Evaluating Rotation Invariant Strategies for Mitosis Detection Through YOLO Algorithms

Dibet Garcia Gonzalez[1(✉)], João Carias[1], Yusbel Chávez Castilla[1], José Rodrigues[1], Telmo Adão[1,2], Rui Jesus[3], Luís Gonzaga Mendes Magalhães[2], Vitor Manuel Leitão de Sousa[4], Lina Carvalho[4], Rui Almeida[4], and António Cunha[5]

[1] Center for Computer Graphics, Guimarães, Portugal
dibet.gonzalez@ccg.pt
[2] Algoritmi Center, Minho University, Guimarães, Portugal
[3] University of A Coruña, BMD Software, A Coruña, Spain
[4] Institute of Anatomical and Molecular Pathology, Faculty of Medicine, University of Coimbra, Coimbra, Portugal
[5] Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

Abstract. Cancer diagnosis is of major importance in the field of human medical pathology, wherein a cell division process known as mitosis constitutes a relevant biological pattern analyzed by professional experts, who seek for such occurrence in presence and number through visual observation of microscopic imagery. This is a time-consuming and exhausting task that can benefit from modern artificial intelligence approaches, namely those handling object detection through deep learning, from which YOLO can be highlighted as one of the most successful, and, as such, a good candidate for performing automatic mitoses detection. Considering that low sensibility for rotation/flip variations is of high importance to ensure mitosis deep detection robustness, in this work, we propose an offline augmentation procedure focusing rotation operations, to address the impact of lost/clipped mitoses induced by online augmentation. YOLOv4 and YOLOv5 were compared, using an augmented test dataset with an exhaustive set of rotation angles, to investigate their performance. YOLOv5 with a mixture of offline and online rotation augmentation methods presented the best averaged F1-score results over three runs.

Keywords: Rotation invariance · deep learning · YOLO · mitosis counting

1 Introduction

Cancer is of great interest to medical specialists and pathologists because of its consequences on people's health. This disease affects populations everywhere and can be fatal if not treated on time. Nowadays, there are several ways to diagnose this pathology, involving the presence of several specialists (clinicians, pathologists) with a high level

A. Cunha et al. (Eds.): MobiHealth 2022, LNICST 484, pp. 24–33, 2023.
https://doi.org/10.1007/978-3-031-32029-3_3

of expertise. For example, there are strategies to characterize certain cancers, such as the Nottingham Grading System [1], which aids in the determination of breast cancer grade, resorting to three biomarkers: nuclear atypia, tubule formation and the mitotic cell count.

One of the known ways to assess tumor proliferation is mitotic counting, widely used by pathologists, who manually perform this task by analyzing biopsy Hematoxylin and Eosin (H&E) stained samples, from high-resolution microscope imagery. This process refers to the number of dividing cells visible in H&E stained histopathology [2], and it is an established step in cancer diagnosis and prognosis procedures [3]. In [4] authors note that "*the mitotic cell count is an important biomarker for predicting the aggressiveness, prognosis, and grade of breast cancer*".

Diagnosis can be influenced by multiple factors intrinsic to the biomaterial under analysis (e.g., small differences between mitotic and normal cells) or by the lack or absence of qualified professionals. Also, it is time-consuming, tiresome and subjective [4] which encourages the search for new diagnosis procedures. Digital decision support systems can help to overcome these problems. Advances in various domains such as image capture, image storage, image visualizing and deep learning algorithms allow the development of more robust and reliable decision support systems, speeding up the diagnosis, reducing the workload and supplying an opportunity to improve diagnosis and patient outcome [5].

Moreover, Whole Slide Images technology known as WSI, which allows scanning of conventional glass slides to produce digital image with high resolution [6], has been providing significant contributes for the development of digital pathology. Today, it is paving the way for DL algorithms (e.g. classification, object detection, segmentation) by capturing the required data (images).

Although deep learning-based algorithms have achieved superior results in recent years, there are challenges to overcome. For example, the digitalization methods output cell with assorted rotation angles but, generally, pathologists are successful at classifying independently of this condition. Thus, it is expected that any ideal mitosis detector would be robust to disturbances such as rotation or flip.

In this work, we present an evaluation of the You Only Look Once (YOLOv4 [7] and YOLOv5 [8]) algorithms in relation with rotation and/or flip transformations for the mitoses detection use-case. We consider the mitotic detection process useful for mitotic counting task, and, thereby, we trained several YOLO models, and evaluated their performances on an exhaustive rotated test dataset. Afterwards, according to our experiments, we proposed a suitable procedure to obtain a performant mitosis detection model regarding rotation invariance.

The main contributions of this work are:

- a rotation invariant strategy for mitosis detection for object detection algorithms, and
- a direct comparison between YOLOv4 and YOLOv5 considering the proposed strategy in the training stage.

The rest of this work is organized as follows: Sect. 2 includes background about the application of deep learning methods for mitotic counting; Sect. 3 presents the strategy used to train the YOLO models and the creation of datasets from two public set of

images; Sect. 4 shows some results; Sect. 5 elaborates on these results and conclusions are drawn in Sect. 6.

2 Previous Work

According to mitosis detection reviews [3, 9], there are three ways to address this problem:

1) using hand-crafted features: where morphology, color and texture features are provided, as input, to machine learning algorithms for pattern recognition or classification (e.g., Support Vector Machine, Artificial Neural Networks);
2) using deep learning methods: which is able to learn sub-visual image features that may not be easily discernible by the human eye [5]. Convolutional Neural Networks (CNN) are the most popular algorithms; and
3) using a combination of previous methods: where this strategy combines the speed of hand-crafted methods and the accuracy of CNN.

In [10], the authors explain how they carry out digital image processing and deep learning techniques in some use cases under the context of Digital Pathology (DP). One of these is mitosis detection, where they apply a blue-ratio segmentation technique combined with morphological operations (e.g., dilate using a 20 pixels disk radial mask) and deep learning for the classification task. In [11], the combination of hand-crafted features (morphology, color, and texture features) and CNN-derived features maximize the performance by leveraging the disconnected feature sets.

In [12], a framework for the analysis and prediction of tumor proliferation from histological images is proposed. The framework includes three modules: i) Whole Slide Image Handling, ii) Deep Convolutional Neural Networks based Mitosis Detection, and iii) Tumor Proliferation Score Prediction. The first module applies the Otsu thresholding and binary dilation method to extract the Region of Interest (ROI). The second uses a pre-trained ResNet model (128×128) to classify into two classes (mitosis or normal). The last includes a Support Vector Machine (SVM) with radial basis function (RBF) kernel using an RBF kernel for tumor proliferation prediction.

You Only Look Once, also known as YOLO, was created in 2015 [13] as an approach for object detection task and over the years it has continued to evolve, constantly improving its performance. This architecture identifies bounding boxes in a single shot regression approach. Version 4 includes improved feature aggregation, mish activation, bag of freebies with augmentation, and other improvements. Version 5 is faster and smaller than version 4, allowing it to be embedded in devices more easily. In training stage, YOLOv5 combines auto learning bounding box anchors with a data loader that make online augmentation (e.g., scaling, color space adjustments and mosaic augmentation).

YOLOv3 [14] was applied for mitosis detection reaching a F1-score equal to 0.8924. YOLOv4 was proposed in [15] to evaluate the efficacy of training a deep learning model. The authors used a private dataset which include 778 images of breast carcinoma slides from Sunshine Coast University Hospital. The model has a sensitivity of 0.92 and a positive predictive value of 0.64 but is frequently wrong for pyknotic cells.

3 Materials and Methods

In this article, the process of mitosis cell detection is considered as object detection task. In this sense, we present a multi-resolution pipeline including: i) a tissue detection step combining the Otsu thresholding technique and mathematical morphology on low resolution image, ii) a slide window approach with size 640 × 640 pixels on high resolution, and iii) application of object detection strategy over each patch on images containing tissue. Next, we present the image dataset used in this work and the experiments carried out to compare the deep learning models.

3.1 Initial Dataset

In this work, we use a dataset annotated at tiled patch level with dimension 640 × 640 pixels. We adjusted two datasets of breast cancer histological images mitos-atypia-14 [16] and TUPAC16 [17] for object detection tasks containing images that include at least one cell undergoing mitosis (See Fig. 1).

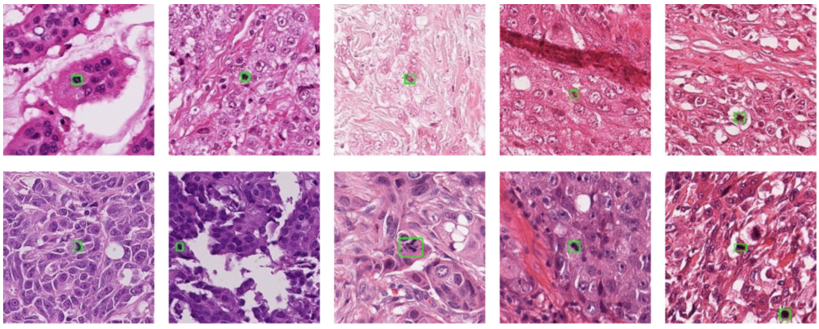

Fig. 1. Dataset samples of breast cancer histological images for object detection tasks. Image dimensions: 640 × 640 pixels. Five top images belongs to mitos-atypia-14 [16]. Five bottom images belongs to TUPAC16 [17].

Both datasets include images from 22 and 500 WSI respectively, which are scanned using two different scanners (Aperio and Hamamatsu). They were previously annotated using the center of the mitotic events. To adapt this dataset for object detection task, it was necessary to transform the center into a rectangular region in a manually way. Finally, the **base** dataset contains a total of 2502 images including one or more cells in mitotic events. The dataset was randomly partitioned into three sets: train (70%), validation (15%) and test (15%).

3.2 Rotation and Flip Augmentation Experiment

The application of automatic algorithms to detect mitosis should be rotation/flip invariant (i.e., the same mitosis should be detected independently from image rotation angle). Generally, to tackle the rotation/flip invariant object detection problems, the training dataset

is augmented with random rotations and flips. In this section, we make experiments to evaluate the rotation and flip response of the YOLOv4 and YOLOv5 algorithms.

The horizontal flip augmentation is a simple procedure that mirrors the pixels along a centered **y** axis. Depending on the object of interest (its symmetry), it may look entirely different to most object detection models so it's a very useful strategy to increase the dataset.

The rotation augmentation, while simple in itself, introduces some challenges. For the image rotation, either the size is changed or it remains the same but with clipping part of it. This means some mitotic events may be clipped or lost altogether. To keep all the identified mitosis and avoid its duplication, for each image and corresponding annotations, we apply the following procedure:

1. According to the dataset specification, apply a horizontal flip transformation and add to the initial dataset;
2. Rotate images by the objective rotation maintaining its original size;
3. If mitosis annotations fit inside the (possibly clipped) rotated image, insert this image into the dataset;
4. Else try to re-center the image fitting annotations considering its bounding boxes;
5. If all annotations are fitted inside the image, add the image and annotations to dataset;
6. Else, different pairs of images/annotations are created in order to have represented all the original mitosis; duplicated mitoses are erased (see Fig. 2).

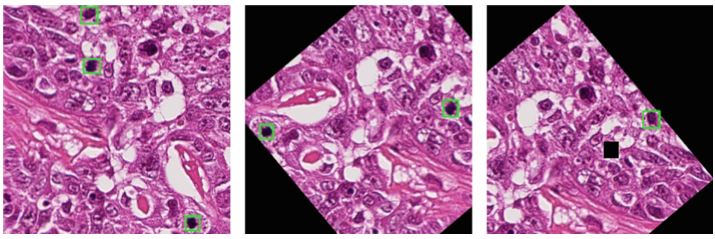

Fig. 2. Rotation augmentation example.

With the application of this procedure, an increase of the performance is expected since the augmentations allow the models to train on complete mitosis events, i.e. without clipping.

According with Fig. 2, the image on the left was flipped and rotated 50º, clockwise. Since a single image wasn't able to contain all the annotations, two images were generated. On the right image, since the left mitosis was already contained in the previous image, a black box was applied over the mitosis bounding box.

To apply the rotation in the previously mentioned step 2, three strategies for the rotation of annotated bounding boxes were considered (see Fig. 3). In the first strategy, the new bounding box is calculated by finding the minimums and maximums $\boldsymbol{x}$ and $\boldsymbol{y}$ coordinates of the corners of the rotated bounding box; in the second one, only the center

point of each side is considered; while in the third, each side is divided by 3 and only the midsection is considered for the new bounding box.

For the first strategy, the area of the bounding box (in some cases) increase significantly. For the second one, the inverse was observed, often clipping relevant areas of the target object. A reasonable middle ground was found in the third strategy keeping the relevant parts of the mitosis inside the new bounding box while minimizing the increase of its area size. In this work, we used the third strategy.

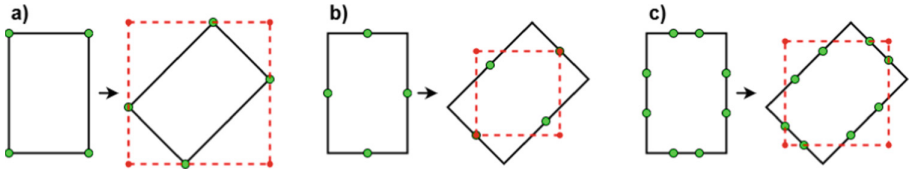

Fig. 3. New bounding box (red) for a 45º angle rotation. a) first strategy, b) second strategy and c) third strategy. (Color figure online)

The training and validation subsets of the dataset were augmented, in a case-by-case basis, with 2 different transformations: horizontal flip (referred as flip) and a set of rotations (see Table 1) following the third strategy. Concretely, for the "**16rots**" dataset, each image in the **base** dataset was rotated by a multiple of 22.5º, completing a 360º turn. For the "**flip/04rots**", each image was flipped and, for both flipped and un-flipped images, a set of a multiple of a 90º rotation was applied; the latter procedure was also applied to create the "**flip/36rots**" dataset but a multiple of 10º was used instead. Finally, the test subset, is transformed according with the "**flip/360rots**" parameters, i.e., flip and an extensive rotation of 1º, containing 270440 images occupying 170 GB.

Table 1. Details of the created datasets.

Dataset	Flip	Angle	Images Count	Goal
base	no	Not apply	1751	Train/Validation
16rots	no	22.5º	28036	Train/Validation
flip/04rots	yes	90º	14008	Train/Validation
flip/36rots	yes	10º	126168	Train/Validation
flip/360rots	yes	1º	270440	Test

3.3 Object Detection Models

From two object detection architectures, YOLOv4 and YOLOv5, we apply transfer learning technique using the pre-trained models: *yolov4-custom* and *yolov5x* respectively. Models trained on several datasets (see Table 1) with different online (i.e., during

training) data augmentation strategies (see Table 2). As the original YOLOv4 framework doesn't have the online rotation augmentation capability, we implement this feature in the context of this paper.

Table 2. Trained models.

Name	Architecture	Dataset	Online Flip	Online Rotation
v4	YOLOv4	**Base**	Not apply	Not apply
v5	YOLOv5	**Base**	Not apply	Not apply
v4/tf/tr180	YOLOv4	**Base**	Yes	180º
v5/tf/tr180	YOLOv5	**Base**	Yes	180º
v4/f/36	YOLOv4	**flip/36rots**	Not apply	Not apply
v5/f/36	YOLOv5	**flip/36rots**	Not apply	Not apply
v4/16/tf	YOLOv4	**16rots**	Yes	Not apply
v5/16/tf	YOLOv5	**16rots**	Yes	Not apply
v4/16/tf/tr11	YOLOv4	**16rots**	Yes	11º
v5/16/tf/tr11	YOLOv5	**16rots**	Yes	11º

When using flip during training, there's a 50% probability of the image being flipped. For both architectures, when online rotation is active, the model trains on a rotated image by a random angle of a given interval - possibly clipping or removing some cells in mitosis. The rotation intervals were chosen in order to cover, approximately, the entire circumference.

All models were trained on a single GPU (RTX 3080, with 10 GB) and thus, their batch and model size were constrained by the graphics card memory. For each architecture, the best training settings were found empirically and, from there, the only variations were the target dataset and the online data augmentations. For YOLOv4, the model weights were saved every 100 training steps; after training, the best version was found by comparing the F1-score calculated on the validation set (augmented the same way as the training set). In YOLOv5, the framework already saves, at every epoch, the best weights of a fitness function that targets the validation set, therefore the fitness function was changed to output the F1-score. Finally, the models were compared by calculating their F1-Score against the "**flip/360rots**" test subset. Each model comprised three instantiations (the same settings trained three times on the same dataset) and the final F1-Score resulting from their average.

4 Results

Table 3 shows a summary of the experiments results where we can observe a direct comparison between YOLOv4 and YOLOv5 considering previously defined datasets (Table 1) and training configurations (Table 2). Throughout the results, YOLOv5 shows better performance than YOLOv4, often requiring more time to train.

Table 3. Models results, inferenced on the "**flip/360rots**" test subset, averaged after 3 runs. The last column in the table (timestamp) represents the time to reach the best model, avoiding overfitting, considering the maximum F1-score on the validation subset.

Name	F1-score	Recall	Precision	mAP@0.50	Timestamp
v4	0.809	0.794	0.824	0.787	7h
v5	0.825	0.829	0.822	0.810	5h
v4/tf/tr180	0.844	0.832	0.857	0.855	4h
v5/tf/tr180	0.883	0.875	0.892	0.891	5h
v4/f/36	0.882	0.904	0.861	0.924	4h
v5/f/36	0.900	0.889	0.912	0.922	13h
v4/16/tf	0.883	0.908	0.860	0.927	4h
v5/16/tf	0.901	0.910	0.893	0.909	8h
v4/16/tf/tr11	0.876	0.870	0.882	0.917	6h
v5/16/tf/tr11	0.907	0.905	0.909	0.921	14h

From the comparison between the models **v4** vs **v4/tf/tr180** and **v5** vs **v5/tf/tr180** we observe that using online data augmentations lead to a better performance when used on a baseline dataset. Also, to evaluate the proposed offline augmentation procedure, we compare the results of the **v4/tf/tr180** vs **v4/f/36** and **v5/tf/tr180** vs **v5/f/36** where models trained on **flip/36rots** dataset show better F1-score results.

In addition, the comparisons between **v4/f/36** vs **v4/16/tf** and **v5/f/36** vs **v5/16/tf** show that a decrease in the offline augmentations (and the corresponding reduction in train dataset size) didn't affect the performance of both models.

Finally, **v5/t16/tf/tr11** achieved the top performance among all the models tested, this being a YOLOv5 model with online flip and $\pm 11^{\circ}$ rotation interval trained on offline augmented dataset with 16 multiples rotations of 22.5º, reaching a good balance between precision (0.909) and recall (0.905).

5 Discussion

YOLOv5, when compared with YOLOv4, has the benefits of having online augmented rotations, a valuable tool for a better performance in mitosis detection. Considering that both algorithms are open source, we decided to implement this functionality to guarantee an impartial comparison between them.

In our experiments, YOLOv5 usually needs more training time than YOLOv4. However, it performs better than YOLOv4 and can be considered a good choice for the mitosis detection use case.

Analyzing the results of applying offline and online data augmentation strategies, we demonstrate that both are beneficial for the use case of this work. An increasing number of offline rotation augmentations usually led to a higher performance on any model, except for YOLOv4 when changing the number of rotations from 16 to 36.

Our proposed offline rotation augmentation procedure is beneficial for mitosis detection, mainly because, otherwise, some mitoses may be clipped or disappear when the training process applies a large enough online random rotation augmentation, as in the cases of **v4/tf/tr180** and **v5/tf/tr180**. As the number of mitotic events that are localized in the periphery increase, the importance of this strategy also increases. On the other hand, having an increase amount of augmented offline training data directly leads to an increasing amount of occupied memory which, for big enough datasets, could be a challenge or even not practical. Moreover, the results obtained with datasets **flip/36rots** and **16rots** shows that a significant reduction in data size (from 126168 to 28036 images) doesn't translate to a relevant decrease in the model's performance (F1-score).

For YOLOv5, combining offline and online rotation augmentations, as in **v5/16/tf/tr11**, showed the best performance from all the models tested.

6 Conclusions

From these experiments, we conclude that the proposed offline augmentation procedure helps YOLO algorithms achieve a better detection response in presence of mitotic events. Moreover, a reasonably small amount of offline rotation augmentations and the complementing online default augmentations are a good combination for maximizing the F1-score. Considering the memory limitations of our system, we conclude, according to our experiments, that YOLOv5 offers the best performance for mitosis detection.

As future work and taking in consideration the best result of our experiments, we will explore if a better compromise of data size and performance can be achieved by combining other offline and online rotations augmentations.

Acknowledgements. This work was financed by the project "iPATH, An Intelligent Network Center for Digital Pathology" (N° POCI-01-0247-FEDER-047069 and CENTRO-01-0247-FEDER-047069), financed by Portugal 2020, under the Competitiveness and Internationalization Operational Program, the Lisbon Regional Operational Program, and by the European Regional Development Fund (ERDF).

References

1. Elston, C.W., Ellis, I.O.: Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology **19**(5), 403–410 (1991). https://doi.org/10.1111/j.1365-2559.1991.tb00229.x
2. Cree, I.A., et al.: Counting mitoses: SI (ze) matters! Mod. Pathol. **34**, 1651–1657 (2021)
3. Mathew, T., Kini, J.R., Rajan, J.: Computational methods for automated mitosis detection in histopathology images: a review. Biocybern. Biomed. Eng. **41**(1), 64–82 (Jan.2021). https://doi.org/10.1016/J.BBE.2020.11.005
4. Mahmood, T., Arsalan, M., Owais, M., Lee, M.B., Park, K.R.: Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and Deep CNNs. J. Clin. Med. **9**(3), 749 (2020). https://doi.org/10.3390/jcm9030749
5. Pati, P., Foncubierta-Rodríguez, A., Goksel, O., Gabrani, M.: Reducing annotation effort in digital pathology: A Co-Representation learning framework for classification tasks. Med. Image Anal. **67**, 101859 (2021). https://doi.org/10.1016/j.media.2020.101859

6. Jahn, S.W., Plass, M., Moinfar, F.: Digital pathology: advantages, limitations and emerging perspectives. J. Clin. Med. **9**(11), 3697 (2020). https://doi.org/10.3390/jcm9113697
7. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: optimal speed and accuracy of object detection (2020)
8. Jocher, G., et. al.: ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCVDNN support. Zenodo (2021). https://doi.org/10.5281/zenodo.5563715
9. Pan, X., et al.: Mitosis detection techniques in H&E stained breast cancer pathological images: a comprehensive review. Comput. Electr. Eng. **91**, 107038 (2021). https://doi.org/10.1016/j.compeleceng.2021.107038
10. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J. Pathol. Inform. **7**(1), 29 (2016). https://doi.org/10.4103/2153-3539.186902
11. Wang, H., et al.: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J. Med. Imaging **1**(3), 34003 (2014). https://doi.org/10.1117/1.jmi.1.3.034003
12. Paeng, K., Hwang, S., Park, S., Kim, M., Kim, S.: A unified framework for tumor proliferation score prediction in BreastHistopathology. CoRR, vol. abs/1612.07180 (2016). http://arxiv.org/abs/1612.07180
13. Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: unified, real-time object detection. CoRR, vol. abs/1506.0, 2015. http://arxiv.org/abs/1506.02640
14. Sreeraj, M., Joy, J.: A machine learning based framework for assisting pathologists in grading and counting of breast cancer cells. ICT Express, **7**(4), 440–444 (2021). https://doi.org/10.1016/j.icte.2021.02.005
15. Clarke, N., Dettrick, A., Armes, J.: Efficacy of training a deep learning model for mitotic count in breast carcinoma using opensource software. Pathology **53**, S23–S24 (2021). https://doi.org/10.1016/j.pathol.2021.06.019
16. Ludovic, R.: Mitos & Atypia 14 Contest (2014). https://mitos-atypia-14.grand-challenge.org/Home/. Accessed 28 Dec 2021
17. Veta, M., et al.: Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. Med. Image Anal. **54**, 111–121 (2019). https://doi.org/10.1016/j.media.2019.02.012