evolution of Morphological Size, Shape, and Absolute Morphological Specialization in a Large Clade of Birds. *The American Naturalist*. https://doi.org/10.1086/701630

Gill, F., Donsker, D., & Rasmussen, P. (2021). IOC world bird list. *International Ornithologists' Union*, *10.1*. https://doi.org/10.14344/IOC.ML.10.1

Gill, F., Donsker, D., & Rasmussen, P. (2025). IOC world bird list. *International Ornithologists' Union*, *15.1*. https://doi.org/10.14344/IOC.ML.15.1

Gommers, R., Virtanen, P., Haberland, M., Burovski, E., Reddy, T., Weckesser, W., Oliphant, T. E., Cournapeau, D., Nelson, A., alexbrc, Roy, P., Peterson, P., Polat, I., Wilson, J., endolith, Mayorov, N., van der Walt, S., Colley, L., Brett, M., … Striega, K. (2025). scipy/scipy: SciPy 1.15.0. *Zenodo*. https://doi.org/10.5281/zenodo.14593523

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, *33*(6), 1635–1638. https://doi.org/10.1093/molbev/msw046

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, *1*(4), 541–551. Neural Computation. https://doi.org/10.1162/neco.1989.1.4.541

Lu, X., Huo, R., Li, Y., Liao, W., & Wang, C. (2011). Breeding ecology of ground tits in northeastern Tibetan pla- teau, with special reference to cooperative breeding system. *Current Zoology*, *57*(6), 751–757.

Pasquet, E., Bourdon, E., Kalyakin, M. V., & Cibois, A. (2006). The fulvettas (Alcippe, Timaliidae, Aves): A polyphyletic group. *Zoologica Scripta*, *35*(6), 559–566. https://doi.org/10.1111/j.1463-6409.2006.00253.x

Sun, J. (2025). OSEA, a deep learning-based bird classification tool, with pre-trained model, mobile and command line applications. *EcoEvoRxiv*. https://doi.org/10.32942/X2FP6T

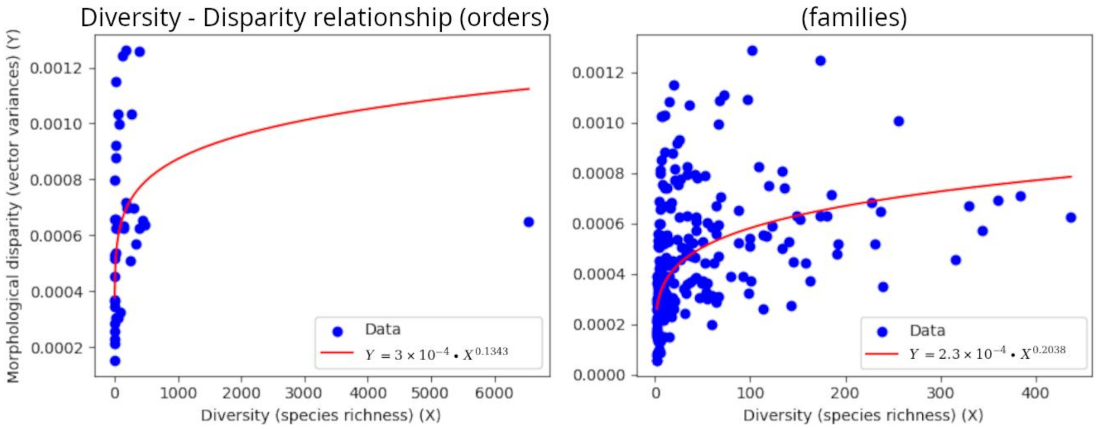
Table 1. All correlation coefficients and corresponding p-values calculated in the analysis, including the Pearson correlation coefficient for linear, log-linear and power-law models, as well as Spearman’s rank correlation coefficient.

Types of correlation coefficient		Order level		Family level	
		coefficient	p-value	coefficient	p-value
Including monospecific taxa	Spearman’s	0.6043	$1.4*10^{-5}$	0.7045	$1.1*10^{-39}$
	Pearson (linear)	0.0810	0.60	0.3770	$4.5*10^{-10}$
	Pearson (log-linear)	0.5757	$4.3*10^{-5}$	0.6502	$3.7*10^{-32}$
	Pearson (power-law)	N/A (logarithm is undefined for 0)			
Excluding monospecific taxa	Spearman’s	0.5110	$6.4*10^{-4}$	0.5164	$3.5*10^{-16}$
	Pearson (linear)	0.0540	0.74	0.2958	$9.3*10^{-6}$
	Pearson (log-linear)	0.4607	$2.4*10^{-3}$	0.4541	$1.9*10^{-12}$
	Pearson (power-law)	0.5144	$5.8*10^{-4}$	0.5190	$2.3*10^{-16}$

Figure 1. Morphological clustering result, taxonomically consistent branches are collapsed.

(See next page.)

Figure 2. The power-law fitted relationship functions between diversity (species richness) and morphological disparity (vector variances) on order and family levels.




Passeriformes (most species)
Struthionidae
Rheiformes
Apterygiformes
Casuariiformes
Galliformes (with Tinamiformes, Mesitornithidae, Menuridae, Cariamidae, Rhynochetidae, Turnicidae, *Coua*, *Carpococcyx*, *Neomorphus*, nested in)
Charadriiformes (Rostratulidae, *Scolopax*, *Gallinago*)
Gruiformes (Rallidae, Psophiidae and Sarothruridae)
Anseriformes(*Aix*), Gruiformes (Heliornithidae), Eurypygiformes
Gruiformes (*Fulica*, most species)
Gaviiformes
Podicipediformes
Anseriformes (Anatidae, most species)
Charadriiformes (most species, with Phoenicopteriformes nested in)
Phoenicopteriformes
Suliformes (Sulidae, Fregatidae)
Pelecaniformes (Pelecanidae)
Suliformes (Phalacrocoracidae, Anhingidae)
Procellariiformes
Sphenisciformes (with *Pinguinus* nested in)
Charadriiformes (Alcidae, with *Pelecanoides* nested in)
Galliformes (*Galloperdix lunulata*)， Columbiformes (*Petrophassa*)
Pterocliformes
Charadriiformes (Glareolidae, Pluvianidae, *Vanellus*, and serverl other species)
Charadriiformes (Burhinidae)
Otidiformes
Pelecaniformes (Ardeidae)
Ciconiiformes
Pelecaniformes (Scopidae)
Pelecaniformes (Scopidae, Threskiornithidae, with *Aramus* nested in)
Gruiformes (Gruidae)
Pelecaniformes (Balaenicipitidae), Anseriformes (Anhimidae, Anseranatidae, *Plectropterus*)
Passeriformes (Hirundinidae)
Apodiformes (Apodidae, Hemiprocnidae)
Apodiformes (Trochilidae)
Caprimulgiformes *s. l.* (including Steatornithiformes, Nyctibiiformes, Podargiformes and Aegotheliformes)
Strigiformes
Accipitriformes (with Falconiformes nested in)
Coraciiformes (Todidae)
Passeriformes (Eurylaimidae)
Passeriformes (Chloropseidae, Vireolanius)
Piciformes (Megalaimidae)
Piciformes (Ramphastidae, Semnornithidae, and part of Capitonidae)
Psittaciformes
Columbiformes
Cuculiformes (72 species)
Some species with no obvious pattern
Musophagidae (*Corythaeola*, *Tauraco*, *Ruwenzorornis*, *Musophaga*)
Columbiformes (5 species)
Cuculiformes (3 species)
Cuculiformes (49 species)
Phoeniculidae
Bucerotiformes (Bucerotidae, Bucorvidae)
Passeriformes (Buphagidae)
Coliiformes
Musophagidae (*Corythaixoides*, *Crinifer*)
Piciformes (Lybiidae, Indicatoridae, Picidae, and part of Capitonidae)
Coraciiformes (most species, with suborder Galbuli nested in)
Trogoniformes
Passeriformes (Pittidae, with Brachypteraciidae nested in)