Attention-driven Spatial Transformer Network for Abnormality Detection in Chest X-Ray Images

Joana Rocha, Aurélio Campilho & Ana Maria Mendonça.
INESC-TEC and Faculty of Engineering, University of Porto.

Introduction

Automated abnormality detection in chest X-ray (CXR) images can be useful to prioritize certain exams. The presence of image artifacts in these images often leads to a harmful bias in the classifier, by increasing false positive results. Consequently, healthcare would benefit from a system that selects the thoracic region of interest prior to deciding whether an image is possibly pathologic.

Methods

Three main experimentation settings were defined and evaluated on the CheXpert dataset. The first two are used as control for comparison purposes with the third and main setting, which is the proposed method.

1. Baseline binary classifier (VGG16), minding the “normal” and “abnormal” classes.
2. Thorax detection model (YOLOv5) followed by the baseline classifier. Models are trained separately.
3. Spatial Transformer Network (STN) composed of an attention-driven ST module and the baseline classifier.

Results

- The STN classifier outperformed both baseline settings, reducing the FP rate more than the YOLO model.
- The YOLO model is trained independently of the classifier, requiring spatial annotations. The STN is trained end-to-end in a spatially unsupervised way.
- The STN does not preserve original image proportions unlike the YOLO, but is more precise at removing artifacts from the images.

Conclusions

- An STN whose attention-driven module and classifier are trained end-to-end to crop and classify the CXR scans.
- A reduced computational cost in comparison to the typical object detection models, which require localization ground truth annotations and larger architectures and training times.

Acknowledgements

This work was funded by the ERDF - European Regional Development Fund, through the Programa Operacional Regional do Norte (NORTE 2020) and by National Funds through the FCT - Portuguese Foundation for Science and Technology, I.P. within the scope of the CMU Portugal Program (NORTE-01-0247-FEDER-045905) and LA/P/0063/2020. The work of J. Rocha was supported by the FCT grant contract 2020-06595/BD.