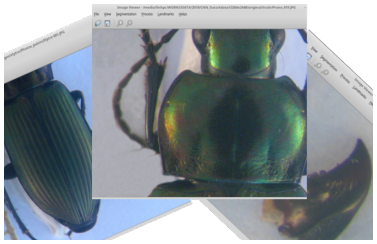
MorphoBoid

MorphoBoid[1] group studies the techniques to **automatically** process images relating to different application domains: biology, medicine, archaelogy,...

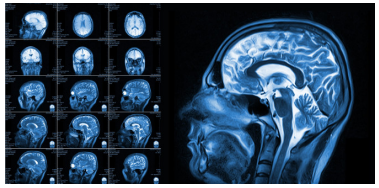
Applied techniques:

- Image processing algorithms
- Deep learning

Projects:

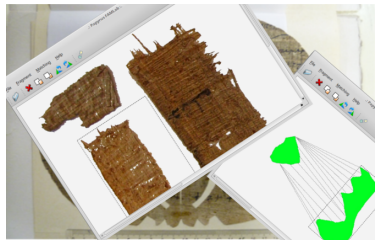


DevMAP Project



Studying MRI brain images

Heterogeneous Data Analysis



Automatisation of assembling ancient document fragments

[1] Website: **https://morphoboid.labri.fr/**

EB-Net for landmark detection on biological images

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

van-linh.le@labri.fr/linhlv@dlu.edu.vn, 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

Image and Sound Day

Talence, 29th September 2019

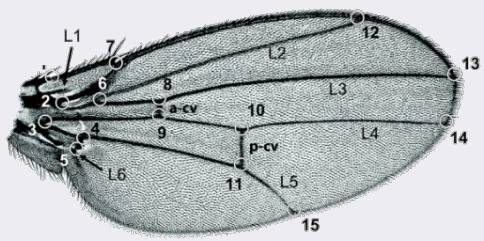
Context

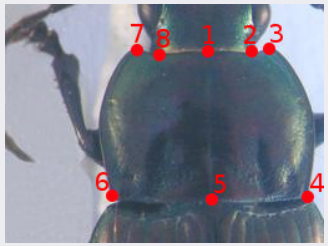
Morphometric analysis

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

Landmark

- A kind of **point of interest**.
- A specific point defined by biologist. For example, intersection of veins on fly wing, the corner of beetle's pronotum shape,. . .

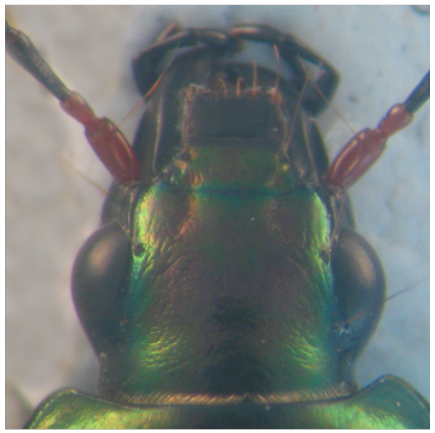




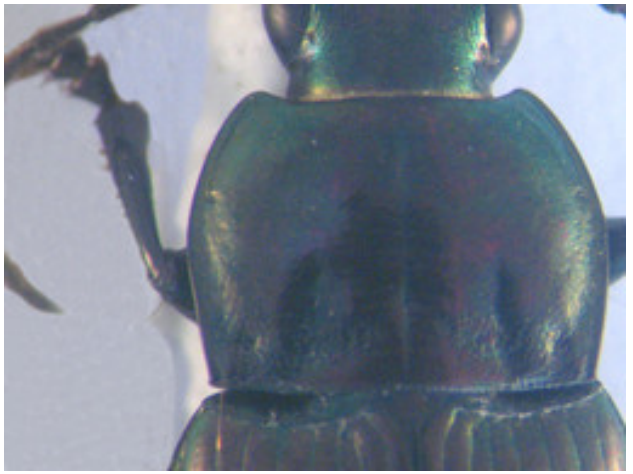
Dataset

- Images have been taken from 293 **beetles**, seperate into 5 parts: *left and right mandibles, pronotum, elytra, and head*.
- Format: 2D in RGB color.
- Focus on three parts. A set of manual landmarks have been provided as ground truth for each part.

Body part/elytra



Head part



Pronotum part/thorax

Landmark setting

Landmark identification

Manual setting:

- Time-consuming.
- Difficult to reproduce.

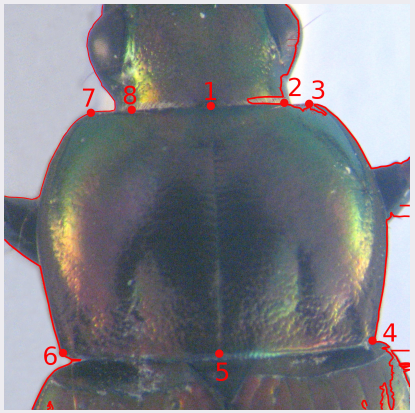
Automatically with IPT[1]:

- Success on mandibles.
- Meet difficulties on other parts

Problems

For example with pronotum image:

- Not precised: contains also a part of head and body.
- Difficult to segment this object.
- The landmarks are set both on the shape and inside the object.



[1] Image Processing Techniques

Landmark setting

Landmark identification

Manual setting:

- Time-consuming.
- Difficult to reproduce.

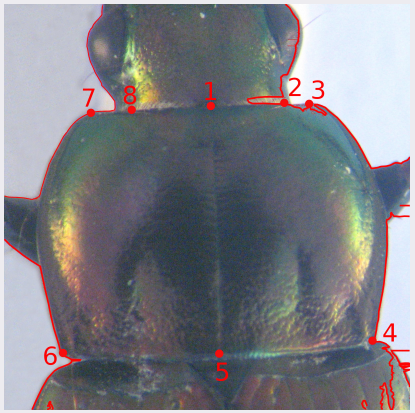
Automatically with IPT[1]:

- Success on mandibles.
- Meet difficulties on other parts

Problems

For example with pronotum image:

- Not precised: contains also a part of head and body.
- Difficult to segment this object.
- The landmarks are set both on the shape and inside the object.



How to automatically predict the landmarks coordinates?

[1] Image Processing Techniques

Content

Deep learning and CNNs

Deep learning[1]

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

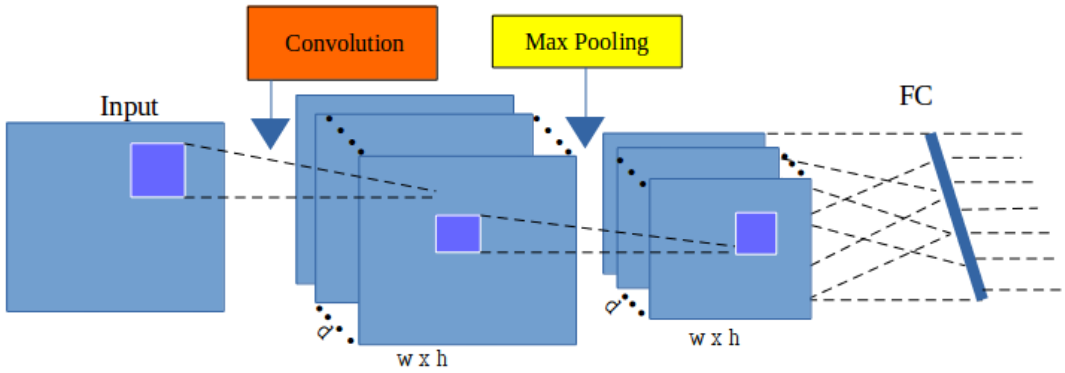
Applications

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

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015

Convolutional neural networks

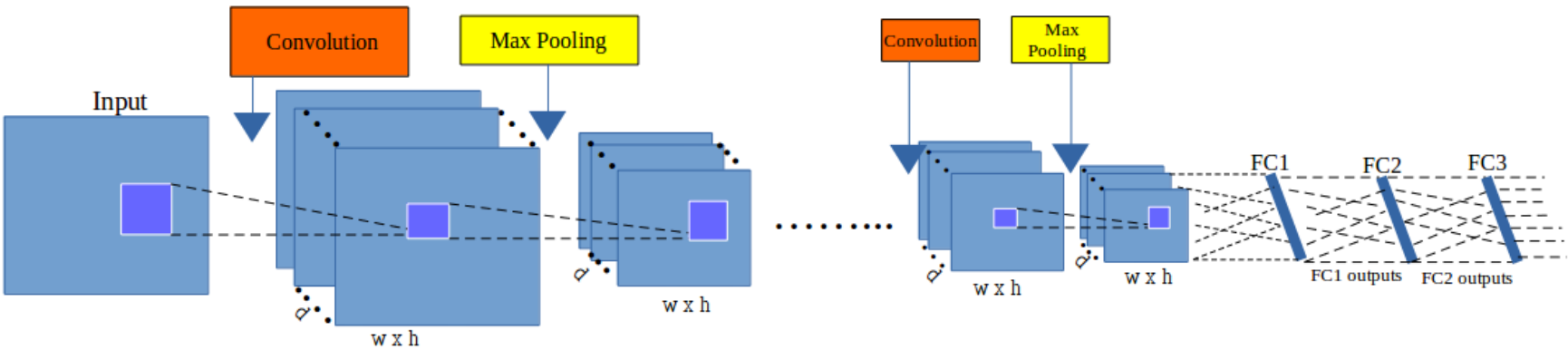
- Consists of an input, an output and multiple hidden layers[1]
- Arranges the data in 3 dimensions: *width, height and depth*
- Classical layers: **convolutional** layers, **pooling** layers, **dropout** layers, **full-connected** layers, . . .



[1] Y. LeCun et al, “Convolutional networks and applications in vision", 2010.

The first applied networks:

- Repeated three times of the pair of convolutional and maximum pooling layers
- Trying to adjust the parameters of the layers

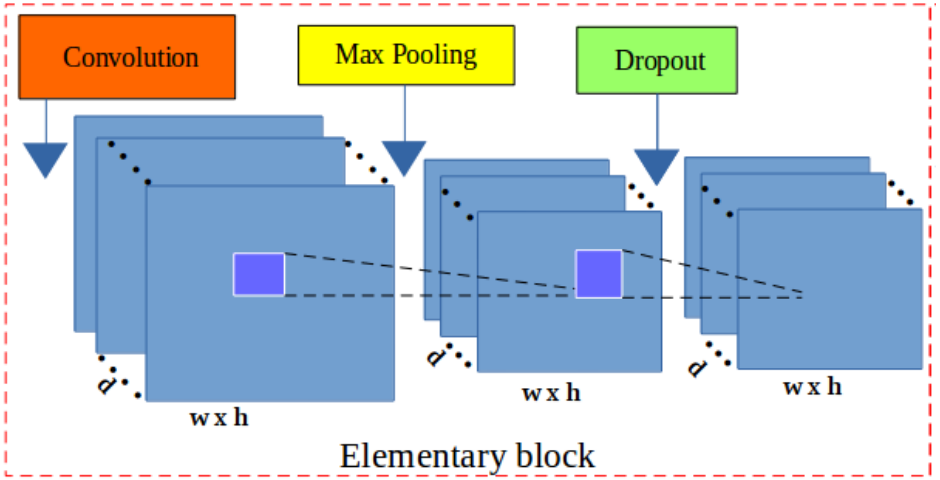


Results are not satisfying

Our proposition
Elementary block

Elementary block:

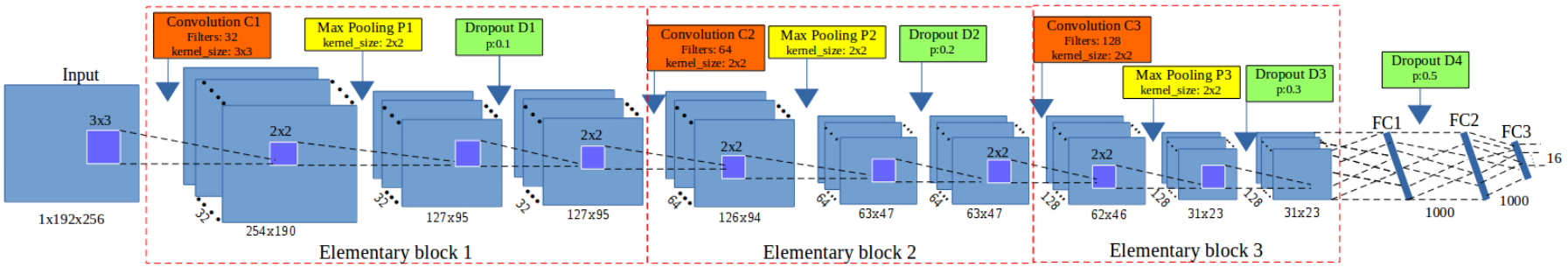
- A **convolutional** layer.
- A **maxixum pooling** layer.
- A **dropout** layer.



Our proposition
Elementary Blocks Network

The proposed model:

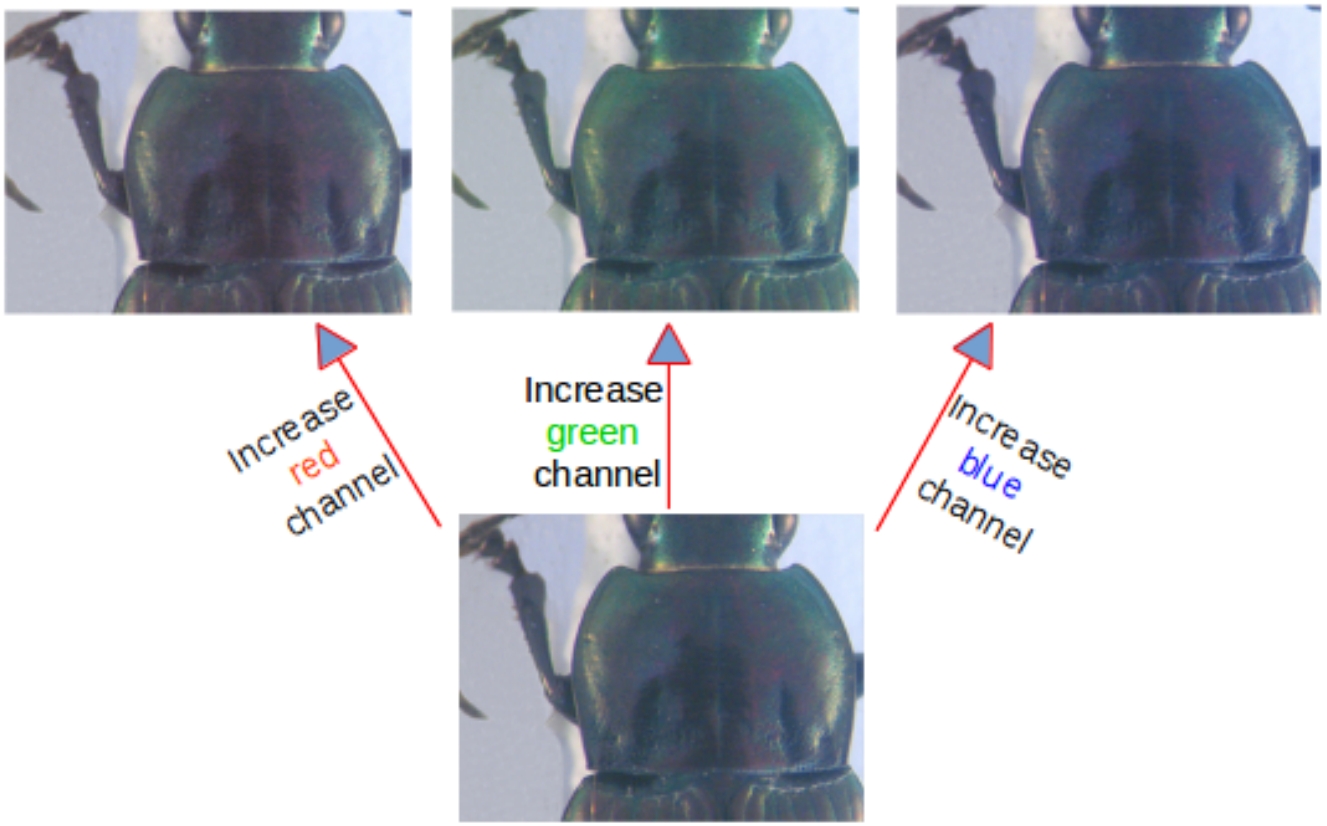
- **Three Elementary Blocks**.
- **Three** full-connected (FC) layers.
- A dropout layer was inserted between the first and the second FC layer.



Data augmentation

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

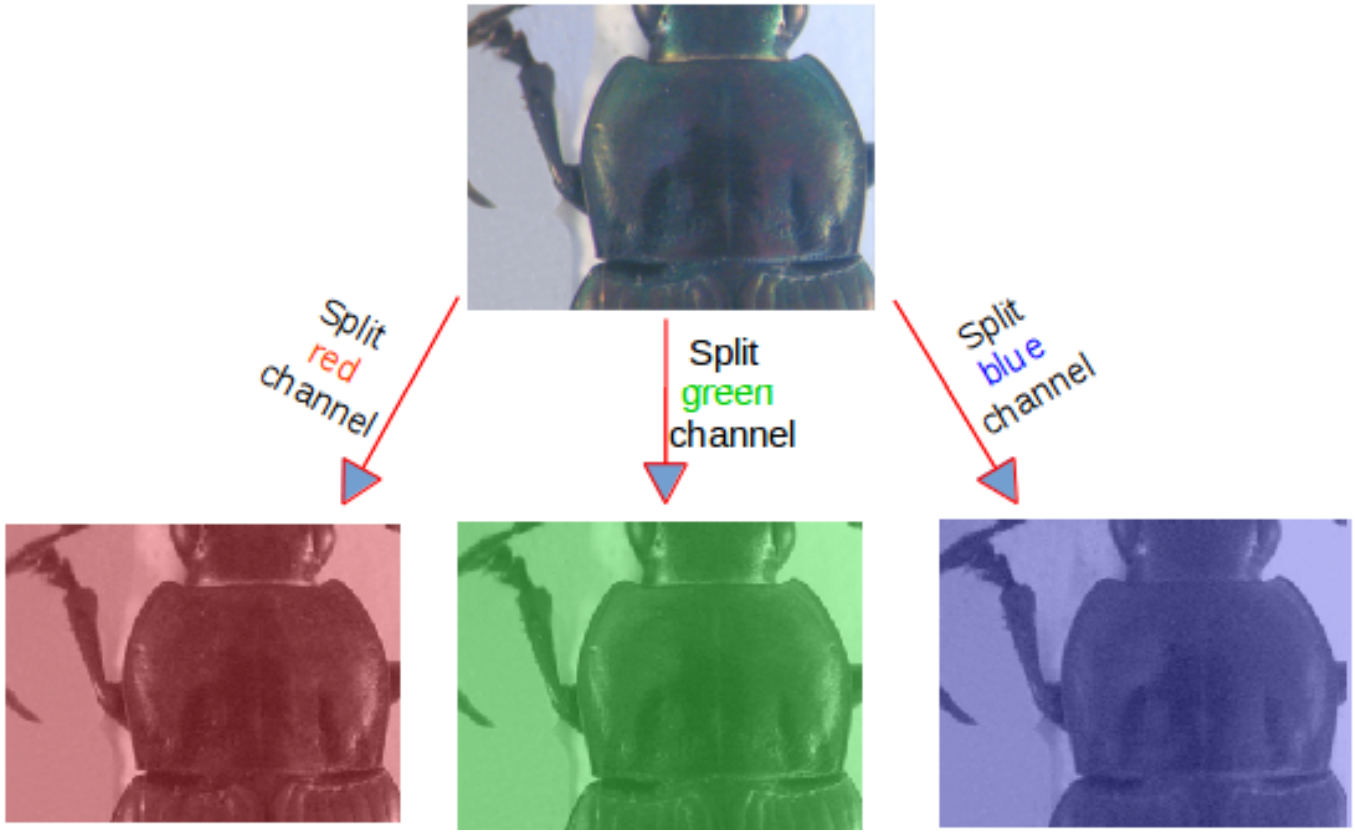
- Increase the value of each channel.



Data augmentation

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

- Split the channels.

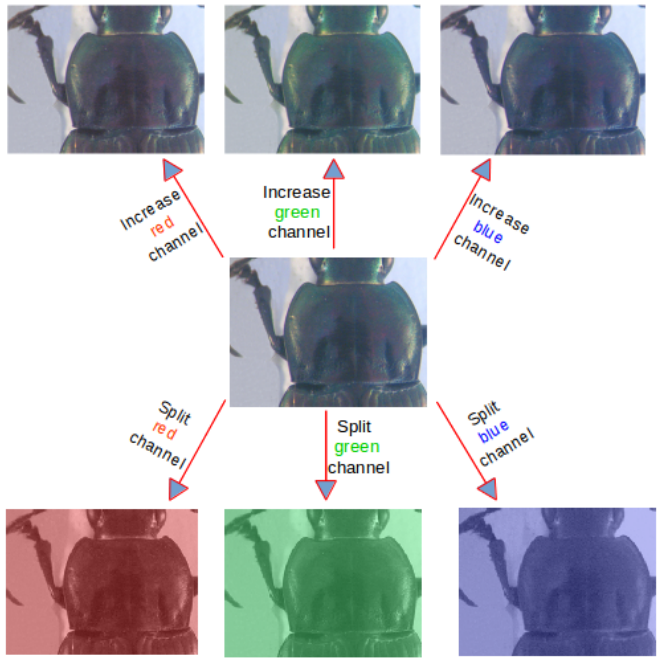


Data augmentation

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

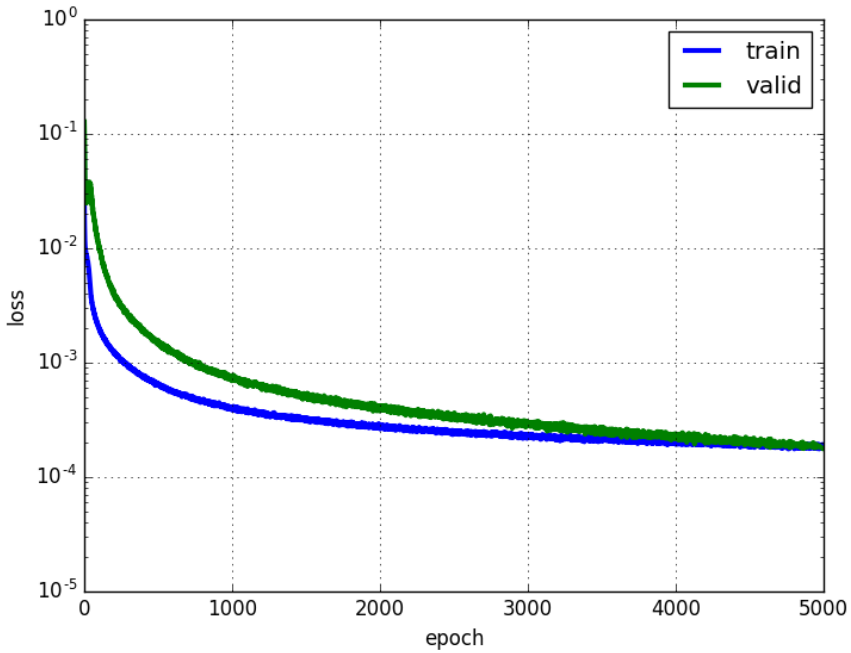
- Increase the value of each channel.
- Split the channels.

$\Rightarrow$ Total: $293 \times 7 = 2,051$ images



Training from scratch

- Training dataset: $1,820$ images (260×7)
- Apply the cross-validation to select training data
- Training parameters: momentum ($0.9 \rightarrow 0.9999$), learning rate ($0.03 \rightarrow 0.00001$), 5000 epochs[1]



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

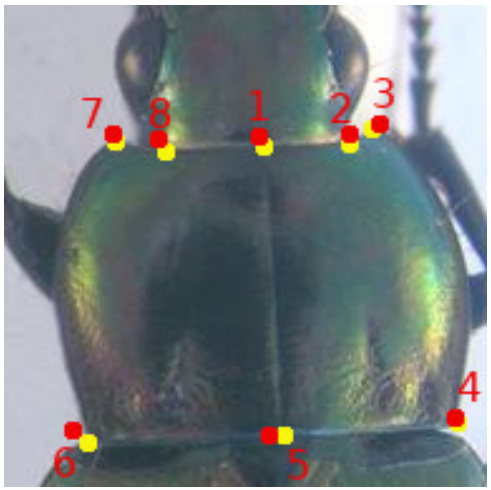
First results

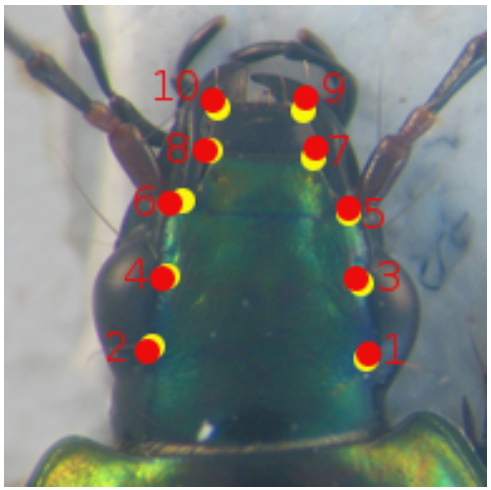
Correlation metrics and landmarks on the images

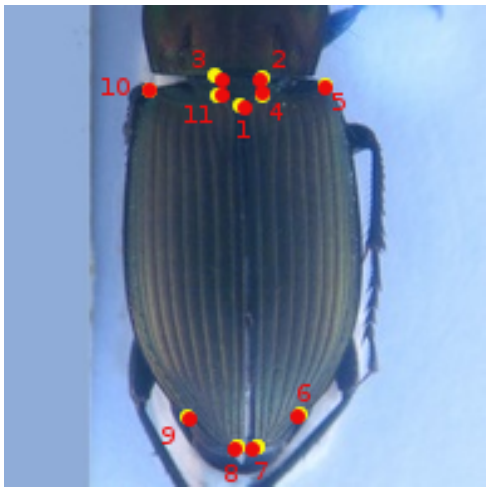
- Pearson correlation scores on each dimension of landmark

Pearson corr. score	X-dimension	Y-dimension
Pronotum	**0.8116**	**0.9438**
Head	**0.9194**	**0.9271**
Elytra	**0.9523**	**0.8743**

- Display the landmarks on the images:







First results
Average distances

- Calculate the distance between predicted landmarks and corresponding manual landmarks.
- Compute the average distance by landmark.

Landmark	Pronotum	Head	Elytra
1	**4.00**	**5.53**	**3.87**
2	4.48	5.16	3.97
3	4.30	5.38	3.92
4	4.39	5.03	**3.87**
5	4.299	4.84	4.02
6	**5.369**	**4.45**	4.84
7	4.64	4.79	5.21
8	4.94	4.53	**5.47**
9	-	5.14	5.27
10	-	5.06	4.07
11	-	-	3.99

Transfer learning/Knowledge transfer

- Re-use model developed for a specific task/dataset to lead another task on another dataset
- **Fine-tuning**: retrain a pretrained model
- **Model Zoo** (Caffe library): people share their pre-trained networks.

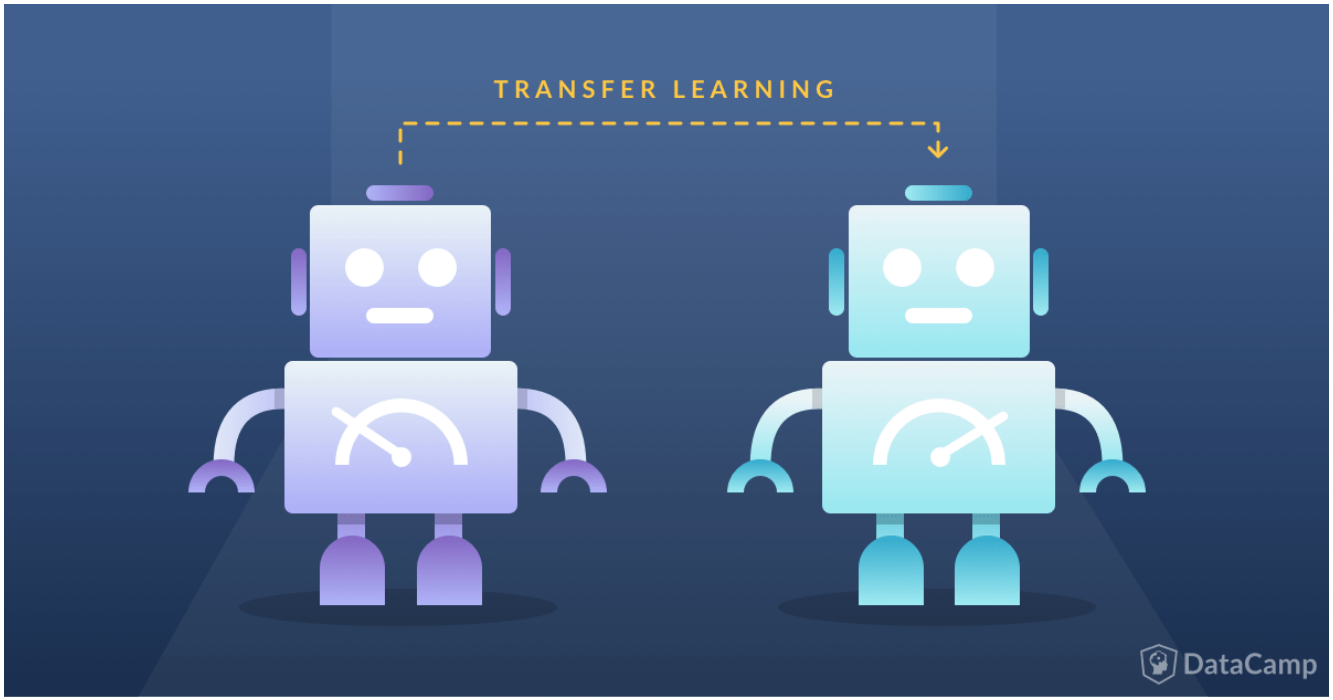
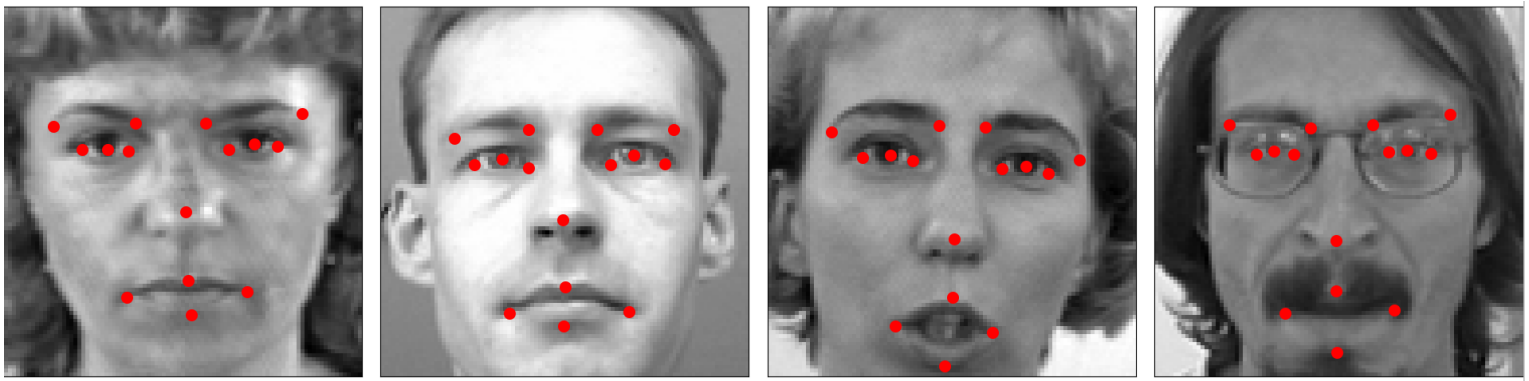


Image source: **DataCamp**

Fine-tuning our model

- Estimated landmarks on pronotum images when fine-tuning on **VGG-16, VGG-19, ResNet50** have not been improved.
- Train the model on a Facial Keypoint dataset[1].
- Fine-tune the pre-trained model on each part.



Images source: from dataset.

[1] https://www.kaggle.com/c/facial-keypoints-detection

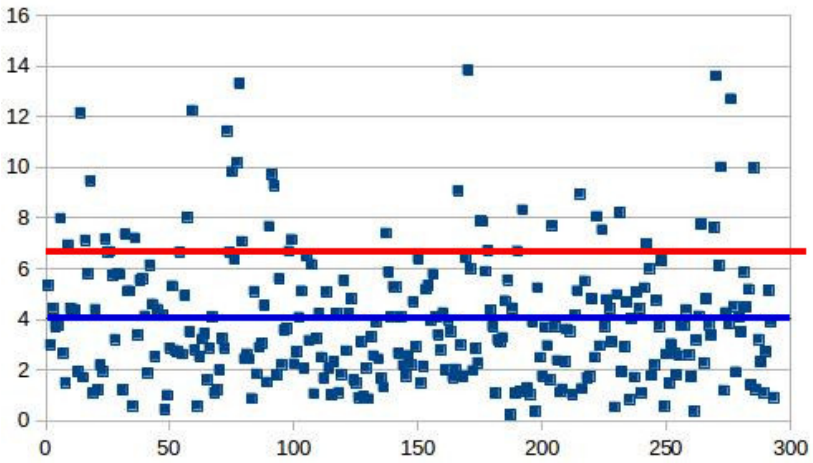
Results
A comparation of average distances

Comparing the average distances between two processes: training from scratch (**F.S**) and fine-tuning (**F.T**).

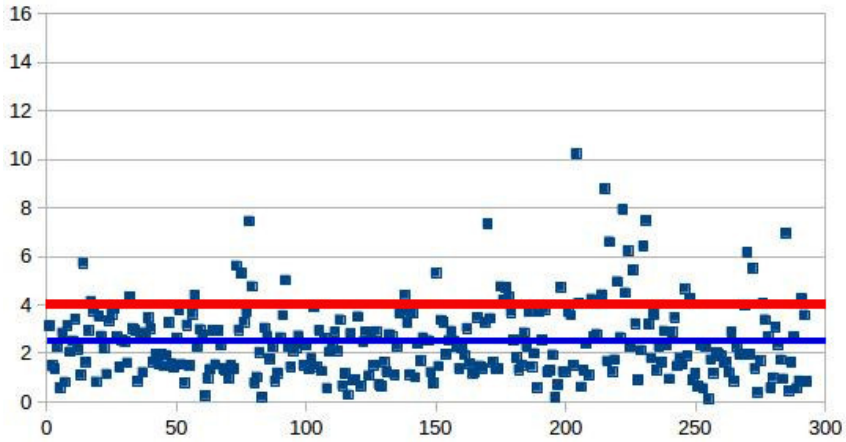
LM	Pronotum		Head		Elytre	
	F.S	F.T	F.S	F.T	F.S	F.T
LM1	**4.00**	2.99	**5.53**	**4.82**	**3.87**	3.21
LM2	4.48	3.41	5.16	4.21	3.97	3.28
LM3	4.30	2.98	5.38	4.73	3.92	**3.20**
LM4	4.39	3.54	5.03	4.11	**3.87**	3.22
LM5	4.29	3.37	4.84	4.18	4.02	3.31
LM6	**5.36**	**4.06**	**4.45**	3.50	4.84	4.21
LM7	4.64	**2.93**	4.79	3.92	5.21	4.54
LM8	4.94	3.64	4.53	**3.40**	**5.47**	**4.76**
LM9	-	-	5.14	4.17	5.27	4.55
LM10	-	-	5.06	3.94	4.07	3.39
LM11	-	-	-	-	3.99	3.29

Results
Distribution of average distances on pronotum

- The distribution of distance of the best result: 1st landmark



(a) Training from scratch

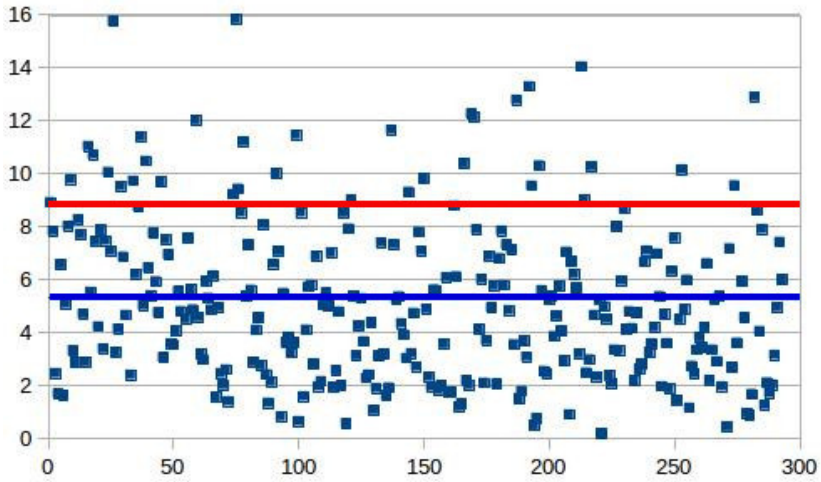


(b) With fine-tuning

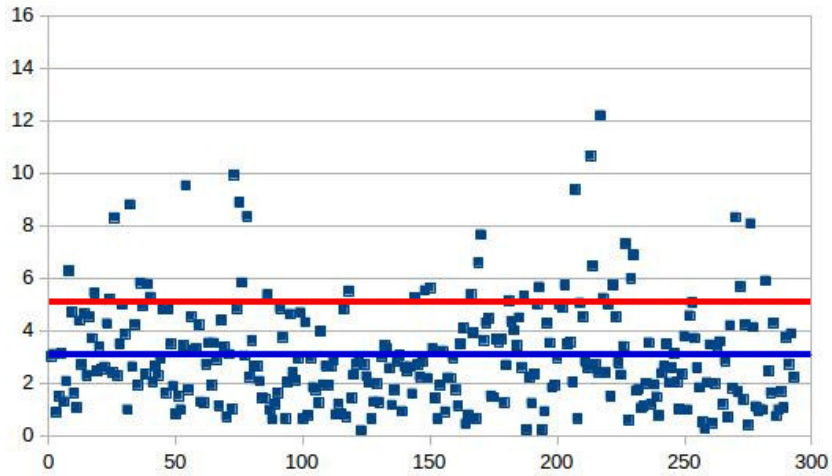
Results

Distribution of average distances on pronotum

- The distribution of distance of the worst result: 6^{th} landmark



(a) Training from scratch



(b) With fine-tuning

Conclusion

Conclusion

- Propose a new CNN architecture with elementary blocks to predict the landmarks on the images of three remaining parts.
- Propose a new procedure to augment the dataset.
- Apply fine-tuning to improve the accuracy of predicted landmarks coordinates.
- The predicted landmarks are abled to replace the manual landmarks without segmentation step.

Conclusion

Conclusion

- Propose a new CNN architecture with elementary blocks to predict the landmarks on the images of three remaining parts.
- Propose a new procedure to augment the dataset.
- Apply fine-tuning to improve the accuracy of predicted landmarks coordinates.
  - The predicted landmarks are abled to replace the manual landmarks without segmentation step.

Future works

- Trying with different compositions of Elementary Blocks.
- Going deeply how to design the right model for landmarking problem.

Thank you for attention!