A ResNet-18 Network to Estimate the Ball Diameter in Basketball Images

Adrien Maglo
adrien.maglo@cea.fr

Astrid Orcesi
astrid.orcesi@cea.fr
Université Paris-Saclay, CEA, List
Palaiseau, France

Quoc-Cuong Pham
quoc-cuong.pham@cea.fr

ABSTRACT

This technical report describes the method we developped to participate to the MMSports 2022 ball 3D localization challenge. Our simple approach uses a ResNet-18 convolutional neural network pretrained on ImageNet to estimate the ball diameter. It is trained by regressing the diameter with a L1 loss. A random rotation is applied to the input images during training to increase the sample variety. Compared to the challenge baseline, our method achieves a mean absolute diameter error about 41 % lower on the test dataset.

1 INTRODUCTION

For the MMSports 2022 workshop, a ball 3D localization challenge for basketball is organized. Given the position of the ball center in the input frame, the aim is to estimate its diameter in pixels.

We describe in this report a simple approach that regresses the ball diameter with a ResNet-18[1] network.

2 PROPOSED METHOD

2.1 Input data augmentation

To estimate the ball diameter, our model takes as input 100 x 100 pixels images centered around the ball. In order to increase the variety of images during the network training, we randomly rotate them with a bicubic interpolation. 50 x 50 pixels center crops of the images are then extracted. They are resized to 224 x 224 pixels with a bilinear interpolation because the ResNet-18[1] layers of the network have been pretrained with this input resolution. Input images are finally normalized before being fed to the network.

2.2 Network design

Our model is based on a ResNet-18 convolutional neural network [1]. It extracts features at the 7 x 7 resolution with a depth dimension of 512. A 2D Adaptive average pooling reduces these features to a vector of dimension 512. A final linear layer converts the output dimension from 512 to 1 to obtain the ball diameter.

2.3 Training

Our model is trained to regress the ball diameter with a L1 loss:

$$l = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{d}_i - d_i \right|$$

where n is the batch size, d_i is the ground truth ball diameter, $\hat{d}_i$ is the estimated diameter.

	MADE	MAPE	MARE
Our method	**1.26**	**168.23**	**0.06**
Challenge baseline	2.12	305.71	0.10

Table 1: Results on the test dataset of the MMSports 2022 ball 3D localization challenge challenge.

3 EXPERIMENTS

3.1 Dataset

The ball 3D localization challenge is based on the DeepSportRadar Basketball Instants Dataset [5]. It contains 322 training images, 42 testing images and 35 challenge images for which the ball is visible. We do not use images for which the ball size is not set.

3.2 Implementation details

We implemented our method with the Pytorch framework. The model is trained during 200 epochs with the AdamW optimizer and a batch size n of 4. The learning rate is set to 10^{-4} for the first 133 epochs and then divided by 10.

The model layers before the 2D Adaptive average pooling have been pretrained on ImageNet [2]. The weights are provided by the Timm library [4]. The other layers are initialized with default Pytorch initialization.

3.3 Metrics

The Mean Absolute Diameter Error (MADE) between the estimated diameter and the ground-truth is used to rank the methods submitted to the challenge. The Mean Absolute Projection Error (MAPE) and the Mean Absolute Relative Error (MARE) are also computed by the evaluation server for information. These metrics are described in [3].

3.4 Results

The results of our method are presented in the table 1. Our model performs better than the challenge baseline for the three metrics. The MADE is reduced by about 41 %.

REFERENCES

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[2] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.

[3] Gabriel Van Zandycke and Christophe De Vleeschouwer. 2022. Ball 3D localization from a single calibrated image. *arXiv preprint arXiv:2204.00003* (2022).

[4] Ross Wightman. 2019. PyTorch Image Models. https://github.com/rwightman/pytorch-image-models. https://doi.org/10.5281/zenodo.4414861
[5] Gabriel Van Zandycke, Vladimir Somers, Maxime Istasse, Carlo Del Don, and Davide Zambrano. 2022. DeepSportradar-v1: Computer Vision Dataset for Sports Understanding With High Quality Annotations. In *5th Int. Workshop on Multimedia Content Analysis in Sports (MMSports'22) @ ACM Multimedia 2022*. https://openreview.net/forum?id=plAwG0ZkbVy