Article

MorphoNet: A Novel Bivalve Images Classification Framework with Convolutional Neural Network

Chanon Dechsupa[1,a,*], Pongpun Prasankok[2,b], Wiwat Vattanawood[3,c], and Arthit Thongtak[3,d]

1 Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
2 School of Biology, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand
3 Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
E-mail: [a,*]chanon.de@chula.ac.th (Corresponding author), [b]Prasankok@sut.ac.th, [c]wiwat@chula.ac.th, [d]arthit.t@chula.ac.th

Abstract. The bivalves' morphometric analysis of the freshwater shell characteristics is based on the shell size, shape, tooth, scars, and texture. We experimented and compared the accuracies of the following popular convolutional neural network architectures: ResNeSt, MobileNet, VGG16, Transfer Learning, and EfficientNet, whose model trainings are based on the bivalve image dataset obtained from a biology laboratory. The MobileNet model that gives the highest accuracy rate by 72% is selected to be a classification model of our framework named MorphoNet. We also applied the YOLO4 object detection in the MorphoNet to detect the teeth and scars on the bivalve image. The framework can identify the bivalve class labels and detect the interesting features on the bivalve images automatically. It is an alternative tool to help the biologists in a preliminary class label identification and support the land-marking creation and morphometric analysis instead of doing it by hand.

Keywords: Image processing, morphometrics, bivalves image, deep learning, MobileNet.

1. Introduction

Morphological variations observed in the study of bivalves arise from the combined influences of genetic inheritance, the environment, and natural selection. Biologists must validate the morphological variability of freshwater bivalves collected within river basins to ascertain the correlation between genetics and morphological forms. Typically, biologists employ morphological analysis tools for measuring aspects such as size, shape, and the relationship between shape, a methodology known as geometric morphometric analysis [1].

Molecular data is increasingly employed to assess the taxonomic status of bivalves. Similar to numerous other bivalves, *Ensidens* species exhibit remarkable morphological similarity, with their differentiation primarily dependent on genetic patterns. This analysis is conducted by biologists who meticulously identify and mark specific features by hand. Subsequently, analysis tools quantify and categorize morphological variations in these bivalve images based on the marked features and associated class labels. These procedures are labor-intensive and time-consuming due to the precise manual marking required on all images. Consequently, the genus *Ensidens* is an ideal candidate for evaluating the potential of convolutional neural networks in image classification applications.

Deep Convolutional Neural Networks (DCNNs) [2, 3] have evolved as pivotal tools in image classification, proficient at extracting multi-scale textural information from a wide array of image sources. They have seen substantial advancements in terms of performance and accuracy, finding applications across diverse domains. An intriguing DCNN is the Residual Attention Network, or ResNet [4], which incorporates a mixed attention mechanism encompassing bottom-up/top-down feed-forward convolutional structures. Additionally, the Split Attention Network architecture, referred to as ResNeSt [5], has emerged, enhancing multi-attention channels within a unified split-attention block. ResNeSt has exhibited promise in improving learned feature representations, thereby elevating image classification performance. We believe that ResNeSt offers potential for effective bivalve image classification. However, the suitability of specific DCNNs for the morphometric analysis of *Ensidens* species remains an open question.

In this paper, we present experimental results obtained by comparing the accuracy of various network models, including ResNeSt, MobileNet, VGG16, Transfer learning, and EfficientNet. Among these models, we selected the one that achieved the highest accuracy for the implementation of a morphometric analysis framework for bivalve image classification. The implemented framework not only enhances performance in learned feature representations but also serves as a means of representing geometric morphometrics. Consequently, the time-consuming processes of feature marking and morphological analysis are significantly reduced when employing this framework due to its streamlined procedures.

The structure of this paper is as follows: In Section 2, we describe background of bivalve features and the associated morphometric procedures, including convolutional neural network techniques. Section 3 presents an overview of related works encompassing image classification and detection networks. Section 4 elaborates on our methodology, experimental results, and the framework implementation. Lastly, Section 5 comprises the discussion and the conclusion of this study.

2. Background

2.1. Bivalve Features and Morphometric Analysis Processes

Unionoida represents a monophyletic order within the class of bivalves primarily inhabiting freshwater environments. Freshwater serves as their primary habitat type, and they play a significant role as natural food sources in nature. However, it's important to note that not all freshwater bivalves are edible. Some bivalves act as intermediate hosts for diseases, including *Opisthorchis viverrini*, Cestoda, and Nematoda, underscoring their relevance in disease ecology. Additionally, bivalves are utilized as monitoring indicators in environmental impact assessments (EIA) [6] for development projects. Taxonomic status of bivalves plays a crucial role in facilitating the planning and execution of missions carried out by the Department of Fisheries and Public Health.

Typically, the classification of bivalves is undertaken through a three-fold approach:

1) DNA analysis, which involves identifying deoxyribonucleic acid (DNA) information by inferring phylogenetic relationships.
2) Morphology analysis, Fig. 1 provides an example of a bivalve image with marked points, where a truss network comprising twenty-six markers is established and measured using the MorPhoJ software [7].
3) Subsequently, a comparative discussion is carried out to establish correlations between the results of DNA analysis and morphometric analysis, shedding light on the relationship between bivalve DNA and its monophyletic classification.

Morphometric analysis heavily relies on the characteristics of the inner texture of a calcareous valve structure in bivalves. Specifically, attention is drawn to the anatomical structure located at the connection point between the left and right valves, referred to as a "tooth". Of particular interest in bivalve morphometric analysis are the pseudocardinal tooth and lateral tooth, which are separated by a "beak". These distinctive features are pivotal in the analysis.
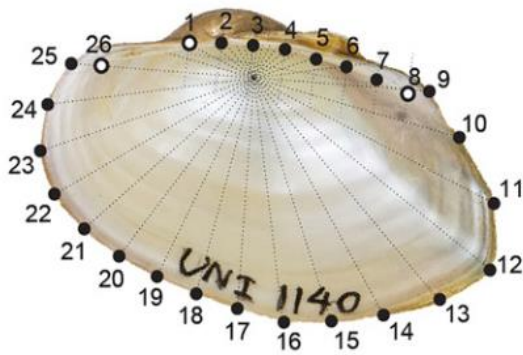
Fig. 1. Example of the determined features with multiple markings on the bivalve image [8].

Furthermore, it is essential to consider a combination of factors, including adductor muscle scars, umbo, shape, periphery, texture, and stripes. This comprehensive approach is necessary because certain bivalve classes may exhibit unclear or overlapping tooth structures. While the pseudocardinal and lateral teeth, along with the beak, serve as the primary landmarks for analysis, the remaining features are used to a lesser extent in morphometric analysis.

Scientists categorize bivalve classes primarily based on the length of the pseudocardinal tooth and lateral tooth. However, analysis using factors such as the adductor scar position, calcareous texture, and periphery can yield varying accuracy rates. This variability arises from the dependence of these dimensions on the bivalve's habitat and age, which are influenced by the physiological processes of growth and development. Some biologists consider muscle scars as an optional feature, often assessed in conjunction with tooth length.

In Fig. 2, the features on the inner right-valve image are illustrated. The teeth, represented by the red block (pseudocardinal tooth) and the green block (lateral tooth), serve as the connection point to the left valve. The
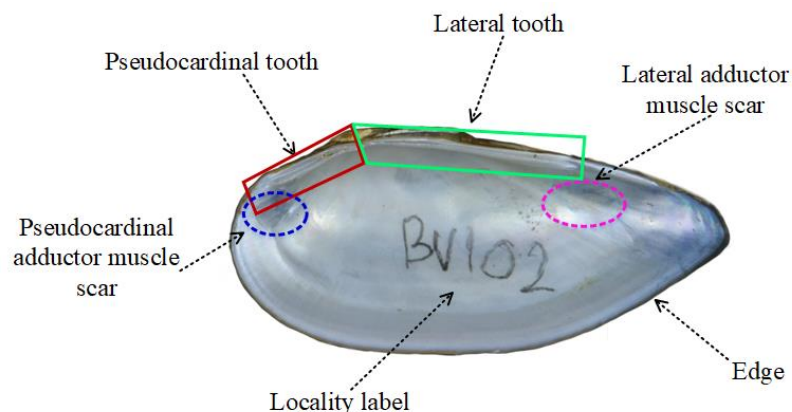
dashed-line circles denote muscle scars, where the point of attachment of the muscle is integral to the bivalve's main muscular system responsible for opening and closing the valves. These features on the bivalve images are examined using a morphometric analysis tool, involving a series of four key steps:

1) Landmark determination: landmarks are generated on all bivalve images using the MakeFan program.

2) Dataset creation: the training and test datasets are formed through the random selection and conversion of instance images, facilitated by the TpsUtil program [9].

3) Digitized landmarks spotting: Biologists manually mark the input images within the TpsDig program [10], making these markings a crucial resource for subsequent morphometric analysis [11, 12].

4) Morphometric analysis: the TPS files are imported into the MorphoJ program for conducting morphometric analysis and subsequently comparing the results with those obtained from DNA analysis.

As illustrated in Fig. 1, the marked bivalve image serves as input data for truss network measurement [13] performed using the MakeFan program. Marking numbers 1, 8, and 26 correspond to specific landmarks of interesting features. Number 1 represents the starting point of the pseudocardinal tooth, number 2 signifies the endpoint of the lateral tooth, and it also denotes the division point between the pseudocardinal tooth and the lateral tooth. These landmarks are critical identity features essential for bivalve classification.

It is evident that many of the landmarks marked in each image are mandatory and necessitate careful attention. The process involves the utilization of multiple tools. Biologists not only invest additional time in marking images but also engage with various tools throughout the morphometric analysis.

Fig. 2. An example of interesting features and a locality label on the inner right valve.

2.2. Convolutional Neural Network for the Object Detection and Image Classification

Convolutional neural networks or CNNs are mainly used for applications in image classification. Currently, there are many convolutional neural networks trained on millions of images to classify thousands of different categories: MobileNet, VGG16, ResNet, ResNeSt, EfficientNet. To achieve results in image classification, advanced image classification techniques have been enhanced by using training feed-forward convolutional neural networks called very deep structure [14, 15, 16]. A convolutional neural network is tuned by using attention mechanisms, and the subjunctive attention modules were built to work intimately with feed-forward network architecture to create attention-aware features. There are popularly effective attention mechanisms in deep neural networks such as ResNet [4] and ResNeSt. The convolutional neural network using attention mechanisms can be implemented, and the attention mechanism structure may be top-down, bottom-up, or bottom-up/top-down. Residual learning [16] with an attention mechanism is used to increase the depth of the feed-forward neuron network to benefit from residual learning. An architecture of the CNN that consists of three parts is shown in Fig. 3.

1) Convolutional layer: A rectangular neuron grid of the convolutional layer grid takes the rectangular neuron grid inputs from the previous layer. The convolutional layer acts as an image convolution, where the weights specify the convolution filter. As in equations (1) and 2, an N*N rectangular neuron layer is followed by a convolutional layer. If an m*m filter (ω) is determined, a size of ((N-m) +1) * ((N-m) + 1) is the convolutional layer output. The processing of the pre-nonlinearity input in some unit (x^l_{ij}) of convolutional layer is to sum up the contributions from the previous layer cells.

$$x^l_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega^l_{ij} y^{l-1}_{(i+a)(j+b)} \qquad (1)$$

Then, the convolutional layer applies its non-linearity:

$$y^l_{ij} = \sigma(x^l_{ij}) \qquad (2)$$

2) Pooling layer: There may be a pooling layer that is imposed after for each convolutional layer. It takes rectangular blocks and produces a single output from that block, where rectangular blocks are obtained from the convolutional layer. Pooling layers may be implemented a learned linear combination of neurons in the block or constituted the average or the maximum. In the case of max-pooling layers implementation, they are simple and are without learning themselves. The maximum is a single value output of the region K*K. For instance, if their input layer is a N*N layer, they will then output a Nk*Nk layer, as each k*k block is reduced to a single value via the max function.

3) Fully connected layer: connected layers are fully connected to all neurons in pooling or convolutional layers, which are the high-level reasoning in the network architecture.

3. Related Work

The taxonomic revision of the genus *Pilsbryoconcha*, a group of freshwater mussels in the Class Bivalvia in Indochina was studied in [17]. Through a comprehensive analysis encompassing shell morphology, biogeography, and molecular data, eight species within the genus were identified, including both previously recognized and newly described species. The research also illuminates the evolutionary history of *Pilsbryoconcha*, tracing its origins to the Khorat Plateau during the middle of the Eocene, with subsequent range expansion and diversification across Indochina. Next, a novel genus of freshwater mussels, *Namkongnaia,* is introduced from the Mekong River Basin in the work of [18]. The genus is distinguished by its unique conchological features, absence of hinge dentition, and elongated shells. Molecular analysis places it within the Pseudodontini tribe, closely related to *Monodontina*, but with conchological similarities to *Pilsbryoconcha*.
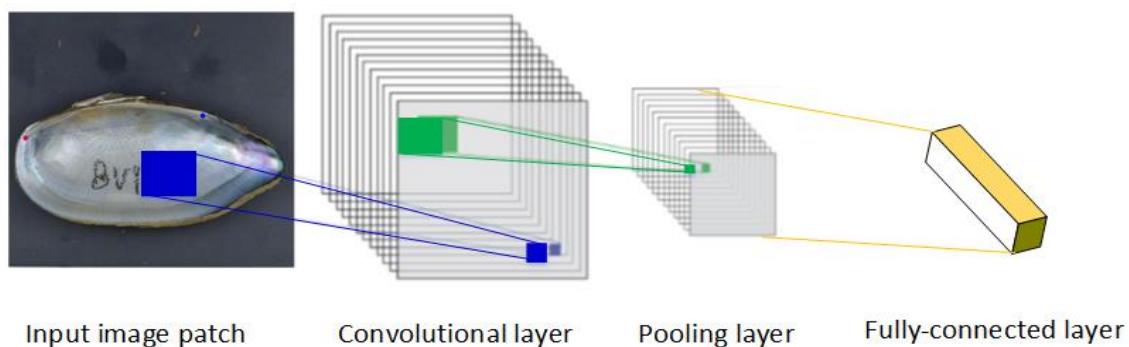

Fig. 3. Convolutional neural network architecture.

The existing species-level taxonomy of bivalves relies primarily on morphological characteristics, specifically shell shape, and size. Nevertheless, bivalve shell morphology exhibits significant resemblance and often overlaps across various populations and species [19, 20]. As a result, molecular data has become increasingly essential for evaluating the taxonomic classification of bivalves. This similarity is particularly notable among *Ensidens* species, where distinctions are primarily based on genetic patterns.

While image processing combined with morphometric analysis is becoming more prevalent, there is still relatively little research that applies CNNs to analyze biological forms in images. Sarma et al. [21] introduces a method, incorporating a KNN classifier, to detect optical distortions, or bubbles, in pork images. By effectively identifying and filtering out images containing these unwanted artifacts, it significantly enhances the accuracy of pork image classification. The proposed bubble detection method combines image pre-processing techniques with morphological and region segmentation operations, aiming to achieve the highest accuracy in detecting distorted images.

Object recognition with deep learning is the field of computer vision used to recognize and detect objects in images to determine what is in an image. The basis of detection with a mark point or mark area can be done in several ways. The area markings are popular techniques, which involve drawing a box around the object (called a "bounding box") or filling in every pixel of that object (called a "segmentation"). The YOLOv4 framework [22] is one of the best architectures to use to train a custom object detector for predicting objects, has been used in many works [23, 24].

Recent advances in image classification training feedforward convolutional neural networks using "very deep" structure are proposed. There are many works using traditional CNN models and enhancing the neural networks for image classification and object detection. The VGG model [25] was proposed by Simonyan et al., and its structure is like that of the AlexNet model [26]. The layer structures of the VGG model have variety, and the VGG-16 which contains 16 weight levels, is used for our experiment. In 2017, Google proposed MobileNet [27] focusing on mobile or embedded devices. The advantages of MobileNet are lower calculation costs and parameters than those of other networks. Google also proposed MobileNetV2 [28] to solve the problem by using inverted residuals and linear bottlenecks. Next, a squeeze and excitation (SE) block [29] is used to build channel-wise attention and neural architecture search (NAS) for MobileNetV3 [30]. They are used for enhancing efficiency and accuracy in MobileNetV3, released in 2019. EfficientNetV1 [31] requires manual tuning to arbitrarily increase the network model's depth or width. It supports higher input image resolution and improves accuracy. Next, the memory usage problem of the EfficientNetV1 in large image size results is resolved, and the network

model is proposed as EfficientNetV2 [32] that conveys a smaller model and a faster training method.

Transferred learning technique [33, 34] is a network that is trained on a large general dataset, and its weights are then used to initialize other models. This technique is suitable for limited time and data. For instance, GoogLeNet [35] and ResNet contribute to image classification purposes, and the model's weights are derived from training on a large general dataset named ImageNet. The ResNet is invented by Wang et al. in 2017, is a residual attention network mixed with an attention mechanism using an encoder-decoder style attention module. The attention-aware features are generated by the attention mechanism, which changes adaptively as layers deepen. The authors reported that their network not only increases performance in terms of consistency but is also robust to noisy inputs. He et al. [36] proposed a pre-activation structure to further improve the performance of the network in which the layout BN and ReLU carry out experiments, and the successful ResNet is trained with the network structure of more than 1000 layers. Whereas Veit et al. [37] customized a trained ResNet by dropping some of the layers in the model and compared its performance with other networks.

Over-fitting caused by insufficient data is the main issue at CNN. Data augmentation [38] such as cropping, flipping, rotation, color space, and translation, including noise injection, can be applied to improve the size of the training dataset and increase its dimensions. The CNN training with learned data augmentation can be performed by using an automated data enhancement [39, 40]. Data augmentation can not only improve accuracy, model robustness, performance, and misidentification but also increase the model's ability to be robust and make the model more stable during training.

4. Methodology

The experiments consist of two parts: 1) the finding and development of a suitable network, and 2) the framework's implementation. The network development part describes the dataset preparation, network architecture selection, and model training results. Whereas the framework implementation portrays the MorphorNet architecture. The details of each part are as follows:

4.1. Finding Suitable Network

An original bivalve image dataset is the heterogeneous species data obtained from the biologists, which contains the six classes based on molecular phylogeny representing *Ensidens* species: *E. ingallsianus, E. telus, E. jaculus, E. spiculus, E. dugasti,* and *E. sagittarius* [20], respectively (the class labels A to F) as shown in Table 1. It comprises 419 images taken in a closed environment in a biology laboratory by using a high-resolution camera, on which the ruler tool used to position the image and the locality label are marked before the image is taken. However, the dataset only includes the images taken on the inner right

valve texture, while the inner left valve images and outer texture were not collected. As the dataset listed in Table 1, there are unbalanced classes, and Classes C and D have a limited number of images due to their rarity. Thus, we neglected the class labels C and D. We also added 75 images for the classes E and F, counterbalancing each class up to 100 test images of the classes A, B, E, and F. The additional images were taken on the right valve inner texture of the raw bivalve specimens that are collected from the same freshwater location. Although certain specimens have never passed the DNA analysis processes to classify their class, the class label of each image was manually re-checked and identified by an expert based on the interesting features and the biologist's experiences once again.

Table 1.   Bivalve image information and dataset preparation.

Label: *Ensidens* species	Origin	Added	Train	Test	Remark
A: *E. ingallsianus*	146	0	100	46	-
B: *E. telus*	140	0	100	40	-
C: *E. jaculus*	8	0	0	0	Neglected
D: *E. spiculus*	3	0	0	0	Neglected
E: *E. dugasti*	50	75	100	25	-
F: *E. sagittarius*	72	75	100	47	-
Total	419	150	400	158	-

Table 2. Training model comparison.

Network models	Params (M)	Accuracy	Precision	Recall	F1-score	Training time (s)
EfficientNetB0	5.3	0.6734	0.6622	0.6797	0.6681	185.82
EfficientNetB5	25.0	0.6512	0.6753	0.6719	0.6974	1,042.51
EfficientNetB7	30.0	0.6479	0.6513	0.6562	0.6431	1,602.01
MobileNet	**5.4**	**0.7254**	**0.7183**	**0.7195**	**0.7074**	**221.35**
RestNet	25.0	0.6911	0.6921	0.6816	0.6808	7,622.20
ResNeSt	27.5	0.6447	0.6645	0.6705	0.6811	9,912.73
VGG16	134.2	0.6514	0.6422	0.6598	0.6552	8,686.74
VGG16 Transfer learning	10.6	0.68.9	0.6719	0.6682	0.6421	4,952.21

Table 3. Image data augmentation parameters.

Parameters	Values
Image rotation	25
Image ship width	0.2
Image ship height	0.1
Image shear	0.2
Image zoom	0.2
Horizontal flip	True
Vertical flip	True
image fill	Nearest
Color	-20
Contrast	AllChannelsHistogramEqualization

Due to computational constraints, we downscaled the image scale from 6000×4000 to a fixed resolution of 224 x 224 pixels. We also used the Python data augmentation module to enhance the training dataset because of the insufficient dataset issue. For making the adequate training dataset, image cropping by removing the ruler tool from the image, vertical and horizontal flipping, noise and blur reduction, affine modifications, and contrast and brightness controlling were performed. Image data augmentation parameters are shown in Table 3 however, the test dataset was still unaugmented. The prepared datasets were used for the network training and testing, with eight network models shown in Table 2. The experimental results show that the MobileNet gives predominant accuracy higher than that of the other network models.

Instead of creating a new CNN network, we customized the existing network models listed in 3. As the comparison information in Table 2, we select the MobileNet network model to be the classifier of the MorphoNet framework and use YOLO4 [41] in the objects detection process to identify the teeth and scars positions. The MobileNet network architecture is illustrated in Table 4. Input of the network is normalized bivalve image patch with unit variance and zero mean. The first layer is a convolutional layer with a kernel size of 3 × 3 pixels and 32 output channels. The network model provides F1 scores greater than 70% in almost all categories. Few remarks regarding the results (stopped after the 40th epoch): the loss dropped from 7.3 to 3.1, the accuracy is around 71%, the val_loss is around 5.9 with an accuracy of around 70.5%. However, the accuracies of classes E and F are less than those of classes A and B because the images in classes E and F have overlapped and unclear tooth structure as the example shown in Fig. 4. From the experimental results of all the explored network models, we found two factors that directly affect accuracy as follows:

1) Biologists have drawn the locality labels on certain instances. These instances may be marked with the same label if they are collected from the same river basins. Thus, most of the varieties of the instances in the same habitat are not different. We experimented by removing the locality label from the instances, taking a photo, and training the network models once again. We found that accuracy is decreasing by approximately 2.75% for all the networks. It indicates that the locality label on the images is an information bias.

2) The varieties of the input image resolution and the image positioning directly affect accuracy. Due to the imbalance in the dataset for classes E and F, we had to add 75 images for each class, so the dataset for classes E and F contains two groups of images taken in different environments. The original high-resolution images are taken in the biology laboratory, where the position of the instance and the camera is inflexible. The added images are arbitrarily taken by a photographer with a different camera, positioning, and background. Thus, the accuracy of the class labels A and B is higher than that of the classes E and F.

4.2. Framework Implementation

A web-based application, the MorphoNet framework has been implemented by using the MobileNet network obtained from Subsection 4.1 which is a web application that acts as the bivalve classifier and collects bivalve images into a bivalve instance repository. The MorphoNet architecture is shown in Table 4, which MobileNet and YOLO4 are the core networks mandated by using a web server and database. Biologists can access the system via an Internet browser. There are two steps in the framework providing for the biologists and the anonymous user. First, the bivalve images are uploaded and resized, and then the system generates a classification report with a predicted class label, whereas their teeth and scars are determined with colored circles and rectangles. The classification and detection are based on the base-line network models that are deployed by the model creators. Second, the users answer three questions and donate their bivalve image if they need it. The questions are as follows:
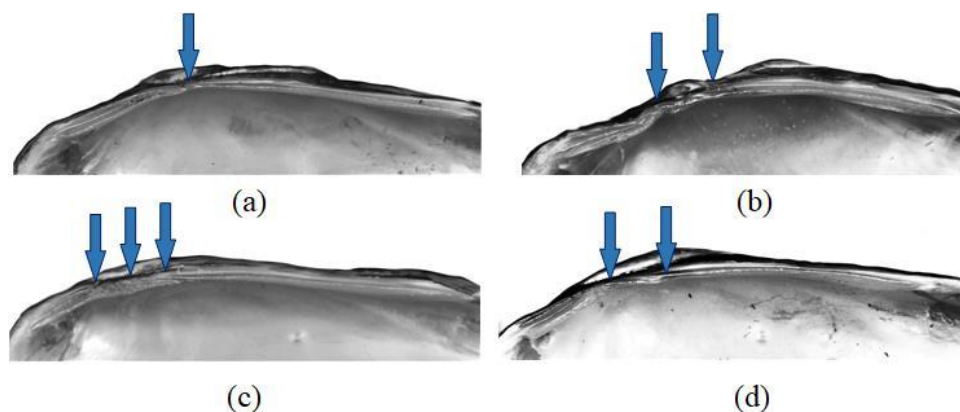

Fig 4. Example of a bivalve tooth: (a) a clearly partitioned tooth with a single delimiter; (b), (c), and (d) an unclearly partitioned tooth containing multiple delimiters.

Table 4. Layers of MobileNet architecture.

| #	Type/Stride	Filter shape	Input size
1	Conv/s2	3 x 3 x 3 x 32	224 x 224 x 3
2	Conv dw/s1	3 x 3 x 32 dw	112 x 112 x32
3	Conv/s1	3 x 3 x 32 x 64	112 x 112 x32
4	Conv dw/s2	3 x 3 x 64 dw	112 x 112 x 64
5	Conv/s1	1 x 1 x 64 x 128	56 x 56 x 64
6	Conv dw/s1	3 x 3 x 128 dw	56 x 56 x 128
7	Conv/s1	1 x 1 x 128 x 128	56 x 56 x 128
8	Conv dw /s2	3 x 3 x 128 dw	56 x 56 x 128
9	Conv/s1	1 x 1 x 128 x 256	28 x 28 x 128
10	Conv dw/s1	3 x 3 x 256 dw	28 x 28 x 256
11	Conv/s1	1 x 1 x 256 x 256	28 x 28 x 256
12	Conv dw /s2	3 x 3 x 256 dw	28 x 28 x 256
13	Conv/s1	1 x 1 x 256 x 512	14 x 14 x 256
14	5 x Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
	Conv/s1	1 x 1 x 512 x 512	14 x 14 x 512
15	Conv dw/s2	3 x 3 x 512 dw	14 x 14 x 512
16	Conv/s1	1 x 1 x 512 x 1024	7 x 7 x 512
17	Conv dw/s2	3 x 3 x 1024 dw	7 x 7 x 1024
18	Conv/s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
19	AVG Pool/s1	Pool 7 x 7	7 x 7 x 1024
20	FC/s1	1024 x 1000	1 x 1 x 1024
21	Softmax/s1	Classifier	1 x 1 x 1000
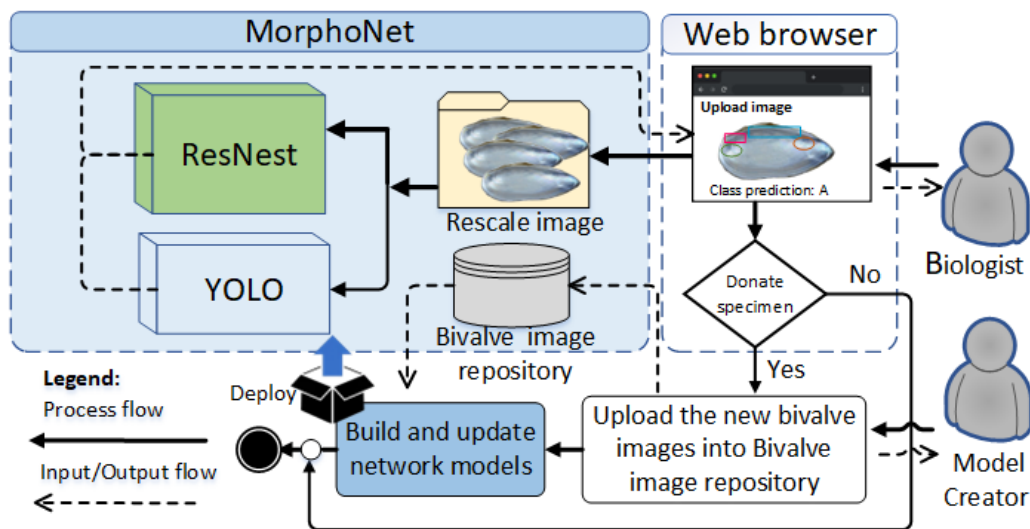
Fig. 5. The bivalve classification system architecture.

1) Do the class labels conform to your expectations (Yes/No)?
2) Please specify the expected class label and give the reasons (if it does not conform to your expectation).
3) Would you like to donate the uploaded image for further experiments?

All the donated bivalve images are collected in the bivalve image repository [42]. They will be manually rechecked on the class label by the biologists once again and taken for training the network model. The deployment of the new-weight network models will be performed in the future by the model creators, and these

processes are executed when the number of images is enough.

5. Discussion and Conclusion

This paper introduces a novel bivalve classification framework based on Convolutional Neural Networks. The primary objective of this research is to identify an appropriate deep-network model for our bivalve image dataset. Experimental results revealed that the ResNeSt network is not well-suited for our dataset due to the small size of the bivalve population. In contrast, the MobileNet network yielded the highest accuracy at 72%.

Two critical factors were found to directly influence accuracy: 1) localized bias in image labeling, and 2) lower-resolution images captured by a camera at an out-of-town biology laboratory, as opposed to those taken in the biology laboratory. These challenges stem from a small and imbalanced dataset. To address these issues, we conducted data augmentation, particularly focusing on the analysis of teeth and scars, which are intricate due to tiny portions and overlapping features.

Furthermore, our findings correlate with the observations made by biologists. Biologists demonstrated proficiency in visually classifying *Ensidens* species such as *E. ingallsianus* (class labeled A) and *E. telus* (class labeled B). In contrast, distinguishing between *Ensidens* species like *E. dugasti* (class labeled E) and *E. sagittarius* (class labeled F) proved notably challenging due to their striking resemblance. It is worth noting that the experiment faced limitations stemming from the relatively small dataset, which impacted the overall classification accuracy.

The framework was implemented with MobileNet for the bivalve classification, and YOLO4 was applied for the interesting object detection. It is an alternative tool for the preliminary morphometric analysis of the bivalve images. Although the accuracy rate of the class prediction is not outstanding, the framework can nonetheless preliminarily classify the label according to the classification of the biologists. The object detection of the interesting features on the bivalve images is reasonable as well. However, the morphometric analysis requires more precise classification, for which the classifier must be trained with a large training dataset.

Therefore, automated land-marking should be better for biologists to be able to decrease the morphometric analysis time. The least it could do is mark landmarks on bivalve images, while the classification should be the responsibility of the analysis tool. Our ongoing work is directed towards implementing the framework enabling to generate the land-marking with multiple spots on the interesting features for each bivalve image automatically and can export and import the land-marked images dataset into the morphometric analysis tools such as MorphoJ and GeoMorph. We will expand the input bivalve dataset so that it covers not only the right-valve images but also the left-valve and inner and outer sides. We will extend the MorhpoNet framework for analyzing the univalve freshwater snail named Pila globosa. The univalve images will be recorded by a video recorder, and the video taken increases the dataset size with multiple frames and heterogeneous image positions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental data

Supplemental data for this article can be accessed here: http://drnadech.com/bivalvedataset
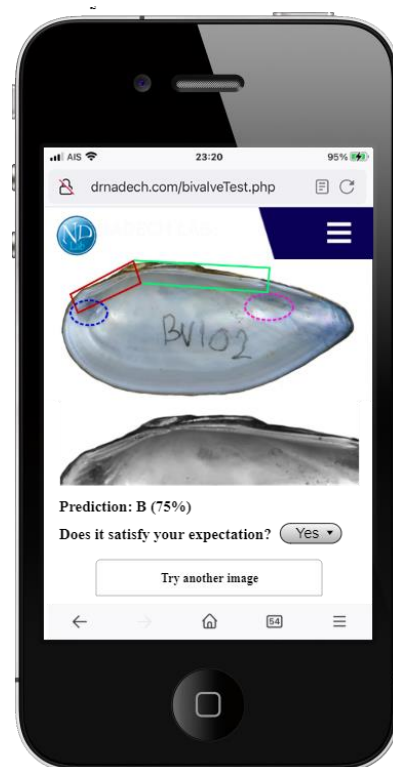

Fig. 6. Screenshot of the MorphoNet application.

References

[1] D. L. S. Zelditch, M. Leah, and H. D. Sheets, *Geometric Morphometrics for Biologists: A Primer.* Academic press, 2012.

[2] W. Rawat and Z. Wang, „Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[3] T. N. Sainath, B. Kingsbury, and A. Mohamed et al., "Deep convolutional neural networks for LVCSR," in *2013 IEEE International Conference on Acoustics*, IEEE, 2013, pp. 8614–8618.

[4] F. Wang, M. Jiang, and C Qian et al., "Residual attention network for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2017, pp. 3156–3164.

[5] H. Zhang, C. Wu, and Z. Zhang et al., "Resnest: Split-attention networks," 2020, *arXiv:2004.08955.*

[6] R. K Morgan, "Environmental impact assessment: The state of the art," *Impact Assessment and Project Appraisal*, vol. 30, no. 1, pp. 5–14, 2012.

[7] C. P. Klingenberg, "Morphoj: An integrated software package for geometric morphometrics," *Molecular Ecology Resources*, vol. 11, no. 2, pp. 353–357, 2011.

[8] S. Phuangphong, E. Jeratthitikul, and S. Panha, "Integrative taxonomy reveals phenotypic plasticity in the freshwater mussel contradens contradens (bivalvia: Unionidae) in Thailand, with a description of a new species," Systematics and Biodiversity, vol. 17, no. 2, pp. 134–147, 2019.

[9] F. James Rohlf, "The tps series of software," *Hystrix*, vol. 26, no. 1, 2015.

[10] F. J. Rohlf, "Tpsdig software, ver. 2.12. Computer program and documentation," Department of Ecology and Evolution, State University of New York, Stony Brook, NY, 2008.

[11] S. X. Cadrin, "Advances in morphometric identification of fishery stocks," *Reviews in Fish Biology and Fisheries*, vol. 10, no. 1, pp. 91–112, 2000.

[12] M. Zelditch, D. L. Swiderski, and H. D. Sheets, *Geometric Morphometrics for Biologists: A Primer.* Academic Pres, 2012.

[13] R. E. Strauss and F. L. Bookstein, "The truss body form reconstructions in morphometrics," *Systematic Biology*, vol. 31, no. 2, pp. 113–135, 1982.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv 1409.1556.*

[15] C. Szegedy et al., "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2015, pp. 1–9.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016, pp. 770–778.

[17] E. Jeratthitikul, S. Chirasak, N. P. Bun, and P. Prasankok, "Molecular phylogeny reveals a new genus of freshwater mussels from the Mekong River Basin (Bivalvia: Unionidae)," *European Journal of Taxonomy*, vol. 775 , no. 2021, pp. 119-142, 2021.

[18] E. Jeratthitikul, S. Paphatmethin, C. Sutcharit, N. P. Bun, K. Inkhavilay, and P. Prasankok, "Phylogeny and biogeography of Indochinese freshwater mussels in the genus Pilsbryoconcha Simpson, 1900 (Bivalvia: Unionidae) with descriptions of four new species," *Scientific Reports*, vol. 12, no. 1, p. 20458, 2022.

[19] S. Muanta, E. Jeratthitikul, S. Panha, and P. Prasankok, "Phylogeography of the freshwater bivalve genus *Ensidens* (Unionidae) in Thailand," *Journal of Molluscan Studies*, vol. 85, pp. 224-231, 2019.

[20] J. M. Pfeiffer, D. L. Graf, K. S. Cummings, and L. M. Page, "Taxonomic revision of a radiation of Southeast Asian freshwater mussels (Unionidae: Gonideinae: Contradentini+Rectidentini)," *Invertebrate Systematics*, vol. 35, pp 394–470, 2021.

[21] D. Sarma, P. Ubolkosold, and W. Patchoo, "Detection of distorted meat image for pork grading system," *Engineering Journal*, vol. 24, no. 5, pp. 237-244, 2020.

[22] A. Bochkovskiy, C. Wang, and H. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934.*

[23] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved yolo-v4," *Sensors*, vol. 21, no. 9, p. 3263, 2021.

[24] Z. Zakria, J. Deng, R. Kumar, M. S. Khokhar, J. Cai, and J. Kumar, "Multiscale and direction target detecting in remote sensing images via modified yolo-v4," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1039–1048, 2022.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556.*

[26] F. N Iandola, S. Han, M. W Moskewicz, K. Ashraf, W. J Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5 mb model size, 2016, *arXiv:1602.07360.*

[27] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M.Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, *arXiv:1704.04861.*

[28] M. Sandler, A. Howard, M. Zhu, Z. Andrey, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[29] J. Hu, L.Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[30] A. Howard, M. Sandler, G. Chu et al, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF*

International Conference on Computer Vision, 2019, pp. 1314–1324.

[31] M. T. and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.

[32] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning*, PMLR, 2021, pp. 10096–10106.

[33] K. Weiss, T. M Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[34] F. Zhuang, Z. Qi, K. Duan et al., "A comprehensive survey on transfer learning," in *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[35] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks, in *European Conference on Computer Vision*, Springer, 2016, pp. 630–645.

[37] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[38] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning, *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[39] E. D. Cubuk, B. Zoph, D. Mane et al., "Autoaugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*.

[40] B. Zoph, E. D. Cubuk, G. Ghiasi et al., "Learning data augmentation strategies for object detection," in *European Conference on Computer Vision*, Springer, 2020, pp. 566–583.

[41] A. Bochkovskiy, C.Wang, and H. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[42] Bivalce Project, 2022, "Bivalve dataset," http://drnadech.com/bivalvedataset, 2022

Chanon Dechsupa received B.S. degree in computer information system from Burapha University, Thailand, and received M.S. degrees in software engineering and Ph.D. degree in computer engineering, Faculty of engineering, Chulalongkorn University. From 2008 to 2015, he was a database programmer and a senior system analyst with many private sectors. His field of interest includes the formal method in software engineering and workflow design. He is currently an assistant professor at college of computing, Khon Kaen University, Thailand.

Pongpun Prasankok received Ph.D. degree in Biological Science from Chulalongkorn University. She is currently an associate professor of Molecular systematics of Mollusc at the Institute of Science, Suranaree University of Technology, Thailand.

Wiwat Vatanawood received Ph.D. degree in computer engineering from Chulalongkorn University, Thailand. He is currently an associate professor of Computer Engineering at Faculty of Engineering, Chulalongkorn University. His research interests include formal specification methods and software architecture.

Arthit Thongtak received Dr. Eng in Electrical & Electronic Engineering from Tokyo Institute of Technology, Japan. He is currently an assistant professor at department of computer engineering, Chulalongkorn University, Thailand. His interests include Asynchronous logic design and verification, Dependable computing, and Computer architecture.