

Historical Biology

An International Journal of Paleobiology

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/ghbi20>

Quantifying plant mimesis in fossil insects using deep learning

Li Fan, Chunpeng Xu, Edmund A. Jarzembski & Xiaohui Cui

To cite this article: Li Fan, Chunpeng Xu, Edmund A. Jarzembski & Xiaohui Cui (2021): Quantifying plant mimesis in fossil insects using deep learning, *Historical Biology*, DOI: [10.1080/08912963.2021.1952199](https://doi.org/10.1080/08912963.2021.1952199)

To link to this article: <https://doi.org/10.1080/08912963.2021.1952199>



Published online: 16 Jul 2021.



Submit your article to this journal 



Article views: 33



View related articles 



View Crossmark data 



Quantifying plant mimesis in fossil insects using deep learning

Li Fan , Chunpeng Xu , Edmund A. Jarzembski  and Xiaohui Cui 

^aKey Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China; ^bState Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology and Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Nanjing, China; ^cUniversity of Chinese Academy of Sciences, Beijing, PR China

ABSTRACT

As an important combination of behaviour and pattern in animals to resemble benign objects, biological mimesis can effectively avoid the detection of their prey and predators. It at least dates back to the Permian in fossil records. The recognition of mimesis within fossil, however, might be subjective and lack quantitative analysis being only based on few fossils with limited information. To compensate for this omission, we propose a new method using a Siamese network to measure the dissimilarity between hypothetical mimics and their models from images. It generates dissimilarity values between paired images of organisms by extracting feature vectors and calculating Euclidean distances. Additionally, the idea of ‘transfer learning’ is adopted to fine-tune the Siamese network, to overcome the limitations of available fossil image pairs. We use the processed Totally-Looks-Like, a large similar image data set, to pretrain the Siamese network and fine-tune it with a collected mimetic-image data set. Based on our results, we propose two recommended image dissimilarity thresholds for judging the mimicry of extant insects (0–0.4556) and fossil insects (0–0.4717). Deep learning algorithms are used to quantify the mimicry of fossil insects in this study, providing novel insights into exploring the early evolution of mimicry.

ARTICLE HISTORY

Received 5 June 2021

Accepted 1 July 2021

KEYWORDS

Mimesis; fossil insects; similarity; deep learning; Siamese network

Introduction

Animals have evolved various strategies to gain advantages among predator-prey interactions, of which mimesis is an effective way to facilitate stealth and survival (Wickler 1968). As a very sophisticated coevolutionary development in animals, mimesis is capable of concealing themselves by mimicking aspects of their surroundings. This adaptation is rather common among insects, the most diverse group of animals on Earth. They usually resemble other objects, especially plants, in their habitats by evolving several different specialised morphologies, e.g. katydids (Tettigonidae) and leaf insects (Phasmatodea) usually mimic leaves with their leaf-like wings or extended tergites (Mugleston et al. 2016), and stick insects (Phasmatodea) normally mimic sticks with their slender abdomen (Footitt and Adler 2009). The plant mimicry indicates the complex associations among insects and plants. Thereby, exploring the original and early evolution of mimesis can provide more information about the coevolution between plants and insects, and reveal more information about the palaeoecology and palaeoclimate. However, the original and evolution of plant-mimesis are little known due to the limitation of fossil data and the incompleteness of fossil preservation. Though fossils have been reported with mimetic features from the Mesozoic (Heads 2008; Wang et al. 2012; Liu et al. 2018; Fang et al. 2020; Yang et al. 2021), these mimicry relationships have been determined without any mathematical analysis. Here, we try to measure mimicry in fossils using deep learning.

Actually, there are already some applications of artificial intelligence in biological mimicry and other fields (Li et al. 2019b). Wham et al. (2019) used a pretrained deep convolutional neural network (DCNN) (Krizhevsky et al. 2017) to quantify the perceptual similarity between bumblebee colour patterns. This neural network, usually for large-scale image recognition, is trained on 1.2 million

images to learn visual features including edges, textures and colours. It transforms the images into digital vector representations, and generates a perceptual distance metric according to the distance between these vectors. Compared with other pixel-based quantitative methods (Williams 2007), a data-abundant and trustworthy metric is proposed to quantify mimicry dynamics in this method, and it is less sensitive to subtle changes in the pattern location. Inspired by Wham et al. (2019), Ezray et al. (2019) calculated perceptual distances between each pair of bumblebee colour pattern graphs. Moreover, they implemented t-distributed stochastic neighbour embedding (t-SNE) (Van der Maaten and Hinton 2008) to visualise the distances. The t-SNE can visualise a high dimensional distance in two-dimension plots. By analysing many geographic distribution data of social bumblebee colour patterns across the United States, they revealed that there is a mimicry complex accompanied by a perceptual continuum in bumblebees. A novel method was used to quantify the fidelity of the bumblebee colour patterns, which provided more possibilities for quantification tasks in the mimicry. Cuthill et al. (2019) used a convolutional triplet neural network (Hoffer and Ailon 2015) to quantify the total visible phenotypic similarity between ventral and dorsal of 38 neotropical butterfly subspecies (*Heliconius erato* and *Heliconius melpomene*). In the training process, three images were used to make up a triplet, two of which are from the same subspecies and the other from different subspecies. When calculating Euclidean distances, the network made those images from the same subspecies closer, and those from different subspecies further away. Prior to this, the testing of evolutionary hypotheses was subjective. This work demonstrated that a phenotypic spatial embedding can be generated by deep learning, and quantitatively verified a crucial prediction of Müllerian mimicry (Müller 1879). De Solan et al. (2020) used deep learning to quantify the snake’s Batesian mimicry behaviour

(Bates 1863) whose actual frequency has remained largely unknown. They trained a DCNN to identify different venomous species plus a ‘foreign’ class in the Western Palearctic. When training the network, they used ‘transfer learning’ in which a pretrained Xception network based on ImageNet (Deng et al. 2009) was fine-tuned on the collected snake images. Then the DCNN was used to classify non-venomous snake images into venomous or ‘foreign’ species.

Currently, much of the researches are focused on studying mimicry behaviours of extant creatures by artificial intelligence. However, no corresponding research has been commenced on the measurement and analysis of mimicry among fossils, though the oldest fossil record of plant mimicry dates back to the Permian (Garrouste et al. 2016) by visual inspection. Our study explores a new method that uses a Siamese network to measure the dissimilarity between hypothetical mimics and their models from images.¹ Siamese network consists of two sub-networks with outputs connected together. During its training, the feature vectors of the two inputs are extracted by sub-networks, and distance between them is calculated by a final fully connected layer. A discriminative loss function is minimised in the training process, which makes the distance smaller for the image pair of the same class but bigger for that of different classes.

Additionally, to overcome the limitations of available fossil image pairs, we adopt the idea of ‘transfer learning’. The network is pretrained on the processed Totally-Looks-Like (TLL) (Rosenfeld et al. 2018) dataset and fine-tuned on part of the Living-Insect-Mimicry-Dataset (LIMD). We test it on extant insects and fossil insects respectively, and obtain two corresponding thresholds of dissimilarity values for their mimicry judgements. We quantify mimicry relationship among fossils in this study and recommend a dissimilarity value threshold (Fossil-threshold) which can be used to assist future studies.

Data and method

Data

For pretraining the network, we use the processed Totally-Looks-Like (TLL) as the pretraining data set. TLL is a data set with 6,016 image pairs which are considered very similar by naked eyes, but usually have low similarity in low-level features. The image pairs include various kinds of objects, scenes, patterns, animals, and

human faces. All images are in three image styles: sketches, cartoons and natural images, providing sufficient diversity and complexity for humans. Similar image pairs, with the same corresponding numbers are placed in left and right folders, respectively. Some examples of TLL in different categories are in Figure 1.

Actually, we refine the TLL as a pretraining data set, not all the image pairs in TLL being used. The face image pairs in TLL were eliminated, in which the similarity is mainly determined by the features of the eyes, nose and mouth. Its variable positions include head pose, occlusion, lighting conditions and facial expressions (Trigueros et al. 2018). However, mimicry in insects is closely related to morphology such as shapes (Mugleston et al. 2016). There are great differences in the two situations. Therefore, to make the network accurately extract the characteristics in images of insects and their plant model better, the face images were removed manually from the TLL data set. Finally, 1836 non-face image pairs, mainly including various styles of objects, animals, etc., were retained.

For fine-tuning the network, we introduce a new data set, called Living-Insect-Mimicry-Data set (LIMD) based on the iNaturalist (Ueda 2020), one of the world’s largest communities dedicated to exploring biodiversity. LIMD consists of photographs of extant insects and their model mimic plants, including four common mimicry-involved insect genera: *Uropyia* Staudinger, 1892, (Valkonen et al. 2014), *Doleschallia* Felder & Felder, 1860 (Suzuki et al. 2014), *Megaphasma* Caudell, 1903 & *Sipyloidea* Brunner von Wattenwyl, 1893 (Carlberg 1981), *Typophyllum* Serville, 1838 (Castner and Nickle 1995), and their model plants (Table 1). *Megaphasma* and *Sipyloidea* considered together because of their same familial attribution and mimicry. We give several classic insect mimicry illustrations in Figure 2 as examples. Among the 2118 image pairs, 1453 image pairs were used for Siamese network fine-tuning and 665 image pairs for testing.

In addition to 665 image pairs of LIMD, we also gather a Fossil-Insect-Mimicry-Data set (FIMD) as another test data set with 57 image pairs. The image pairs of fossil insects and their plant models are extracted from literatures. Three fossil mimic species are involved: *Juracimbrophlebia ginkgofolia* Wang et al. (2012), *Phyllochrysa huangi* Liu et al. (2018), and *Lichenipolystoechoes angustimaculatus* Fang et al. (2020) (Wang et al. 2012; Liu et al. 2018; Fang et al. 2020) (Table 2). We give examples of the FIMD in Figure 2.



Figure 1. Some examples of image pairs in different categories in the TLL data set.

Table 1. Image pairs of Living-Insects-Mimicry-Data set (LIMD).

Genus	For fine-tuning			For testing		
	Insects	Plants	Image pairs	Insects	Plants	Image pairs
<i>Uropyia</i>	14	29	406	9	18	162
<i>Doleschallia</i>	21	21	441	14	15	210
<i>Megaphasma</i> and <i>Sipyloidea</i>	24	13	312	17	9	153
<i>Typophyllum</i>	14	21	294	10	14	140

Image pairs of LIMD and FIMD are cut out from their background in Adobe Photoshop CC to avoid the influence of their surroundings and make the results more reliable before inputting. Additionally, parts of images are cut out to exclude the disturbance of internal irrelevant body parts and external irrelevant debris. Images of lichen thallus are reshaped to keep consistence with

wings of *Lichenipolystoechotes angustimaculatus* (Figure 2(l)). Comprehensively, we undertake some unified processing on the whole data set. Firstly, all images are converted to greyscale using Pytorch to ignore the colour features. The original colourations of fossils usually absent due to the preservation and unable to be an element for calculating dissimilarity (Martinez-Delclos et al. 2004; Vinther 2015), so we mainly consider their morphologies, e.g. shape and texture features. Meanwhile, the colouration patterns are remained after converted to greyscale, which also provide information for fossil mimicry. Secondly, to avoid overfitting, some data augmentation operations are implemented to enhance data diversity. We use Pytorch's built-in function (transforms) to perform a horizontal flip operation and random rotation of 15 degrees on images of the processed TLL and part of LIMD (images for fine-tuning). By adding different angles and directions, the network can learn more comprehensive features. Finally, each image is resized to

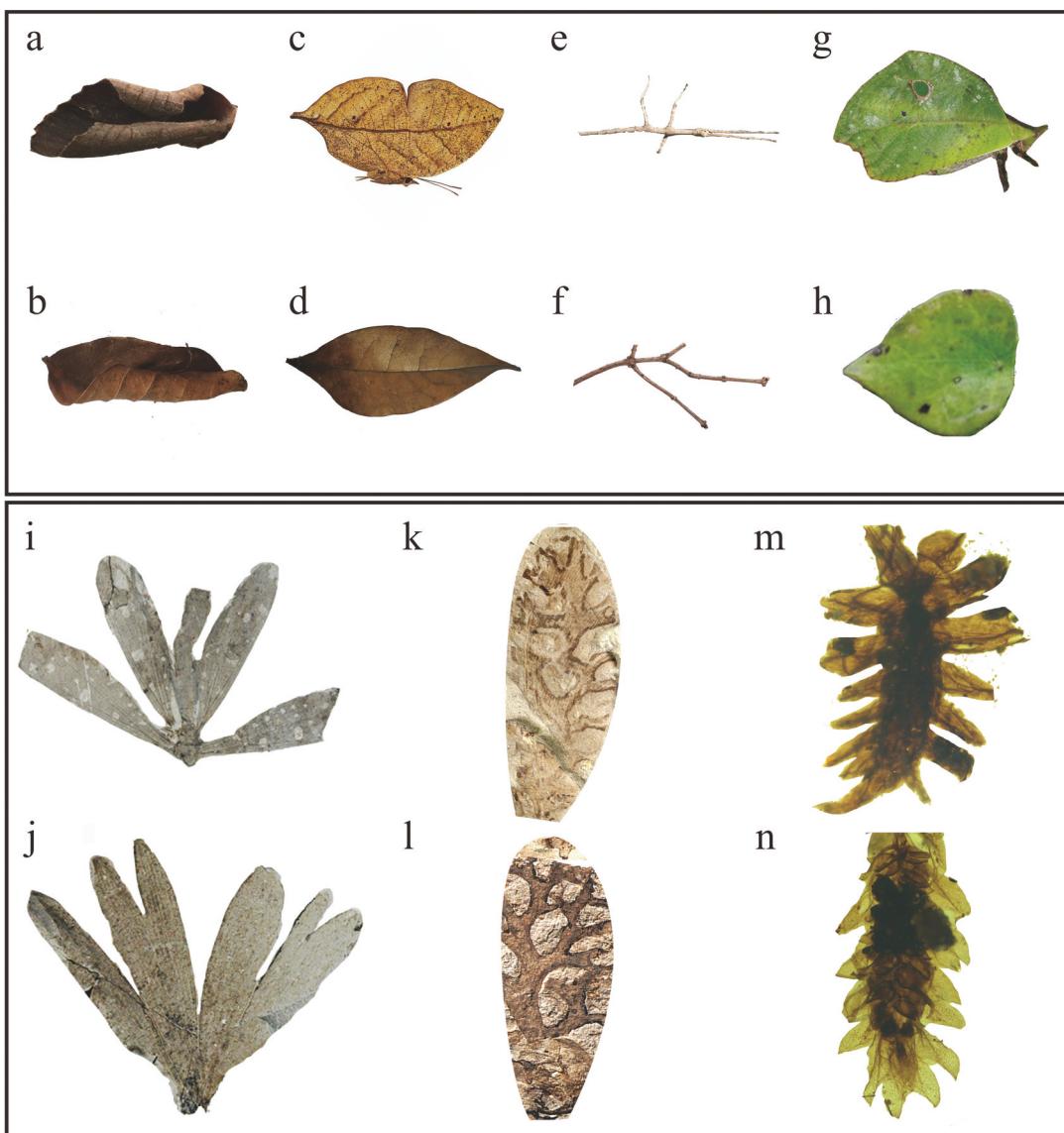


Figure 2. Image pairs of Living-Insects-Mimicry-Data set (LIMD), (a–h); and image pairs in Fossil-Insect-Mimicry-Data set (FIMD), i–n. (a, b) Lepidopteran genus *Uropyia* Staudinger, 1892 and its mimicry model plant. (c, d) Lepidopteran genus *Doleschallia* Felder & Felder, 1860 and its mimicry model plant. (e, f) Phasmatoidean genus *Megaphasma* Caudell, 1903 and *Sipyloidea* Brunner von Wattenwyl, 1893 and its mimicry model plant. (g, h) Orthopteran genus *Typophyllum* Serville, 1838 and its mimicry model plant. Images of fossil insects and their mimicry model plants. (i, j) Neuropteran species *Juracimbrophlebia ginkgofolia* Wang et al. 2012 and its mimicry model plant. (k, l) Neuropteran species *Lichenipolystoechotes angustimaculatus* Fang et al. 2020 and its mimicry model plant. (m, n) Neuropteran species *Phyllochrysa huangi* Liu et al. 2018 and its mimicry model plant.

Table 2. Image pairs of Fossil-Insect-Mimicry-Data set (FIMD).

Genus	Insects	Plants	Image pairs
<i>Juracimbrophlebia ginkgofolia</i>	3	5	15
<i>Phyllochrysa huangi</i>	4	9	36
<i>Lichenopolystoechotes angustimaculatus</i>	2	3	6

224 × 224 to fix the network input size and reduce the amount of calculation as well as the computer CPU load.

Transfer learning

Fine-tuning and transfer learning are crucial tools to solve the problem of insufficient training data in deep learning. In transfer learning (Yosinski et al. 2014; Agrawal et al. 2014), a CNN is firstly trained on a large data set for some general tasks and the trained network parameters are saved. After that, the network is transferred to another CNN model and trained on a smaller data set, which is usually domain-specific. Yin et al. (2017) studied the visualisation of CNN fine-tuning and transfer learning process. Their research reported a common phenomenon among the layers of many CNN models: the layer closer to the input data learns more common features, and learned features become specific with the raising of layers.

In this work, the constructed Siamese network is pretrained on the processed TLL data set and saved its network parameters. Then, the network is fine-tuned on the part of LIMD to obtain the similarity measurement model for our research. We test two different fine-tuning strategies: (1) fine-tuning the whole model; (2) training only the top new layer. We find that strategy (2) gets the best performance consistently in the most resource-efficient situation. And some studies (Yosinski et al. 2014; Agrawal et al. 2014) show that fine-tuning the whole model with less data causes overfitting easily. Therefore, similar to the traditional transfer learning process, there are two main steps. Firstly, a new top model² is initialised with random weights and trained by inputting the

features extracted from the pretrained CNN basic model. At this point, the filter of the convolutional layer is not updated. The second step is to combine the pretrained basic model with the top model and update the entire network weights during the training process.

Siamese network

We try to find a mapping function which maps the inputs to a low-dimensional target space. This is convenient to approximate the ‘semantic’ distance in the original input space with simple distance. Additionally, the natural distance in low-dimensional target space is not affected by the input irrelevant distortion. Therefore, we could easily estimate each new category’s probability model from very few samples.

Herein, we use Siamese network, which is the most commonly used network in similarity measurement (Koch et al. 2015; Melekhov et al. 2016; Appalaraju and Chaoji 2017). Its process of calculating distance is shown in Figure 3. The reasons for choosing Siamese network in this work are as follows:

- (1) The network takes sample pairs as input. The matching process of samples increases the size of data set, which makes it suitable for training on small data sets.
- (2) The network weakens labels by only considering whether the sample pairs are similar or not. This enhances the extensibility of the network to deal with untrained categories.
- (3) The contrastive loss function used in the network helps to process paired data. In this paper, the mimicry is simplified as is the relationship between insect-plant image pairs.
- (4) In this network, two CNNs with shared parameters are used to extract image feature vectors. Compared with using complex networks such as ResNet (He et al. 2016) or GoogleNet (Szegedy et al. 2015) etc. alone, two basic three-layer CNNs

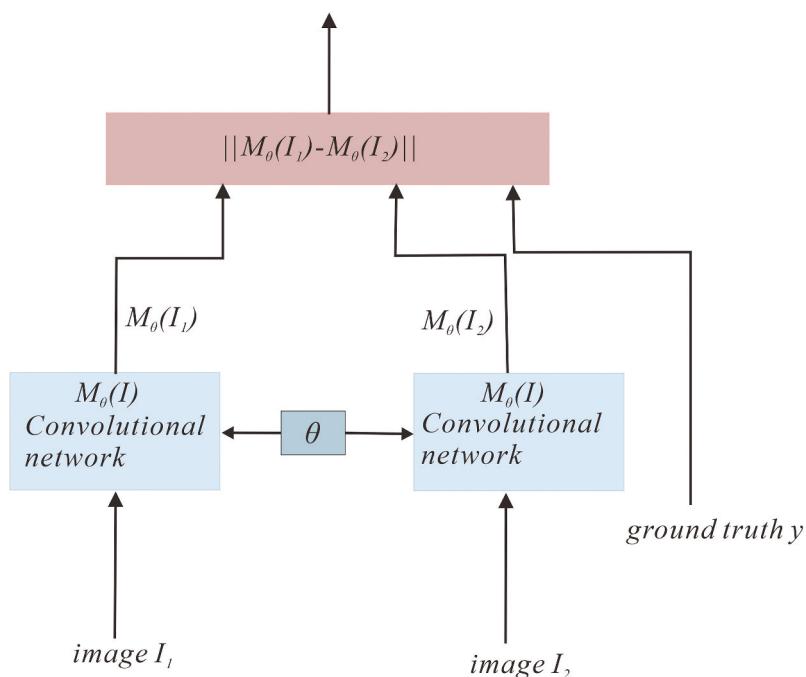


Figure 3. The distance calculating of Siamese network. $M_\theta(I_1)$ and $M_\theta(I_2)$ are the low-dimensional embedding vectors got by projection; Loss is based on Euclidean distance of the embedding vector; when $y = 1$, I_1 and I_2 are similar image pair; when $y = 0$, I_1 and I_2 are dissimilar image pair. The specific structure of mapping process $M_\theta(I)$ is shown in Figure 4.

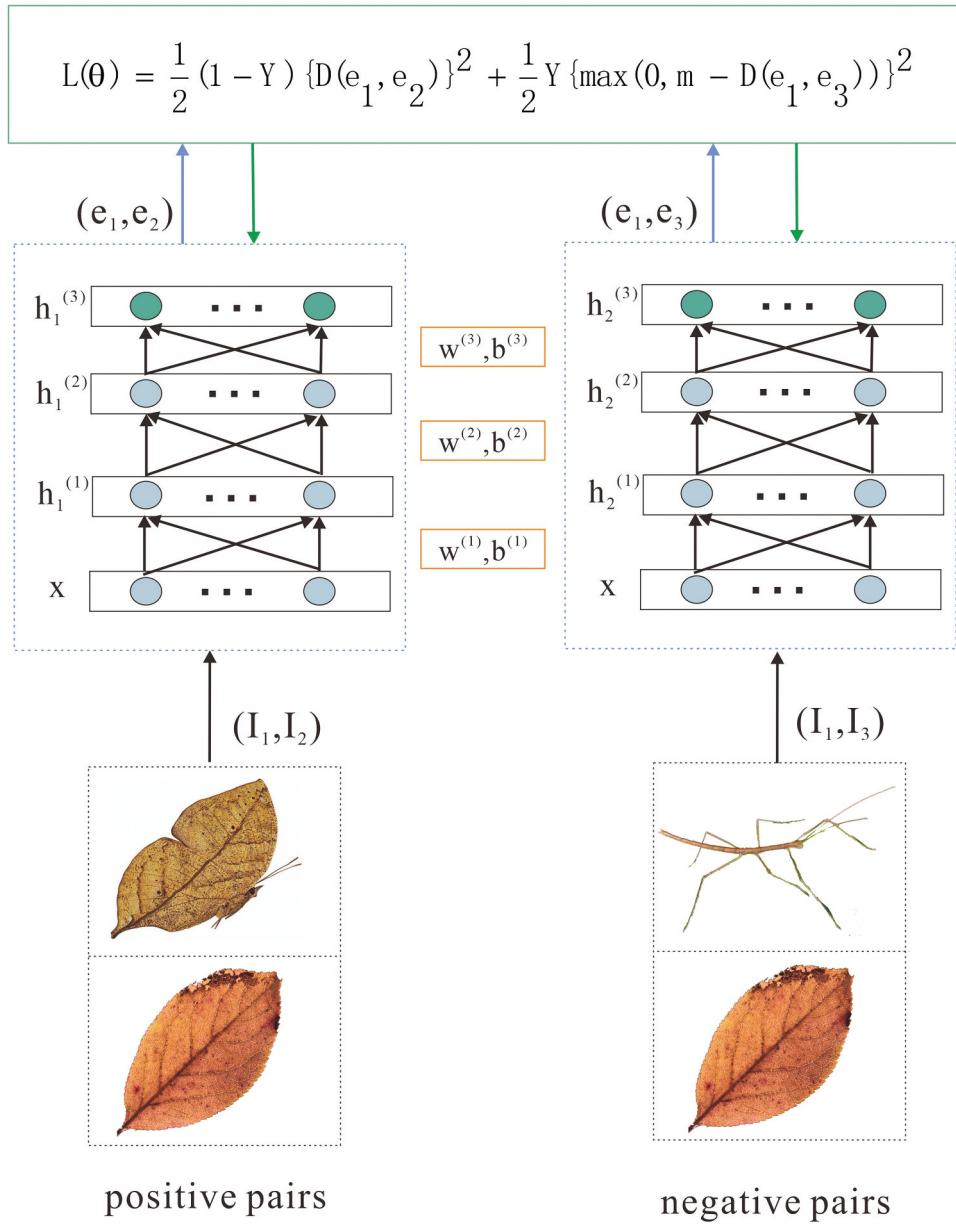


Figure 4. Specific architecture of mapping process for generating image embedding vectors based on Siamese network. The top is calculation method of the loss function.

sharing parameters can greatly reduce the number of parameters.

In neural networks, the goal of learning is to find the parameter θ (Usually contains weight matrix W and bias vector b) which minimises the loss function. Following the previous work (Chopra et al. 2005), the discriminative learning framework of the energy-based model (EBM) is used to derive the loss function. If we only consider minimising the energy $E_\theta(e_1, e_2)$ of the image pairs from the same category, when the mapping function f is a constant function, it may happen that the energy and loss of the input from the same category become 0 and the different category is very large. Therefore, a contrastive loss function can not only ensure that energies for pairs of inputs from different categories are high but also that energies for pairs from same category are small. In this paper, we fix a specific loss function on this basis (Hadsell et al. 2006). See Equation (1).

$$L(\theta) = \frac{1}{2}(1-y)\{D(e_1, e_2)\}^2 + \frac{1}{2}y\{\max(0, m - D(e_1, e_3))\}^2 \quad (1)$$

where y is the label of the image pairs, e is the extracted feature vector or image embedding, and when they are similar (e_1, e_2), $y = 0$, the right-hand addition part disappears; the loss function then becomes the distance between two similar embedded images. When they are dissimilar (e_1, e_3), $y = 1$, the left-hand addition part disappears; the loss function becomes the hinge loss. The idea of contrastive loss is also applied in generative adversarial network, which is a hot topic in data synthetics (Li et al. 2019a; Li et al. 2021). The m is the margin between similar and dissimilar images whose value is empirically decided. A larger m can push dissimilar and similar images further apart. In our work, we have used $m = 2$.

As shown in Figure 4, the inputs of this network are image pairs. They may be positive image pairs (I_1, I_2) which are obtained through data augmentation or two different variations of the same category image. Or they may be negative image pairs (I_1, I_3) which

are from different categories. The two convolutional neural networks are used to map I_1, I_2, I_3 to the latent space to get the image embedding e_1, e_2, e_3 . The network is multilayered (M) and each layer has multiple neurons (n). For a given image, the specific process of mapping is as follows: Assuming that the input x of the projection function f is d-dimensional, the m-th hidden layer outputs $h^m = f(w^m x + b^m)$, where w^m is the weight matrix of the m-th layer, and b^m is the bias vector. Here, f is a non-linear activation function that projects a d-dimensional image to a p-dimensional subspace. In this p-dimensional subspace, the similar images are closer, and dissimilar images are further away.

Our model is implemented with Python 3.6 based on Pytorch as a Jupyter Notebook. A standard CNN architecture is applied and batch normalisation is used after each convolution layer. The network structure includes three convolutional layers with a convolution kernel size of five and three fully connected layers. Though the Siamese network is composed of two symmetrical shared parameter networks, their weights are restricted to be identical. Therefore, we use one network in the actual training process, referring to the previous work (Koch et al. 2015). This can greatly save memory without affecting the accuracy and other indicators. Specifically, two images are inputted to one model in succession. They are used for calculating the loss function value, which can be back-propagated to optimise the network. The loss function used in the training process is identical to Equation (1), where D is defined as the Euclidean distance of the network output e for two images. See Equation (2)

$$D(e_1, e_2) = \sqrt{e_1 - e_2} \quad (2)$$

Since our network structure inputs an image pair, along with their labels (similar or dissimilar), the training data needs to be imported. The images are loaded by reading from folders, which is easy to generalise to any other data sets. To prevent the imbalance between similar image pairs and dissimilar image pairs in the imported data, we force 50% of the image pairs to come from the same category and the others from different categories. In other words, the ratio of positive and negative pairs is 1 in training.

When applying ‘transfer learning’, there are two steps: pretraining and fine-tuning. During pretraining, the network trains 100 epochs on the loaded training set, using Adam optimiser, whose learning rate is 0.001. The batch_size of pretraining is 64. After training, we draw a graph of the loss function decline curve of the training process and save the network parameters. When fine-tuning, the network parameters are imported firstly, with 30 epochs trained on the fine-tuning data set later. To lessen the damage to the saved basic network parameters, we reduce the learning rate to 0.0005. The batch_size is 32 in fine-tuning.

Results and discussions

Quantification of extant insect mimesis

Extant insects in the Living-Insects-Mimicry-Data set (LIMD) are definite mimicry insects, and their mimicry behaviour is recognised in the biological community (Carlberg 1981; Castner and Nickle 1995; Suzuki et al. 2014; Valkonen et al. 2014). Therefore, we use them as a test data set, to test the dissimilarity measurement of our model on mimicry behaviour.

We use 665 image pairs of the LIMD to test our model for the quantification of extant insect mimicry (Table 1). We calculate the dissimilarity values of all possible image pairs to measure the mimicry of each insect genus separately (Table 3). After that, all dissimilarity values of each insect genus are averaged to obtain four reference values for their mimicry judgement: *Uropyia*, 0.7953;

Table 3. Test results of extant insects.

Genus	Segmented statistics			
	[0,0.5]	[0.5,1]	[1,+∞]	Average value
<i>Uropyia</i>	59	50	53	0.7953
<i>Doleschallia</i>	111	79	20	0.5142
<i>Megaphasma</i> and <i>Sipyloidea</i>	103	48	2	0.3970
<i>Typophyllum</i>	123	17	0	0.2481
Median (Threshold)			0.4556	

Segment statistics is to count the number of dissimilarity values corresponding to each interval.

Doleschallia, 0.5142; *Typophyllum*, 0.2481; *Megaphasma* and *Sipyloidea*, 0.3970. From a statistical point of view, the median can reflect the general situation of a set of numbers. To get a comprehensive similarity threshold for the mimicry judgements, the median of four average values is taken as the final threshold. Thus, our recommended threshold of dissimilarity value for extant insets (Extant-threshold) is 0–0.4556. It will also serve as a basis for determining the mimesis of fossil insects in the next part.

Additionally, to grasp potential mimicry relationship as more as possible, we also recommend a generalised threshold of dissimilarity value for extant insets, 0–1. In Table 3, to more intuitively reflect the distribution range of dissimilarity values within each insect genus, we make a segmented statistics. We take 0.5 as an interval and unified those higher (over 1 into an interval), and calculate numbers of dissimilar values in each interval. From the distribution of the dissimilarity value of each insect genus, over 88% of them are less than 1 (Table 3). Within the four taxa, *Uropyia*, *Doleschallia*, *Typophyllum*, *Megaphasma* and *Sipyloidea*, about 65%, 90%, 100% and 98% of dissimilarity values are less than 1. Therefore, to generalise it to all organisms better, we propose that the biological image pairs with a dissimilarity value less than 1 can be further considered for mimicry. To prove the applicability of our recommend generalised threshold, we calculate the dissimilarity values between each insect genus and any other irrelevant plant. The dissimilarity values within mimicry image pairs are less than 1 (Figure 5(a)); however, dissimilarity values within irrelevant image pairs are greater than 1 (Figure 5(b)). The model can thus numerically distinguish similar or dissimilar image pairs by extracting the feature vectors and calculating the distance between the vectors, thereby tentatively determining whether there is a mimicry relationship between the two imaged organisms.

Quantification of fossil insect mimesis

All image pairs from FIMD are fossil insects and their potential model plants, which are artificially identified with mimicry relationship based on palaeontologists’ experience (Wang et al. 2012; Liu et al. 2018; Fang et al. 2020). We quantify the mimicry relationship among fossils and recommend a dissimilarity value threshold (Fossil-threshold) that can be used to assist future studies.

First, we test all image pairs (57) of the FIMD by our model for the quantification of fossil insect mimicry (Table 2). We calculate the dissimilarity values of all image pairs to measure the mimicry of each insect species separately (Table 4). After that, the average dissimilarity values of each fossil insect species are calculated: *Juracimbrophlebia ginkgofolia*, 5936; *Phyllochrysa huangi*, 0.3684; *Lichenopolystoechotes angustimaculatus*, 0.4531. Taking account of the incompleteness of fossils, together with previous palaeontological research, all these fossil species can be considered as plant mimicry. The differences between each average dissimilarity value and the Extant-threshold are calculated as: *Juracimbrophlebia ginkgofolia*,

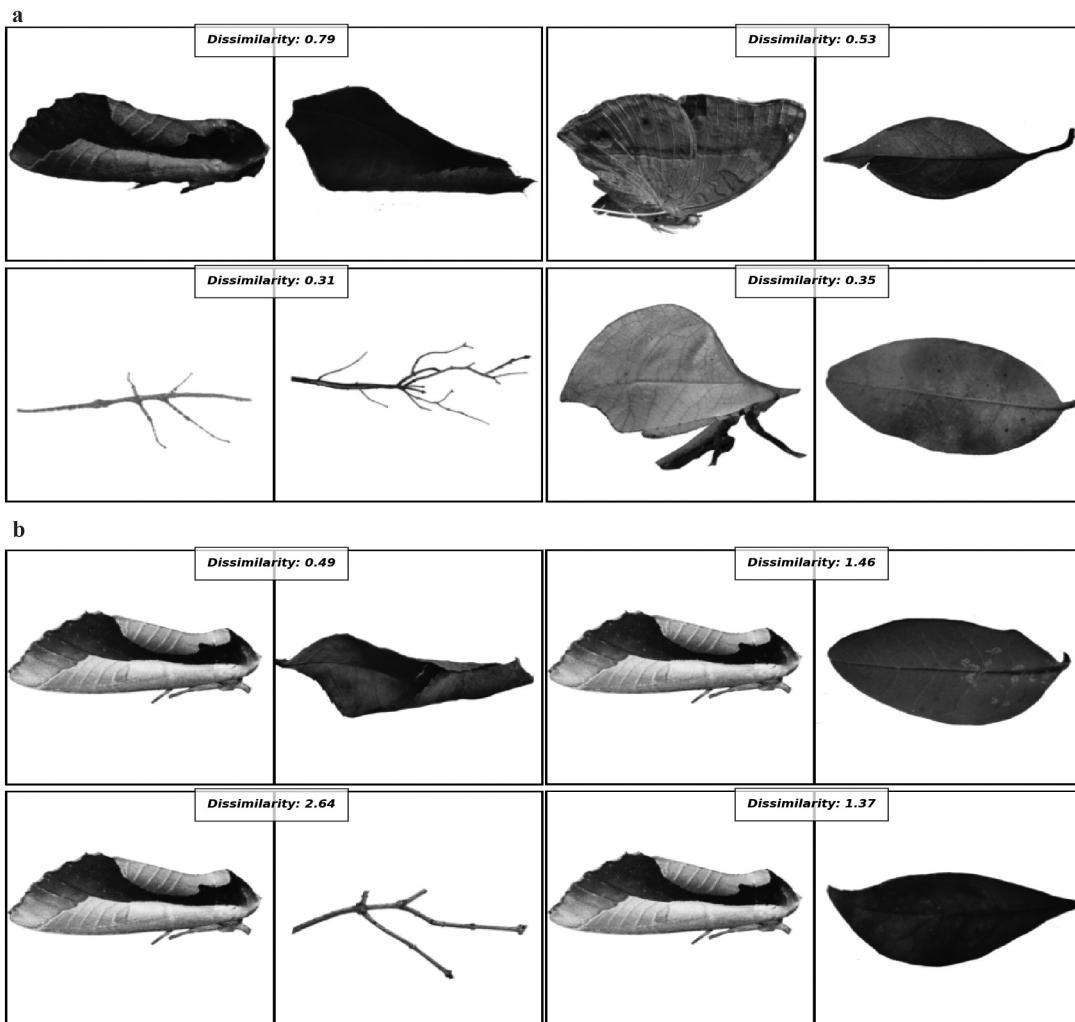


Figure 5. (a) Dissimilarity values between each insect (left) and its mimicked plant (right), whose dissimilarity values are about the average values. From left to right, from top to bottom, there are *Uropyia*, *Doleschallia*, *Megaphasma* and *Sipyloidea*, *Typophyllum*. (b) Dissimilarity values between *Uropyia* (left) and its mimicry plant and model plants (right).

Table 4. Test results of fossil insects.

Genus	Segmented statistics				Difference
	[0,0.5]	[0.5,1]	[1, + ∞]	Average value	
<i>Juracimbrophlebia ginkgofolia</i>	7	5	3	0.5936	0.1380
<i>Phyllochrysa huangi</i>	25	10	1	0.3684	-0.0872
<i>Lichenipolystoechotes angustimaculatus</i>	4	1	1	0.4531	-0.0025
Average difference	0.0161				
Threshold	0.4717				

The difference is obtained by subtracting the dissimilarity threshold of extant insects from the average value of fossils.

0.1380; *Phyllochrysa huangi*, -0.0872; and *Lichenipolystoechotes angustimaculatus*, -0.0025. Finally, we derive the final threshold for judging the mimicry of fossil insects (Fossil-threshold), 0–0.4717, by adding the average of above differences (0.0161) to Extant-threshold (0–0.4556). The detailed statistical analysis results are shown in Table 4. Figure 6 shows examples of the dissimilarity measurement results within mimicry fossils and irrelevant fossils.

Finally, all the experimental results are analysed quantitatively by box scatter plot (Figure 7). Box scatter plot can clearly reflect the discrete distribution of data and the outliers. There is no significant

difference between the average dissimilarity of fossil insects and that of living insects. The data points of each insect are relatively concentrated, and there are almost no outliers.

Conclusion

In this study, we develop a symmetric CNN model-Siamese network for the quantification of extant and fossil insect mimicry. It achieves the consistency with the artificial judgement based on biological morphologies, both involving mimicry in extant and fossil insects. We draw the dissimilarity threshold for judging extant insect mimicry, 0–0.4556. Additionally, images of fossil insects are also tested and our CNN model-Siamese network concurs with previous researches. The recommended threshold for fossil insect mimicry is 0–0.4717, slightly higher than extant insects due to the incompleteness of fossils. However, our model still has certain errors and its generalisation is not good enough due to the limitation of data sets. As long as enough data are available, it should be possible to achieve a good model fine-tuning effect. Therefore, as data become available, we will enlarge the range of the calculated recommended threshold to minimise the omission of biological mimicry behaviour.

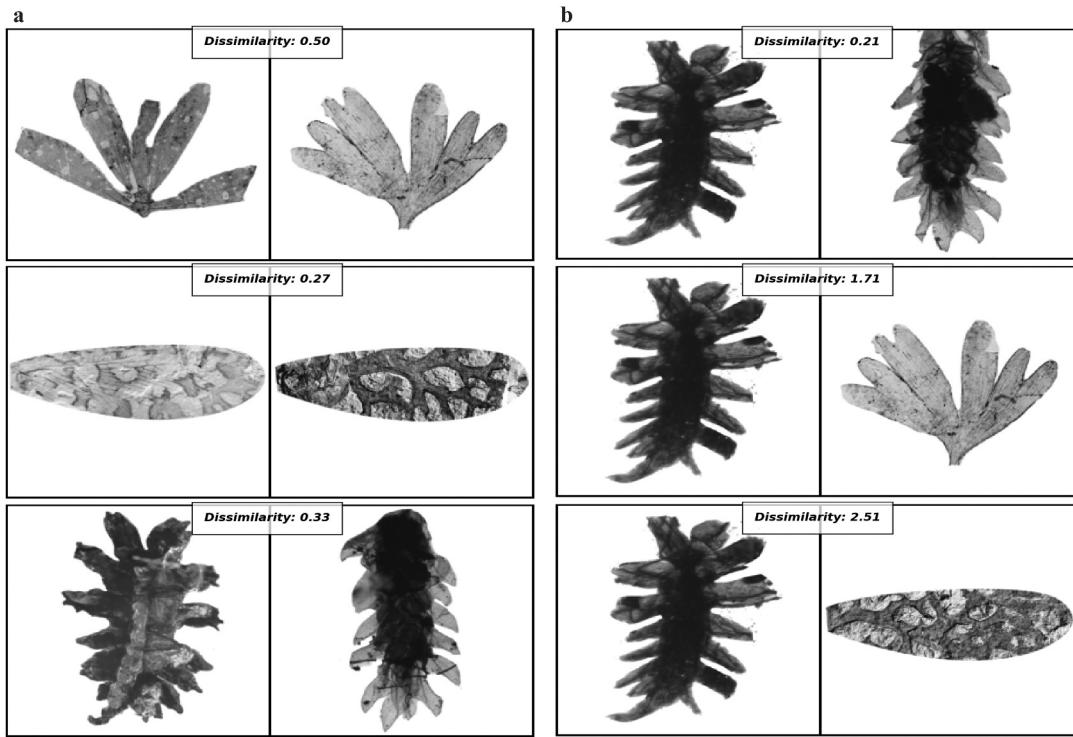


Figure 6. (a) Dissimilarity values between each fossil insect mimic (left) and its model plant (right), whose dissimilarity values are about the average values. From left to right, from top to bottom, the insects are *Juracimbrophlebia ginkgofolia*, *Lichenipolystoechotes angustimaculatus*, and *Phyllochrysa huangi*. (b) Dissimilarity values between *Phyllochrysa huangi* (left) and its model plant and irrelevant plants (right).

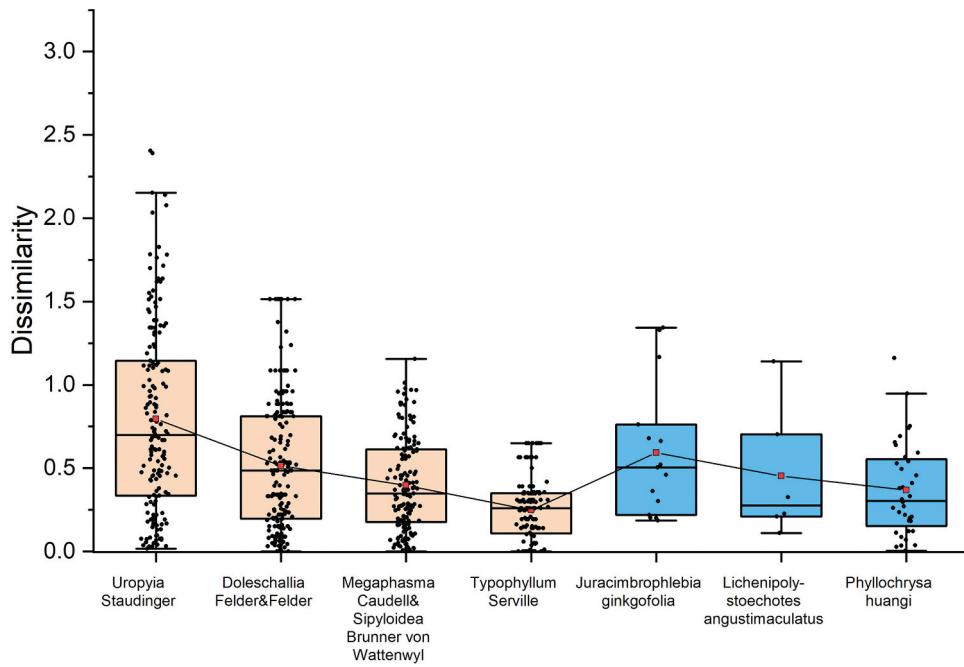


Figure 7. Box scatter plots of dissimilarity data display lower and upper extremes, lower and upper quartile, medians (black lines in the box) and averages (red dots) of dissimilarities of all the insects. (Black dots represent the dissimilarity values between insects and plants).

The definition of mimesis within fossils always relies on subjective judgements. Our model quantifies the mimesis among fossil insects and their model plants and calculates a tentative threshold for the judgement of mimicry of fossil insects. The threshold achieves the transformation from qualitative to quantitative

judgment and provides a starter for future studies. Compared with biologists' time-consuming and exhausting judgement of mimicry, our model can provide a preliminary judgement quickly and effectively to assist palaeontologists' work, by measuring the mimicry similarity from suggestive images.

Notes

1. Data and code are available on <https://github.com/fanliaveline/Siamese-Network-Fossil-Mimicry.git>
2. We refer to 'basic model' as a pretrained architecture normally; 'top model' as a similarity calculation network and fully connected layer; and 'convolutional layer' as the initial modules.

Acknowledgments

We appreciate the Nanjing Institute of Geology and Palaeontology for the data support and the Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, Wuhan University, China for the experimental equipment support.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB26000000), the Second Tibetan Plateau Scientific Expedition and Research (2019QZKK0706), and the National Natural Science Foundation of China (41688103). E.A.J. thanks the Chinese Academy of Sciences for the support under the President's International Fellowship Initiative (PIFI). This is a Leverhulme Emeritus Fellowship Contribution for E.A.J.

ORCID

Li Fan  <http://orcid.org/0000-0002-2647-9019>

References

- Agrawal P, Girshick R, Malik J 2014. Analyzing the performance of multilayer neural networks for object recognition. European conference on computer vision; Sep 6-12; Zurich, Switzerland, Springer. p. 329–344.
- Appalaraju S, Chaoji V. 2017. Image similarity using deep CNN and curriculum learning. arXiv preprint arXiv:170908761.
- Bates HW. 1863. Contributions to an insect fauna of the amazon valley Coleoptera: longicornes. Ann Mag Nat Hist. 12(71):367–381. doi:[10.1080/0022936308681538](https://doi.org/10.1080/0022936308681538).
- Carlberg U. 1981. Defensive behaviour in females of the stick insect *Styloidea siyulus* (Westwood) (Phasmida). Zool Anz. 207:177–180.
- Castner JL, Nickle DA. 1995. Notes on the biology and ecology of the leaf-mimicking katydid *Typophyllum bolivari* vignon (Orthoptera: Tettigoniidae: Pseudophyllinae: Pterochrozinii). J Orth Res. 4:105–109. doi:[10.2307/3503465](https://doi.org/10.2307/3503465).
- Chopra S, Hadsell R, LeCun Y 2005. Learning a similarity metric discriminatively, with application to face verification. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); Jun 20-25; San Diego, CA, USA; Vol. 1. IEEE. p. 539–546.
- Clune J, Bengio Y, Lipson H. 2014. How transferable are features in deep neural networks? Advances in neural information processing systems; Dec 8-13; Montreal, Quebec, Canada; p. 3320–3328
- Cuthill JFH, Guttenberg N, Ledger S, Crowther R, Huertas B. 2019. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. Sci Adv. 5(8):eaaw4967. doi:[10.1126/sciadv.aaw4967](https://doi.org/10.1126/sciadv.aaw4967).
- De Solan T, Renault JP, Geniez P, David P, Crochet PA. 2020. Looking for mimicry in a snake assemblage using deep learning. Am Nat. 196(1):74–86. doi:[10.1086/708763](https://doi.org/10.1086/708763).
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L 2009. Imagenet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; Jun 20-25; Miami, FL, USA; Ieee. p. 248–255.
- Ezray BD, Wham DC, Hill CE, Hine HM. 2019. Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum. Proc R Soc B. 286(1910):20191501. doi:[10.1098/rspb.2019.1501](https://doi.org/10.1098/rspb.2019.1501).
- Fang H, Labandeira CC, Ma Y, Zheng B, Ren D, Wei X, Liu J, Wang Y. 2020. Liche mimesis in mid-mesozoic lacewings. eLife. 9:e59007. doi:[10.7554/eLife.59007](https://doi.org/10.7554/eLife.59007).
- Footit RG, Adler PH. 2009. Insect biodiversity. Wiley-Blackwell. Oxford.
- Garrouste R, Hugel S, Jacquelin L, Rostan P, Steyer JS, Desutter-Grandcolas L, Nel A. 2016. Insect mimicry of plants dates back to the Permian. Nat Commun. 7(1):1–6. doi:[10.1038/ncomms13735](https://doi.org/10.1038/ncomms13735).
- Hadsell R, Chopra S, LeCun Y 2006. Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06); Jun 17-22; New York, NY, USA; IEEE. Vol. 2. p. 1735–1742.
- He K, Zhang X, Ren S, Sun J 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; Jun 17-19; San Juan, PR, USA; IEEE. p. 770–778.
- Heads SW. 2008. The first fossil Proscopiidae (Insecta, Orthoptera, Eumastacoidea) with comments on the historical biogeography and evolution of the family. Palaeontology. 51(2):499–507. doi:[10.1111/j.1475-4983.2008.00756.x](https://doi.org/10.1111/j.1475-4983.2008.00756.x).
- Hoffer E, Ailon N 2015. Deep metric learning using triplet network. International Workshop on Similarity-Based Pattern Recognition. OCT 12-14; Copenhagen, Denmark; Springer. p. 84–92.
- Koch G, Zemel R, Salakhutdinov R 2015. Siamese neural networks for one-shot image recognition. Proceedings of the 32nd International Conference on Machine Learning, ICML deep learning workshop; Jul 6-11; Lille, France; Vol. 2.
- Krizhevsky A, Sutskever I, Hinton GE. 2017. Imagenet classification with deep convolutional neural networks. Commun ACM. 6:84–90. doi:[10.1145/3065386](https://doi.org/10.1145/3065386).
- Li W, Ding W, Sadasivam R, Cui X, Chen P. 2019a. His-GAN: a histogram-based GAN model to improve data generation quality. Neural Netw. 119:31–45. doi:[10.1016/j.neunet.2019.07.001](https://doi.org/10.1016/j.neunet.2019.07.001).
- Li W, Liang Z, Ma P, Wang R, Cui X, Chen P. 2021. Hausdorff GAN: improving GAN Generation Quality with Hausdorff Metric. IEEE Trans Cybern. doi:[10.1109/TCYB.2021.3062396](https://doi.org/10.1109/TCYB.2021.3062396)
- Li W, Liu X, Liu J, Chen P, Wan S, Cui X. 2019b. On improving the accuracy with auto-encoder on conjunctivitis. Appl Soft Comput. 81:105489. doi:[10.1016/j.asoc.2019.105489](https://doi.org/10.1016/j.asoc.2019.105489).
- Liu X, Shi G, Xia F, Lu X, Wang B, Engel MS. 2018. Liverwort mimesis in a cretaceous lacewing larva. Curr Biol. 28(9):1475–1481. doi:[10.1016/j.cub.2018.03.060](https://doi.org/10.1016/j.cub.2018.03.060).
- Martinez-Delclos X, Briggs DE, Penalver E. 2004. Taphonomy of insects in carbonates and amber. Palaeogeogr Palaeoclimatol Palaeoecol. 203 (1v2):19–64. doi:[10.1016/S0031-0182\(03\)00643-6](https://doi.org/10.1016/S0031-0182(03)00643-6).
- Melekho I, Kannala J, Rahtu E 2016. Siamese network features for image matching. 2016 23rd International Conference on Pattern Recognition (ICPR). Dec 4-8; Cancún, Mexico; IEEE. p. 378–383.
- Mugleston J, Naegle M, Song H, Bybee SM, Ingley S, Suvorov A, Whiting MF. 2016. Rein-venting the leaf: multiple origins of leaf-like wings in katydids (Orthoptera: Tettigoniidae). Invertebr Syst. 30(4):335–352. doi:[10.1071/IS15055](https://doi.org/10.1071/IS15055).
- Müller F. 1879. Ituna and thyridia; a remarkable case of mimicry in butterflies (transl. byralph meldola from the original German article in kosmos, may 1879, 100). Tra Entomol Soc Lond. 1879: xx-xxix.
- Rosenfeld A, Solbach MD, Tsotsos JK 2018. Totally looks like-how humans compare, compared to machines. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Jun 18-22; Salt Lake City, UT, USA; p. 1961–1964.
- Suzuki TK, Tomita S, Sezutsu H. 2014. Gradual and contingent evolutionary emergence of leaf mimicry in butterfly wing patterns. BMC Evol Biol. 14 (1):1-13. doi:[10.1186/s12862-014-0229-5](https://doi.org/10.1186/s12862-014-0229-5).
- Szegedy C, Liu W, Jia SP, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. 2015. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. Jun 7-12; Boston, MA, USA; IEEE Computer Society; p. 1–9.
- Trigueros DS, Meng L, Hartnett M. 2018. Face recognition: from traditional to deep learning methods. arXiv preprint arXiv:181100116.
- Ueda K. 2020. Occurrence dataset; [2020 inaturalist research-grade observations.]. [accessed 2020 Dec 10]. doi:[10.15468/ab3s5x](https://doi.org/10.15468/ab3s5x).
- Valkonen JK, Nokelainen O, Jokimäki M, Kuusinen E, Paloranta M, Peura M, Mappes J. 2014. From deception to frankness: benefits of ontogenetic shift in the anti-predator strategy of alder moth *acronicta alni* larvae. Curr Zool. 60 (1):114–122. doi:[10.1093/czoolo/60.1.114](https://doi.org/10.1093/czoolo/60.1.114).
- Van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. J Mach Learn Res. 9(Nov):2579–2605.
- Vinther J. 2015. A guide to the field of palaeo colour: melanin and other pigments can fossilise: reconstructing colour patterns from ancient organisms can give new insights to ecology and behaviour. BioEssays. 37 (6):643–656. doi:[10.1002/bies.201500018](https://doi.org/10.1002/bies.201500018).
- Wang Y, Labandeira CC, Shih C, Ding Q, Wang C, Zhao Y, Ren D. 2012. Jurassic mimicry between a hangingfly and a ginkgo from China. Proc Natl Acad Sci U S A. 109(50):20514–20519. doi:[10.1073/pnas.1205517109](https://doi.org/10.1073/pnas.1205517109).

- Wham DC, Ezray BD, Hines HM. 2019. Measuring perceptual distance of organismal color pattern using the features of deep neural networks. *bioRxiv*. 736306.doi:10.1101/73606
- Wickler W. 1968. Mimicry in plants and animals. New York: McGraw Hill.
- Williams P. 2007. The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. *Biol J Linn Soc*. 92(1):97–118. doi:[10.1111/j.1095-8312.2007.00878.x](https://doi.org/10.1111/j.1095-8312.2007.00878.x).
- Yang H, Shi C, Engel MS, Zhao Z, Ren D, Gao T. 2021. Early specializations for mimicry and defense in a Jurassic stick insect. *Natl Sci Rev*. 8(1):nwaa056. doi:[10.1093/nsr/nwaa056](https://doi.org/10.1093/nsr/nwaa056).
- Yin X, Chen W, Wu X, Yue H 2017. Fine-tuning and visualization of convolutional neural net-works. 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). Jun 18–22; Siem Reap, Cambodia. p. 1310–1315.
- Yosinski J, Clune J, Bengio Y, Lipson H. 2014. How transferable are features in deep neural neural networks? In: Advances in neural information processing systems; Dec 8–13; Montreal, Quebec, Canada ; 3320–3328.