Cover Page

1) Title of the paper:

Two-step Neural Networks for an Automated Cerebrovascular Landmark Detection along the Circle of Willis

2) authors' affiliations and address:

Nantes Université, LTeN, CNRS, rue Ch. Pauc, 44306 Nantes, France.

Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France.

3) e_mail address:

Rafic.Nader@univ-nantes.fr

Vincent.LAllinec@chu-angers.fr

Romain.Bourcier@chu-nantes.fr

Florent.Autrusseau@univ-nantes.fr

4) Conference & Publisher information:

IEEE Journal of Biomedical and Health Informatics - JBHI
https://www.embs.org/jbhi/

5) bibtex entry:

```
@Article{IEEE-JBHI2025,
  author       = {R. Nader and V. {L'Allinec} and R. Bourcier and F. Autrusseau},
  date         = {2025},
  journaltitle = {the IEEE Journal of Biomedical and Health Informatics, (JBHI)},
  title        = {Two-step Neural Networks for an Automated Cerebrovascular
Landmark Detection Along the Circle of Willis},
}
```

Two-Steps Neural Networks for an Automated Cerebrovascular Landmark Detection along the Circle of Willis

Rafic Nader, Vincent L'Allinec and Romain Bourcier, Florent Autrusseau

***Abstract*— Intracranial aneurysms (ICA) commonly occur in specific segments of the Circle of Willis (CoW), primarily, onto thirteen major arterial bifurcations. An accurate detection of these critical landmarks is necessary for a prompt and efficient diagnosis. We introduce a fully automated landmark detection approach for CoW bifurcations using a two-step neural networks process. Initially, an object detection network identifies regions of interest (ROIs) proximal to the landmark locations. Subsequently, a modified U-Net with deep supervision is exploited to accurately locate the bifurcations. This two-step method reduces various problems, such as the missed detections caused by two landmarks being close to each other and having similar visual characteristics, especially when processing the complete MRA Time-of-Flight (TOF). Additionally, it accounts for the anatomical variability of the CoW, which affects the number of detectable landmarks per scan. We assessed the effectiveness of our approach using two cerebral MRA datasets: our In-House dataset which had varying numbers of landmarks, and a public dataset with standardized landmark configuration. Our experimental results demonstrate that our method achieves the highest level of performance on a bifurcation detection task.**

***Index Terms*— Vascular Bifurcations, Landmarks detection, Circle of Willis, Heatmaps regression**

I. Introduction

THE detection of cerebral vascular bifurcations landmarks is important for multiple clinical applications, including enhanced diagnostic precision, surgical planning, and customized therapeutic interventions. A precise identification of these bifurcation points enables accurate vascular mapping, which is essential for diagnosing cerebrovascular diseases such as intracranial aneurysms, arterial malformations, and stenosis [1]–[3]. Consistent identification of landmarks is important to monitor any vascular modification. Such a practice is critical to follow the evolution of various diseases and evaluate the effectiveness of therapies over time [4], [5]. Developing an automated method to identify brain landmark locations could greatly assist in clinical applications, given the increasing workload and the intricate nature of the detection procedure required by radiologists. The localization of anatomical landmarks has been proposed across various imaging modalities and organs [6]–[8]. Traditional approaches include rule-based methods [9], [10], statistical shape (SSMs)-based methods [11], [12], or conventional machine learning algorithms [13]–[15]. All these methods depend on the quality of handcrafted features and become too complex to manage as image complexity increases. Recently, deep learning algorithms have achieved a great success in medical image analysis including anatomical landmark detection predominantly through two main approaches. The first strategy employs convolutional neural networks (CNNs) for coordinate regression [7], [16]–[19], where the models analyze local pixel-level patterns to directly predict the landmarks coordinates. Remarkable success was achieved using fully convolutional network (FCN) based on heatmaps regression approaches [8], [20]–[26]. This technique constructs heatmaps that represent the probability distribution of landmarks at each pixel, thereby enhancing detection robustness against noise and improving localization accuracy through probabilistic rather than deterministic outputs. Unlike the first technique, which utilizes CNN for coordinate/heatmap regression, the second strategy leverages deep reinforcement learning [27], [28]. In this process, an agent continuously improves its method to precisely identify landmarks in images. It achieves this by learning via a reward-based system that reinforces good identifications and penalizes mistakes. Nevertheless, the intricate nature of these multi-agent systems makes it difficult to identify many landmarks at the same time.

Other researchers have adapted object detection and image segmentation networks to perform landmark detection. For example, [29] implemented Faster R-CNN integrated with a multitask loss function to identify landmarks. The authors in [30] proposed a coarse-to-fine deep learning method for detecting cranio-maxillofacial landmarks from cone-beam computed tomography (CBCT) using Faster R-CNN. On the other hand, Zhang et. al [31] employed a semantic segmentation network for bone structure segmentation, followed by a top-down heatmap-based approach with unbiased heatmap encoding and a distribution-prior coordinate representation method to accurately perform landmark detection in craniomaxillofacial regions.

In the context of vascular landmark detection, some works have been proposed [32]–[38] for arterial classification. The

This work was supported by the French RHU-ANR project "eCAN" # ANR-23-RHUS-0013 and INSERM CoPoC # MAT-PI-22155-A-01

All authors are with the Institut du Thorax, Quai Moncousu, 44007 Nantes, France (e-mails: FirstName.LastName@univ-nantes.fr), F.A. is also with the LTeN, Polytech, rue Ch. Pauc 44306 Nantes; author V.L. is also with CHU Angers (Vincent.LAllinec@chu-angers.fr).

aim of these approaches is to produce an anatomical atlas of the various arteries making up the circle of Willis. The majority of these methods, primarily rely on graph matching or predefined segmentation techniques. These approaches necessitate significant preprocessing steps, including the semi-automatic extraction of the vascular centerline [39] or the assumption of an existing centerline-plus-diameter model. Consequently, these methods cannot be directly applied to raw imaging data. However, our approach here is significantly different, as we focus solely on the classification of bifurcations; we do not seek to recognize arteries. Our main aim is to provide neuroradiologists with a means of rapidly scanning bifurcations at risk of intracranial aneurysm. In a recent work [40], we have proposed a deep learning approach for an automatic detection and classification of vascular bifurcations along the circle of Willis ; although this approach showed significant progress, it relies on the performance of a pre-segmentation step, possibly leading to a sub-optimal result. Moreover it uses local patches to perform bifurcation classification without regressing the intersection points between the merging/splitting branches. Recently, the authors in [41] proposed a multi-task UNet based network which accomplishes landmark heatmap regression, vascular semantic segmentation and orientation field regression in order to detect 19 cerebral landmarks. To the best of our knowledge, the work in [41] is the only one dealing with arterial bifurcation classification using a similar approach (based on landmark detection) that we can compare to.

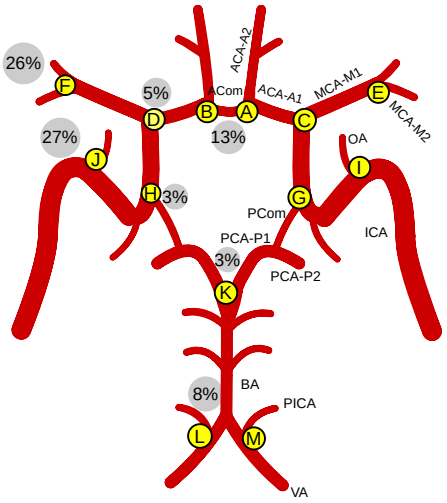


Fig. 1: Schematic representation of the CoW. The yellow labels (from A to M) depict the particular bifurcations we are interested in for this study. The percentages within the gray discs represent the risk of aneurysm formation.

As in our former works, in this paper, we focus on 13 landmark points (bifurcations labeled A to M in Fig. 1) corresponding to the bifurcations representing the highest risk of aneurysm occurrence. This work was mainly motivated by certain needs expressed by neuroradiologists in the detection and follow-up of intracranial aneurysms. Thus, this study follows on from our previous work [40], in which we intended to improve performance, particularly with regard to the E and F labels. It is widely accepted that most aneurysms occur along the Circle of Willis [42]–[45]. The work proposed in [41] considers several bifurcations away from the circle of Willis. However, it is important to note that for certain uses (stenosis detection for example), it may be useful to locate these bifurcations (but not for aneurysmal pathology).

This network exhibits a variety of configurations, characterized by differences in the number, shape, and size of the constituent arteries [46], [47]. The authors in [48] show an exhaustive overview of the many possible CoW configurations, we can see that the CoW is very rarely complete. It is estimated that around 70% of people have an incomplete Circle of Willis [49]. It is therefore crucial to develop tools capable of locating bifurcations on any incomplete configuration. Identifying anatomical landmarks is difficult because of the inconsistent structure of the Circle of Willis. Missing segments or deformations of the vessels, can modify the local appearance around landmarks, resulting in ambiguity during their recognition. Additionally, the size of 3D medical images often exceeds the memory capacity of GPUs, necessitating either a resampling or the design of a patch-based method. Furthermore, the approximate locations of certain bifurcations can cause heatmap predictions to have strong responses at several nearby positions, which may result in detection errors. Besides, the current techniques for identifying landmark bifurcations, as described in [8], [41], have practical constraints. They generally presume that different subjects have the same number of landmarks. However, this assumption is invalid because of the previously mentioned anatomical alterations in the Circle of Willis, as well as disparities in imaging technology and acquisition procedures. To circumvent this problem, we introduce a two-step neural networks [30] approach that is particularly effective in managing scenarios where the number of landmarks differs across images, enabling more flexible and precise detections.

The rest of the paper is organized as follows. Section II describes the pipeline of the proposed method including the neural networks architectures and the evaluation protocol. In Section III, we present the experimental results. We conduct a comprehensive evaluation of the method's performance and compare it with two of the state-of-the arts methods in landmark detection. Finally, Section IV concludes this work.

II. Materials and Methods

In this work, we propose a two-step landmark prediction system for CoW bifurcations. Fig. 2 illustrates the workflow of the proposed pipeline. In the first step, we formulate the problem as an object classification and bounding box regression problem. Landmarks are represented by fixed-size bounding boxes that are centered around them. The 3D nnDetection network predicts candidate regions of interest (ROIs) along with corresponding labels and scores. In the second step, the precise coordinates of each landmark are determined via heatmap regression using an encoder-decoder segmentation model. The second phase improves the initial detection, guaranteeing precise localization of landmarks and minimizing the identification of false positive detection.

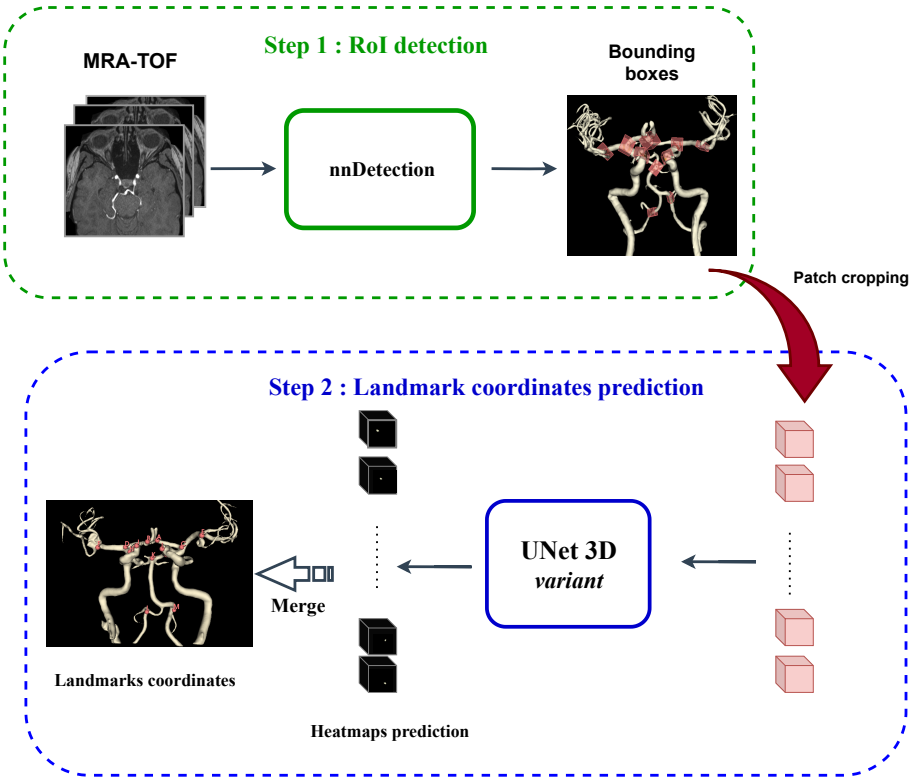


Fig. 2: Schematic representation of the proposed landmark detection pipeline. The pipeline consists of 2 steps: (1) detection of region of interests (RoI) delimiting the landmarks and (2) landmark coordinates prediction via heatmap regression.

A. Landmark position estimation

Let us now delve into the first process: the initial geographical localization estimation. This step provides a coarse estimation of the bifurcation positioning (via a nnDetection), which will later on be refined (using a UNet).

1) Deep learning model: We used the state-of-the-art self-configuring framework for volumetric medical object detection, nnDetection [50], as our object detection network in this study. nnDetection autonomously selects the optimal configuration for a given dataset by following a set of interdependent principles: 1) fixed parameters, which remain constant regardless of the data (architecture, loss function, augmentation); 2) rule-based parameters which utilizes a set of heuristics based on the relevant properties of the training data (patch size, anchor optimization); and 3) empirical parameters, which are optimized only during test time (IoU threshold for the Non Maximum Suppression, model selection, etc..). nnDetection follows the topology of Retina U-Net [51] (Fig. 3) combining the detection capabilities of the popular RetinaNet detector with the segmentation strengths of the U-Net.

Although we have used the rule based parameters and the empirical parameters without modifications, we did adapt the augmentation parameters to fit our task. We deliberately avoid applying horizontal flipping due to the symmetric nature of our classes, which could introduce bias and lessen the detection and classification accuracy.

2) Ground-Truth annotations: nnDetection uses a hybrid neural network model combining aspects of RetinaNet detector and U-Net. Indeed, for the training step, bounding boxes and segmentation annotations are necessary. In the development of nnDetection for a landmark detection task, we consider each landmark as an object to be located and classified within the image. Initially, fixed-size bounding boxes are established around each landmark point, which standardizes the detection area and simplifies the identification process. When detailed pixel-level annotations are available (In-House dataset), the segmentation is performed within these bounding boxes to precisely delineate arteries at the landmark. Conversely, when pixel-level annotations are unavailable (public dataset), the bounding box is used as an alternative to the segmented arteries. This method, although less comprehensive, still enables accurate landmarks identification by considering the enclosed area as the desired region, providing a practical alternative in cases where thorough segmentation data is unavailable [50].

When approaching landmark detection as an object detection problem, it is reasonable to choose a uniform size for creating ground truth bounding boxes. The chosen size must strike a balance by considering various aspects: the bounding box (BB) should be large enough to capture contextual information around the landmark, providing the network with relevant spatial information. On the other hand, if the bounding box is very large, it may contain multiple landmarks, which might confuse the network during its classification step, and hence, decrease the localization accuracy. This can happen in particular for pairs of bifurcations A & B, C & G and D & H, the latter being very close to each other. Considering these factors, we choose two distinct bounding boxes configurations:

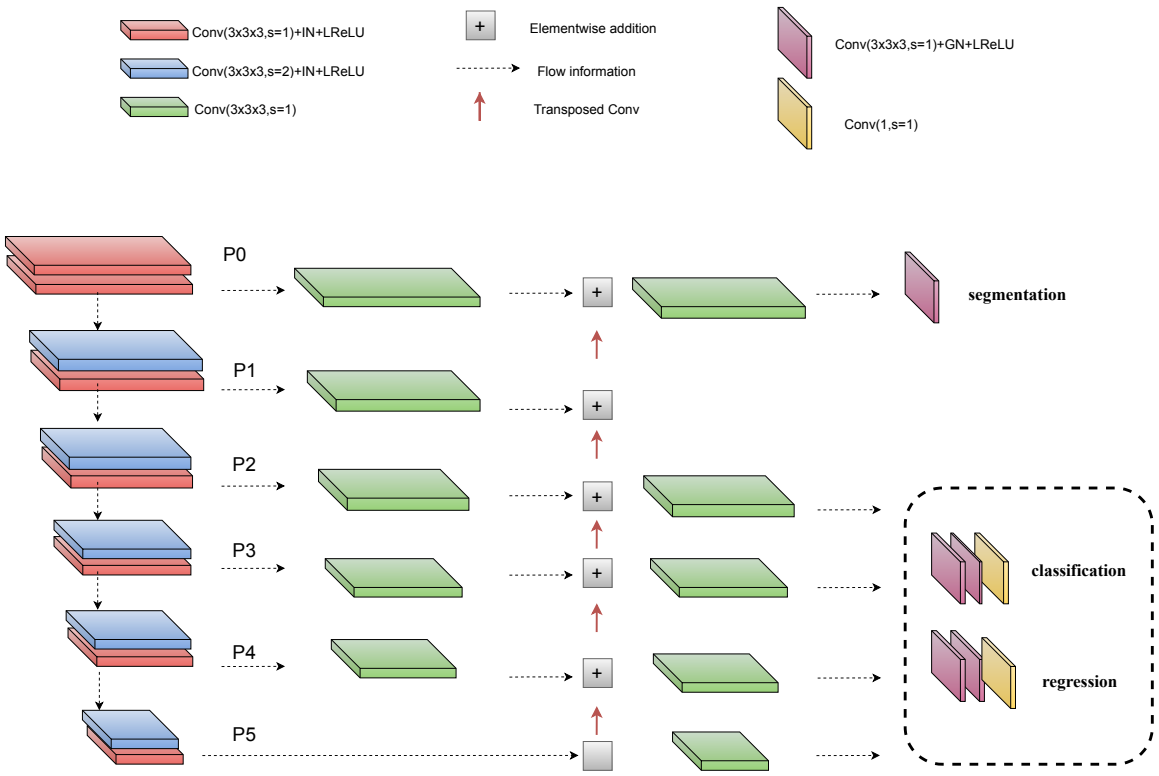


Fig. 3: Retina U-Net uses a backbone network (stacked convolutions) that extracts features at multiple scales. This backbone feeds into a Feature Pyramid Network (FPN) that is essential for detecting objects of different sizes. The pyramid includes several layers (P0 to P5), each one representing a different scale of feature representation. The features maps (P2-P5) are used for the object detection and classification. Additional layers (P0-P1) are integrated into the FPN to support segmentation tasks.

1) The BB size for landmarks E, F, I, J, K, L, and M is defined as 32 voxels. These landmarks are frequently separated from other bifurcations. A 32-sized bounding box, centered at the landmark coordinates, is used to enclose the branching structures related to the landmark. This gives the network enough context for an accurate identification.
2) The ground truth BB locations for landmarks A and B, C and G, and D and H are determined in such a way to encompass both landmarks simultaneously. These landmarks frequently appear in close proximity to one another, making it beneficial to establish a single bounding box that encompasses both landmarks, rather than separate boxes. In order to adequately include both landmarks, we begin by generating bounding boxes (BBs) with a size of 32 around each unique landmark. A unified bounding box, is then generated to contain both of these initial boxes.

Overall, we define 10 classes corresponding to 10 regions of interest. These are determined as follows: #R1 (encompasses landmarks A and B), #R2 (encompasses landmarks C and G), #R3 (encompasses landmarks D and H), #R4 (corresponds to landmark E), #R5 (corresponds to landmark F), #R6 (for landmark I), #R7 (landmark J), #R8 (landmark K), #R9 (landmark L), and #R10 (landmark M).

3) Training: The model undergoes training for a total of 60 epochs, with each epoch including 2500 batches. The training techniques are standardized using a batch size of 4 to ensure stability. For optimization, Stochastic Gradient Descent is employed with a Nesterov momentum term of 0.9. Additionally, a Poly learning rate schedule is implemented to vary the learning rate throughout epochs. The detection branch utilizes Binary Cross-Entropy as a loss function for object classification and Generalized IoU Loss for improving anchor box predictions. The segmentation branch is trained using the Cross-Entropy and Dice loss functions.

B. Landmarks position refinement

Once the approximate localization obtained through the previous network configuration, we need to pinpoint the target bifurcations; and especially for those bounding boxes enclosing two separate bifurcations of interest. To do so, we use a second neural network as explained in the following.

1) Neural Network: To find the exact location of each landmark within the ROI patch obtained from the first step, we used a UNet [52] following the encoder-decoder architecture enhanced with deep supervision. The detailed structure of the model is depicted in Fig. 4. The model takes as inputs 3D patches of size ($32 \times 32 \times 32$). Each level of the encoder consists of a residual block [53] followed by a max pooling layer for downsampling. Each residual block is composed by

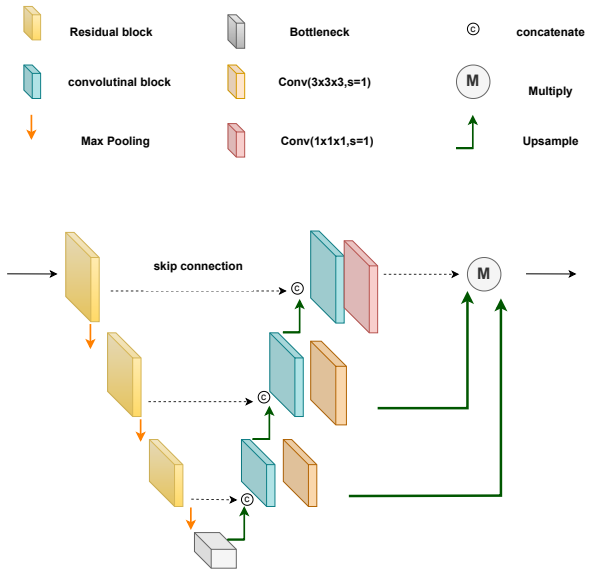


Fig. 4: UNet variant for heatmap regression.

two 3D convolution layers with batch normalisation and ReLU activation. A residual connection is created between the input and output in order to alleviate the vanishing gradient problem. For the decoder, each block contains an upsampling layer, concatenated with the features from the encoder through skip connections and then processed using a convolutional block. Moreover, the deep supervision mechanism is incorporated by generating intermediate outputs at various stages in the upsampling path [54], [55]. Each of these outputs is intended to predict the landmark heatmaps, which are then upscaled to match the output dimensions. In the final stages of the network, outputs from multiple levels are merged by multiplication, focusing on areas of agreement across the scales to enhance confidence and suppress noise in the predictions. To prevent over-fitting, we apply dropout. The last layer of each output is a $1 \times 1 \times 1$ convolutional layer with 13 channels, each one being a predicted heatmap for a specific landmark. To train the neural network, we used $32 \times 32 \times 32$ voxel patches, centered around the specific landmarks. The proposed network was trained for 500 epochs using Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and a learning rate of 0.0001.

2) Ground-Truth heatmaps: The ground-truth heatmaps are generated by applying a Gaussian distribution centered on the landmark coordinates. The intensity of the heatmap at any given voxel x indicates the probability that this voxel represents the i^{th} landmark. This probability decreases with the distance from the voxel to the landmark position x_i, modulated by the standard deviation σ, which determines the spatial dispersion of the landmark signal. The formulation for the heatmap associated with N landmarks can be defined as:

$$G_i(x) = \theta \times \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\sigma}\right)^2\right), \quad i = 1, 2, \ldots, N,$$

where θ is a scaling factor to avoid numerical instabilities due to small values of the Gaussian. We apply the mean squared error as a loss function between the predicted heatmaps P_i and the Ground-Truth heatmap G_i.

3) Inference: During the test phase, the landmark coordinates are estimated by the argmax of the relevant heatmap determined by the bounding box label. For a BB encompassing two landmarks, the precise label for the heatmap is obtained by using a threshold that is determined by peak detection during the validation process. A landmark is considered missing if the peak intensity in a heatmap channel falls below this threshold.

To summarize the entire inference procedure:

- In the first stage, bounding boxes and confidence scores are predicted using nnDetection. As multiple bounding boxes may be detected for the same label, only the region of interest (ROI) with the highest probability is retained for each label, and the rest are discarded.
- During the second stage, patches of size 32^3 are cropped around the center of each predicted bounding box. These patches are processed through the UNet network to obtain the predicted heatmaps.
- The argmax operation is performed on the associated channel(s) of the predicted heatmap for each label patch. Landmarks are kept if the pixel intensity exceeds a chosen threshold. As an illustration, in the case of category #R1, we assess the highest intensity of a voxel in two heatmap channels (0 and 1) corresponding to landmarks A & B, respectively. These landmarks are grouped together inside the same region of interest (ROI). If the highest pixel intensity in either or both heatmaps exceeds the threshold, the coordinates of the corresponding landmarks are identified.

III. Experimental Results

A. Datasets

In this study, we evaluate our method on two datasets with different vascular structures. We exploit our In-House MRA dataset with possible missing landmarks and a public MRA dataset with full landmarks.

1) In-House dataset: We have gathered an heterogeneous dataset of 145 images from the ICAN database [40]. These images were issued from 28 French institutions and acquired from various MRI scanners. All the scans were spatially normalized (voxel spacing was set to $0.4 \times 0.4 \times 0.4\ mm^3$). This dataset is split into a training data set of 97 images, and a test set of 48 images. For each volume, 13 bifurcations of interest (please refer to Fig. 1) were annotated by a trained operator and validated by a neuroradiologist. It is worth noting that the In-House dataset is not composed of a balanced amount of landmarks. This variability is caused by the MRA acquisition protocol where in certain cases the posterior cerebral tree is cut. Moreover, the CoW have naturally many different configurations among patients [48] explained by aplasia (missing arteries) or hypoplasia (very thin arteries). We consider that when an artery is missing, the bifurcation is not present and thus no landmark can be provided.

2) MRA public dataset: The public MRA dataset contains 104 scans taken from the UNC dataset (`https://public.kitware.com/WIKI/TUBETK/DATA`).

The images were acquired by a Siemens Allegra head-only 3T MR system. Each image has a voxel spacing of $(0.5 \times 0.5 \times 0.8mm)$. The public dataset includes annotations for all 19 theoretical bifurcation landmarks, regardless of whether the corresponding vascular segments are anatomically present. While a few scans were discarded in [41] due to the complete absence of vascular structures, many cases remain in which certain bifurcations are labeled despite the absence of key arteries (e.g., missing PCom or PCA-P1 segments). Consequently, we treat the public dataset as one containing all theoretical bifurcation landmarks. The annotation of the posterior landmarks (M and L) are absent for this dataset. We select the remaining 11 landmarks from the annotated 19 landmarks. This dataset was split into 77 training images and 27 test images.

We can observe that there are significant differences between these two datasets, firstly, we can note the significant variety of image acquisition devices in our in-house dataset compared to the single device used in the public dataset. Another important difference lies in the choice of the bifurcations of interest. Indeed, the authors of [41] propose a detection of 19 bifurcations, whereas we only focus on 13. Let's see how our selection differs.

We discarded bifurcations 3 and 4 from [41], as they are unfortunately under-represented in our dataset. We have also discarded bifurcations 10 to 13, as intracranial aneurysms occur very rarely here. Furthermore, bifurcations 16 and 18 are in fact very close to each other (as are bifurcations 17 and 19). A bounding box around bifurcations 16 and 19 will almost certainly encompass bifurcations 17 and 18, so it was not deemed worthwhile to separate such close bifurcations. Finally, we have included the L and M bifurcations as they carry a non-negligible risk of aneurysm formation (Fig. 1).

Moreover, it is important to point out that in [41], the authors opted for a method that seeks to determine the locations of bifurcations of interest, even in the case of a missing artery. However, their work consider two distinct cases: aneurysms and stenoses. It is possible for a stenosis to occur at the original location of a missing bifurcation, i.e. where a bifurcation should be, if the artery were present. In our work, we focus solely on the possible locations of typical intracranial aneurysms of bifurcations, the vast majority of which occur indeed, at bifurcations (more precisely between the daughter arteries). In others words, unlike stenosis or intracranial dissection, a typical intracranial aneurysms of bifurcations do not occur if there is a missing bifurcation.

B. Evaluation

Following the configuration described in nnDetection, we used stratified fivefold cross-validation during the training of both networks (nnDetection and UNet) when applied to our in-House dataset. Briefly, the development set was partitioned into five folds, with four of the folds used to train the model, while the remaining fold served as the validation data to select the best-performing model through the epochs. After the training, five different models were obtained, and the ensemble of these five models was used to predict the coordinates of the landmarks of the test set. When applied to the public MRA dataset, we used 61 images for training and 16 for validation for both networks. We test on the remaining 27 MRA-TOFs. To evaluate our method, we use the two metrics used in the *ISBI2015* Challenge: the Mean radial error (MRE) in mm and the Successful Detection Rate (SDR). The MRE computes the euclidean distance between the predicted locations and the ground truth locations of the landmarks. Whereas, SDR measures the percentage of successful detection within a given neighborhood. In our case, we used three different neighborhood dimensions (3 mm, 4 mm and 5 mm). For instance, in the following SDR-3 corresponds to a detection percentage when the predicted bifurcation center falls within a radius of 3 mm from the ground truth bifurcation center.

For the In-House dataset, we also report the false positive landmarks and the false negative ones. A false positive (FP) is a landmark that was predicted but doesn't exist whereas a false negative (FN) is a true landmark that was not predicted. During Step 1, when predicting candidate bifurcations, the 3D nnDetection network provides a confidence score per class. For each class, only the bounding box presenting the highest score is retained. The FP and FN scores are thus simply computed based on the presence/absence of a given bounding box class (the distance separating an actual Ground Truth bifurcation and it's allocated bounding box class is not taken into account). It is worth noting that FPs and FNs are not considered when calculating the MRE.

C. Comparison with heatmap regression methods

To demonstrate the effectiveness of simultaneously employing object detection alongside heatmap regression compared to using only heatmap regression for the entire image, we evaluate our approach against two advanced deep learning methods for heatmap regression: SCN [8] and FARNet [26]. We have chosen to use these two methods because they have shown the best performance in terms of the reduction of MREs in [41]. The Spatial Configuration Network (SCN) integrates constraints on landmark relative positioning directly within the architecture, optimizing both local appearance and spatial configuration in a unified, end-to-end trainable framework. Whereas, the method described in [26] introduces a novel deep learning architecture called FARNet, which is enhanced by a feature aggregation module for multi-scale feature fusion and a feature refinement module tailored to fine-tune high-resolution heatmap regression. We have re-implemented these two methods with their default settings and applied them to our datasets. The input images for both datasets were initially resampled with a voxel size of $0.6 \times 0.6 \times 0.8$ and then automatically cropped to dimensions of $192 \times 160 \times 160$ based on the mean landmark distributions.

D. Results

As previously explained, in this section, we provide a thorough evaluation of the proposed method, which we compared with two state-of-the-art methods of heatmap regression: SCN [8] and FARNet [26].

1) In-House dataset: Fig. 5 presents the distributions of maximum voxel intensities within heatmap predictions on the validation data of our dataset. For each identified landmark, we have collected the voxel intensities from two distinct groups : 1) the voxel values from the heatmap channel specifically associated with the landmark, and 2) the voxel values from all remaining channels. This approach allowed us to directly compare the specific responses of the true positive channel—those that are precisely correlated with the landmark with the non-specific responses exhibited by the other channels. The intersection point ($th = 40$), serve as a natural threshold. It can be applied to unseen test data to determine the landmark point present in a patch. Here, the plot "KDE 2" has been rescaled for a better visualization, but obviously the intersection point has been computed on unscaled data.

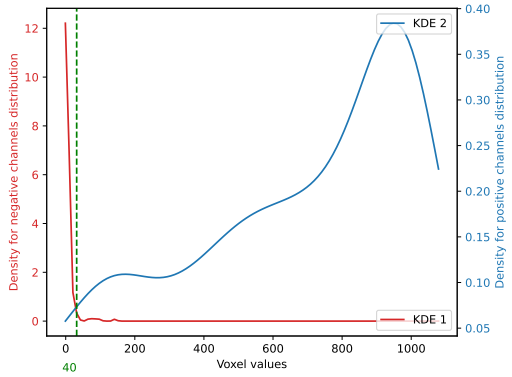


Fig. 5: Voxel intensity distribution analysis of heatmaps in true positive channels *vs.* other channels. We implement Kernel Density Estimation to convert the voxel intensities into continuous probability density functions for each type of heatmap and then we collect the curves crossing point.

TABLE I: Detection comparison on the In-House MRA dataset, quantified using MRE and SDR for three neighborhood sizes (3 mm, 4 mm and 5 mm).

Metric/Method	FARNet	SCN	OURS
MRE (mm)	2.73	2.38	**1.97**
SDR(%)-3	72.17	81.73	**82.31**
SDR(%)-4	82.36	88.69	**91.24**
SDR(%)-5	90.08	92.69	**95.27**
#FP	47	47	**9**
#FN	/	/	3

For the In-House dataset, the results on the test set are shown in Table I. Our method achieves an average mean radial error of 1.97 mm, which shows improvements of 0.39 mm over SCN and 0.76 mm over FARNet. While SCN performs similarly for small distance thresholds, our method gradually outperforms the others as the threshold increases, peaking at 95.27% at 5 mm. This improvement is especially remarkable in situations where landmarks have similar local appearances, a situation that often results in localization errors with heatmap-based techniques like SCN and FARNet. These approaches primarily concentrate on local characteristics within each voxel, resulting in a significant number of incorrectly detected landmarks since they are not able to efficiently distinguish between significant and insignificant features. This can be seen if we compute the number of failures cases for each method. Following [21], we consider that a prediction landmark is a failure case if the distance from predicted to ground-truth position is larger than 10 mm. While we report 15 failure cases for SCN, 13 for FARNet, our method only presents 6 failure cases. Consequently, not only does our method reduce the MRE, but it also ensures a lower incidence of failure cases. Moreover, our methodology includes an early phase of object detection, which, in our specific situation, results in the omission of 3 landmarks. Nevertheless, the significant occurrence of false positives (47 for both SCN and FARNet) compared to our proposed method (9 FPs) highlights the constraints of heatmap-based approaches that solely rely on continuous heatmap predictions and lack mechanisms to effectively handle missed landmarks. For applications requiring a balance between accuracy, low failure cases and low false positive detection, our method is distinctly advantageous. Furthermore, it can be used for datasets with a varying number of landmarks while the other methods have been designed for acquisitions comprising all bifurcations. For time efficiency evaluation, we report the average inference processing time (in seconds) required to obtain the final output using one fold of our in-house dataset. The time measurement is conducted on a machine equipped with 128GB RAM, an Intel Xeon CPU, and an Nvidia RTX A4500 GPU. This entire inference process per image, including preprocessing, takes a total of 10.67 seconds. In contrast, SCN and FARNet achieve a faster inference times of 1.63 seconds and 1.28 seconds, respectively. As for the training process, the first step of our process (the nnDetector) takes about 52 hours to process one fold, whereas the second step (the 3D U-Net) takes about 3.6 hours per fold (the nnDetector network contains 19.7 million trainable parameters, whereas the UNet network has 7.7 million parameters). However, despite the increased processing time, our method provides higher accuracy and fewer false positives, making it a more reliable choice for precise predictions. Fig. 6 depicts the results of the MRE for each bifurcation label. Some bifurcations are more difficult to locate than others due to the intricate structure of the branches and the anatomical variations in the Circle of Willis. Notably, the uncertainty associated with division of the middle cerebral artery (MCA) into its M1 and M2 segments leads to a greater error (MRE) of 2.4 and 2.8 mm for landmarks E and F, respectively. Furthermore, landmarks L and M exhibit elevated MREs due to the limited availability of data and their inconsistent presence in the dataset. However, for such landmarks, compared to state-of-the-art approaches, our method demonstrates significantly reduced detection error (M with an MRE of 2.12 mm using our method compared with 5.68 mm and 5.12 mm for FARNet and SCN) primarily due to the step 1 detection algorithm that focuses on the most relevant regions of interest (ROI) of labels #R9 and #R10.

For a complete evaluation, Table II also presents the mean Euclidean distance between the centers of the ground truth bounding boxes and the predicted bounding boxes obtained

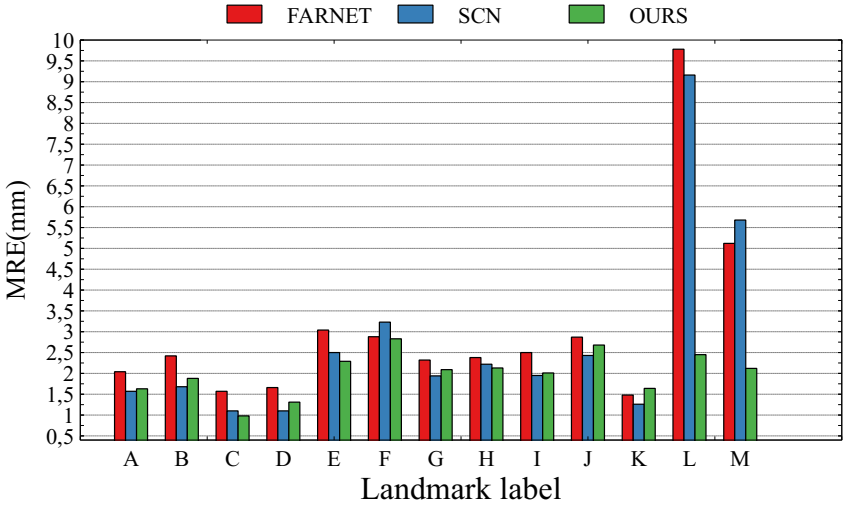


Fig. 6: Mean radial errors (mm) for 13 landmark labels for the In-House dataset.

TABLE II: Mean Euclidean distance between the centers of the ground-truth and predicted bounding boxes obtained from Step 1 for each class.

Classes	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
Distance (mm)	**1.99**	**1.53**	**1.69**	**2.51**	**2.85**	**2.03**	**2.68**	**1.64**	**2.46**	**3.46**

from Step 1 for each class (#R1 to #R10). The computed mean Euclidean distance is 2.21 mm across all classes, which aligns with our overall findings. This results confirm that the first stage effectively provides a reliable coarse localization of landmarks, while the second stage refines the coordinates by concentrating on a more restricted search space. Finally, qualitative comparisons are presented in Fig. 7. By way of comparison, the diameters of arteries measured in the literature vary between 1 and 4 mm [57]–[59], which implies that the estimation errors we obtain are minimal, and fall in the direct vicinity of bifurcations (the said error being often smaller than the width of an artery).

2) Public MRA dataset: In our investigation using the public MRA dataset, which encompasses a complete set of landmarks, we present the comparative results in Table III. The results of our method show a MRE of $1.47\ mm$,

TABLE III: Detection comparison on the public MRA dataset, quantified using MRE and SDR for three distinct neighborhood sizes.

Metric/Method	FARNET	SCN	OURS
MRE (mm)	2.21	1.72	**1.47**
SDR(%)-3	82.49	91.91	**92.59**
SDR(%)-4	89.89	93.26	**94.94**
SDR(%)-5	**96.28**	93.60	95.62

indicating moderate improvements to the SCN and FARNet approaches, with an increase of $0.24\ mm$ and $0.77\ mm$ respectively. A smaller MRE indicates a higher level of accuracy in determining the true positions of landmarks. This is particularly important for applications with strong requirements on the spatial accuracy. Our approach surpasses the performance of previous methods when it comes to detecting landmarks inside tighter boundaries, specifically for the 3 and $4\ mm$ neighborhood radius, achieving SDRs of 92.59% and 94.94% respectively. When considering the $5\ mm$ threshold, even though FARNet achieves the maximum SDR of 96.28%, our technique still achieves a competitive performance of 95.62%.

As illustrated in Fig. 8, the detection performance per label demonstrates a pattern consistent with that observed for the In-House dataset (Fig. 6). The bifurcation landmarks E and F exhibit larger detection errors compared to other bifurcations (excluding L and M). However, our method outperforms SCN and FARNet in detecting landmarks E and F. To further analyze this, we present boxplots of the mean radial errors for landmarks E and F in Fig. 9. Our method presents the lowest median and whiskers for landmarks E and F. This implies that it generally predicts landmarks that are closer to the true location, demonstrating a good level of accuracy for most landmarks. The outliers that can be generated by our method represent occasional mistakes which can be attributed to misclassification occurrences (proximal bounding boxes) in step 1 due to similarities in appearance with the target landmarks of MCA-M1 and M2 branches. SCN and FARNet present larger whiskers and fewer outliers compared to our proposed method. This indicates a wider range of distances, resulting in less consistency. They often have significant difficulty in discriminating between characteristics associated within the tree structure along the MCA.

E. Detection of aneurysm bearing bifurcations

To better emphasize the impact of aneurysm presence on the detection of a landmark, we gathered data from the ICAN database comprising 134 MRA-TOF images, each one presenting an aneurysm, with various volumes going from 2 mm^3 to 1050 mm^3. We annotated the landmark coordinates and labeled the specific locations where the aneurysms occurred. The distribution of these aneurysms across the bifurcations is detailed in Table IV. We calculated the detection rates of landmarks within five different distance

(a) GT (b) GT (c) GT

(d) SCN (e) SCN (f) SCN

(g) FARNet (h) FARNet (i) FARNet

(j) Proposed (k) Proposed (l) Proposed

Fig. 7: Qualitative Comparison. Red marks refer to GT landmarks or landmarks predicted within 2 mm. Green marks refer to predictions within $(2-5\ mm)$ from the GT. Purple marks refer to failure cases ($>10\ mm$). Blue marks refer to FP predictions.

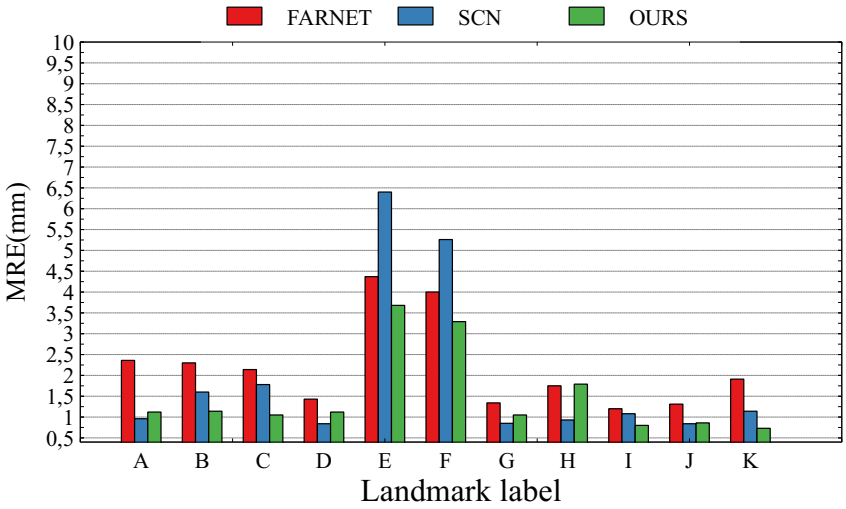


Fig. 8: Mean radial errors (mm) for 11 landmarks labels for the public dataset.

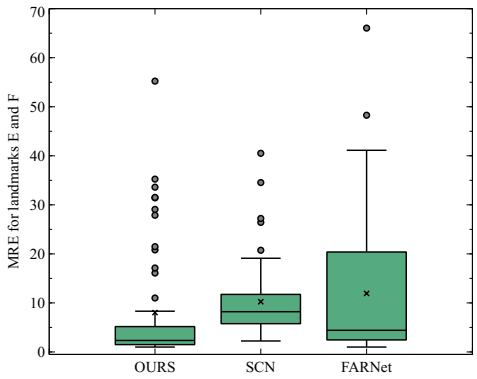


Fig. 9: Distance distribution between the predicted bifurcations and the Ground Truth (Labels E & F).

thresholds: 2 mm, 3 mm, 4 mm, 5 mm, and 10 mm. The findings are as follows: 60% of landmarks were detected within a 2 mm radius, 73% within 3 mm, 87% within 4 mm, 90% within 5 mm, and 97% within a 10 mm radius.

TABLE IV: Distribution of aneurysm for each bifurcation label. The landmarks H and K were missing in the dataset

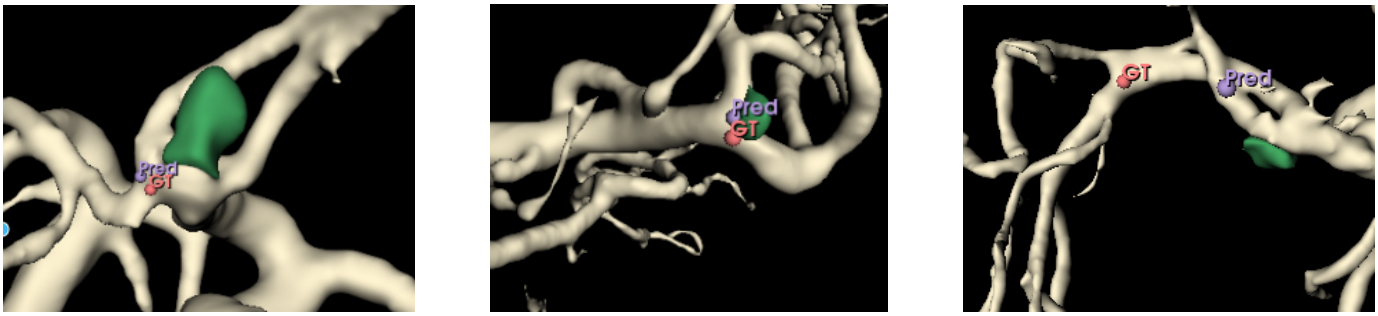
Bifurcation	A	B	C	D	E	F	G	I	J
# of ICAs	7	11	6	4	19	39	3	28	17

In a comparative analysis of landmark detection rates with and without the presence of aneurysms, we observed a similar SDR of approximately 95% for a 5 mm threshold. This suggests that the presence of aneurysms does not significantly impact the accuracy of detection. However, it is noteworthy that four landmarks were predicted within a distance greater than 10 mm. The failure cases are linked to bifurcations E and F and are caused by the inaccurate prediction, by nnDetction network, of the Region of Interest (ROI) of labels #R4 and #R5 due to the resemblance in the MCA branching patterns. Fig. 10 shows qualitative results of landmark detection with the presence of aneurysms.

IV. Conclusion

In this paper, we propose a two-steps deep learning-based framework for landmark detection of bifurcations along the circle of Willis. Initially, the challenge is framed as a task of object classification paired with bounding box regression. Each landmark is enclosed in a fixed-size bounding box that is positioned centrally around the landmark. Using a 3D nnDetection network, we detect and classify potential regions of interest (ROIs), assigning each ROI a specific label and score. In the second step, the precise locations of the landmarks are determined using heatmap regression via a UNet model. The experimental results demonstrate that our method achieves location accuracy that is comparable to the leading heatmap regression algorithms [8], [26], as seen by the minimal mean radial error (MRE). Our method leads to a reduction in misidentification errors of similar bifurcations seen in the proximal branches (landmarks E and F). In addition, our two-step approach is specifically designed to handle datasets that have a varying number of landmarks. This differentiates our proposed technique from heatmap regression methods, which are not designed for such an eventuality. This property is particularly important given that the circle of Willis is incomplete in approximately 60–75% of individuals [56], [57]). For our In-house dataset, we train the neural network to detect only the bifurcations that are actually present in each image. However, the same CNN can be easily retrained to predict all theoretical bifurcation locations—including those that are missing due to anatomical variation. This flexibility was demonstrated using the public dataset (see Fig. 8), where all 11 landmarks are annotated regardless of anatomical presence.

In order to enhance performance, particularly in accurately identifying similar bifurcations within the MCA branch, it would be beneficial to augment the training dataset with a more diverse assortment of MCA landmarks. Expanding the application of the detection model to a broader spectrum of anatomical differences should improve its ability to precisely recognize and distinguish closely situated and visually similar



(a) Detection of Landmark B within 2 mm with the presence of a large aneurysm (green shaded volume).

(b) Detection of Landmark E within 2 mm with the presence of a small aneurysm.

(c) Misdetection of Landmark F; Failure case related to MCA arterial tree and not related to the presence of the ICA.

Fig. 10: Qualitative results on landmark detection with the presence of aneurysms

bifurcations, thus boosting the overall accuracy of the detection process.

ACKNOWLEDGEMENTS

The authors would like to thank Nesrin Mansouri for her help, including additional experiments conducted during manuscript revision.

REFERENCES

[1] T. van Seeters, J. Hendrikse *et al.*, "Completeness of the circle of willis and risk of ischemic stroke in patients without cerebrovascular disease," *Neuroradiology*, vol. 57, pp. 1247–1251, 2015.

[2] L. Rinaldo, B. A. McCutcheon *et al.*, "Relationship of a1 segment hypoplasia to anterior communicating artery aneurysm morphology and risk factors for aneurysm formation," *J. Neurosurg.*, vol. 127(1), 2016.

[3] K. Kayembe, M. Sasahara, and F. Hazama, "Cerebral aneurysms and variations in the circle of willis." *Stroke*, vol. 15(5), pp. 846–850, 1984.

[4] J. Boucherit, B. Kerleroux *et al.*, "Bifurcation geometry remodelling of vessels in de novo and growing intracranial aneurysms: a multicenter study," *Journal of NeuroInterventional Surgery*, 2022.

[5] W. Jeong, M. Han, and K. Rhee, "The hemodynamic alterations induced by the vascular angular deformation in stent-assisted coiling of bifurcation aneurysms," *Computers in Biology and Medicine*, vol. 53, pp. 1–8, 2014.

[6] C.-W. Wang, C.-T. Huang *et al.*, "A benchmark for comparison of dental radiography analysis algorithms," *Medical image analysis*, vol. 31, pp. 63–76, 2016.

[7] H. Wu, C. Bailey *et al.*, "Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet," in *MICCAI 2017: 20th Intl Conf., Quebec, Canada, Sept. 11-13, 2017*, 2017, pp. 127–135.

[8] C. Payer, D. Štern *et al.*, "Integrating spatial configuration into heatmap regression based cnns for landmark localization," *Med. Image Anal.*, vol. 54, pp. 207–219, 2019.

[9] K. Kim and S. Lee, "Vertebrae localization in ct using both local and global symmetry features," *Computerized Medical Imaging and Graphics*, vol. 58, pp. 45–55, 2017.

[10] V. Grau, M. Alcaniz *et al.*, "Automatic localization of cephalometric landmarks," *J. of Biomedical Informatics*, vol. 34(3), pp. 146–156, 2001.

[11] T. Norajitra and K. H. Maier-Hein, "3D statistical shape models incorporating landmark-wise random regression forests for omni-directional landmark detection," *IEEE Trans. on Medical Imaging*, vol. 36(1), 2016.

[12] R. Kafieh, A. Mehri, and S. Sadri, "Automatic landmark detection in cephalometry using a modified active shape model with sub image matching," in *Int. Conf. on Machine Vision*. IEEE, 2007, pp. 73–78.

[13] O. Oktay, W. Bai *et al.*, "Stratified decision forests for accurate anatomical landmark localization in cardiac images," *IEEE Trans. on Medical Imaging*, vol. 36(1), pp. 332–342, 2016.

[14] D. Štern, T. Ebner, and M. Urschler, "From local to global random regression forests: exploring anatomical landmark localization," in *MICCAI : 19th Intl. Conf., Athens, Greece, Oc. 17-21, 2016*, 2016.

[15] M. Urschler, T. Ebner, and D. Štern, "Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization," *Med. Image Anal.*, vol. 43, pp. 23–36, 2018.

[16] H. Lee, M. Park, and J. Kim, "Cephalometric landmark detection in dental x-ray images using convolutional neural networks," in *SPIE Medical imaging 2017*, vol. 10134, 2017, pp. 494–499.

[17] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," *IEEE Trans. on Image Proc.*, vol. 26(10), pp. 4753–64, 2017.

[18] J. M. Noothout, B. D. De Vos *et al.*, "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE Trans. on medical imaging*, vol. 39(12), pp. 4011–4022, 2020.

[19] M. Zeng, Z. Yan *et al.*, "Cascaded convolutional networks for automatic cephalometric landmark detection," *Medical image analysis*, vol. 68, 2021.

[20] Y. Zheng, D. Liu *et al.*, "3D deep learning for efficient and robust landmark detection in volumetric data," in *MICCAI : 18th International Conference, Munich, Germany, Oct. 5-9, 2015*, 2015, pp. 565–572.

[21] C. Payer, D. Štern *et al.*, "Regressing heatmaps for multiple landmark localization using cnns," in *Intl Conf. on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 230–238.

[22] R. Chen, Y. Ma *et al.*, "Cephalometric landmark detection by attentive feature pyramid fusion and regression-voting," in *MICCAI : 22nd Intl. Conf., Shenzhen, China, Oct. 13–17, 2019*, 2019, pp. 873–881.

[23] J. Xu, H. Xie *et al.*, "Hip landmark detection with dependency mining in ultrasound image," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3762–3774, 2021.

[24] K. Oh, I.-S. Oh *et al.*, "Deep anatomical context feature learning for cephalometric landmark detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 806–817, 2020.

[25] Y. Wang, L. Huang *et al.*, "Multi-input adaptive neural network for automatic detection of cervical vertebral landmarks on x-rays," *Computers in Biology and Medicine*, vol. 146, 2022.

[26] Y. Ao and H. Wu, "Feature aggregation and refinement network for 2d anatomical landmark detection," *Journal of Digital Imaging*, vol. 36, no. 2, pp. 547–561, 2023.

[27] F. C. Ghesu, B. Georgescu *et al.*, "An artificial agent for anatomical landmark detection in medical images," in *MICCAI : 19th Intl. Conf., Athens, Greece, Oct. 17-21, 2016*, 2016, pp. 229–237.

[28] A. Alansary, O. Oktay *et al.*, "Evaluating reinforcement learning agents for anatomical landmark detection," *Medical image analysis*, vol. 53, pp. 156–164, 2019.

[29] J. Qian, M. Cheng *et al.*, "Cephanet: An improved faster r-cnn for cephalometric landmark detection," in *IEEE 16th Intl. symposium on biomedical imaging (ISBI)*, 2019, pp. 868–871.

[30] X. Chen, C. Lian *et al.*, "Fast and accurate craniomaxillofacial landmark detection via 3d faster r-cnn," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3867–3878, 2021.

[31] R. Zhang, B. Jie *et al.*, "Craniomaxillofacial bone segmentation and landmark detection using semantic segmentation networks and an unbiased heatmap," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 427–437, 2024.

[32] H. Bogunovic, J. M. Pozo *et al.*, "Anatomical labeling of the circle of willis using maximum a posteriori probability estimation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1587–1599, 2013.

[33] D. Robben, E. Türetken *et al.*, "Simultaneous segmentation and anatomical labeling of the cerebral vasculature," *Medical Image Analysis*, vol. 32, pp. 201–215, 2016.

[34] M. Bilgel, S. Roy *et al.*, "Automated anatomical labeling of the cerebral arteries using belief propagation," in *Medical Imaging 2013: Image Processing*, vol. 8669. SPIE, 2013, pp. 340–345.

[35] I. Gurobi *et al.*, "Gurobi optimizer reference manual," *Optimization*, 2015.

[36] T. Dunås, A. Wåhlin *et al.*, "A stereotactic probabilistic atlas for the major cerebral arteries," *Neuroinformatics*, vol. 15(1), pp. 101–10, 2017.

[37] X. Wang, Y. Liu *et al.*, "Automatic labeling of vascular structures with topological constraints via HMM," in *MICCAI*, 2017, pp. 208–215.

[38] D. Alblas, I. Vos *et al.*, "Deep-learning-based extraction of circle of willis topology with anatomical priors," *scientific reports*, vol. 14, no. 1, p. 31630, Dec 2024.

[39] A. Chowriappa, Y. Seo *et al.*, "3D vascular skeleton extraction and decomposition," *IEEE J. of Biomedical and Health Informatics*, vol. 18(1), pp. 139–147, 2014.

[40] R. Nader, R. Bourcier, and F. Autrusseau, "Using deep learning for an automatic detection and classification of the vascular bifurcations along the circle of willis," *Medical Image Analysis*, p. 102919, 2023.

[41] Z. Tan, J. Feng *et al.*, "Multi-task global optimization-based method for vascular landmark detection," *Computerized Medical Imaging and Graphics*, vol. 114, p. 102364, 2024.

[42] V. Kumar, A. Abbas, and J. Aster, *Robbins Basic Pathology (10th edition)*. elsevier, 2017.

[43] J. Brisman, J. Song, and D. Newell, "Cerebral aneurysms," *New England Journal of Medicine*, vol. 355(9), pp. 928–939, 2006.

[44] W. Schievink, "Intracranial aneurysms," *N Engl J Med*, vol. 336(1), pp. 28–40, 1997.

[45] C. Sadasivan, D. Fiorella *et al.*, "Physical factors effecting cerebral aneurysm pathophysiology," *Ann Biomed Eng*, vol. 41(7), pp. 1347–65, 2013.

[46] S. Iqbal, "A comprehensive study of the anatomical variations of the circle of willis in adult human brains," *Journal of clinical and diagnostic research: JCDR*, vol. 7, no. 11, p. 2423, 2013.

[47] H. Şahin and Y. Pekçevik, "Anatomical variations of the circle of willis: evaluation with ct angiography," *Anatomy*, vol. 12(1), pp. 20–26, 2018.

[48] L. Hindenes, A. Haberg *et al.*, "Variations in the circle of willis in a large population sample using 3d tof angiography: The Tromsø study," *PlosOne*, vol. 15, no. 11, Nov 3 2020.

[49] M. van Kammen, J. M. Charles *et al.*, "Heritability of circle of willis variations in families with intracranial aneurysms," *PLoS One*, vol. 29, no. 13, 2018.

[50] M. Baumgartner, P. F. Jäger *et al.*, "nnDetection: a self-configuring method for medical object detection," in *MICCAI : 24th Intl. Conf., Strasbourg, France, Sept. 27–October 1, 2021*, 2021, pp. 530–539.

[51] P. F. Jaeger, S. A. Kohl *et al.*, "Retina u-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Machine Learning for Health Workshop*. PMLR, 2020, pp. 171–183.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI: 18th Intl. Conf., Munich, Germany, Oct.5-9, 2015*, vol. 9351, 2015, pp. 234–241.

[53] K. He, X. Zhang *et al.*, "Deep residual learning for image recognition," in *IEEE Conf. Computer Vision and Pattern Recog.*, 2016, pp. 770–778.

[54] J. Stawiaski, "A multiscale patch based convolutional network for brain tumor segmentation," *arXiv preprint arXiv:1710.02316*, 2017.

[55] A. Hilbert, V. I. Madai *et al.*, "Brave-net: Fully automated arterial brain vessel segmentation in patients with cerebrovascular disease," *Frontiers in Artificial Intelligence*, vol. 3, 2020.

[56] H. Zhou, J. Sun *et al.*, "Correlation between the integrity of the circle of willis and the severity of initial noncardiac cerebral infarction and clinical prognosis," *Medicine*, vol. 95(10), p. e2892, 2016.

[57] M. Krabbe-Hartkamp, J. van der Grond *et al.*, "Circle of willis: morphologic variation on three-dimensional time-of-flight mr angiograms," *Radiology*, vol. 207(1), pp. 103–11, 1998.

[58] J. Shatri, D. Bexheti *et al.*, "Influence of gender and age on average dimensions of arteries forming the circle of willis study by magnetic resonance angiography on kosovo's population," *Open Access Maced J Med*, 2017.

[59] S. Mujagic, M. Moranjkic *et al.*, "The inner diameter of arteries of the circle of willis regarding gender and age on magnetic resonnance angiography," *Acta Medica Saliniana*, 2013.