Deep Learning for Combined Lesion Detection and Classification in Prostate MRI

Lewis Smith¹
Rohan Mirchandani¹
Russel Demos¹
Elliot Smith¹
Becks Simpson¹

¹ AI Development, Maxwell Plus

Editors: Under Review for MIDL 2019

Abstract

Prostate cancer cases increase every year and multi-parametric MRI is becoming commonplace for diagnosis and staging, meaning that a computer-aided detection system is more important than ever to assist in clinical decision support. Recent advances in deep learning architectures for object detection have been applied to identifying abnormalities in medical imaging to achieve such an aim. This study presents a model based on the 3D context enhanced deep convolutional neural network adapted to process prostate multi-parametric MRI and also predict the clinical significance of candidate lesions. Combining both steps into a single model reduces complexity and simplifies integration to clinical workflow. Results indicate this is a successful approach, with an AUC of 0.89 and 0.891 for malignant and benign lesion classification.

Keywords: prostate cancer, magnetic resonance imaging, detection, classification

1. Introduction

Prostate cancer is the second most commonly occurring cancer in men, with an approximated 1.3 million new cases worldwide (Alhosseini and Analoui, 2006). Globally, it is the third leading cause of cancer-related deaths in men in developed countries and sixth for those in developing countries. Although the survival rate is higher if the cancer is found early, the traditional methods of screening such as the prostate specific antigen (PSA) blood test and the digital rectal examination (DRE) are fraught with issues such as high rate of false positives (for PSA) and likelihood of missing critical lesions depending on their location in the prostate (DRE) (Brawer and Kirby, 1999). Increasingly, MRI is being used as an integral part of prostate cancer diagnosis to more accurately detect and stage prostate lesions prior to biopsy. However, multi-parametric MRIs are difficult to read, requiring the expertise of a radiologist specifically trained to identify lesions (Alhosseini and Analoui, 2006). As such, there has been increasing effort to assist clinicians to automate the process of lesion detection and classification of clinical significance (Alhosseini and Analoui, 2006; Yan et al., 2018; Tsehay et al., 2017; Liu et al., 2017).

With improvements in computation power and deep learning advancements in object detection, these methods have been applied to lesion detection in prostate multi-parametric MRI. In particular, (Tsehay et al., 2017) used a convolutional neural network architecture
to detect prostate lesions to present to clinicians as a reference for their own analysis. Compared with previous machine learning based approaches which used hand-engineered features to distinguish between cancerous and non-cancerous abnormalities, their model achieved higher detection rate at 94% compared to 85% (both with a rate of 10 false positives per patient). This demonstrates the utility of deep learning for this task, however the analysis was performed on a small cohort of 52, focused on 2D image slices rather than including 3D context and relied on segmented lesions making it difficult to obtain a large enough dataset.

Instead of requiring fully annotated lesions, another approach is to use object detection methods with bounding boxes as inputs, as in (Yan et al., 2018). The approach of using a 3D context enhanced deep convolutional neural network also used slices preceding and following a lesion-positive slice to incorporate the 3D nature of medical imaging. This study was performed on the DeepLesion dataset containing 32k CT scans with a sensitivity of 84.37% (with 4 false positives per image). Given the success of this method as a universal lesion detection algorithm, it can also be extended for use on multi-parametric MRI and to predict whether a lesion is clinically significant to aid clinical decision support.

2. Methodology

This section covers the dataset, pre-processing, model architecture and metrics.

2.1. Dataset and Pre-Processing

The model was trained on 2626 multi-parametric MRI cases from a combination of the PIE-AAPM-NCI Prostate MR Classification Challenge (ProstateX) and a private clinical collaborator, containing 3042 lesions with at least one lesion per case. Of those lesions, 291 were labelled as clinically insignificant, 379 as clinically significant (Gleason Score greater than 6 (Yan et al., 2018)) and 2372 as unknown. Lesions labelled as unknown were only used for detection training and were not included in the validation stage. Labels also included lesion centre points and regional information used to estimate a bounding box. The model required a combination of a T2W axial series and an ADC series. Pre-processing included center cropping scans around the prostate using a segmentation map, registering the ADC series onto the T2W and extracting cubes of three slices prior and following a slice with a lesion present. Online image augmentation was used to improve the variance of the dataset. Inputs where randomly flipped and deformed at train time.

2.2. Metrics

The performance of the model was measured with several metrics. In addition to the standard AUC, precision and recall metrics, an additional term called the Intersection Over Union (IOU) or Jaccard Index was used as a measure of similarity between predicted bounding box of a given class to the ground truth, as in Equation (1). A prediction was counted as a true positive if it had an IOU greater than 0.5 with a ground truth box of the same class.

\[
J(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cup B|}
\]
2.3. Architecture

The model was initially implemented based on the architecture presented in (Yan et al., 2018) but was modified to improve suitability for prostate MRI images. The reason for selecting the region-based fully convolutional network is, as the authors in (Yan et al., 2018) also note, because it is faster, more accurate and more memory efficient than the so-called ‘Faster R-CNN’ (Ding et al., 2017; Ren et al., 2015). Our model utilized two truncated VGG heads, for T2 and ADC series respectively. As in (Yan et al., 2018) each modality was stacked in the batch and channels creating a $[7,256,256,3]$ input centered around a slice with a lesion - suitable for input to a VGG model. The outputs of two heads where concatenated to create one unified feature map. This was done to try extract unique features from each modality. Moreover, the output of the model changed to a two class detection problem in an attempt to separate malignant and benign lesions. The updated architecture is detailed in Figure 1.

![Figure 1: Model architecture modified from (Yan et al., 2018)](image)

3. Results

From the combined data set, 550 cases were selected for testing and excluded from the training data. On this test set, the model achieves an average AUC of 0.89 over both classes, which shows that the model is able to generalize well over different data in terms of classification. At IOU >0.5 the mean average precision (mAP) was calculated at 0.76 with an average of 0.175 false positive per scan, demonstrating that the large dataset helped immensely with the detection task.

4. Conclusion

In conclusion, the performance of this model shows that a combined approach of detection and classification is suitable for the problem of prostate cancer identification and clinical decision support for diagnosis. Combining both steps into one model reduces the model complexity and the pipeline for integration into clinical workflow. Despite the lack of formal labels for the bounding boxes, use of radiological annotations and centre locations was sufficient to provide information for the model to train from. Future work will involve training and testing on a larger dataset, investigating the effects of combining additional sequences and extending the model to predict Gleason Scores rather than Gleason Bands.
References


