Fraunhofer
PORTUGAL

RESEARCH CENTER FOR ASSISTIVE INFORMATION AND COMMUNICATION SOLUTIONS

CLUSTER-BASED ANCHOR BOX OPTIMISATION METHOD FOR DIFFERENT OBJECT DETECTION ARCHITECTURES

INTRODUCTION

Many deep learning detection architectures propose object candidates based on anchor boxes – bounding box templates extracted at specific locations of the feature map of the convolutional neural networks (CNNs) [1]. As the anchors' properties determine the object shapes and scales recognized by the model, they must be carefully defined and adjusted to the dataset used.

METHODOLOGY

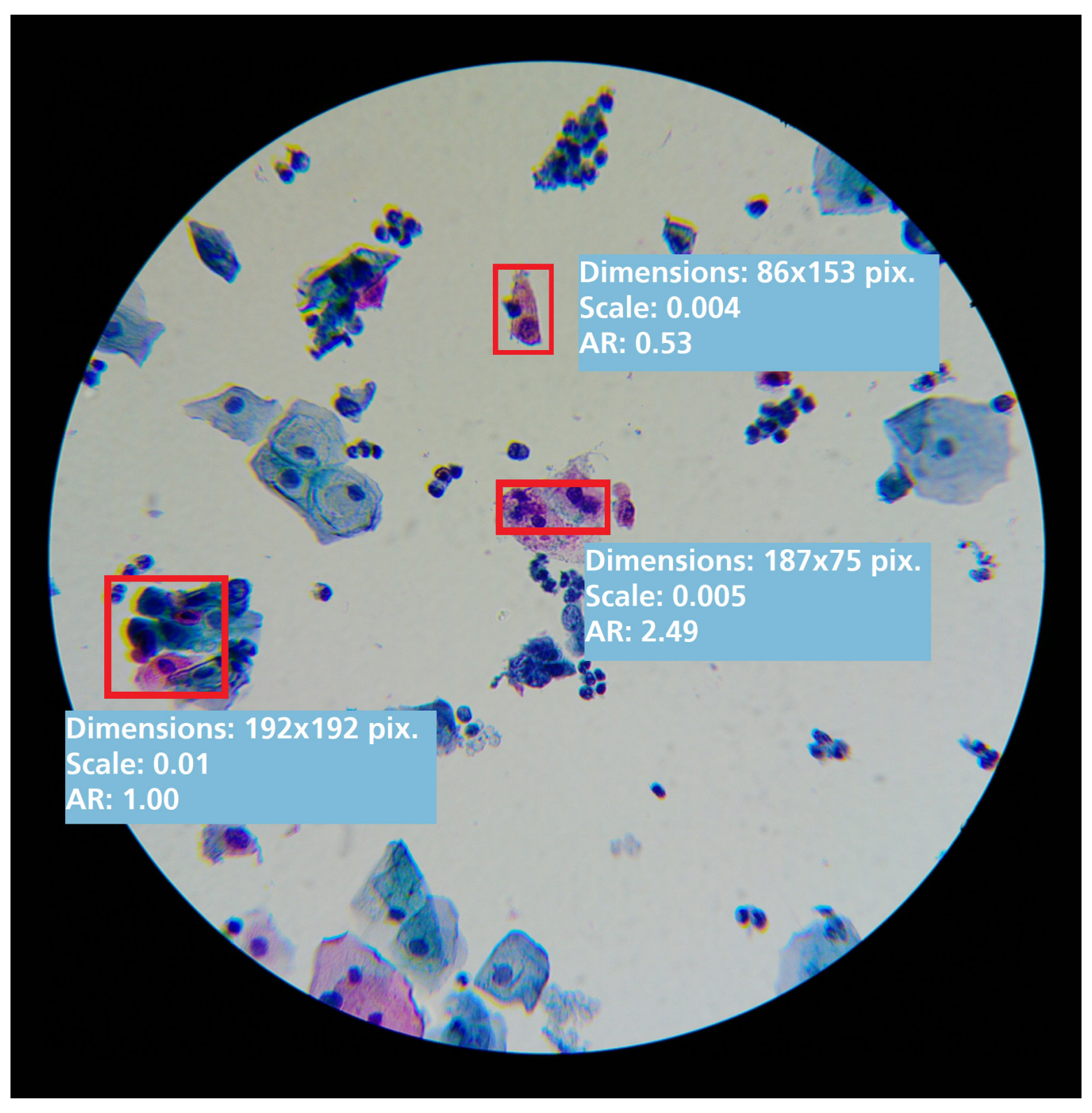
This work presents a methodology to adjust the anchor properties to the type of objects existing in a specific dataset. Clustering is used to identify the most representative bounding box sizes and shapes present in the dataset, which are mapped to the parameters of four state-of-the-art object detection CNNs.

Considering the design differences of those networks, in particular the amount of feature maps used for anchor extraction, this approach applies *k*-means clustering in 3 domains:

1. **Width and height dimensions:** the main width/height combinations are directly used to define the anchor boxes in the YOLO model, and the average intersection over union (IoU) between the cluster centres and the dataset's objects is maximized to find the optimal number of anchors.
2. **Scales:** computed as the area ratio between the annotated objects and the input image, they establish the size of the anchors extracted in the SSD, Faster R-CNN and RetinaNet algorithms; the optimal number of scales is found by minimizing the within-cluster sum-of-squares distance, representative of the intra-cluster variability.
3. **Aspect ratios (ARs):** obtained by dividing the bounding box widths by their heights, the ARs determine the main object shapes detected by the SSD, Faster R-CNN and RetinaNet architectures. The ARs are selected by simultaneously minimizing the intra-cluster variability and maximizing the inter-cluster separation, to ensure that the model is able to detect a sufficiently diverse set of object shapes

The chosen **number of clusters** is established considering the **trade-off between the optimization of error metric selected and the implied computational burden of the model.**

RESULTS



The proposed approach was validated using a private dataset comprised of 1489 microscopic images acquired from liquid-based cervical cytology samples of 21 patients with a µSmartScope device [2]. This dataset includes 2436 bounding box annotations of abnormal regions (indicative of cervical lesions), illustrated in Fig. 1.

Fig 1. Image from the cervical cancer dataset with the bounding boxes of abnormal cells outlined (in red) and the information of the respective dimensions, scales and aspect ratios.

The results for the tested *k* values are presented in Fig. 2, and the final anchor dimensions, scales and ARs - obtained from the coordinates of the corresponding cluster centers for the selected *k* – are included in Table 1.

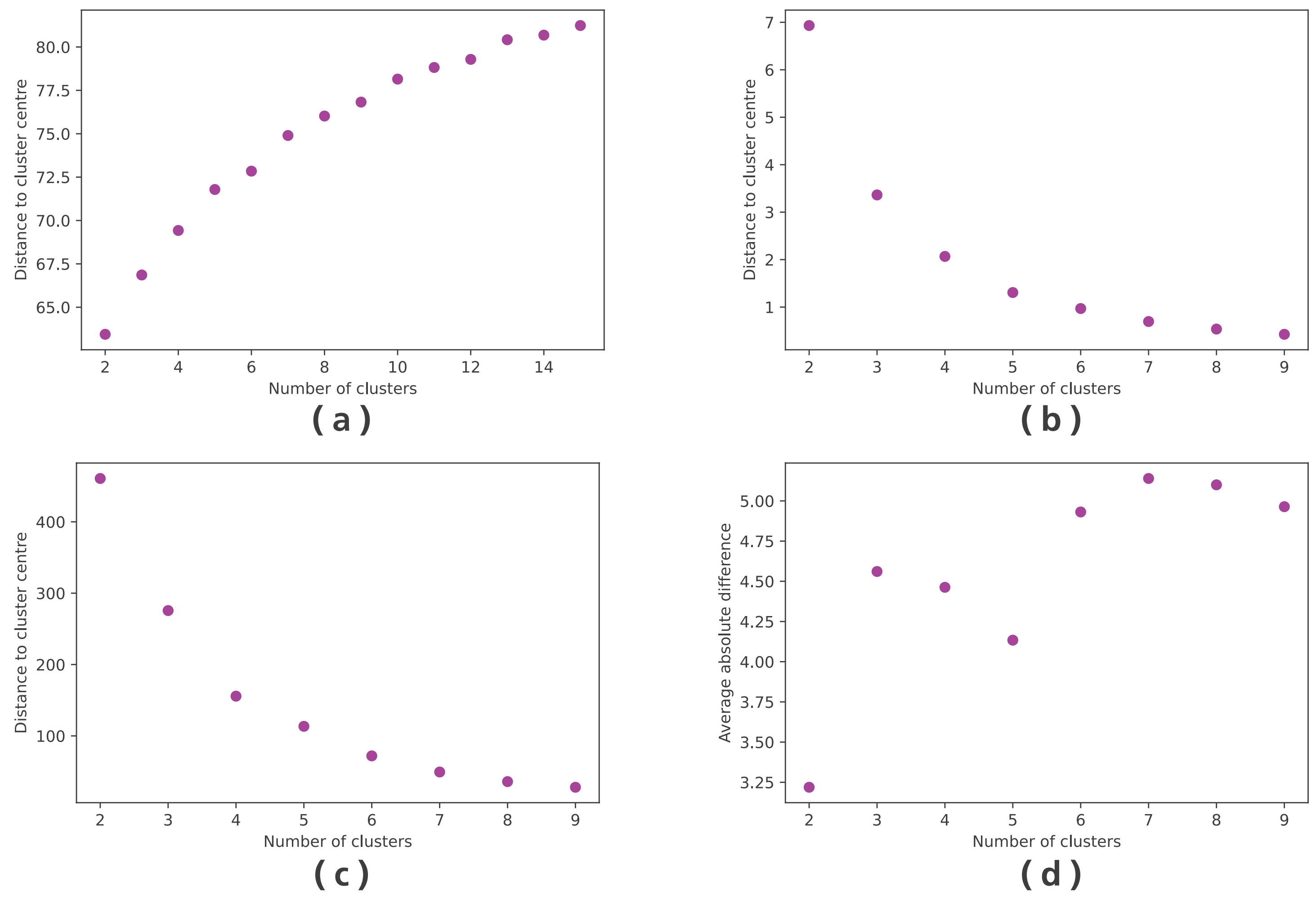


Fig 2. Graphical representation of the optimised metrics according to the *k* value used in the experiment: (a) average IoU, for the width/height clustering; within-cluster sum-of-squares distance for the (b) scale and (c) aspect ratio values; (d) average absolute difference among the aspect ratio values in each set.

Table 1. Final anchor dimensions, scales and aspect ratios for the selected number of clusters

CLUSTERED VARIABLE	NUMBER OF CLUSTERS (K)	FINAL VALUES
ANCHOR DIMENSIONS	9	(0.36,0.38), (0.30,0.20), (0.15,0.39), (0.19,0.19), (0.3,0.58), (0.30,0.29), (0.23,0.28), (0.67,0.67), (0.57,0.28)
SCALES	6	0.06, 0.13, 0.25, 0.44, 0.66, 0.91
ASPECT RATIOS	6	0.68, 1.18, 1.90, 3.63, 7.47, 14.55

CONCLUSIONS AND FUTURE WORK

This work presents a method to enable a more targeted object localisation in detection networks, achieved through the adjustment of the anchor boxes to the properties of the dataset used.
Nonetheless, the experiments reported still correspond to exploratory work. Future tests should include the examination of the impact of the anchors' setup in the final detection performance through the comparison of the adjusted anchor settings with the default ones, as well as a characterization of the computational burden yielded by some of the possible anchor configurations. Different clustering approaches, as well as more informative distance metrics for cluster validation, should also be explored.

Ana Filipa Sampaio[1], João Gonçalves[1], Luís Rosado[1], Maria João Vasconcelos[1]

1 Fraunhofer Portugal Research Center for Assistive Information and Communication Solutions, Rua Alfredo Allen, 455/461, Porto, Portugal - {ana.sampaio, joao.goncalves, luis.rosado, maria.vasconcelos}@fraunhofer.pt

ACKNOWLEDGEMENTS

This work was done under the scope of "CLARE: Computer-Aided Cervical Cancer Screening", project with reference POCI-01-0145-FEDER-028857 and financially supported by FEDER through Operational Competitiveness Program – COMPETE 2020 and by National Funds through Fundation for Science and Tecnhology FCT/MCTES.

FCT Fundação para a Ciência e a Tecnologia
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

[1] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review"
[2] L. Rosado, P. T. Silva, J. Faria, J. Oliveira, M. J. M. Vasconcelos, D. Elias, J. M. C. da Costa, and J. S. Cardoso, "µSmartScope: Towards a fully automated 3d-printed smartphone microscope with motorized stage," in Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science, pp. 19–44,Springer International Publishing