Automated Morphometrics using Deep Neural Networks : Case Study on a Beneficial Insect Species

Le Van Linh[b,c,*], Zemmari Akka[b], Marie Alexia (?)[a], Beurton-Aimar Marie[b,1], Parisey Nicolas[a,1]

[a] *UMR 1349 IGEPP, BP 35327, 35653 Le Rheu, France*
[b] *University of Bordeaux, 351, cours de la Libration, 33405 Talence*
[c] *Dalat University, Dalat, Lamdong, Vietnam*

Abstract

Aims: ... *Methods:* ... *Conclusions:* ... Landmark is one of the important concepts in morphometry analysis. Finding landmarks is not only used to measure the shape of the object but also applied to analyze the inter-organisms variations. Currently, the landmarks are mostly determined manually by the biologist. In this work, we propose a method to automatic predict the landmarks on biological images: Deep Learning, more specific is Convolutional Neural Network (CNN). We proposed a CNN architecture which was built from the "elementary blocks". Each block is made up of some popular layers of CNN. The network then trained and tested on a dataset includes five parts of beetle (head, elytra, pronotum, left and right mandibles). These works have also introduced another procedure to augment the dataset which can see a little bit small in our case. In the experiments, we apply two strategies to evaluate the network and to improve the obtained results: training from scratch and applying a fine-tuning step. The predicted landmarks from the network have been compared with the manual landmarks which provided by the biologists. The obtained results have proved that the predicted landmarks are considered to be statistically good enough to replace the manual landmarks.

Keywords: Landmarks, morphometry, deep learning, CNN

1. Introduction

In the context of ecosystem services, there is an interest in studying complex interactions between evolution of insect populations and environmental factors affecting their functions. In order to assess specifically pest-regulating services and in line with studies pointing to shape traducing function [1], there are more and more research about beneficial insect morphometrics [2, 3]. In such morphometric studies, it is common to analyze subject's shape independently of their poses and sizes [4]. Since the late 20^{th} century [5], rooted in a strong statistical background, geometric morphometrics addresses the study of such biological shapes [6]. It is an effective set of methods with several specialised softwares readily available [7, 8]. Classical geometric morphometrics uses a set of landmarks to describe shape, a landmark being a two-dimensional anatomically-relevant point. In order to investigate the possibility of automated morphometric geometrics on beneficial insects, we chose to focus on one of the most common and ubiquitous beneficial insect of north-western France, *Poecilus cupreus* (Carabidae). It is considered a polyphagous

*Corresponding author
Email address: van-linh.le@labri.fr (Le Van Linh)
[1] both authors contributed equally to this work.

predator [9] beneficial to agriculture, being able to consume a large variety of agricultural pests including weed seeds, slugs and aphids [10]. As a Coleoptera, its morphological variability is usually measured on exoskeleton structures such as the head, pronotum and elytra [11].

Of course, the first step in any morphometric geometrics study is the digital imaging of the biological specimens, usually with controlled illumination and contrasting background. As such, morphometric landmark detection and positioning can be though as a particular problem of features detection and solved using robust digital image processing [12]. In the recent years, the term "deep learning" emerged describing class of computational models composed of multiple processing layers learning representations of data with multiple levels of abstraction [13]. Each layer extracts the representation of the input data from the previous layer and computes a new representation for the next layer. In the hierarchy of model, higher layers of representation enlarge aspects of the input that is important for the computational task (classification, regression, ...) and suppress irrelevant variations. As supervised learning algorithms, they use gradient descent optimization method to update the learnable parameters via backpropagation. Deep learning algorithms have proved to be very efficient in a wide variety of domains, notably image recognition and classification [14, 15, 16], speech recognition [17, 18**?**], question answering [19] and language translation [20, 21]. Within deep learning, Convolutional Neural Networks (CNNs) are well known for their success in many computer vision tasks such as image classification [14, 15] and objets recognition [22**?**]. Recent success of this algorithm in human biometry [23] lead us to believe in its potential for insect morphometrics.

1.1. Related works

Landmark or point of interest is a specific point that may contain the useful information. For example, the tip of the nose or the corners of the mouth are landmarks on human face. In image processing, we can consider two kinds of cases: the object of interest can or not be segmented. Setting landmarks can not be achieved in the same way depending on which situation we are. When segmentation can be applied, Lowe et al. [24] have proposed SIFT method to find the corresponding keypoints between two images. Palaniswamy et al. [25] have proposed a method based on probabilistic Hough Transform to automatically locate the landmarks in digital images of Drosophila wings. In our work [26], we have proposed a method which have been extended from Palaniswamy's method, to determine landmarks on mandibles of beetles. The mandibles of beetle have the simple shape and easy to segment. We have obtained good enough results about determining the landmarks automatically on mandibles. Unfortunately, this method can not be applied to other parts of beetles than the pronotum seems is segmentation has too many noises.

In recent years, deep learning is known as a solution in computer vision. Using convolutional network to determine the landmarks on 2D images has achieved better results and it seems that good solutions for the images that can not segment. Yi Sun et al. [27] have proposed cascaded convolutional neural networks to predict the facial points of interest on the human face. Zhanpeng Zhang et al. [28] proposed a *Tasks-Constrained Deep Convolutional Network* to optimize facial landmarks detection. The model determines the facial landmarks with a set of related tasks such as head pose estimation, gender classification, age estimation, face recognition, or facial attribute inference. In biology field, Cintas et al. [23] has introduced a network to predict the landmarks on human ears. After training, the network has the ability to predict 45 landmarks on human ears. In this way, we have

applied CNN computing to work with pronotum landmarks.

1.2. Contributions

We present an approach to predict morphometric landmarks based on standardized digital pictures of a coleoptera anatomical parts. For each anatomical parts, we train a convolutional neural network and statisticaly assess the suitability of the predicted landmarks to replace manual landmarks in further geometric morphometric studies.

2. Material and Methods

In the context of this section, we first present the dataset that we have used in this study, as well as the strategies to pre-process the data. Then, we show the network architecture model that we have designed to predict the landmarks in the beetle's images.

2.1. Dataset and preprocessing

In order to provide the experiment data, we have selected the Brittany lands (North-West of France) to collect the samples. After collecting in three months, a collection of 293 beetles has been established (147 males and 146 females/ 155 organic and 138 conventional) (Figure 1). As usual, images of beetles have been chosen to be studied instead of using real objects for pratical reasons. The pictures of each body parts were captured under a trinocular magnifier at ≈ 300 pixels/mm for elytra, ≈ 600 pixels/mm for pronotum and head, 1500 pixels/mm for mandibles. One can note that the head, pronotum, and elytra parts have been captured before dissection. The Lleft and right mandibles have been separated from the beetle's body before taking the photos. All the images have been taken with the same camera under same conditions to release in the RGB color mode with the size of 3264×2448.

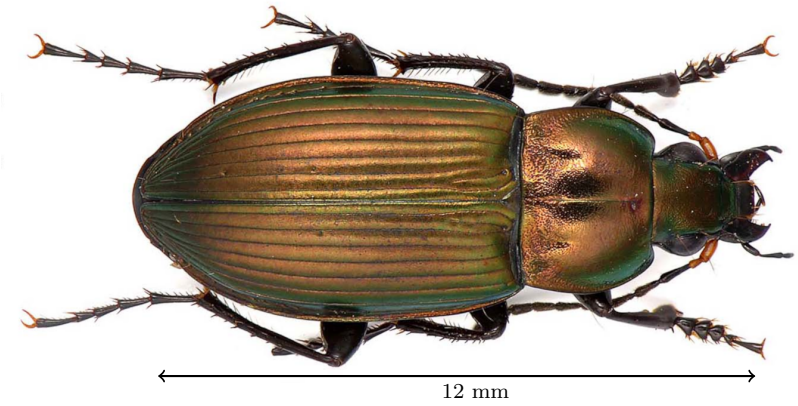
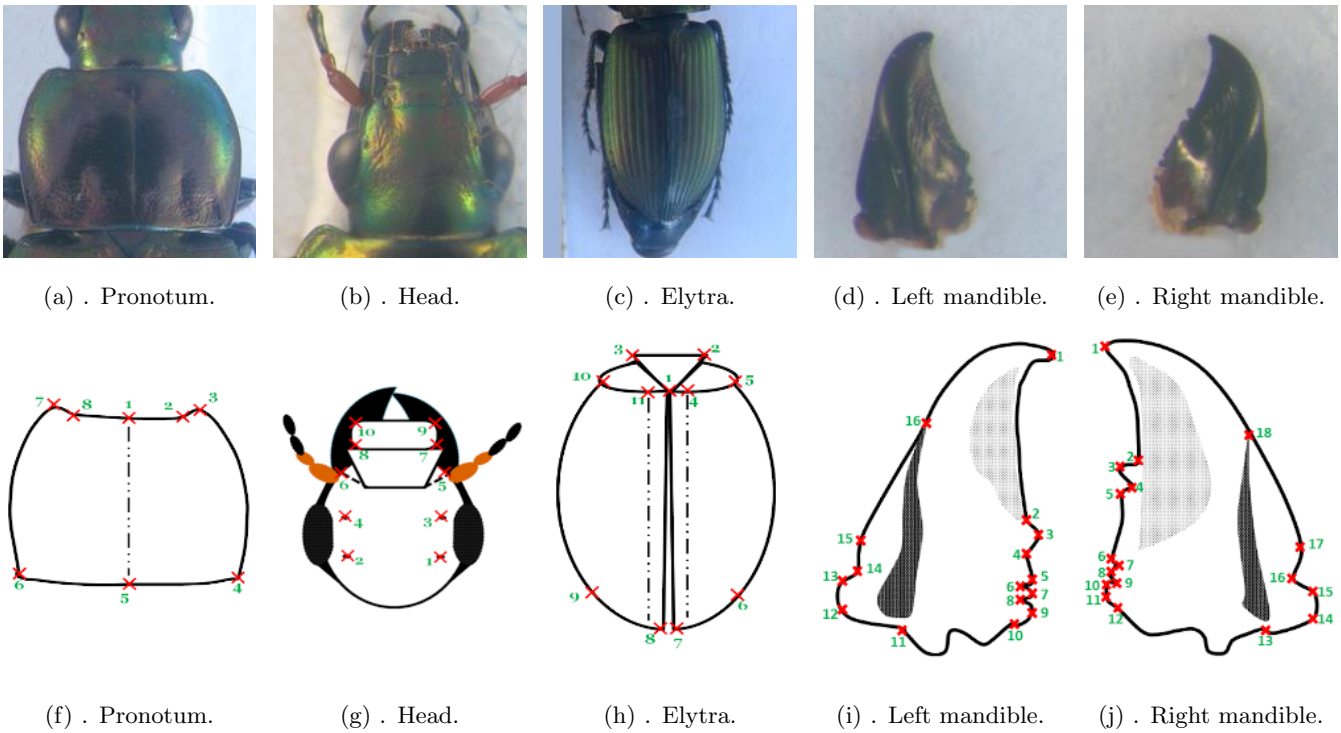


Figure 1: An illustration of the beetle.

In the next step, morphological landmarks were first set manually on the dorsal views of each body part of the beetles (head, pronotum, elytra, right and left mandibles). The morphology of each body part was processed and analyzed separately in order to limit variation resulting from their relative positions due to articulation. Landmarks were chosen according to the ease and the precision of their location on each specimen (Figure 2). Replicability analyses were performed to confirm the accuracy of landmarks positioning. They were positioned on each picture

with TPSDig2 software (version 2.17) (Rohlf, 2013a). In some individuals, mandibles could not be processed because they were lacking or broken. For each specific part, a set of number of landmarks has been provided, for example, 8 *landmarks for pronotum,* 10 *landmarks for head,* 11 *landmarks for elytra,* 16 *and* 18 *landmarks for left and right mandibles, respectively* (Figure 2). In the context of this study, these manual landmarks have been used as ground truth to evaluate the output of our method.



(a) . Pronotum. (b) . Head. (c) . Elytra. (d) . Left mandible. (e) . Right mandible.

(f) . Pronotum. (g) . Head. (h) . Elytra. (i) . Left mandible. (j) . Right mandible.

Figure 2: The sample images in our dataset (top row) and manual landmarks on each part defined by biologists (bottom row).

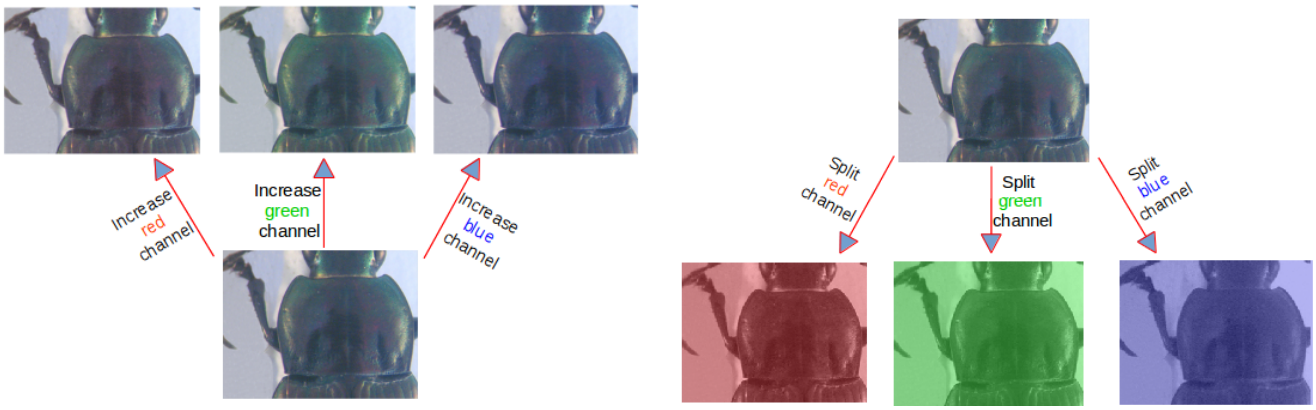
The success stories [] have proved that CNN models have been trained on a large dataset with an enormous number of data samples before using it to perform on testing data. Training the model with a big dataset can help the model able to learn more different cases and to improve the learning ability of the network. Unfortunately, providing a large dataset is too costly in several domains, e.g., in biology, medical. A solution to deal with this problem is to create the misshapen data from real data and to add them to the dataset. In our case, we have only 293 images for each part of the beetles. This number is large from the point of view of manual operations, but it is not enough to apply deep learning methods. So, we have augmented the number of images in each set of images.

Most often in deep learning applications, dataset augmentation uses operations such as translation, rotation, or scaling, which are well-known efficient to generate the new version of existing images []. However, these operations can be invariant in some cases []. We have done some tests by moving the object in the picture. In each time, we have quickly gone to the over-fitting in the training step (more detail in Section X). Consequently, we have preferred different ways to produce misshapen images by operating on the image's color channels. We have proposed two strategies to augment the number of images in our dataset.

The first strategy was applied to change the value of each channel in the original image. According to this, a constant have been added to a channel of RGB image for each time. For example, if we add a constant $c = 10$ to

the red channel from an original RGB image, we will obtain a new image with the values at red channel by greater than the red channel of original image a value of 10. By this way, we can generate three new RGB images from a RGB image.

The second procedure was split the channels of RGB images to create three gray-scale images. This work seems promising because the network model on single-channel images. At the end, we have generated six versions from an image. In total, we have obtained $293 \times 7 = 2051$ images for each set of images. Figure 3 illustrates the two strategies that we have described.



(a) . Add a constant to each channel.

(b) . Split the channels of image.

Figure 3: The two strategies to augment the number of images in our dataset.

In the content of this study, we work on pronotum part of beetle. The provided dataset contains 293 images, each image with 8 landmarks provided by biologists. The dataset was split into a training set with 260 images (training and validation) and a testing set of 33 images. During the training, the network learned the information through a pair of *(image, landmarks)* in training set. At the testing phase, the image without landmarks was given to the trained network and the predicted landmarks will be given at the output. Fig. 4 shows an example of pronotum image with its manual landmarks.

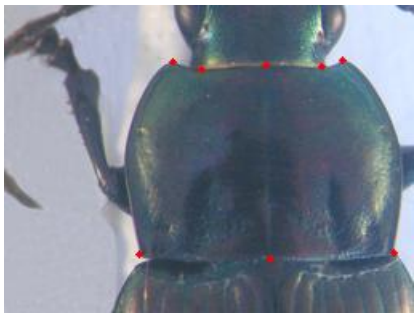


Figure 4: An example of pronotum with manual landmarks

In some succeed networks [14][27][23], the maximum size of the inputs is not over 256 pixels. In our case, the resolution of the image is large, it becomes a difficulty for the network. During training and testing, the images are down-sampling to the new resolution of 256×192. Certainly, the landmark coordinates of the image are also scaled to suit their new resolution.

The proposed network has a large number of learnable parameters. In addition, the size of the dataset is limited, this means that overfitting will occur during the training process. Therefore, we need to enlarge the size of the dataset. In image processing, we usually apply transform procedure (i.e rotate, translate) to generate a new image but in fact, when we compute the value of the pixels, it does not change while CNN computes the values of the pixels.Therefore, we have applied two other procedures to increase the number of images in the dataset. To address this problem, we have applied two procedures to enlarge the size of the dataset.

The first procedure was applied to change the value of each channel in the original image. According to this, a constant is added to a channel of RGB image and for each time, we just change the value of one of three channels. For example, from an original RGB image, if we add a constant $c = 10$ to the red channel, we will obtain a new image with the values at red channel by greater than the red channel of original image a value of 10. By this way, we can generate three new RGB images from a RGB image.

The second procedure is splitting the channels of RGB images. It means that we separate the channels of RGB into three gray-scale images. This work seems promising because the network works on single-channel images. At the end, we can generate six versions from an image, the total number of images used to train and validate is $260 \times 7 = 1820$ images (six versions and original image). The number of images that used for training and validation is splitted randomly by a ratio (training: 80%, validation: 20%) that has been set during the network setup.

In practical, when we work with CNN, convergence is usually faster if the average of each input variable over the training set is close to zero. Moreover, when the input is set closed with zero, it will be more suitable with the sigmoid activation function [29]. According to [29], the brightness of the image is normalized to $[0, 1]$, instead of $[0, 255]$ and the coordinates of the landmarks are normalized to $[-1, 1]$, instead of $[0, 256]$ and $[0, 192]$ before giving to the network.

2.2. Network architecture and training

2.3. Measuring similarities between predicted and observed landmarks

3. Results

3.1. Automated landmarks prediction for different anatomical parts

The dataset has been built by the biologists. It includes the images and manual landmarks. So, we can use the manual landmarks coordinates as ground truth to evaluate the coordinates of predicted landmarks. In the context of deep learning, landmark prediction can be seen as a regression problem. Therefore, the quality metric is used to evaluate the results. In particular, we use root mean square error (RMSE) to compute the accuracy of the implemented architecture.

Fig.5 and 6 show the training errors and the validation errors of a training time on the first and the third model, respectively. The blue curve presents RMSE on training set, the green curve presents the validation error. Clearly, the overfitting has appeared in the first model. In Fig.5, we can see that if the training is able to decrease with the number of epochs[2], it is not the case of validation loss. At the opposite in the third model, we can see some

[2]An epoch is a single pass through the full training set.

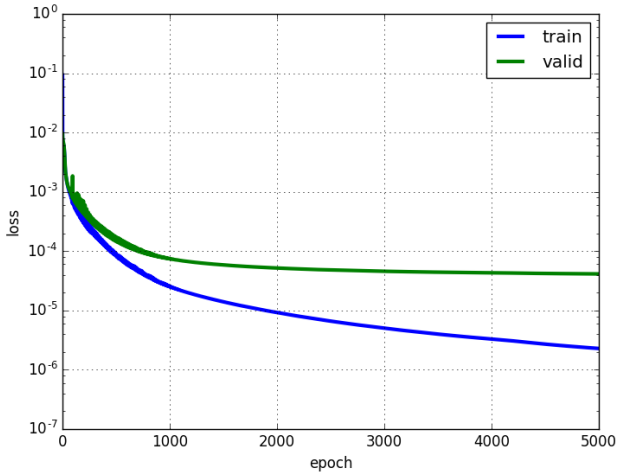


Figure 5: Learning curves of the first model.

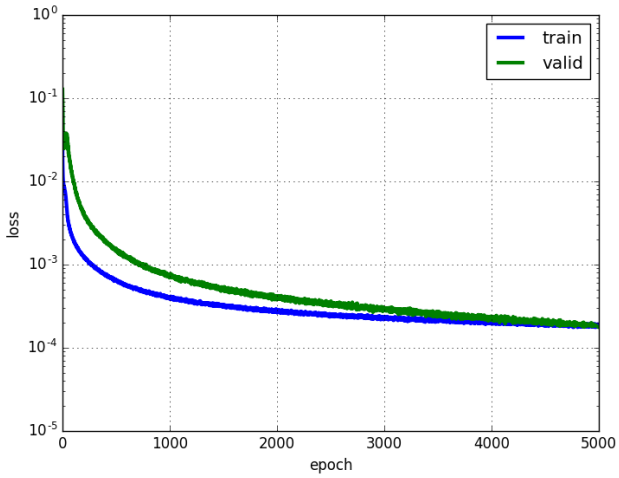


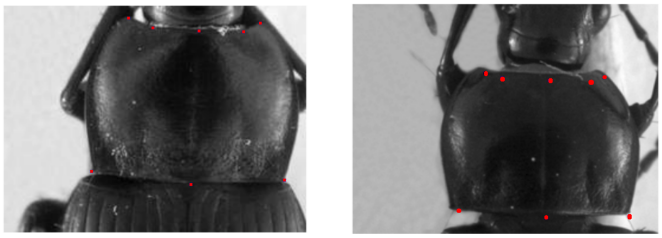
Figure 6: Learning curves of the last model.

different values for the two losses at the beginning but after several epochs, these values become more proximate and the overfitting problem has been solved.

Fig.7 shows the predicted landmarks on test images set by the thrid model. When we consider the distance between the predicted and manual landmarks, the accuracy on coordinates of predicted landmarks on Fig.7a is 99%. The propotion on Fig.7b is 80%.

Besides the losses of training, the distance from predicted landmarks to manual landmarks of the test images deserve attention also. Firstly, the distance between them is calculated. Then, the standard deviation [30] is used to quantify the dispersion of a set of distances. Table.1 shows the average error distance given on each landmark.

Fig.8 shows the distribution of the distances on the first landmarks of all images. The accuracy based on the distance of each image can be separated into three spaces: the images have the distance less than average value (4 pixels): **56.66**%; the images have the distance from average value to 10 pixels: **40.27**%; and the images have the distance greater than 10 pixels: **3.07**%. The network has enabled to detect the landmark on pronotum



(a) Image with well-predicted landmarks

(b) Image with inaccuracy landmarks

Figure 7: The predicted landmarks on an image in test set (red points)

Table 1: The average distance per landmark

#Landmark	Distance
1	4.002
2	4.4831
3	4.2959
4	4.3865
5	4.2925
6	5.3631
7	4.636
8	4.9363

automatically.

Fig.9 shows the proportion of acceptable landmarks. In our case, a predicted landmark is acceptable if the distance between it and corresponding manual landmarks is less than the average distance plus a value of standard deviation. Most of the landmarks have been detected with the accuracy greater than 70%.

At the test phase, the trained network is used to predict the landmarks on a set of test images. The program outputs the predicted-landmarks of the images as TPS files; in additional, it also fills and displays the predicted-landmarks on sixteenth firstly images of test data. With the outputs are TPS files, the user can use MAELab [26] framework[3] to display the landmarks on the images.

4. Conclusion and future works

With beetle mandibles images, the object is easy to segment and we have succeeded to determine the landmarks automatically. In opposite, the pronotum images are difficult to segment. Methods which not suppose to be based on segmentation are necessary. In this paper, after testing several models, we have presented a convolutional neural

[3]MAELab is a free software written in C++. It can be directly and freely obtained by request at the authors.

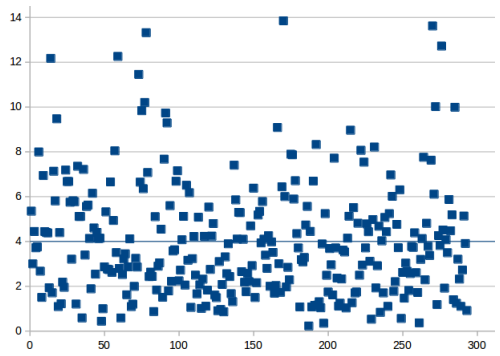


Figure 8: The distribution of the distances on the first landmark. The blue line is the average value of all distances.

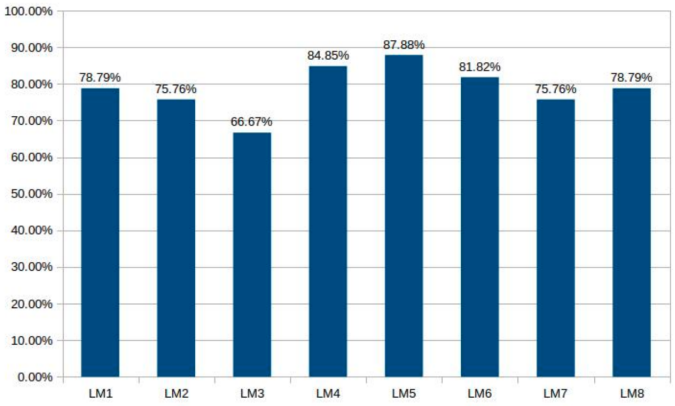


Figure 9: The proportion of acceptable predicted landmarks

network for automatic detection landmarks on the pronotum. It includes three times repeated structure which consists of a convolutional layer, a max pooling layer, and a dropout layer, followed by the connected layers. During the training phase, suitable techniques are used to prevent overfitting, a common issue of the neural networks. The network was trained several times in different selections of training data. After training with the manual landmarks given by the biologist, the network is able to predict the landmarks on the set of unseen images.

The results from the test set have been evaluated by calculating the distance between manual landmarks and corresponding predicted-landmarks. The average of distance errors on each landmark has been also considered. Using the convolutional network to predict the landmarks on biological images is promising good results in the case that the image can not be segmentation. The quality of prediction allows using automatic landmarking to replace manual landmarks in some aspects. In our case, the training dataset is limited. As a result, the accuracy of the network is acceptable. However, when we expect more about the accuracy of predicted landmarks (coordinates of predicted landmarks), the result of this work is still needed to improve (for example using a larger training dataset). Therefore, future research in landmarking identification appears as an improved of the worth exploring.

4.1. Evalution metrics for further predicted landmarks usage

TODO

5. Discussion

TODO : a biological part, a computational part

[1] C. P. Klingenberg, Evolution and development of shape: integrating quantitative approaches, Nature Reviews Genetics 11 (9) (2010) 623–635. doi:10.1038/nrg2829.
URL https://www.nature.com/articles/nrg2829

[2] K. Sasakawa, Utility of geometric morphometrics for inferring feeding habit from mouthpart morphology in insects: tests with larval Carabidae (Insecta: Coleoptera), Biological Journal of the Linnean Society 118 (2) (2016) 394–409. doi:10.1111/bij.12727.
URL https://academic.oup.com/biolinnean/article/118/2/394/2194832

[3] L. Raymond, A. Vialatte, M. Plantegenest, Combination of morphometric and isotopic tools for studying spring migration dynamics in Episyrphus balteatus, Ecosphere 5 (7) (2014) 1–16. doi:10.1890/ES14-00075.1.
URL http://onlinelibrary.wiley.com/doi/10.1890/ES14-00075.1/abstract

[4] D. G. Kendall, The diffusion of shape, Advances in Applied Probability 9 (3) (1977) 428–430. doi:10.1017/S0001867800028743.
URL https://www.cambridge.org/core/journals/advances-in-applied-probability/article/diffusion-of-shape/7CFF1175D4DCCF6063E403847120BE7B

[5] F. L. Bookstein, Foundations of Morphometrics, Annual Review of Ecology and Systematics 13 (1) (1982) 451–470. doi:10.1146/annurev.es.13.110182.002315.
URL http://www.annualreviews.org/doi/10.1146/annurev.es.13.110182.002315

[6] F. J. Rohlf, On Applications of Geometric Morphometrics to Studies of Ontogeny and Phylogeny, Systematic Biology 47 (1) (1998) 147–158. doi:10.1080/106351598261094.
URL https://academic.oup.com/sysbio/article-lookup/doi/10.1080/106351598261094

[7] D. C. Adams, E. Otrola-Castillo, geomorph: an r package for the collection and analysis of geometric morphometric shape data, Methods in Ecology and Evolution 4 (4) (2013) 393–399. doi:10.1111/2041-210X.12035.
URL http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12035/abstract

[8] C. P. Klingenberg, MorphoJ: an integrated software package for geometric morphometrics, Molecular Ecology Resources 11 (2) (2011) 353–357. doi:10.1111/j.1755-0998.2010.02924.x.

[9] A. Larochelle, The Food of Carabid Beetles:(coleoptera: Carabidae, Including Cicindelinae), Sillery: Association des entomologistes amateurs du Qubec, 1990.

[10] B. Kromp, Carabid beetles in sustainable agriculture: a review on pest control efficacy, cultivation impacts and enhancement, Agriculture, Ecosystems & Environment 74 (1) (1999) 187–228. doi:10.1016/S0167-8809(99)00037-7.
URL http://www.sciencedirect.com/science/article/pii/S0167880999000377

[11] T. Eldred, C. Meloro, C. Scholtz, D. Murphy, K. Fincken, M. Hayward, Does size matter for horny beetles? A geometric morphometric analysis of interspecific and intersexual size and shape variation in <Emphasis Type="Italic">Colophon haughtoni</Emphasis> Barnard, 1929, and <Emphasis Type="Italic">C. kawaii</Emphasis> Mizukami, 1997 (Coleoptera: Lucanidae), Organisms Diversity & Evolution 16 (4) (2016) 821–833. doi:10.1007/s13127-016-0289-z.
URL https://link.springer.com/article/10.1007/s13127-016-0289-z

[12] R. C. Gonzalez, R. E. Woods, Digital Image Processing (3rd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[15] D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 3642–3649.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[17] T. Mikolov, A. Deoras, D. Povey, L. Burget, J. Černockỳ, Strategies for training large scale neural network language models, in: Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on, IEEE, 2011, pp. 196–201.

[18] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Processing Magazine 29 (6) (2012) 82–97.

[19] A. Bordes, S. Chopra, J. Weston, Question answering with subgraph embeddings, arXiv preprint arXiv:1406.3676.

[20] I. Sutskever, O. Vinyals, Q. V. Le, Sequence to sequence learning with neural networks, in: Advances in neural information processing systems, 2014, pp. 3104–3112.

[21] S. Jean, K. Cho, R. Memisevic, Y. Bengio, On using very large target vocabulary for neural machine translation, arXiv preprint arXiv:1412.2007.

[22] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, A convolutional neural network cascade for face detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5325–5334.

[23] C. Cintas, M. Quinto-Sánchez, V. Acuña, C. Paschetta, S. de Azevedo, C. C. S. de Cerqueira, V. Ramallo, C. Gallo, G. Poletti, M. C. Bortolini, et al., Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks, IET Biometrics 6 (3) (2016) 211–223.

[24] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.

[25] S. Palaniswamy, N. A. Thacker, C. P. Klingenberg, Automatic identification of landmarks in digital images, IET Computer Vision 4 (4) (2010) 247–260.

[26] V. L. LE, M. BEURTON-AIMAR, A. KRÄHENBÜHL, N. PARISEY, MAELab: a framework to automatize landmark estimation, in: WSCG 2017, 25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2017, Plzen, Czech Republic, 2017.
URL https://hal.archives-ouvertes.fr/hal-01571440

[27] Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial point detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3476–3483.

[28] Z. Zhang, P. Luo, C. C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, in: European Conference on Computer Vision, Springer, 2014, pp. 94–108.

[29] Y. A. LeCun, L. Bottou, G. B. Orr, K.-R. Müller, Efficient backprop, in: Neural networks: Tricks of the trade, Springer, 2012, pp. 9–48.

[30] J. M. Bland, D. G. Altman, Statistics notes: measurement error, Bmj 313 (7059) (1996) 744.