Landmarks Detection by Applying Deep Networks

Van-Linh LE[1,3], Marie BEURTON-AIMAR[1],
Akka ZEMMARI[1], Nicolas PARISEY[2]

linhlv@dlu.edu.vn/van-linh.le@labri.fr, beurton@labri.fr
akka.zemmari@labri.fr, nicolas.parisey@inra.fr

[1]LaBRI-CNRS 5800, Bordeaux University, France
[2]IGEPP, INRA 1349 Rennes, France
[3]ITDLU, Dalat University, Vietnam

MAPR Conference

Ho Chi Minh City, 5-6 April, 2018

Context

Morphometry analysis

- Used to study the complex interaction between the evolution of insect and environmental factors.
- Characterize the common information of biological shape, such as, shape, sizes, or **landmarks**,. . . .

Landmark

- A kind of **point of interest**
- A specific point defined by biologist. For example, intersection of viens on fly wing, the tip of beetle's mandible,. . .

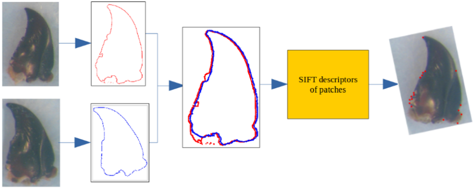
Dataset

- Images have been taken from 293 **beetles**, seperate into 5 parts (images),
- Format: 2D in RGB color,
- Focus on **pronotum** images.

(a) Left mandible (b) Right mandible (c) Body (d) Head

Problems

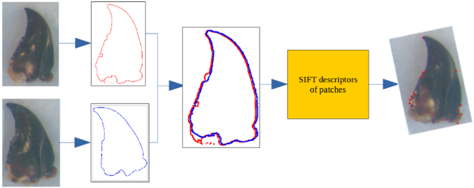
With segmentable images:[1]



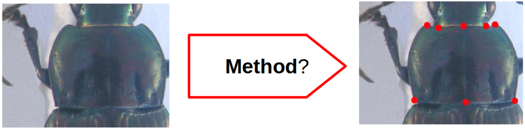
[1] Van-Linh Le, Marie Beurton-Aimar, Adrien Krähenbühl, and Nicolas Parisey. “MAELab: a framework to automatize landmark estimation." WSCG 2017.

Problems

With segmentable images:[1]

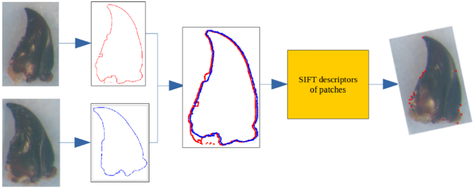


With un-segmentable images:

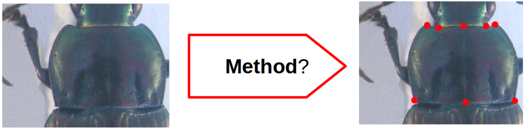


[1] Van-Linh Le, Marie Beurton-Aimar, Adrien Krähenbühl, and Nicolas Parisey. “MAELab: a framework to automatize landmark estimation." WSCG 2017.

With segmentable images:[1]



With un-segmentable images:



How to predict the landmarks coordinates?

[1] Van-Linh Le, Marie Beurton-Aimar, Adrien Krähenbühl, and Nicolas Parisey. "MAELab: a framework to automatize landmark estimation." WSCG 2017.

Content

Deep learning

Definition

- A class of machine learning,
- Use a cascade of multiple layers for feature extraction and transformation,
- Learn multiple levels of representation in supervised or unsupervised.

Deep learning

Definition

- A class of machine learning,
- Use a cascade of multiple layers for feature extraction and transformation,
- Learn multiple levels of representation in supervised or unsupervised.

Applications

- Computer vision (image recognition and classification)
- Speech recognition
- Question answering [], language translation[]

CNNs

- Consists an input, an output and multiple hidden layers
- Arranges the data in 3 dimensions: *width, height and depth*
- Classical layers: convolutional layers (**CONV**), pooling layers (**POOLING**), dropout layers (**DROPOUT**), full-connected layers (**FC**), . . .

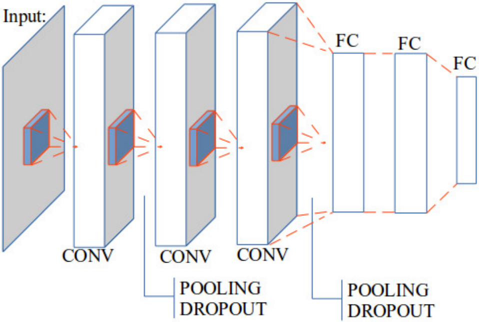


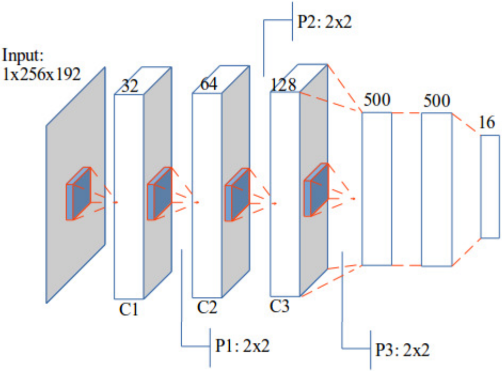
Figure: An example of CNN

The first model includes:

- An gray-scale input,
- 3 CNN layers,
- 3 POOLING layers,
- 3 FC layers.

Problems:

- Output is not good enough,
- Overfitting.

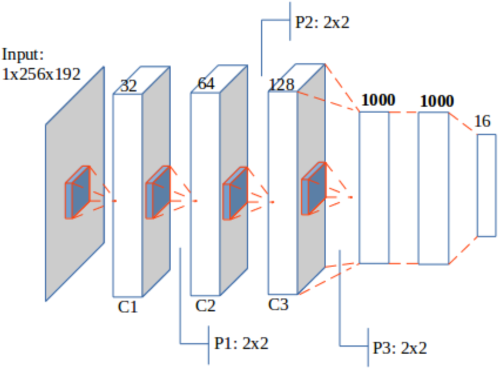


Proposed method

Network architecture

The second model:

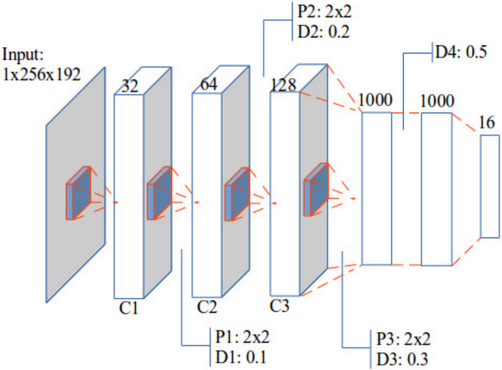
- Has the same architecture with the first one,
- Modify the output of FC layers,
- Result is not improved.



Proposed method
Network architecture

The **third** model includes:

- An gray-scale input,
- 3 CNN layers,
- 3 POOLING layers,
- 4 **DROPOUT layers,**
- 3 FC layers.



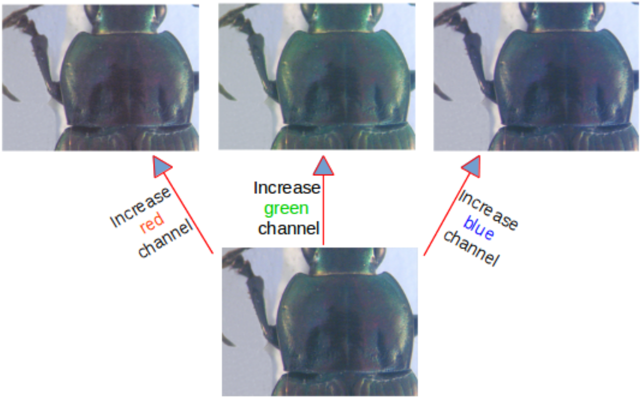
Proposed method

Data augmentation

Dataset: 293 pronotum images in RGB format.

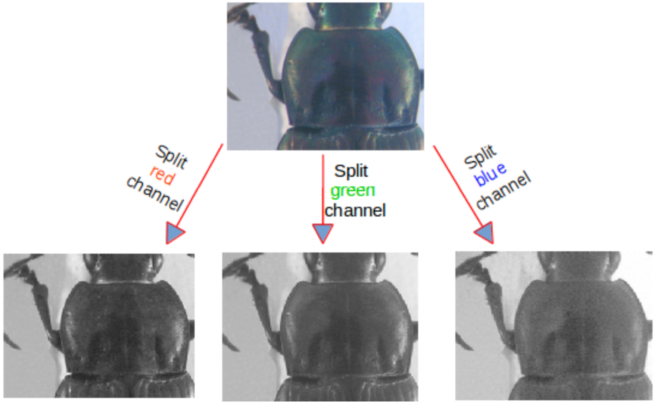
Augmentation methods:

- Increase the value of each channel,



Augmentation methods:

- Split the channels.

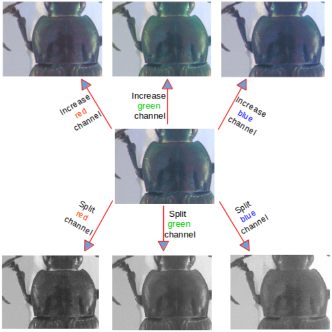


Proposed method
Data augmentation

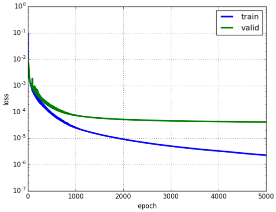
Dataset: 293 pronotum images in RGB format.
Augmentation methods:

- Increase the value of each channel,
- Split the channels.

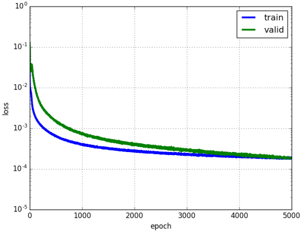
Total: $293 \times 7 = 2051$ images



- Model: the third model in 5000 epochs[2]
- Training dataset: 1820 images (260 × 7)
- Testing set: 33 images
- Images shows training and validation losses of models.
  Blue curves are training losses, green curves are validation losses.



(a) The first architecture



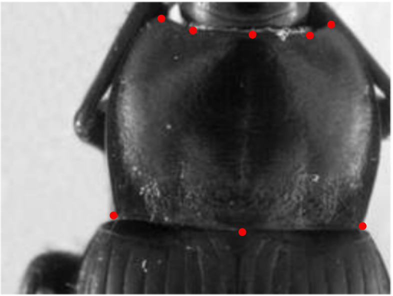
(b) The third architecture

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

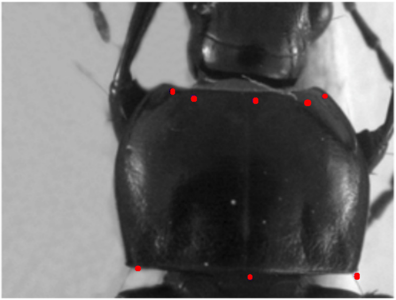
- Run the trained model to predict the landmarks on testing images,
- Calculate the distance between predicted landmarks and corresponding manual landmarks,
- Compute the average distance of all images per landmark.

#**Landmark**	**Distance** (in pixels)
1	4.002
2	4.4831
3	4.2959
4	4.3865
5	4.2925
6	5.3631
7	4.636
8	4.9363

Images show the result on testing images.



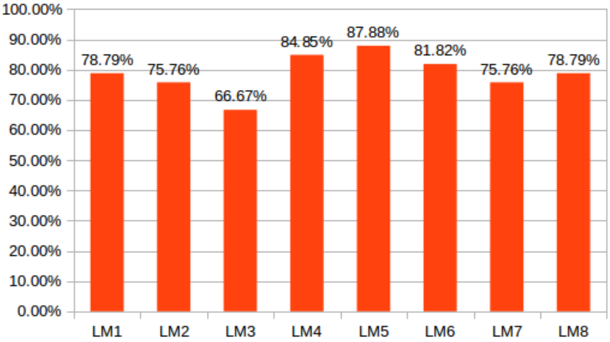
(a)



(b)

Chart shows the propotion of acceptable predicted landmarks

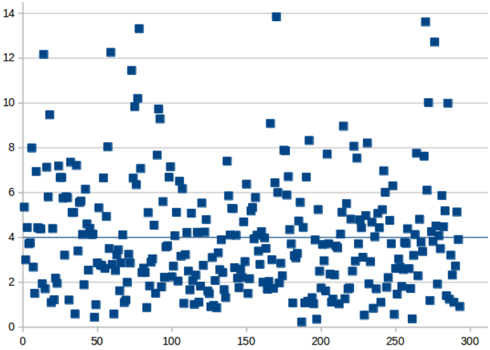
- Average accuracy: $\sim$ 75%
- Highest accuracy: 87.88%
- Lowest accuracy: 66.67%



Result

Distribution of distance on the first landmark

- Good prediction: 56.66%
- Acceptable prediction: 40.27%
- Bad prediction: 3.07%



Result
Comparing with related works

Quality metrics: coefficient of determination (r^2), explained variance (EV), Pearson correlation.

Metric	$\mathbf{r^2}$	**EV**	**Pearson**
Cintast et al. [**?**]	0.884	0.951	0.976
Proposed architecture	**0.9952**	**0.9951**	**0.9974**

Conclusion

Conclusion

- Proposed a CNN to predict the landmarks on pronotum images.
- Proposed procedure to augment the dataset.
- The location of the predicted landmarks are acceptable with high accuracy ($\sim$ 75%). It allows to replace manual landmarks.

Conclusion

- Proposed a CNN to predict the landmarks on pronotum images.
- Proposed procedure to augment the dataset.
- The location of the predicted landmarks are acceptable with high accuracy ($\sim$ 75%). It allows to replace manual landmarks.

Future works

Continue improving the landmarks coordinates by continuing on deep learning, *for example*, using transfer learning.

Thank you for attention!