

Towards landmarks prediction with Deep Network

Van-Linh LE^{1,3}, Marie BEURTON-AIMAR¹,
Akka ZEMMARI¹, Nicolas PARISEY²

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

¹LaBRI-CNRS 5800, Bordeaux University, France

²IGEPP, INRA 1349 Rennes, France

³ITDLU, Dalat University, Vietnam

ICPRS Conference

Valparaíso, 22-24 May, 2018

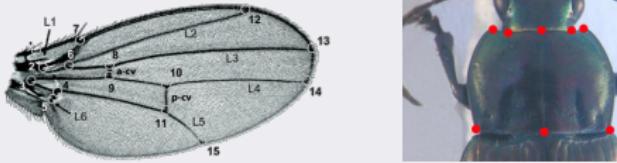


Morphometry analysis

- ▶ 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 veins on fly wing, the corner of beetle's pronotum,....



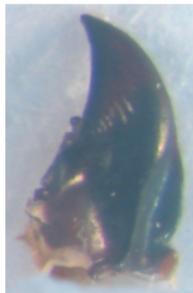
Dataset



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



(a) Left mandible



(b) Right mandible



(c) Body

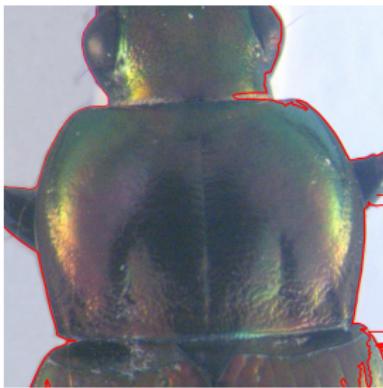


(d) Head

Problems



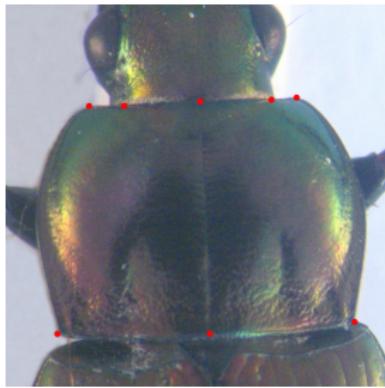
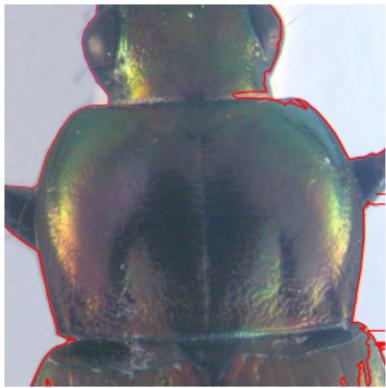
Un-segmentable images:



Problems



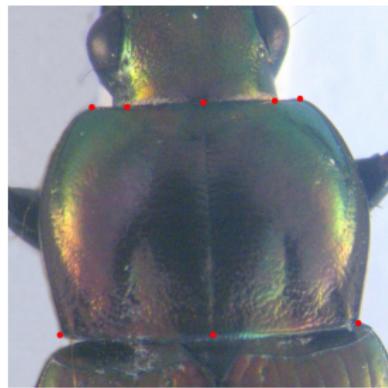
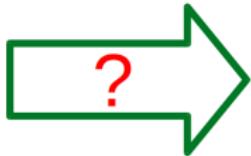
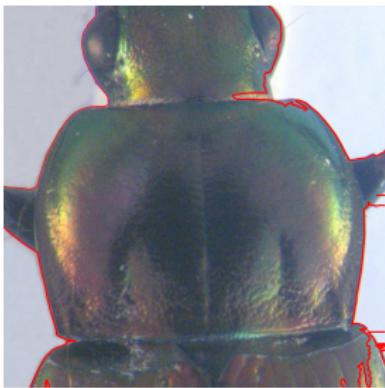
Un-segmentable images:



Problems



Un-segmentable images:



How to predict the landmarks coordinates?

Content



Deep learning and Convolutional Neural Networks

Deep learning

Convolutional neural networks (CNNs)

Proposed method

Network architectures

Data augmentation

Training from scratch

Fine-tuning

Results

Conclusion



Definition¹

- ▶ 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.

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



Definition¹

- ▶ 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.

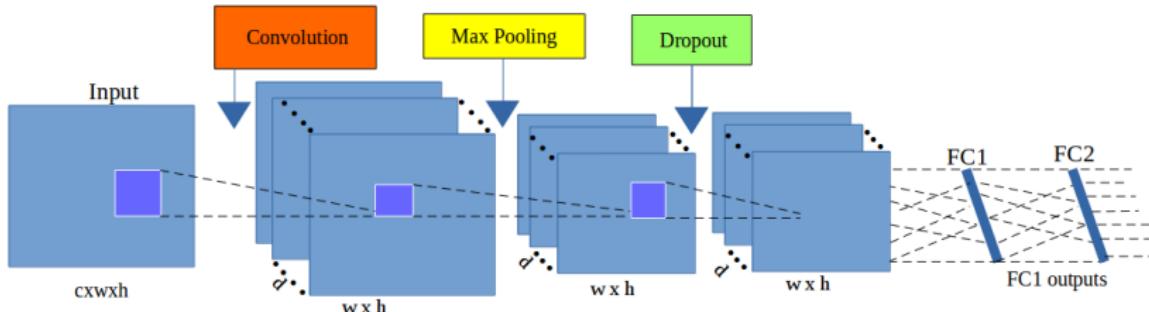
Applications¹

- ▶ 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



- ▶ Consists an input, an output and multiple hidden layers¹
- ▶ 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**), ...



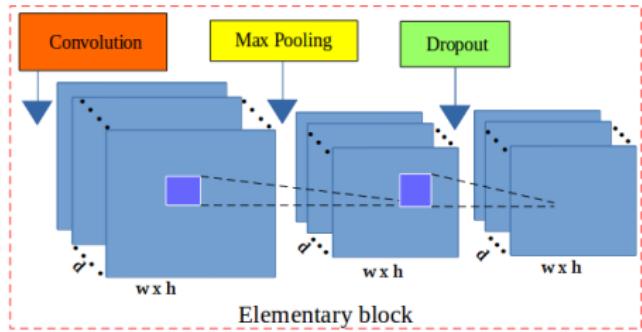
¹ Y. LeCun et al, "Convolutional networks and applications in vision", 2010.

Network architecture



Elementary block:

- ▶ A CONV layer,
- ▶ A maximum POOLING layer,
- ▶ A DROPOUT layer

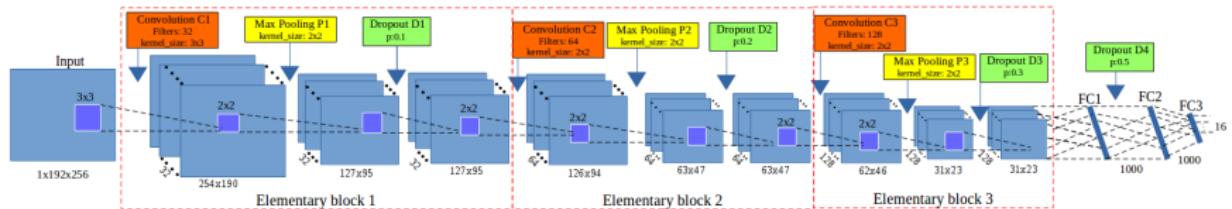


Network architecture



The proposed model:

- ▶ **Three** elementary blocks,
- ▶ **Three** full-connected (FC) layers
- ▶ A dropout layer was inserted between the first of two FCs



Data augmentation



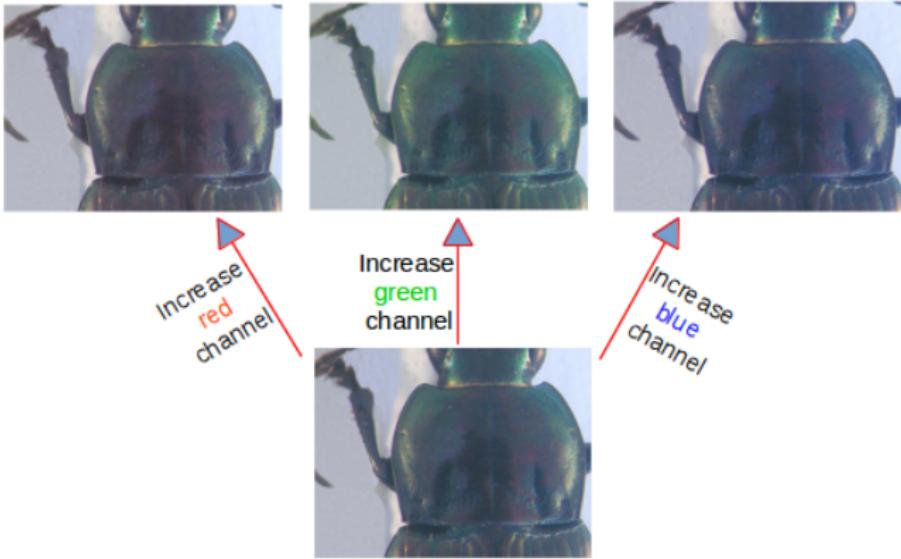
Dataset: 293 pronotum images in RGB format.

Data augmentation



Augmentation methods:

- ▶ Increase the value of each channel,

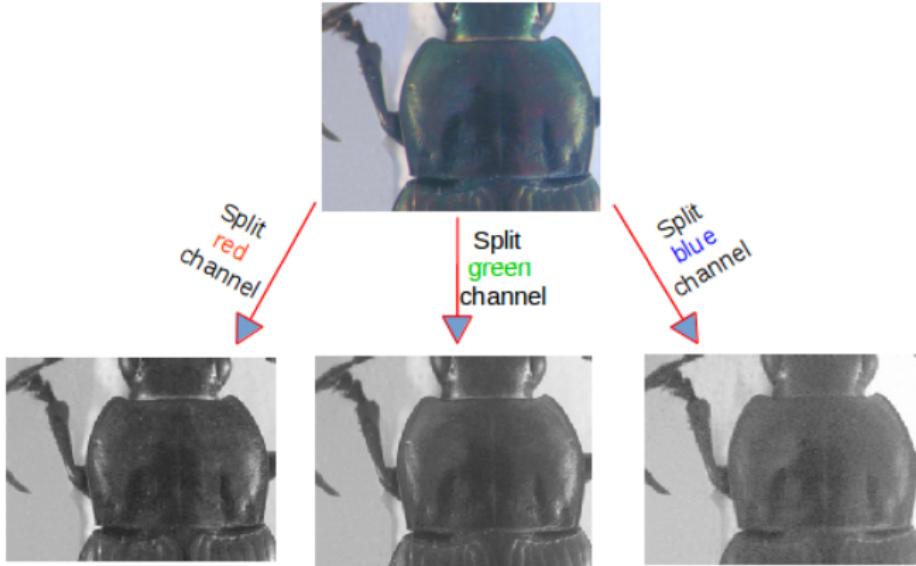


Data augmentation



Augmentation methods:

- ▶ Split the channels.



Data augmentation

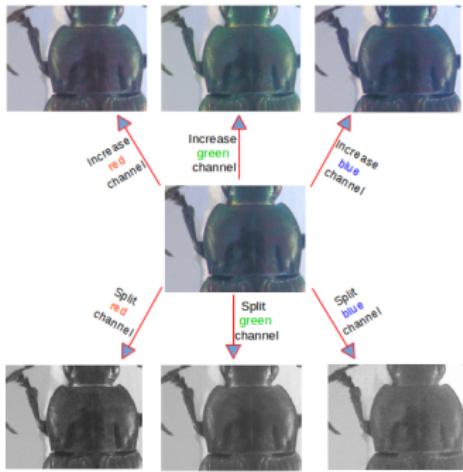


Dataset: 293 pronotum images in RGB format.

Augmentation methods:

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

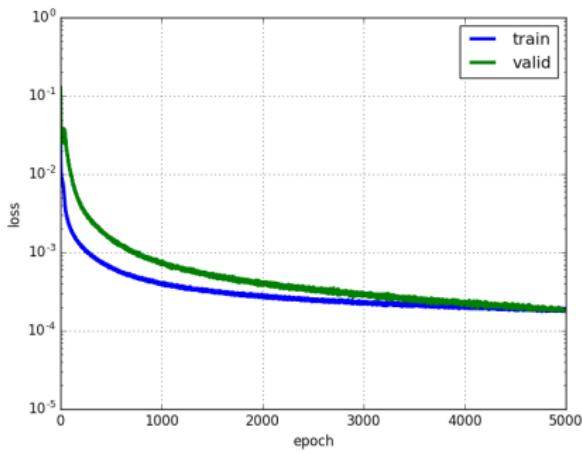
Total: $293 \times 7 = 2,051$ images



Training



- ▶ Training parameters: momentum ($0.9 \rightarrow 0.9999$), learning rate ($0.03 \rightarrow 0.00001$), 5,000 epochs¹
- ▶ Training dataset: 1,820 images (260×7)
- ▶ Image shows training and validation losses of the model.
Blue curve is training losses, green curve is validation losses.
- ▶ Training time: 3 hours using NVIDIA TITAN X card.



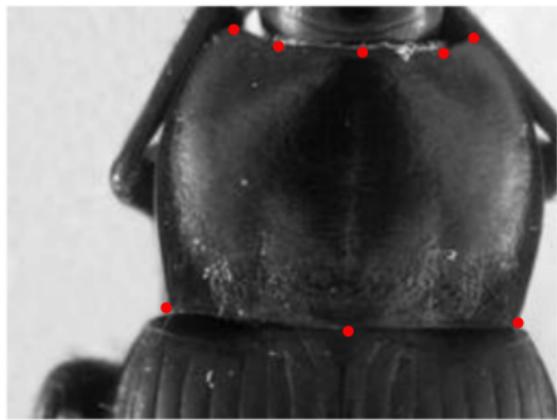
¹An epoch is a single pass through the full training set.

First result

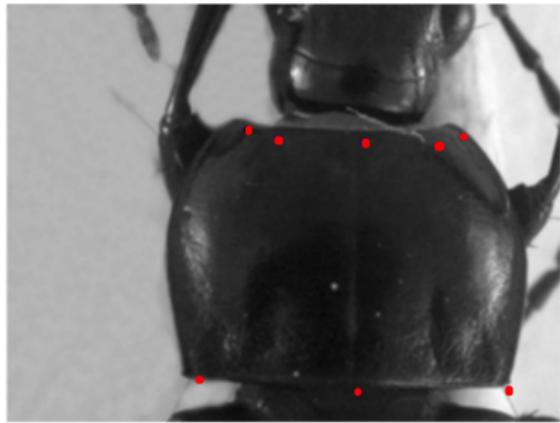
Landmarks on images



- ▶ Run the trained model to predict the landmarks on testing images,
- ▶ Calculate the distance between predicted landmarks and corresponding manual landmarks,



(a)



(b)

First result

Correlation metrics and average distances



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

Metric	r^2	EV	Pearson
Proposed architecture	0.9952	0.9951	0.9974

First result

Correlation metrics and average distances



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

Metric	r^2	EV	Pearson
Proposed architecture	0.9952	0.9951	0.9974

- ▶ The average distances 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

Transfer learning



- ▶ Re-uses model developed for a specific task/dataset to lead another task with another dataset
- ▶ **Fine-tuning:** retrain a pretrained model
- ▶ **Model Zoo** (Caffe library): people share their network weights.

TRANSFER OF LEARNING



The application of skills, knowledge, and/or attitudes that were learned in one situation to another **learning** situation (Perkins, 1992)

Fine-tuning our model

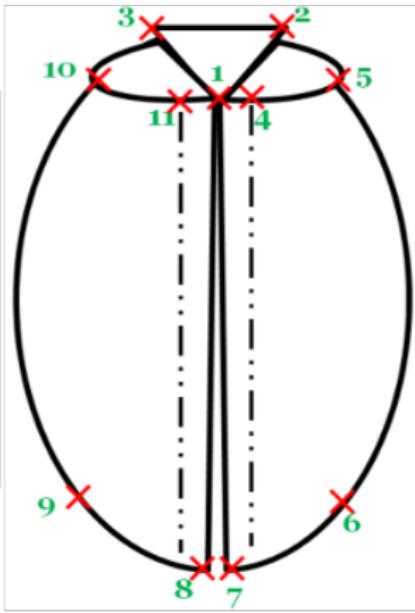
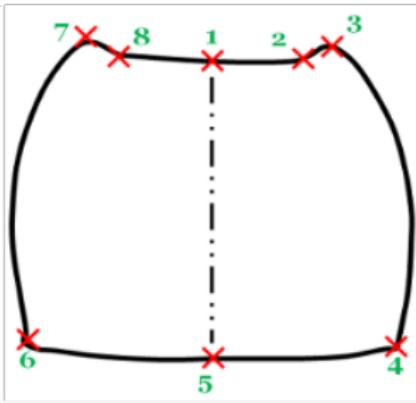
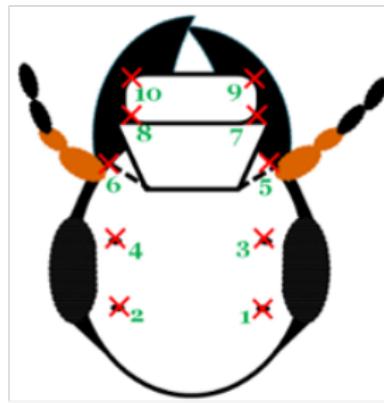


- ▶ Train the model on a dataset including the images of 3 parts of beetles (5,460 images)
- ▶ Fine-tune pretrained model on pronotum dataset

Fine-tuning our model



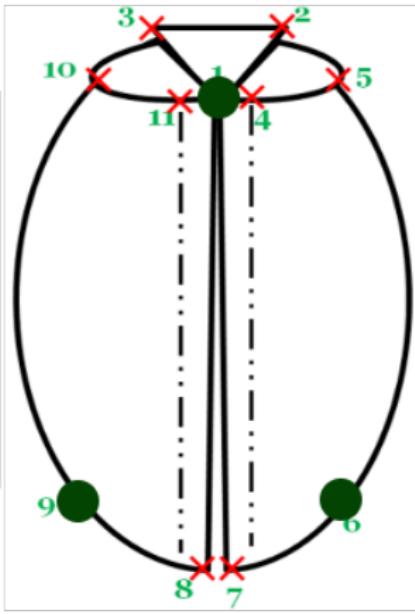
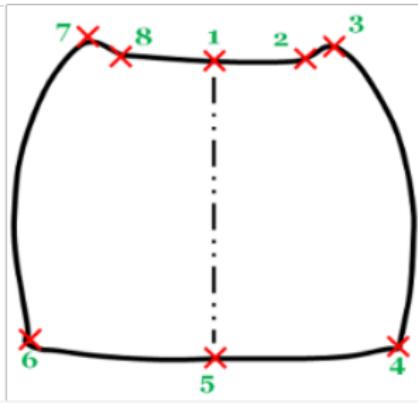
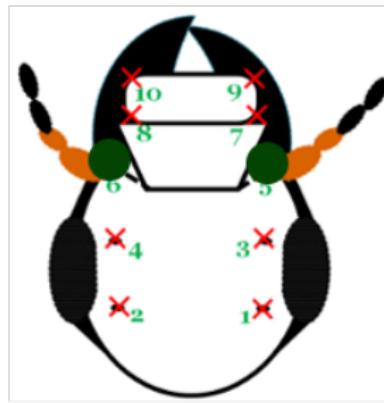
- ▶ Train the model on a dataset including the images of 3 parts of beetles (5,460 images)
- ▶ Fine-tune pretrained model on pronotum dataset



Fine-tuning our model



- ▶ Train the model on a dataset including the images of 3 parts of beetles (5,460 images)
- ▶ Fine-tune pretrained model on pronotum dataset



Results

Average distance comparing



The average distances of all images per landmark.

Landmark	Distance (in pixels)
1	2.486
2	2.7198
3	2.6523
4	2.7709
5	2.4872
6	3.0492
7	2.6836
8	2.8709

Results

Average distance comparing



Comparing the average distances between two processes (training from scratch and fine-tuning).

Landmark	From scratch		With fine-tuning	
	Average	SD	Average	SD
LM1	4.002	2.5732	2.486	1.5448
LM2	4.4831	2.7583	2.7198	1.7822
LM3	4.2959	2.7067	2.6523	1.8386
LM4	4.3865	3.0563	2.7709	1.9483
LM5	4.2925	2.9086	2.4872	1.6235
LM6	5.3631	3.4234	3.0492	1.991
LM7	4.636	2.8426	2.6836	1.7781
LM8	4.9363	3.0801	2.8709	1.9662

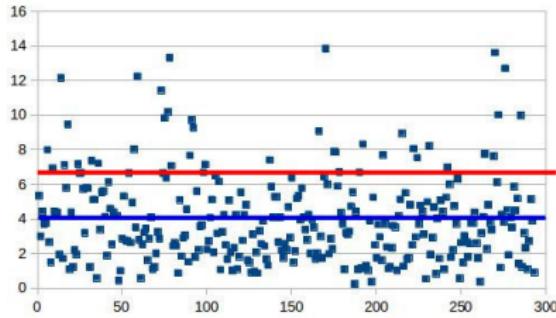
Results

Distribution of average distances

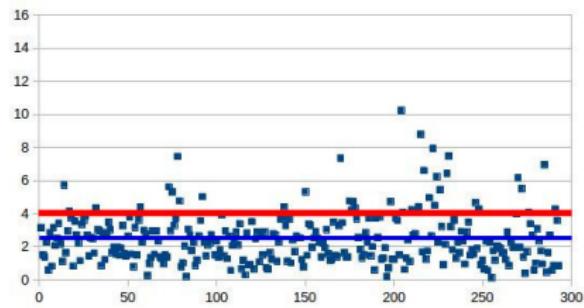


15

- The distribution of distance of the best result (1^{st} landmark)



(a) Training from scratch



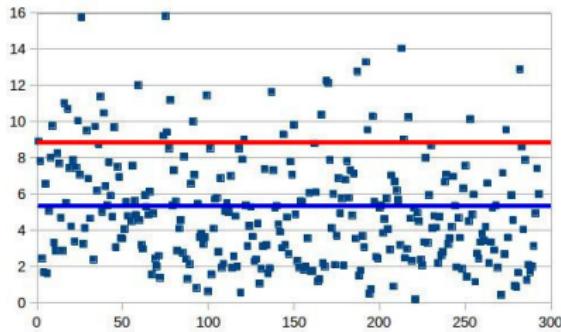
(b) With fine-tuning

Results

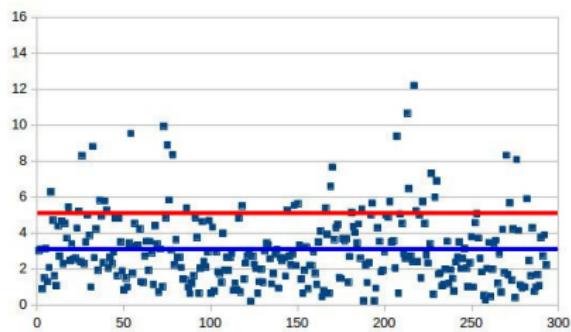
Distribution of average distances



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



(a) Training from scratch



(b) With fine-tuning

Conclusion



Conclusion

- ▶ Propose a CNN to predict the landmarks on pronotum images.
- ▶ Propose procedure to augment the dataset.
- ▶ Apply fine-tuning to improve the quality of predicted landmarks.
- ▶ The last of the predicted landmarks leads to satisfying results without a need for segmentation step. It allows replacing manual landmarks.

Conclusion



Conclusion

- ▶ Propose a CNN to predict the landmarks on pronotum images.
- ▶ Propose procedure to augment the dataset.
- ▶ Apply fine-tuning to improve the quality of predicted landmarks.
- ▶ The last of the predicted landmarks leads to satisfying results without a need for segmentation step. It allows replacing manual landmarks.

Future works

Continue improving the landmarks coordinates by going deeply how to characterize the learning problem to design the right pre-training set.



Thank you for attention!