Training model on others input of pronotum datasets

LE Van Linh

April 2, 2018

Abstract

In this study, we use the network model that we had submitted to ICPRS-18 to find the results on the new size of the input images. Instead of using the images with the size of 256×192, we have chosen a "square" size for the images. The new sizes of images which used to study are 224×224 and 96×96. The model will be trained on each dataset in 10000 epochs. The experiments will be compared the losses during training, statistic analysis and time-consuming. In the context of this study, we used term **"the last result"** to mention the result what we have submitted to ICPRS-18.

1 Introduction

In the last result, we have proposed a network to predict the landmarks on pronotum images. It receives an image of $1 \times 256 \times 192$ as the input. The network was constructed from 3 *"elementary block"* following by 3 full-connected layers. An elementary block is defined as a sequence of convolution (C_i), pooling (P_i) and dropout (D_i) layers. The parameters for each layers are as below, the list of values follows the order of elementary blocks:

- CONV layres:
  - Number of filters: 32, 64 and 128,
  - Kernel filters size: $(3 \times 3), (2 \times 2)$, and (2×2)
  - Stride values: $1, 1, 1$
  - No padding is used for CONV layers
- POOL layers:
  - Kernel filters size: $(2 \times 2), (2 \times 2)$, and (2×2)
  - Stride values: $2, 2, 2$
  - No padding is used for CONV layers
- DROP layers:
  - Propabilities: $0.1, 0.2$ and 0.3.

In the last full-connected layers (FC), the parameters are: FC1 output: 1000, FC2 output: 1000, FC3 output: 16. As usual, a dropout layer is inserted between FC1 ond FC2 with a probability equal to 0.5. Fig.1 illustrate the order of the layers in the network.

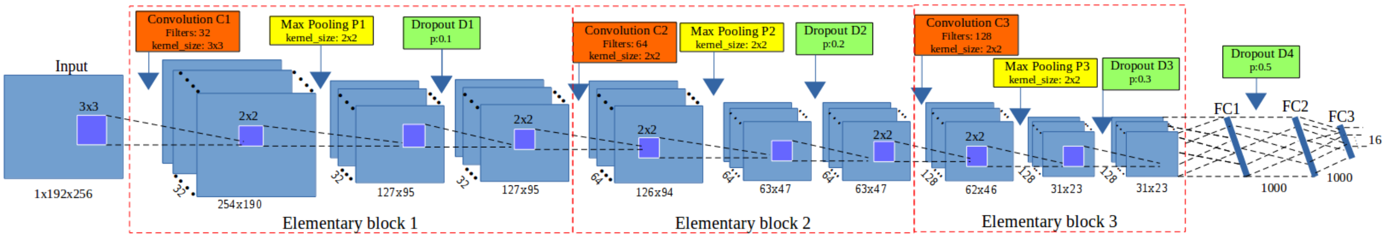


Figure 1: Network architecture using 3 *elementary blocks*. Convolution layer in red, pooling in yellow and dropout in green color.

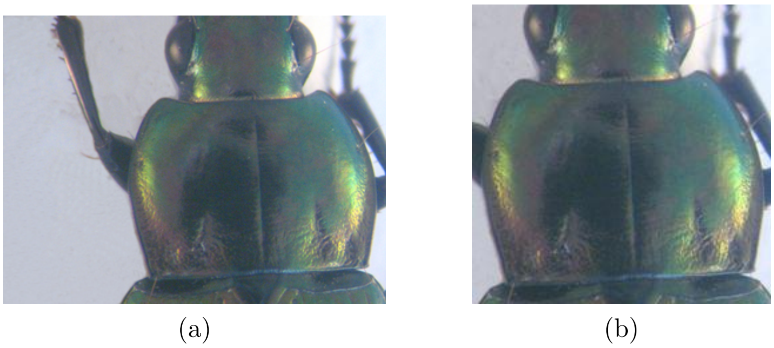
In our study, instead of using a *"square"* image, we have used a *"rectangle"* image as the input. The result shows that the network has the ability to work well on the input that size of width and height are different. Considering as a different of workflows, we would like to see how the network works on "square" input. In this study, we continue train the proposed network () on 2 new dataset with different size: 224 × 224 and 96 × 96.

2 Dataset

The dataset includes 293 RGB-images of beetle's pronotum. The images were taken by the same camera with the same conditions of resolution of 3264×2448. The images in the dataset were divided into two subsets: training (including 260 images) and testing (including 33 images). For each image, a set of 8 manual landmarks have been set by biologists. In this section, we introduce the process to down-sample the original images to the new size of images. Then, we augment the images for training and validation (it have been presented in ICPRS-18).

2.1 DATA_1: size of 224 × 224

To obtain a new size of the input image (224×224) from the original image, we have applied procedure as followed: Firstly, the images are down-sampled to a resolution of 326 × 245 and the coordinates of the manual landmarks are also re-scaled. Secondly, the centroid point of manual landmarks is calculated for each image. The centroid point is considered as the center of the new image that has been cropped from the down-sampled image. The size of the new image is fixed as 224 × 224. From the centroid point, we expand in four directions of the image until satisfying the size. Then, the coordinates of manual landmarks are re-calculated to fit with the new image. Fig. 2 presents an example in dataset after down-sampling and crop the image to fit with the input size of the neural network.



(a)

(b)

Figure 2: An example in dataset. *a)* presents the image after down-sampling. *b)* presents the cropped image from down-sampling image which used as the input of CNN

2.2 DATA_2: size of 96 × 96

In the other side, to obtain the images size of 96 × 96, we have applied the procedure following:

1. Left crop image to obtain the new size of 2448 × 2448,
2. Down-sample the image to size of 96 × 96,
3. Scale the coordinates of the manual landmarks to adapt with the new size of images.

3 Experiments

The proposed network has been trained on two datasets with different of input size: 224 × 224 and 96×96 in 10000 epochs. We have applied cross-validation to obtain fully the prediction of landmarks of all images. Table.1 shows losses during training on each round. For dataset of 224 × 224 images, the average loss of validation stage (RMSE) is $sqrt(0.0000767) \times 112 = 0.981$ pixels; while, this value for dataset of 96 × 96 is x pixels.

Round	224 × 224		96 × 96	
	Train loss	Validation loss	Train loss	Validation loss
1	**0.00012**	**0.00009**	**2.486**	**1.5448**
2	0.00012	0.00006	2.7198	1.7822
3	0.00012	0.00007	2.6523	1.8386
4	0.00012	0.00009	2.7709	1.9483
5	0.00013	0.00010	2.4872	1.6235
6	**0.00012**	**0.00006**	**3.0492**	**1.991**
7	0.00013	0.00008	2.6836	1.7781
8	0.00012	0.00006	2.8709	1.9662
9	0.00013	0.00008	2.8709	1.9662

Table 1: A comparing of losses during training between the datasets.

Table.2 shows the average distances in pixels and standard deviation (SD) when we calculate the distance from the coordinates of predicted landmarks to coordinates of manual landmarks.

Landmark	224 × 224		96 × 96	
	Avg distance	SD	Avg distance	SD
1	**3.2974**	**2.2689**	**0**	**0**
2	3.9845	2.6562	0	0
3	3.4676	2.3424	0	0
4	3.8779	2.7883	2.7709	1.9483
5	3.5482	2.5824	2.4872	1.6235
6	**4.2235**	**3.2746**	**3.0492**	**1.991**
7	3.3834	2.3783	2.6836	1.7781
8	4.0352	2.8116	2.8709	1.9662

Table 2: A comparing of average distance and SD on two dataset.

4 Conclusions

In this study, we have trained the model on pronotum dataset but the size of images are changed to 224 × 224 and 96 × 96. This study shows that the losses during training are the same with the lass one, but the average distance between still higher. However, when we training with small input, the training time had improved than with the large input.