Spatio-temporal deep learning for automatic detection of intracranial vessel perforation in digital subtraction angiography during endovascular thrombectomy

Su, Ruisheng; van der Sluijs, Matthijs; Cornelissen, Sandra A.P.; Lycklama, Geert; Hofmeijer, Jeannette; Majoie, Charles B.L.M.; Niessen, Wiro J.; van der Lugt, Aad; van Walsum, Theo; More Authors

DOI
10.1016/j.media.2022.102377

Publication date
2022

Document Version
Final published version

Published in
Medical Image Analysis

Citation (APA)
Su, R., van der Sluijs, M., Cornelissen, S. A. P., Lycklama, G., Hofmeijer, J., Majoie, C. B. L. M., Niessen, W. J., van der Lugt, A., van Walsum, T., & More Authors (2022). Spatio-temporal deep learning for automatic detection of intracranial vessel perforation in digital subtraction angiography during endovascular thrombectomy. *Medical Image Analysis*, *77*, Article 102377. https://doi.org/10.1016/j.media.2022.102377

Medical Image Analysis 77 (2022) 102377

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Spatio-temporal deep learning for automatic detection of intracranial vessel perforation in digital subtraction angiography during endovascular thrombectomy

Ruisheng Su[a,1,*], Matthijs van der Sluijs[a], Sandra A.P. Cornelissen[a], Geert Lycklama[b], Jeannette Hofmeijer[c,d], Charles B.L.M. Majoie[e], Pieter Jan van Doormaal[a], Adriaan C.G.M. van Es[f], Danny Ruijters[g], Wiro J. Niessen[a,h], Aad van der Lugt[a], Theo van Walsum[a]

[a] Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, The Netherlands
[b] Department of Radiology, Haaglanden Medical Center, The Hague, The Netherlands
[c] Clinical Neurophysiology, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands
[d] Department of Neurology, Rijnstate Hospital, Arnhem, The Netherlands
[e] Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands
[f] Department of Radiology, Leiden UMC, Leiden, The Netherlands
[g] Philips Healthcare, Best, The Netherlands
[h] Faculty of Applied Sciences, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 27 September 2021
Revised 17 January 2022
Accepted 19 January 2022
Available online 29 January 2022

Keywords:
Stroke
Vascular system injuries
X-Rays
Treatment outcome
Decision making
Object detection
Endovascular procedures

ABSTRACT

Intracranial vessel perforation is a peri-procedural complication during endovascular therapy (EVT). Prompt recognition is important as its occurrence is strongly associated with unfavorable treatment outcomes. However, perforations can be hard to detect because they are rare, can be subtle, and the interventionalist is working under time pressure and focused on treatment of vessel occlusions. Automatic detection holds potential to improve rapid identification of intracranial vessel perforation. In this work, we present the first study on automated perforation detection and localization on X-ray digital subtraction angiography (DSA) image series. We adapt several state-of-the-art single-frame detectors and further propose temporal modules to learn the progressive dynamics of contrast extravasation. Application-tailored loss function and post-processing techniques are designed. We train and validate various automated methods using two national multi-center datasets (i.e., MR CLEAN Registry and MR CLEAN-NoIV Trial), and one international multi-trial dataset (i.e., the HERMES collaboration). With ten-fold cross-validation, the proposed methods achieve an area under the curve (AUC) of the receiver operating characteristic of 0.93 in terms of series level perforation classification. Perforation localization precision and recall reach 0.83 and 0.70 respectively. Furthermore, we demonstrate that the proposed automatic solutions perform at similar level as an expert radiologist.

1. Introduction

1.1. Clinical background

Ischemic stroke remains a world leading cause of death and long-term disability (WHO, 2018). Recent studies have demonstrated the efficacy of endovascular therapy (EVT) in improving acute ischemic stroke outcome by mechanically reopening occluded vessels (Goyal et al., 2016b; Berkhemer et al., 2015). However, EVT occasionally leads to peri-procedural hemorrhagic or ischemic complications (e.g., vessel perforation or dissection, vasospasm, intracerebral hemorrhage, subarachnoid hemorrhage).

Intracranial vessel perforation refers to vessel wall rupture and blood extravasation typically caused by catheter/guidewire movement or stent retriever retraction. Fig. 1 visualizes some vessel perforation examples in X-ray digital subtraction angiography (DSA) acquired during EVT procedures. Vessel perforation is a life-

* Corresponding author.
E-mail address: r.su@erasmusmc.nl (R. Su).
[1] Code is available at https://gitlab.com/radiology/igit/q-maestro/perforationdetection

https://doi.org/10.1016/j.media.2022.102377

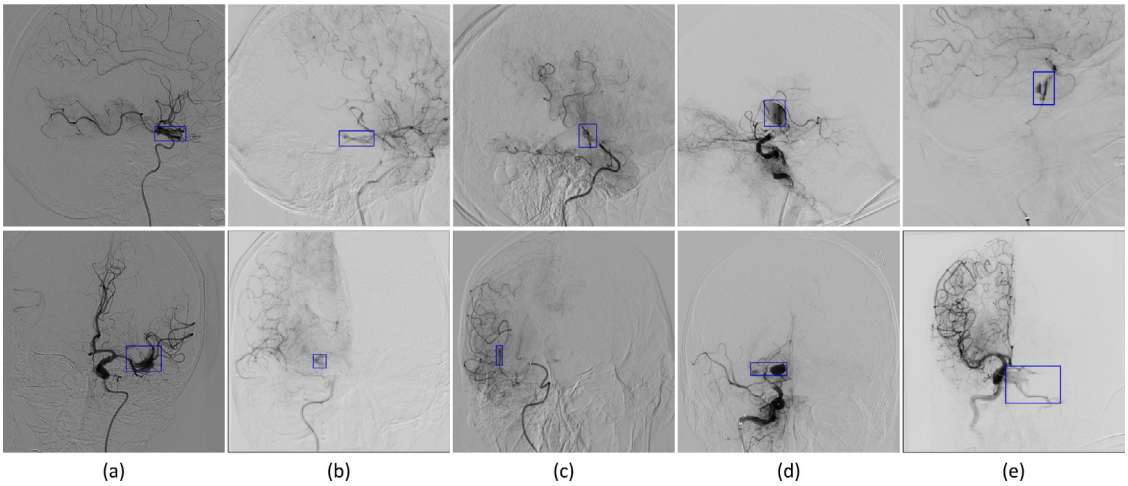


Fig. 1. Five perforation examples in digital subtraction angiography annotated in blue bounding boxes by an experienced neuroradiologist. Top row: lateral view. Bottom row: anteroposterior view. (a)-(c): subarachnoid; (d): parenchymal; (e): AV-fistula.

threatening procedure-related complication that is strongly associated with poor functional treatment outcome (Akins et al., 2014; Nogueira et al., 2012). When vessel perforation happens during an EVT procedure, prompt and appropriate therapeutic actions must be considered to prevent worsening of the clinical status of the patient. The effectiveness of such actions reduces rapidly over time; thus early recognition of vessel perforation is essential.

In current clinical practice, vessel perforations are only recognized by interventionalists by carefully inspecting the DSA images during EVT procedures. EVT is often a challenging procedure which is done under time pressure. As a result, perforations can potentially be missed. An automatic approach may facilitate fast and accurate perforation detection.

Automatic perforation detection, to the best of our knowledge, has not been studied yet. We believe that the reasons largely lie in the difficulties in collecting sufficient and representative data. First, the reported incidence of vessel perforation is within the range of 1%-5% in EVT, which is rather low (Akins et al., 2014; Mokin et al., 2017; Jovin et al., 2015). Besides, considering hospital data collection restrictions and variations in data quality and collection procedures, it is non-trivial to obtain sufficient perforation cases for developing and validating algorithms. Third, there exists large heterogeneity in the appearance of vessel perforations in DSA images. Based on the type of contrast extravasation, perforations can be categorized into several types, e.g., subarachnoid, AV-fistula, intramural, and parenchymal (Fig. 1). Even within the same type, considerable diversity exists with regard to perforation location, size, direction and contrast density.

In this work, we therefore investigate automated intracranial vessel perforation using deep learning, and establish the first benchmark on a large clinical perforation dataset. Specifically, we aim to answer two questions from the clinical perspective: 1) whether there is a perforation event in this DSA run; and 2) where the perforation is located in the image. The answer to the former question may serve as an alarm to interventionalists during EVT, while automation of the latter helps to save time in locating the actual perforation.

1.2. Related work

Perforation detection is intrinsically a pattern recognition task in medical imaging. Object detection using convolutional neural networks (CNN) has been well studied in the computer vision community. General object detection methods can be summarized into several categories: single-stage (e.g., RetinaNet (Lin et al., 2017), SSD (Liu et al., 2016), YoLo (Redmon et al., 2016)), two-stage (e.g., Faster R-CNN (Ren et al., 2015), Mask R-CNN (He et al., 2017)) and multi-stage (e.g., Cascade R-CNN (Cai and Vasconcelos, 2018), Hybrid Task Cascade (Chen et al., 2019a)) methods. In recent years, such object detection methods have been increasingly adopted in medical imaging applications (Shen et al., 2017; Zhao et al., 2021). For example, Liu et al. tailored Faster R-CNN for polyp detection from colonoscopic images (Liu et al., 2021). Wollmann and Rohr proposed to customize RetinaNet for object detection in microscopy images (Wollmann and Rohr, 2021). To avoid reinventing the wheel, we adapt representative methods per category to our task with application-tailored loss functions and further propose to leverage temporal information of DSA series to pursue expert-level performance.

Temporal feature learning models have been proposed typically in processing natural languages, audio and time series, and are also being actively explored for medical image sequence processing. Qin et al. exploited the temporal correlations of MR sequences for dynamic image reconstruction using bidirectional recurrent neural networks (RNN) (Qin et al., 2018). In stroke imaging, Neves et al. converted DSA series into 2D images by averaging over time and employed 2D CNNs to classify DSA into low- and high-grade treatment outcomes (Neves et al., 2021). Su et al. proposed to split DSA series into three temporal phases (arterial, parenchymal, and venous) using a CNN in a four-step automated TICI scoring pipeline (Su et al., 2021). Nielsen et al. utilized gated recurrent units (GRU) for end-to-end automatic TICI scoring (Nielsen et al., 2020; 2021). As an alternative to RNNs, temporal convolution networks (TCN) were employed for stroke lesion outcome prediction in 4D CT perfusion images (Amador et al., 2021). Besides, recently proposed attention networks have also been adopted for temporal feature representation (Li et al., 2020).

1.3. Contributions

In this work, we study the feasibility of automated perforation detection and localization. For this purpose, we developed and extensively evaluated several spatial and spatio-temporal approaches, equipped with application-tailored loss functions and series level post-processing. The main contribution of this work is two-fold:

- we propose to use spatio-temporal deep learning, i.e., adapted 2D object detectors equipped with temporal modules, for au-

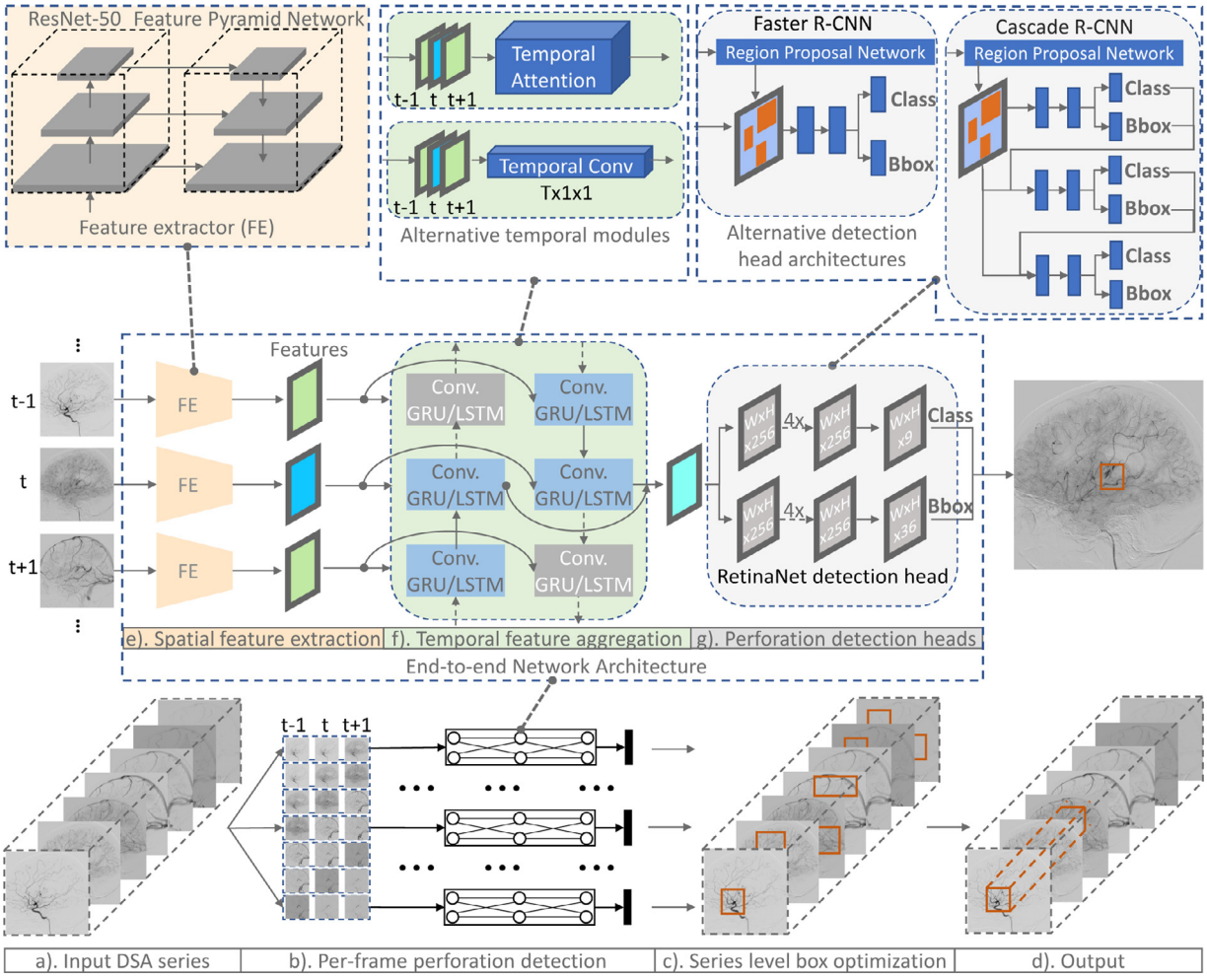


Fig. 2. Overview of the perforation detection architecture. Given an input DSA series (a), perforations are first detected on each frame by taking neigboring frames as context (b), and then optimized on the series level based on temporal consistency (c) (Section 2.4). The mostly likely perforation extravasation trajectory is selected as the final output (d). The end-to-end network architecture for per-frame perforation detection consists of a feature extractor (e) (feature pyramid network based on ResNet50), a temporal feature aggregator (f) (e.g., BiGRU, BiLSTM, temporal attention, Temporal convolution, see Section 2.3), and a perforation detector (g) (e.g., RetinaNet, Faster R-CNN, Cascade R-CNN, see Section 2.1). Note that the two gray GRU/LSTM blocks do not play a role in the produced feature map.

tomatic intracranial vessel perforation detection in DSA image series during EVT;
- we assess the proposed method variants on the largest clinical perforation dataset so far collected from multiple national and international clinical trials and registries, demonstrating expert-level performance, and hence the feasibility for use during interventions.

The remainder of this paper is organized as follows. First, Section 2 details the proposed method components for perforation detection. Next, Section 3 describes the datasets for experiments, including data statistics and the selection process. Experimental results and analyses are reported in Section 4, followed by qualitative analyses and further discussions in Section 5. Finally, Section 6 summarizes the main conclusions of this work.

2. Methods

Due to large variations in perforation size ranging from small perforations of a few mm^2 to larger perforations of more than 100 mm^2, direct image level classification using CNN based models may not well capture small relevant image parts, and hence may be vulnerable to noise and irrelevant features. In this work, we address perforation detection with a localize-to-classify strategy.

As shown in Fig. 2, the proposed approach consists of two steps: per-frame perforation detection using an end-to-end spatio-temporal network, and series level result optimization based on temporal consistency. The end-to-end network comprises three modules: spatial feature extraction, temporal feature aggregation, and perforation detection based on spatio-temporal features. First, the model takes as input the current frame together with a number of sequential frames as context. Subsequently, spatial textural features are extracted independently using a shared backbone CNN. Next, the spatial features of sequential frames are aggregated via RNNs. Last, an object detection head, which consists of a classification subnet and a regression subnet, is applied on the aggregated image features to obtain a set of perforation bounding boxes on the current frame. On the series level, the obtained per-frame bounding boxes are further optimized based on temporal consistency of the perforation contrast extravasation. In this section, we describe the tailored single-frame spatial networks, multi-frame spatio-temporal networks, and the series optimization mechanism.

2.1. Single frame spatial networks

We first study the feasibility of automatic perforation detection on DSA images using single-frame spatial networks. Representative single-, two-, and multi-stage methods are adapted for

this purpose. Additionally, we propose a Distance-ReLu Intersection over Union (DR-IoU) regression loss to cope with imperfectness in ground truth bounding box annotation due to irregular and ambiguous perforation borders.

2.1.1. Faster R-CNN: A two-stage strategy

The region-based CNN family (R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2015)) refers to a set of two-stage object detection meta-architectures, which can build on various feature extraction backbones. Faster R-CNN enables end-to-end training by introducing a region proposal network (RPN) for generating candidate bounding boxes (i.e., anchors) that may contain an object. Spatial feature patches are then cropped based on the region proposals, and fed into two head subnets for object classification and bounding box refinement. The network yields a set of object bounding box coordinates and their corresponding probability scores. In contrast to traditional sliding window based approaches, Faster R-CNN is not only much more efficient by greatly reducing the number of patches per image, but also more accurate by allowing various anchor ratios and sizes and a controlled foreground-background sample ratio (Ren et al., 2015).

2.1.2. RetinaNet: A single-stage strategy

Different from two-stage solutions, single-stage methods do not use region proposal networks and directly perform object classification and regression on feature maps in a fully convolutional manner. Although more time efficient, single-stage methods usually exhibit inferior accuracy in comparison to two-stage methods. Based on the hypothesis that the reduction in accuracy can be attributed to the class imbalance, Lin et al. proposed RetinaNet, an end-to-end meta-architecture equipped with a focal loss for focused training on a sparse set of hard training samples (Lin et al., 2017). Focal loss has been demonstrated especially effective in many medical image analysis applications (Lotter et al., 2021; Zhou et al., 2020), where positive and negative sample imbalance is often naturally inherent. This imbalance also holds true in perforation detection. In each DSA series, only one perforation is generally expected while many non-perforation bounding boxes can be extracted during training.

2.1.3. Cascade R-CNN: A multi-stage method

Cascade R-CNN has been proposed as a multi-stage meta-architecture with a set of sequentially connected detectors (Cai and Vasconcelos, 2018). It is designed to address the dilemma of noisy detection when training with low IoU thresholds and overfitting when using high IoU thresholds for training. Cascade R-CNN trains detectors progressively with gradually increased IoU thresholds using previous detector outputs as input. The high-level architecture is shown in Fig. 2.

2.2. Distance-ReLu IoU regression loss (DR-IoU)

For perforation bounding box regression during training, existing loss functions are mostly based on either ℓ_n-norm functions or IoU. Compared to ℓ_n-norm functions, which are sensitive to scale variance, IoU loss is scale invariant. However, IoU loss suffers from gradient vanishing in case of no overlap between target and ground truth boxes. In 2019, Rezatofighi et al. proposed Generalized IoU (GIoU), which overcomes this drawback while still suffering from slow convergence and inaccurate regression issues (Rezatofighi et al., 2019). Later, Distance-IoU (DIoU) loss was proposed with superior regression accuracy and optimized convergence (Zheng et al., 2020).

IoU-based losses can be generally expressed as in Eq. 1a. For DIoU, the penalty term $\mathcal{R}(B,B)$ is based on the Euclidean distance $d(b_{pd},b_{gt})$ as in Eq. 1b, where b_{pd} and b_{gt} are the center of bounding box B_{pd} and B_{gt} respectively, and c is a normalization factor defined as the diagonal of the smallest box containing B_{pd} and B_{gt}. The subscripts "pd" and "gt" refer to "predicted" and "ground truth" respectively.

$$\mathcal{L}_{\text{IoU}} = 1 - \text{IoU} + \mathcal{R}(B_{pd}, B_{gt}) \tag{1a}$$

$$\mathcal{R}(B_{pd}, B_{gt}) = \frac{d^2(b_{pd}, b_{gt})}{c^2} \tag{1b}$$

Unlike other object detection tasks where the object border is obvious (e.g., pedestrian, head, tumor), the perforation border is ambiguous and irregular when visually inspecting the extravasated contrast in DSA. As a result, a certain level of uncertainty at the borders is expected in the annotated ground truth bounding boxes. Therefore, pursuing a zero DIoU loss would lead to overfitting during bounding box regression. To overcome this issue, we propose DR-IoU to suppress small distance loss values ($\mathcal{L}_D <= \beta$) by a rectified linear unit (ReLu) function in case of sufficiently good predictions as expressed as

$$\mathcal{L}_{\text{DR-IoU}} = \text{ReLu}(1 - \text{IoU} + \frac{d^2(b_{pd}, b_{gt})}{c^2} - \beta)). \tag{2}$$

2.3. Multi-frame spatio-temporal networks

To exploit the temporal characteristics of DSA image series, we propose to incorporate a temporal module in the end-to-end network. This idea is inspired by the visual perforation inspection process, where radiologists rely on previous and subsequent frames to confirm the existence of contrast extravasation. With the goal of detecting a perforation directly after a DSA run (rather than during acquisition), several temporal information aggregation methodologies are studied in this work.

2.3.1. Convolutional recurrent units

Recurrent neural networks have been widely used in learning sequential dependencies. State-of-the-art examples are GRU and Long Short Term Memory (LSTM), which were proposed specifically to handle both long and short term memories using a gate mechanism to regulate the information flow. The recent concept of convolutional recurrent units (Ballas et al., 2015; Shi et al., 2015) leads to a module that can easily be combined with fully convolutional neural networks. These units inherit the strength of both convolutional layers and recurrent layers. That is to represent spatial and temporal features simultaneously while keeping the entire network end-to-end trainable.

The architectures of LSTM and GRU are illustrated in Fig. 3. For detailed information, please refer to Shi et al. (2015) and Ballas et al. (2015) respectively. In this work, we tailor both the convolutional LSTM and GRU modules to be bi-directional (Fig. 2), in which the outputs of forward and backward layer are concatenated along the channel dimension. In such a way, the model learns both the past and future temporal contrast extravasation characteristics. The output of bidirectional GRU/LSTM module is five feature maps of size $T \times 2C \times \frac{H}{R} \times \frac{W}{R} \times C$, where the number of channels $C = 256$ and the down-sampling factor $R \in [2, 4, 8, 16, 32]$. H, W, T denote the frame height, frame width, and frame length respectively.

2.3.2. Temporal convolution

A temporal convolution network (Lea et al., 2016) (TCN) is an alternative to RNN networks for learning sequential data characteristics in a fully convolutional manner. Examples of successful application of TCNs have been presented in (Yan et al., 2020; Dai

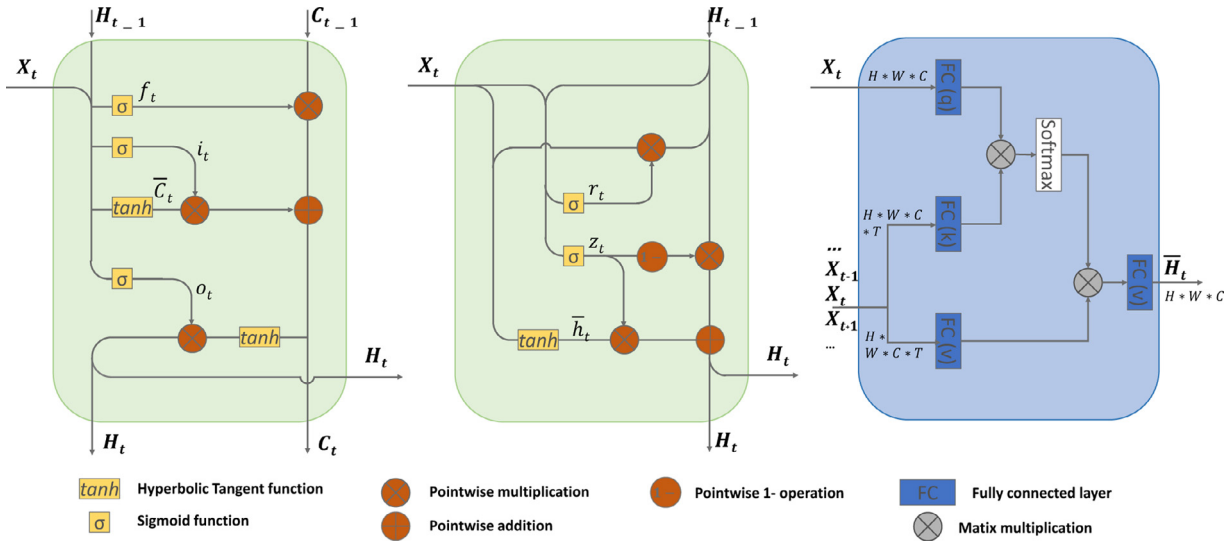


Fig. 3. Illustration of a LSTM block, a GRU block and an attention block. X_t: input feature maps; H_t: output feature maps; f_t: forget gate; i_t: input gate; o_t: output gate; $\bar{C}_t$: candidate cell state; C_t:cell state; r_t: reset gate; z_t: update gate; $\bar{h}_t$: candidate hidden state; $*$: convolution; $\circ$: dot product; H: feature map height; W: feature map width; C: number of channels; T: number of sequential frames; q: queries; v: values; k: keys; [..., X_{t-1}, X, X_{t+1}, ...]: context feature maps.

et al., 2020; Bai et al., 2018). The architecture of the temporal convolution module is shown in Fig. 2. Unlike the classical TCN, which uses causal convolutions via masking, we perform temporal convolutions on both previous and future frames. Via a Tx1x1 (T: frame length) convolution kernel, the model outputs features aggregated along the temporal dimension (i.e., the time intensity curve of each pixel).

2.3.3. Temporal attention

Temporal attention can similarly be an effective way to distill perforation progression in time. It allows the network to focus more on relevant frames in determining the presence of a vessel perforation. Fig. 3 illustrates the details of our implementation of a visual temporal attention module. The module takes as input the feature maps of the current frame (X_t) and neighboring frames ($X_{t\pm i}, i \in [1, k]$), and aggregates temporal context features of each feature point across all frames along the time axis. The resulting feature map is an aggregated feature map with the same size as a single frame feature map, which is then fed into the head subnets for bounding box classification and regression.

2.4. Temporal consistency based series level optimization

With the fully end-to-end deep learning network, perforation bounding boxes can be detected on each frame of a DSA series, each assigned with a probability score. To further exploit the temporal consistency of a vessel perforation, we propose to use a post-processing step to integrate the per-frame detection results by enforcing temporal consistency on the DSA series level. Similar to the concept of non-maximum suppression (NMS) and Seq-NMS (Han et al., 2016) for suppressing overlapping bounding boxes, we propose an application tailored step to determine the most likely bounding box trajectory awarded by temporal consistency (see Fig. 4), based on the assumptions that there is maximum one vessel perforation in a DSA series and that the vessel perforation should be temporally persistent.

Given a series of DSA frames with detected bounding boxes and corresponding probability scores, all possible temporal trajectories across the series are determined. Two bounding boxes in consecutive frames are on the same trajectory if the center of one box falls into another. Subsequently, the accumulated probability score is calculated along each trajectory, which is further multiplied by the temporal duration of the trajectory; this is to allocate extra rewards to temporally persistent detections. Ultimately, only the bounding boxes along the trajectory with the highest score survive. Via this series level optimization step, a set of temporally inconsistent false positives can be suppressed, leading to further improved perforation detection precision.

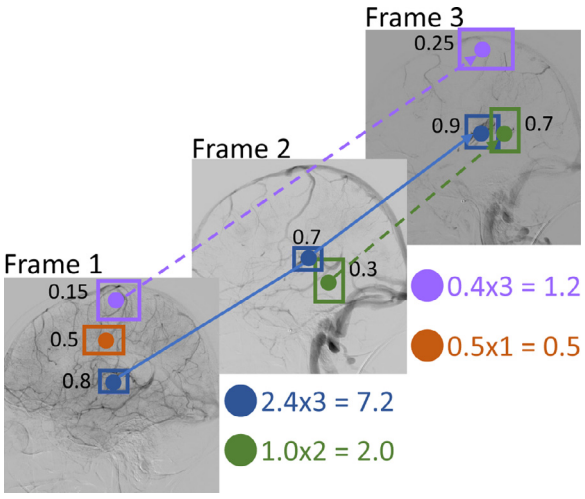


Fig. 4. Illustration of temporal consistency weighted series level bounding box optimization. Blue indicates the trajectory with the highest weighted accumulative score. The scores along the blue trajectory are then reassigned to the trajectory average score (0.8).

3. Data

We collected the largest perforation DSA dataset so far from three clinical studies, namely the MR CLEAN registry (Jansen et al., 2018), the HERMES collaboration (Goyal et al., 2016a), and the MR CLEAN NoIV Trial (Treurniet et al., 2021)[2]. The data selection process and statistics are summarized in Fig. 5.

[2] For data sharing policies, please refer to Jansen et al. (2018); Goyal et al. (2016a); Treurniet et al. (2021)

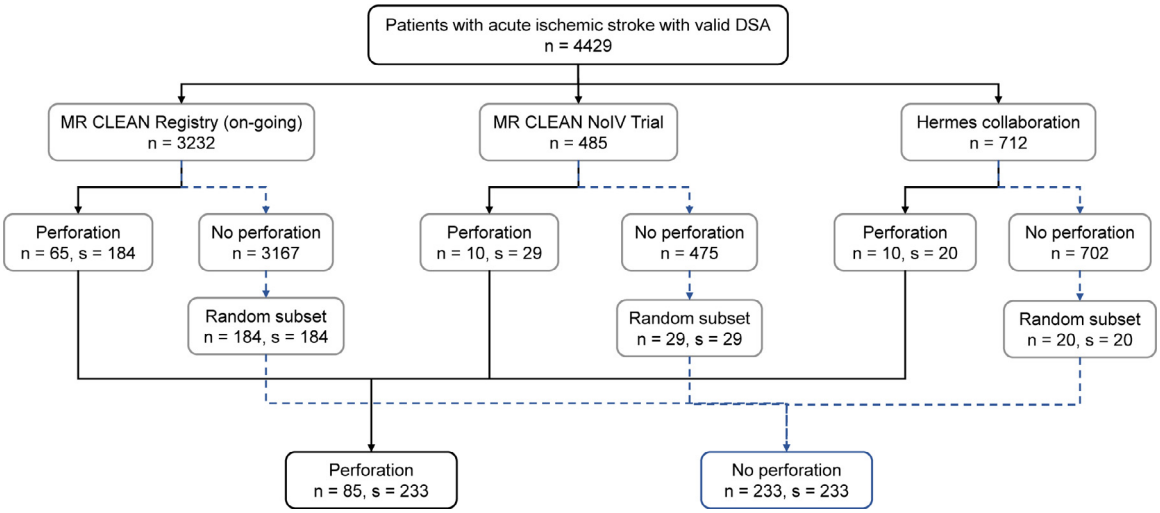


Fig. 5. An overview of data selection and distribution. n denotes the number of patients; s refers to the number of DSA series.

The MR CLEAN Registry is an observational cohort study which included patients with acute ischemic stroke from sixteen centers in the Netherlands between March 2014 and December 2018. For the current work, we used data from all patients registered until November 2017.

The HERMES (Highly Effective Reperfusion Using Multiple Endovascular Devices) collaboration (Goyal et al., 2016a) pooled data from seven major randomized trials: MR CLEAN (Berkhemer et al., 2015), ESCAPE (Goyal et al., 2015), REVASCAT (Jovin et al., 2015), SWIFT PRIME (Saver et al., 2015), EXTEND IA (Campbell et al., 2015), PISTE (Muir et al., 2017) and THRACE (Bracard et al., 2016), which were conducted between 2010 and 2017.

The MR CLEAN-NO IV trial (Treurniet et al., 2021) is a national multicenter randomized clinical trial conducted by the Collaboration of New Treatments of Acute Stroke (CONTRAST) consortium in the Netherlands. The main purpose of the trial is to study the added benefit of intravenous alteplase prior to intra-arterial thrombectomy in stroke patients with an intracranial occlusion of the anterior circulation.

From the above databases, we pooled a total of 4429 patients, out of which 105 patients with intracranial vessel perforation were initially identified based on the annotations from either the intervention operators or an imaging core laboratory. For this study, an experienced radiologist (hereafter referred to as expert 1) thoroughly reviewed the DSA images for all perforation cases and annotated the perforation locations with bounding boxes (see examples in Fig. 1). Besides, patients with a reported subarachnoid hemorrhage (SAH) were also checked for visible perforations in DSA by expert 1. As a result, a total of 85 patients (233 series) were confirmed to include perforations, out of which approximately 80% were subarachnoid perforations. This ground truth is based on the knowledge of intervention operators, core-lab neuroradiologists and expert 1. To obtain a balanced negative data sample counterpart, a comparable subset of patients was randomly selected from the non-perforation patients of each dataset. No stratification was performed. Patients were excluded if they are labelled with perforation according to intervention operators, core-lab neuroradiologists or expert 1. From the randomly selected negative patient subset, one DSA series was randomly chosen from each patient. In summary, 85 perforation patient cases (233 DSA series, 2193 perforation bounding boxes) and 233 non-perforation patient cases (233 DSA series) were used in our experiments.

The DSA series were obtained during EVT by various acquisition systems, including Philips, GE and Siemens, with an average of 13 (standard deviation: 9) series per patient. These images are of size 1024×1024 pixels with time length varying from 1 to 50 frames. The temporal resolution varies from 0.5-4 frames per second both during acquisition and across acquisitions and the spatial resolution is approximately $0.18 \times 0.18\, mm^2$.

4. Experiments and results

4.1. Implementation details

The described models were developed in Python using PyTorch (Paszke et al., 2019). The spatial and spatio-temporal models were trained on an NVIDIA 2080 Ti with 11 GB of memory.

The same feature pyramid network (FPN) backbone based on ResNet50 (He et al., 2016) pre-trained on ImageNet data (Deng et al., 2009) was used as feature extractor for all models. The FPN outputs five feature maps of size $\frac{H}{R} \times \frac{W}{R} \times C$, where the down-sampling factor $R \in [2, 4, 8, 16, 32]$ and the number of channels $C = 256$. H and W denote the image height and width respectively. The usage of an FPN facilitates multi-level feature extraction for various sized object detection. The backbone architecture is illustrated in Fig. 2. For bounding box regression, DR-IoU was adopted in all the networks, unless otherwise stated. For box classification, focal loss was utilized in RetinaNets and cross entropy (CE) loss was used in Faster R-CNN and Cascade R-CNN. A number of adaptations were made from Faster R-CNN (Ren et al., 2015), Cascade R-CNN (Cai and Vasconcelos, 2018), and RetinaNet (Lin et al., 2017) based on the application needs. The number of object classes was set to one as the only foreground object is perforation. Based on the assumption that maximum one perforation would exist in a DSA series, the maximum number of detections per image was also set to one. Non-perforation series were included in the training process for robust negative sample learning and false positive reduction. All spatio-temporal networks were based on RetinaNet due to its superior performance over Faster R-CNN and Cascade R-CNN. We adopted the default setting with 256 channels in the detection head of RetinaNet. For a fair comparison, this setting was kept consistent for RetinaNet and all spatio-temporal models.

The models were trained with a batch size of 2 images with the SGD optimizer (momentum: 0.9, weight decay: 0.0001). The networks configurations followed the standard setting of MMDetection (Chen et al., 2019b) and Detectron2 (Wu et al., 2019) with 12 epochs. Each epoch runs over the entire training set, result-

ing in approximately 27,000 iterations. A cosine annealing learning rate scheduler was used with an exponential warm up period of 1500 iterations and a minimum learning rate being 0.001 of the base learning rate 5×10^{-4}. The β for DR-IoU regression loss was set to 0.25. Besides, the input images were down-sampled to 640x640 pixels, converted to RGB, and normalized using ImageNet (Deng et al., 2009) mean ([123.675, 116.28, 103.53]) and standard deviation ([58.395, 57.12, 57.375]) for both training and inference. While training, we randomly applied the following augmentation techniques with a probability of 0.5: horizontal flipping, horizontal/vertical shifting (ratio $\in (0, 0.0625]$), scaling (factor $\in [-0.1, 0.1]$), and rotation (angle $\in [-10^{\circ}, 10^{\circ}]$).

Our quantitative and qualitative analyses were based on a stratified ten-fold cross-validation procedure with all patients randomly grouped into ten even subsets as shown in Fig. 5, ensuring balanced positive and negative DSA series per subset. Each of the subsets served in turns as validation set for models trained on the remaining subsets. Such a data split was kept consistent when evaluating various models.

4.2. Evaluation metrics

The model performance can be studied from two aspects: perforation/non-perforation classification on DSA series level and perforation localization. Accordingly, evaluation metrics are defined to address both aspects.

4.2.1. DSA series level perforation classification

The area under the curve (AUC) of the receiver operating characteristic (ROC) curve is reported as the main metric for evaluating the overall classification accuracy. Besides, we report the average specificity and sensitivity, where true positives refer to DSA series in which both the algorithm and radiologist identify a perforation. As the value of these three metrics vary along the probability threshold of the classification model, we report the values at the threshold with maximized Youden's J statistic (Youden, 1950):

$$\mathcal{J} = sensitivity + specificity - 1. \tag{3}$$

Due to the low occurrence rate of perforation, low false alarm rate is rather important in clinical practice. Therefore, the average sensitivity at 95% specificity (mSens_{95}) is also reported, which reflects the true alarm ratio while ensuring $\leq 5\%$ false alarm ratio.

4.2.2. Perforation localization

The performance of perforation localization is benchmarked using the average precision (AP) and average recall (AR). As perforations are localized in rectangular boxes, true positives were defined in two ways: 1) IoU $\geqslant 0.5$ and 2) center-overlapping (i.e., the prediction center falls in the ground truth box or vice versa). As reported in Table 1, AP50 and AR50 define the average precision and recall based on the former definition of true positions; AP_cin and AR_cin are based on the latter. The rationale of introducing the second true positive definition is illustrated in Fig. 7, where the detections (in orange box) should be considered as true positives despite their small IoU with reference annotations (in blue boxes). Considering the ambiguous perforation borders and their large variations in appearance, we introduce the center-overlapping criterion to emphasize the correctness of detection rather than the accuracy of boundary matching.

4.3. Quantitative evaluation

In this section, we assess the performance of both spatial (single-frame) and spatio-temporal (multi-frame) models using stratified ten-fold cross-validation. The performance of all model variants is visualized in Fig. 6, and quantitatively summarized in Table 1. Fig. shows the precision-recall curves for perforation localization, while the ROC curves for series level perforation classification is shown in Fig.

Among the three single-frame based models, RetinaNet showed superior performance over Faster R-CNN($P<0.001$) and Cascade R-CNN ($P<0.001$) in a DeLong test (DeLong et al., 1988), achieving a series level classification AUC of $0.862(\pm 0.007)$.

For the spatio-temporal models, we integrated the temporal modules as discussed in Section 2.3 into the architecture of RetinaNet. All the spatio-temporal models demonstrated statistically superior performance ($P<0.001$, DeLong test) to single-frame RetinaNet, archiving an AUC of 0.91 when followed by series level box optimization (SO). The performance difference between the spatio-temporal models were not statistically significant ($P=0.986$, one-way ANOVA). In principle, these models are similar in a sense that they all aggregate a sequence of feature maps along the time axis into one feature map with a fixed temporal window size. The inference speed of each model was above 5 frames per second, which is fast enough in clinical practice. Note that in the following experiments, a spatio-temporal model is randomly chosen if not all are reported.

4.4. Choosing the number of sequential frames

The reported spatio-temporal models in Table 1 take five frames as input: the current frame (t) and four neighboring frames ($t-2, t-1, t+1, t+2$). We hypothesize that the temporal progression is better modelled with more sequential frames. To validate this and determine the optimal number of input frames, we compared the performance of spatio-temporal models with regard to a different number of input frames. Table 2 confirms this hypothesis that a larger number of frames improves performance. A higher frame number could not be investigated due to the GPU memory limit.

4.5. Ensemble modeling

Ensemble methods strategically combine diverse models to improve the performance of a machine learning task, such as classification, prediction. Similar to the idea of accepting manuscript submissions based on reviews from multiple domain experts, ensemble methods may help in reducing single model errors based on majority voting. In this work, we combined the classification results of the four ST+SO models by taking the median value of series level classification probabilities. No information assembling was performed on the bounding box level. Similar to all other models, the reported classification performance was also based on the threshold with maximized Youden's J statistic (Youden, 1950). As shown in Table 1, the ensemble approach significantly outperforms all individual ST+SO models by 2% ($P<0.05$, DeLong test) in terms of series level classification, demonstrating its effectiveness in suppressing detection errors from single models.

4.6. Method versus human expert

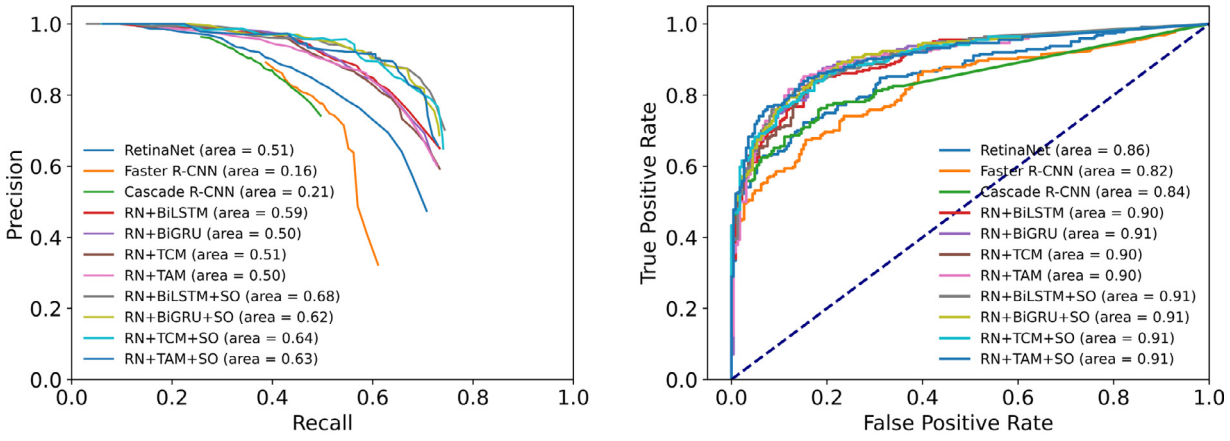
To evaluate the performance of the proposed automatic methods against human expert raters, a second experienced radiologist (hereafter referred to as expert 2) annotated a randomly selected subset (63 patients, 97 DSA series out of which 52 series from 18 patients are with perforations). This subset was also tested using a model trained on the remaining part of the whole perforation dataset.

The method-vs-human-expert experiment, as the name suggests, aims to evaluate the performance of the methods against an expert radiologist, rather than to evaluate interobserver variability. To this end, expert 1 and 2 performed annotation with different purposes and thus under different settings. Expert 1 aimed

Table 1
Performance overview on perforation localization and classification. AP_cin, AR_cin and AUPRC are the average precision, recall and precision-recall AUC under the center-overlapping criterion; AP50 and AR50 are based on the IoU $\geqslant$ 0.5 criterion. mSens and mSpec are the mean sensitivity and specificity for series level classification. All these values are reported based on maximized Youden's J statistic. $mSens_{95}$ refers to the mean sensitivity at the specificity of 95%. The reported performance numbers are based on five runs of each experiment. ST: spatio-temporal; RN: RetinaNet; TCM: temporal convolution module; TAM: temporal attention module; BiGRU: bidirectional convolutional GRU; BiLSTM: bidirectional convolutional LSTM; SO: series level bounding box optimization based on temporal consistency; std: the standard deviation of AUCs in five runs. The ensemble modeling performance is based on all ST+SO models.

Analysis	Method		Perforation localization					Series level classification			
			AP_cin	AR_cin	AUPRC	AP50	AR50	mSpec	mSens	$mSens_{95}$	**AUC(±std)**
Ten-fold	Spatial	Faster R-CNN	0.86	0.41	0.17	0.53	0.26	0.78	0.72	-	**0.819**(±0.006)
Cross	models	Cascade R-CNN	0.87	0.41	0.23	0.55	0.26	0.84	0.73	0.56	**0.832**(±0.010)
validation*		RetinaNet (RN)	0.88	0.45	0.50	0.57	0.29	0.83	0.74	0.59	**0.862**(±0.007)
	ST models	RN+TCM	0.84	0.60	0.52	0.54	0.38	0.83	0.83	0.63	**0.900**(±0.008)
		RN+TAM	0.86	0.58	0.51	0.53	0.36	0.85	0.83	0.63	**0.904**(±0.009)
		RN+BiLSTM	0.83	0.61	0.59	0.47	0.34	0.81	0.86	0.62	**0.901**(±0.008)
		RN+BiGRU	0.85	0.58	0.52	0.53	0.37	0.85	0.81	0.64	**0.900**(±0.011)
	ST+SO models	RN+TCM+SO	0.82	0.69	0.63	0.48	0.41	0.86	0.80	0.67	**0.906**(±0.011)
		RN+TAM+SO	0.84	0.69	0.63	0.49	0.39	0.88	0.81	0.69	**0.911**(±0.007)
			0.83	0.70	0.67	0.43	0.37	0.86	0.82	0.67	**0.909**(±0.008)
		RN+BiLSTM+SO									
		RN+BiGRU+SO	0.83	0.66	0.63	0.49	0.39	0.88	0.78	0.66	**0.905**(±0.014)
	Ensemble modeling		-	-	-	-	-	0.88	0.83	0.70	**0.929**(±0.006)
Generalizability*	RN+BiLSTM+SO		0.88	0.70	0.68	0.46	0.37	0.87	0.82	0.64	0.908(±0.007)
Method	RN+TCM+SO		0.86	0.76	0.64	0.49	0.44	**0.92**	**0.85**	0.70	0.922(±0.005)
vs human*	Ensemble modeling		-	-	-	-	-	**0.93**	**0.88**	0.73	0.935(±0.003)
	Expert 2		0.96	0.68	-	0.63	0.45	**1.00**	**0.71**	-	-

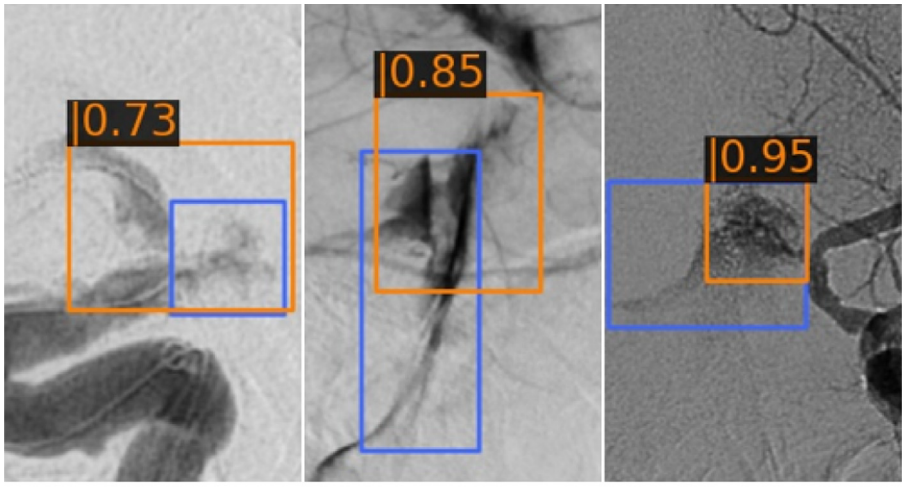
* Note that different data splits are used for cross validation (Section 4.3), generalizability test (Section 4.8), and the method-versus-human test (Section 4.6).



(a) Precision-recall curve for perforation localization with center overlapping metric.

(b) ROC curve for series level perforation classification.

Fig. 6. Performance overview of all model variants using cross validation. Note that values in this figure and in Table 1 sightly differ. Table 1 reports average values of five runs of cross validation, while this figure show results of one run. We picked one run of which the AUCs are close to the average numbers reported in Table 1.



(a) (b) (c)

Fig. 7. Examples of true detections according to the center-overlapping criterion with IoU < 0.5. Blue: reference annotation; orange: detected perforation box.

Table 2
Series level classification AUC of spatio-temporal models with respect to number of sequential frames. The p-values were obtained via DeLong test (DeLong et al., 1988).

ST models	Number of input frames		
	1	3	5
RN+TCM	0.862 (±0.007)	0.875(±0.010) 3vs1: P=0.325	0.900(±0.008) 5vs3: P<0.001 5vs1: P<0.001
RN+TAM		0.888(±0.006) 3vs1: P<0.001	0.904(±0.009) 5vs3: P<0.05 5vs1: P<0.001
RN+BiLSTM		0.885(±0.007) 3vs1: P<0.05	0.901(±0.008) 5vs3: P<0.05 5vs1: P<0.001
RN+BiGRU		0.892(±0.008) 3vs1: P<0.001	0.900(±0.011) 5vs3: P=0.087 5vs1: P<0.001

at defining the reference standard for perforation detection while expert 2 served as a competitor to the proposed methods. Expert 1 was provided only with patients that were known to include perforations or SAH (according to intervention operators and core-lab neuroradiologists) and was made aware of this fact. Non-perforation cases were not assessed by expert 1. In contrast, expert 2 was provided with patients out of which approximately 50% were with perforations and was only told that the dataset included mixed patients.

Fig. 8 and Table 1 demonstrate the comparison between the automated methods and expert 2 versus the ground truth. The annotated perforation boxes from expert 2 (AP50: 0.63, AR50: 0.45) show better alignment to the ground truth than the detected boxes (AP50: 0.49, AR50: 0.44) based on the IoU $\geqslant 0.5$ criterion, they are nevertheless comparable under the center-overlapping criterion (AP_cin: 0.96 versus 0.86, AR_cin: 0.68 versus 0.76). With respect to series level perforation classification, the automated method (i.e., RN+TCM+SO) is on par with expert 2 (specificity: 0.92 vs 1.00, sensitivity: 0.85 vs 0.71). With ensemble modeling using the four ST+SO models, the automated method achieve 0.93 and 0.88 in specificity and sensitivity respectively.

4.7. Ablation study

The proposed solution consists of three components: a spatial module (RetinaNet) for frame textural feature representation, a temporal module (e.g., BiGRU) for temporal feature aggregation, and a series level box optimization module (SO) for further false

Table 3
Ablation study on the proposed methods measured by perforation localization precision-recall AUC under the center overlapping criterion (AUPRC), average sensitivity at 95% specificity ($mSens_{95}$) and AUC for series level classification. The reported values are averaged over five runs. RN: RetinaNet; DR-IoU: Distance-ReLu regression loss; BiLSTM: Bidirectional LSTM; SO: series level bounding box optimization.

Components				Metrics		
RN	DR-IoU	BiLSTM	SO	AUPRC	$mSens_{95}$	AUC
✓	×	×	×	0.504(±0.006)	0.562(±0.035)	0.855(±0.007)
✓	✓	×	×	0.510(±0.009)	0.590(±0.028)	0.862(±0.013)
✓	✓	×	✓	0.640(±0.016)	0.642(±0.038)	0.875(±0.014)
✓	✓	✓	×	0.588(±0.007)	0.619(±0.032)	0.901(±0.008)
✓	✓	✓	✓	0.671(±0.015)	0.671(±0.027)	0.909(±0.008)

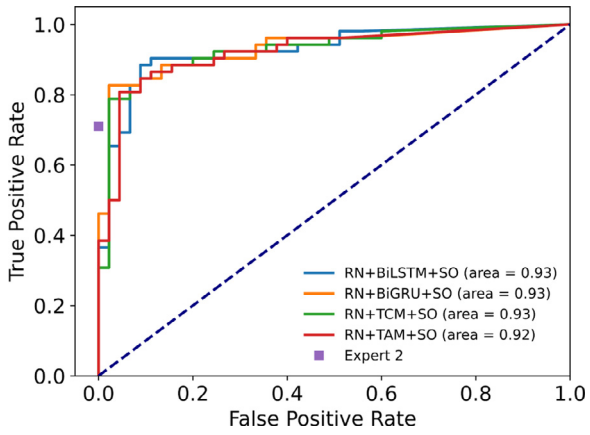


Fig. 8. Method versus human expert: ROC curve of perforation classification on series level.

positive reduction. To assess the added value of each component, we report the results of various module combinations using cross validation in Table 3. The added value of DR-IoU regression loss is also reported in comparison to DIoU loss.

The results demonstrate consistent performance gain when adding any of the components. Specifically, it can be seen that RN+DR-IoU is slightly better than RN+DIoU with $mSens_{95}$ of 0.59 vs 0.56 and AUC of 0.862 vs 0.855 (P=0.6, DeLong test). RetinaNet (RN) equipped with a temporal module (e.g., BiLSTM) outperforms the RetinaNet alone, with an average margin of 4% in terms of AUC (P<0.001, DeLong test). Besides, with the add-on of SO, the $mSens_{95}$ gains a consistent boost of 2%-5% and the AUC increases by approximately 1% (P<0.001, DeLong test) for both spatial and spatio-temporal models, revealing its effectiveness in false positive reduction.

4.8. Generalizability test

With the goal of investigating the generalization capability of the proposed solution, we assess the performance across different clinical trials with the model trained on the remaining datasets. As shown in Table 1, an average AUC of 0.908 (±0.007) is achieved by the RN+BiLSTM+SO model in term of series level classification, which is in line with the cross validation AUC (0.909±0.008). More specifically, Table 4 shows separated cross dataset testing results. When tested on NoIV or HERMES, the series level AUC tested on NoIV is slightly higher than that of the HERMES dataset (0.929 vs 0.879). This is likely due to the fact that NoIV is a national trial similar to MR CLEAN while HERMES is a heterogeneous mixture of multiple international trials. This strong cross dataset performance demonstrates the applicability of the methods in generalized clinical scenarios. Note that MR CLEAN was not selected as testset due to limited size of remaining training data.

Table 4
Model performance of the generalizability test using RN+BiLSTM+SO measured by perforation localization precision-recall AUC under the center overlapping criterion (AUPRC), average sensitivity at 95% specificity ($mSens_{95}$) and AUC for series level classification. The reported values are averaged over five runs.

Data	Train Test	MR CLEAN+HERMES NoIV	MR CLEAN+NoIV HERMES
Metrics	AUPRC	0.724(±0.039)	0.604(±0.031)
	$mSens_{95}$	0.717(±0.029)	0.660(±0.082)
	AUC	0.929(±0.017)	0.879(±0.018)

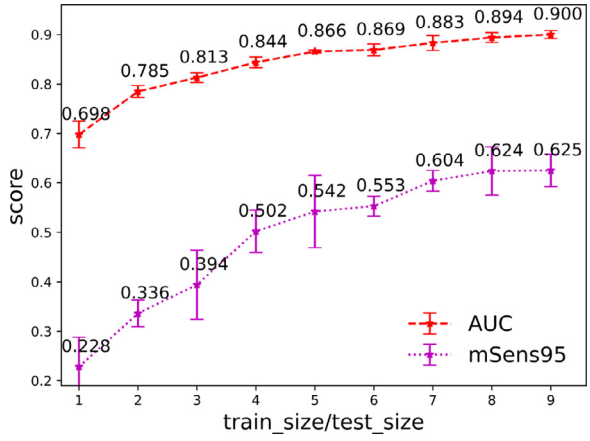


Fig. 9. Series level classification AUC and $mSens_{95}$ with respect to training/test data size ratio. $mSens_{95}$: average sensitivity at 95% specificity over five runs.

4.9. Impact of data size

We identified perforation cases in multiple large clinical databases, resulting in 85 perforation patient cases (233 DSA series). Due to low occurrence rate of perforation and its high heterogeneity, a further extended dataset may help deep learning models better represent visual perforation characteristics. To study the impact of training data size, we assessed the performance of the RN+TCM model trained with randomly reduced number of patients in training set by various ratios in $[1/9, 9/9]$, which approximately correspond to $[1, 9]$ times the size of test set in ten-fold cross validation. The test set was kept consistent during the experiment. Fig. 9 shows that both the AUC and the sensitivity (at 95% specificity) of series level classification gradually increase as the train data size grows, approaching saturation at train/test ratio of 9. The method may still benefit from a further enlarged dataset, but a substantial increase would be required for likely only a minor gain. However, given the variation in perforation appearances, it may

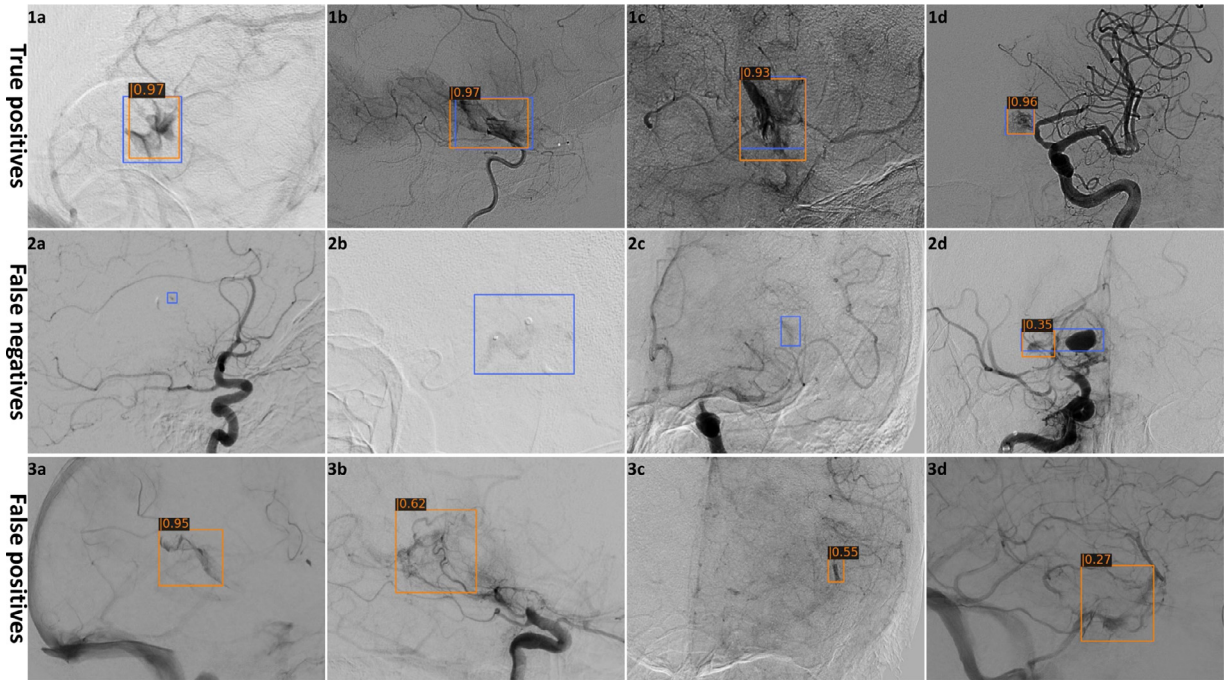


Fig. 10. Visualization of perforation detection results. Blue: ground truth annotations; orange: detected objects by the spatio-temporal model RN+TAM.

still be worthwhile to construct a larger dataset, as it would lower the chance of certain perforation characteristics not being seen by the model (see example 2d in Figure. 10).

5. Discussion

In this work we proposed spatio-temporal networks for fast automatic detection of vessel perforations during EVT, establishing the first benchmark using data from multiple national and international clinical trials and registries. Furthermore, we showcased the model generalizability and its competitive performance via cross-dataset testing and comparisons against a human expert.

To delve deeper into the results, we qualitatively assessed the detected perforations. Fig. 10 contains some true positive (TP), false positive (FP) and false negative (FN) examples using the RN+TAM model. The TPs in the top row showcase the capability of the proposed method in capturing relevant perforation appearance features. The second row exemplifies scenarios where the automated solution fails to locate perforations (FNs), including small object size (2a), inadequate contrast between perforation and its surrounding pixels (2b, 2c), and uncommon perforation types (2d). While successfully identifying true perforations, the model also generates some false positives, shown in the bottom row. The appearance of such false positives is hardly distinguishable from true perforations. Wide venous structures may appear similar to extravasated contrasts; thus they occasionally trigger false alarms (3d).

The benefit of temporal modules is demonstrated in Fig. 11 via visual examples of the spatial model (RetinaNet) and one of the spatio-temporal models (RN+TAM). It is noticeable that the temporal feature aggregation helps in stabilizing detections in consecutive frames in terms of both bounding box localization and classification scores in case (a) and (b).

It may be valuable to gain insights into the model performance with respect to perforation types. The number of patients in this study is however insufficient to provide a meaningful analysis at the perforation type level. Approximately 80% of the perforations are subarachnoid.

In comparison with the automated perforation detection solution, expert 2 achieved a high specificity in identifying DSA series with perforations. It is expected that although human experts are confident on recognized perforations (high specificity), they may overlook perforations (low sensitivity) due to time pressure and the low occurrence rate of perforation. In addition, it is worth to note that this performance does not fully represent the clinical practice due to the fact that expert 2 was explicitly asked to look for perforations and was not under time pressure as in an endovascular intervention.

Unlike online video processing tasks, temporal causality is not necessary in this application. Perforation detection from DSA series is intrinsically an offline action, as DSA images are acquired between retrieval attempts to check the intervention results, in which a perforation could have occurred, rather than is happening. Therefore, this application aims to detect perforations promptly after a DSA acquisition. If integrated in Angio Suites, this application can provide fast feedback to operators between retrieval attempts in stroke thrombectomy.

Due to the low occurrence rate of intracranial vessel perforation, high specificity is desired in clinical practice to avoid unwanted false positives. Additional false positive reduction techniques may further improve the model specificity. We noticed that some venous vessel structures can occasionally fool the model. Due to the fact that the occurrence of perforation is in the arterial phase, it would be practical to automatically exclude the venous phase frames (Su et al., 2021) for perforation detection. Another potential improvement is to embed the prior knowledge of the catheter location in the detection model, which can guide the model attention and suppress false positives in irrelevant brain regions.

The clinical value of the proposed automatic perforation detection can be reflected in the following aspects: 1) **improved detection**: intra-operative vessel perforations can be overlooked. This is evidenced from our experiment in Section 4.6. Such missed perforations may have detrimental clinical consequences (Salsano et al., 2020). An automatic diagnostic solution can help in reducing such false negatives; 2) **time is brain**: time is key to the rescue of

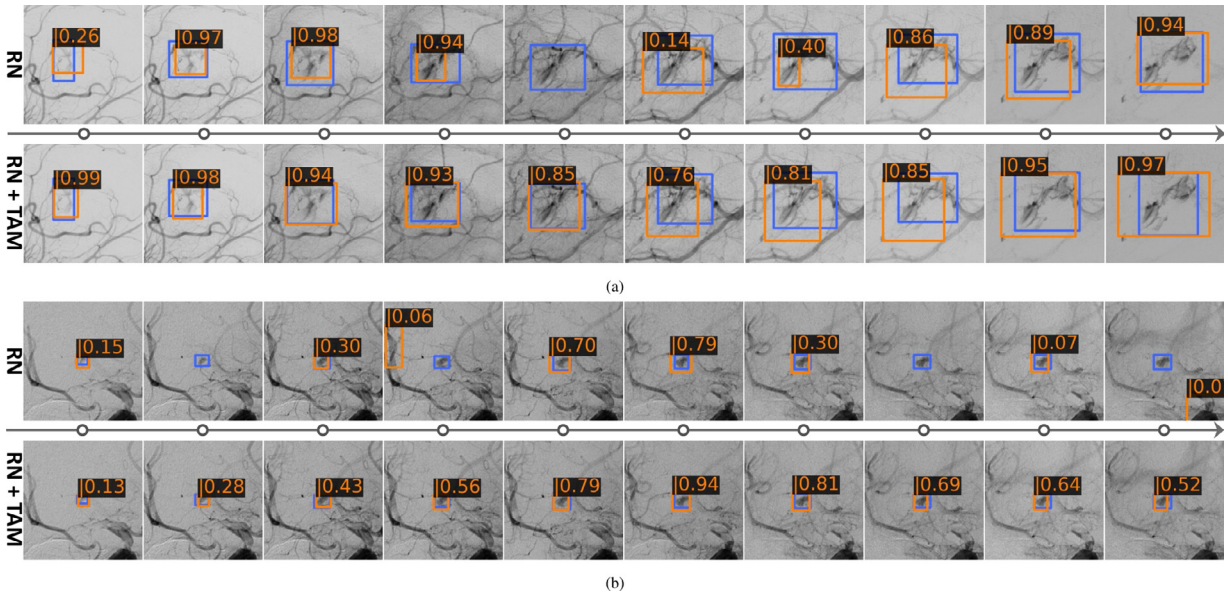


Fig. 11. Visual comparisons between spatial network RetinaNet (RN) and spatio-temporal network (RN+TAM) on two DSA series. Blue: ground truth annotations; orange: model predictions. The numbers are the detection probability scores.

brain tissues in stroke (Powers et al., 2019; Saver, 2006), also in case of a perforation (Jadhav et al., 2018). Early notice of the occurrence would avoid further peri-procedural manipulation of the affected vessel and allow to assess the need of therapeutic rescue actions (Jadhav et al., 2018), e.g., balloon tamponade, immediate reversal of anticoagulants, and lowering of arterial blood pressure (Ryu et al., 2011; Akpinar and Yilmaz, 2016; Leishangthem and Satti, 2014), to prevent clinical deteriorating. Automated solutions can detect perforation events directly after DSA acquisition, thus enabling optimal intra-procedural decision making. 3) **clinical research**: an automatic approach would also facilitate large scale clinical dataset processing and perforation related analyses. Such clinical insights could help in studying the effects of therapeutic decisions on functional outcomes.

6. Conclusion

In this work we presented the first study which addresses automatic intracranial vessel perforation detection during EVT using spatio-temporal networks. Evaluated on three multi-center clinical databases, we demonstrated the generalization capability of the model and established the benchmark for this task. Moreover, leveraging the temporal progression feature of perforation in sequential DSA frames, the proposed solution achieves expert-level performance with an AUC of 93% for series level classification, revealing its potential value in assisting therapeutic decision making in clinical practice.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Charles B.L.M. Majoie received funds from TWIN Foundation (related to this project, paid to institution); and from CVON/Dutch Heart Foundation, Stryker, European Commission, Health Evaluation Netherlands (unrelated; all paid to institution). Charles Majoie is shareholder of Nico.lab, a company that focuses on the use of artificial intelligence for medical imaging analysis.

Danny Ruijters is an employee of Philips Healthcare.

Wiro J. Niessen is founder, scientific lead, and shareholder of Quantib BV.

Acknowledgments

The authors would like to thank the MR CLEAN Registry investigators, MR CLEAN NoIV investigators, and the HERMES collaboration investigators (including MR CLEAN, ESCAPE, REVASCAT, SWIFT PRIME, THRACE, EXTEND-IA, and PISTE randomized controlled trials) for sharing the DSA image data. The MR CLEAN Registry was funded and carried out by the Erasmus University Medical Centre, Amsterdam UMC location AMC, and Maastricht University Medical Centre. The study was additionally funded by the Applied Scientific Institute for Neuromodulation (Toegepast Wetenschappelijk Instituut voor Neuromodulatie). MR CLEAN NoIV study was performed in the framework of the CONTRAST consortium which acknowledges the support from the Netherlands Cardiovascular Research Initiative, an initiative of the Dutch Heart Foundation (CVON2015-01: CONTRAST), and from the Brain Foundation Netherlands (HA2015.01.06). The collaboration project is additionally financed by the Ministry of Economic Affairs by means of the PPP Allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships (LSHM17016). This work was funded in part through unrestricted funding by Stryker, Medtronic and Cerenovus.

The current work on perforation detection was supported by Health-Holland (TKI Life Sciences and Health) through the Q-Maestro project under Grant EMCLSH19006 and Philips Healthcare (Best, The Netherlands).

References

Akins, P., Amar, A., Pakbaz, R., Fields, J., 2014. Complications of endovascular treatment for acute stroke in the swift trial with solitaire and merci devices. American Journal of Neuroradiology 35 (3), 524–528.

Akpinar, S.H., Yilmaz, G., 2016. Periprocedural complications in endovascular stroke treatment. Br J Radiol 89 (1057), 20150267.

Amador, K., Wilms, M., Winder, A., Fiehler, J., Forkert, N., 2021. Stroke lesion outcome prediction based on 4d CT perfusion data using temporal convolutional networks. In: Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., Ernst, F. (Eds.), Proceedings of the Fourth Conference on Medical Imaging with Deep Learning. PMLR, pp. 22–33. https://proceedings.mlr.press/v143/amador21a.html

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.
Ballas, N., Yao, L., Pal, C., Courville, A., 2015. Delving deeper into convolutional networks for learning video representations. arXiv preprint arXiv:1511.06432.
Berkhemer, O.A., Fransen, P.S., Beumer, D., Van Den Berg, L.A., Lingsma, H.F., Yoo, A.J., Schonewille, W.J., Vos, J.A., Nederkoorn, P.J., Wermer, M.J., et al., 2015. A randomized trial of intraarterial treatment for acute ischemic stroke. n Engl J Med 372, 11–20.
Bracard, S., Ducrocq, X., Mas, J.L., Soudant, M., Oppenheim, C., Moulin, T., Guillemin, F., et al., 2016. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (thrace): a randomised controlled trial. The Lancet Neurology 15 (11), 1138–1147.
Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6154–6162.
Campbell, B.C., Mitchell, P.J., Kleinig, T.J., Dewey, H.M., Churilov, L., Yassi, N., Yan, B., Dowling, R.J., Parsons, M.W., Oxley, T.J., et al., 2015. Endovascular therapy for ischemic stroke with perfusion-imaging selection. N top N. Engl. J. Med. 372 (11), 1009–1018.
Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., et al., 2019. Hybrid task cascade for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4974–4983.
Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al., 2019. Mmdetection: open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155.
Dai, R., Xu, S., Gu, Q., Ji, C., Liu, K., 2020. Hybrid spatio-temporal graph convolutional network: Improving traffic prediction with navigation data. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3074–3082.
DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 837–845.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp. 248–255.
Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 1440–1448.
Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587.
Goyal, M., Demchuk, A.M., Menon, B.K., Eesa, M., Rempel, J.L., Thornton, J., Roy, D., Jovin, T.G., Willinsky, R.A., Sapkota, B.L., et al., 2015. Randomized assessment of rapid endovascular treatment of ischemic stroke. N top N. Engl. J. Med. 372 (11), 1019–1030.
Goyal, M., Menon, B.K., van Zwam, W.H., Dippel, D.W., Mitchell, P.J., Demchuk, A.M., Dávalos, A., Majoie, C.B., van der Lugt, A., De Miquel, M.A., et al., 2016. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. The Lancet 387 (10029), 1723–1731.
Goyal, M., Menon, B.K., van Zwam, W.H., Dippel, D.W., Mitchell, P.J., Demchuk, A.M., Dávalos, A., Majoie, C.B., Van Der Lugt, A., De Miquel, M.A., et al., 2016. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. The Lancet 387 (10029), 1723–1731.
Han, W., Khorrami, P., Paine, T.L., Ramachandran, P., Babaeizadeh, M., Shi, H., Li, J., Yan, S., Huang, T.S., 2016. Seq-nms for video object detection. arXiv preprint arXiv:1602.08465.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Jadhav, A.P., Molyneaux, B.J., Hill, M.D., Jovin, T.G., 2018. Care of the post-thrombectomy patient. Stroke 49 (11), 2801–2807.
Jansen, I.G., Mulder, M.J., Goldhoorn, R.-J.B., 2018. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (mr clean registry). BMJ 360.
Jovin, T.G., Chamorro, A., Cobo, E., de Miquel, M.A., Molina, C.A., Rovira, A., San Román, L., Serena, J., Abilleira, S., Ribó, M., et al., 2015. Thrombectomy within 8 hours after symptom onset in ischemic stroke. N top N. Engl. J. Med. 372 (24), 2296–2306.
Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation. In: European Conference on Computer Vision. Springer, pp. 47–54.
Leishangthem, L., Satti, S.R., 2014. Vessel perforation during withdrawal of trevo provue stent retriever during mechanical thrombectomy for acute ischemic stroke: case report. J. Neurosurg. 121 (4), 995–998.
Li, J., Liu, X., Zhang, W., Zhang, M., Song, J., Sebe, N., 2020. Spatio-temporal attention networks for action recognition and detection. IEEE Trans Multimedia 22 (11), 2990–3001.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, pp. 21–37.
Liu, X., Guo, X., Liu, Y., Yuan, Y., 2021. Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images. Med Image Anal 102052.
Lotter, W., Diab, A.R., Haslam, B., Kim, J.G., Grisot, G., Wu, E., Wu, K., Onieva, J.O., Boyer, Y., Boxerman, J.L., et al., 2021. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat. Med. 27 (2), 244–249.
Mokin, M., Fargen, K.M., Primiani, C.T., Ren, Z., Dumont, T.M., Brasiliense, L.B., Dabus, G., Linfante, I., Kan, P., Srinivasan, V.M., et al., 2017. Vessel perforation during stent retriever thrombectomy for acute ischemic stroke: technical details and clinical outcomes. J Neurointerv Surg 9 (10), 922–928.
Muir, K.W., Ford, G.A., Messow, C.-M., Ford, I., Murray, A., Clifton, A., Brown, M.M., Madigan, J., Lenthall, R., Robertson, F., et al., 2017. Endovascular therapy for acute ischaemic stroke: the pragmatic ischaemic stroke thrombectomy evaluation (piste) randomised, controlled trial. Journal of Neurology, Neurosurgery & Psychiatry 88 (1), 38–44.
Neves, G., Warman, P., Bueso, T., Duarte-Celada, W., Windisch, T., 2021. Identification of successful cerebral reperfusions (mtici $\geq$ 2b) using an artificial intelligence strategy. Neuroradiology 1–7.
Nielsen, M., Waldmann, M., Frölich, A.M., Flottmann, F., Hristova, E., Bendszus, M., Seker, F., Fiehler, J., Sentker, T., Werner, R., 2021. Deep learning–based automated thrombolysis in cerebral infarction scoring: a timely proof-of-principle study. Stroke 52 (11), 3497–3504.
Nielsen, M., Waldmann, M., Sentker, T., Frölich, A., Fiehler, J., Werner, R., 2020. Time matters: Handling spatio-temporal perfusion information for automated tici scoring. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 86–96.
Nogueira, R.G., Lutsep, H.L., Gupta, R., Jovin, T.G., Albers, G.W., Walker, G.A., Liebeskind, D.S., Smith, W.S., et al., 2012. Trevo versus merci retrievers for thrombectomy revascularisation of large vessel occlusions in acute ischaemic stroke (trevo 2): a randomised trial. The Lancet 380 (9849), 1231–1240.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Advances in neural information processing systems, pp. 8026–8037.
Powers, W.J., Rabinstein, A.A., Ackerson, T., Adeoye, O.M., Bambakidis, N.C., Becker, K., Biller, J., Brown, M., Demaerschalk, B.M., Hoh, B., et al., 2019. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the american heart association/american stroke association. Stroke 50 (12), e344–e418.
Qin, C., Schlemper, J., Caballero, J., Price, A.N., Hajnal, J.V., Rueckert, D., 2018. Convolutional recurrent neural networks for dynamic mr image reconstruction. IEEE Trans Med Imaging 38 (1), 280–290.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.
Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658–666.
Ryu, C.-W., Lee, C.-Y., Koh, J.S., Choi, S.K., Kim, E.J., 2011. Vascular perforation during coil embolization of an intracranial aneurysm: the incidence, mechanism, and clinical outcome. Neurointervention 6 (1), 17.
Salsano, G., Pracucci, G., Mavilio, N., Saia, V., di Poggio, M.B., Malfatto, L., Sallustio, F., Wlderk, A., Limbucci, N., Nencini, P., et al., 2020. Complications of mechanical thrombectomy for acute ischemic stroke: incidence, risk factors, and clinical relevance in the italian registry of endovascular treatment in acute stroke. International Journal of Stroke. 1747493020976681
Saver, J.L., 2006. Time is brain-quantified. Stroke 37 (1), 263–266.
Saver, J.L., Goyal, M., Bonafe, A., Diener, H.-C., Levy, E.I., Pereira, V.M., Albers, G.W., Cognard, C., Cohen, D.J., Hacke, W., et al., 2015. Stent-retriever thrombectomy after intravenous t-pa vs. t-pa alone in stroke. N top N. Engl. J. Med. 372 (24), 2285–2295.
Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu Rev Biomed Eng 19, 221–248.
Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional lstm network: a machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst 2015, 802–810.
Su, R., Cornelissen, S.A.P., van der Sluijs, M., van Es, A.C.G.M., van Zwam, W.H., Dippel, D.W.J., Lycklama, G., van Doormaal, P.J., Niessen, W.J., van der Lugt, A., van Walsum, T., 2021. Autotici: automatic brain tissue reperfusion scoring on 2d dsa images of acute ischemic stroke patients. IEEE Trans Med Imaging 40 (9), 2380–2391. doi:10.1109/TMI.2021.3077113.
Treurniet, K.M., LeCouffe, N.E., Kappelhof, M., Emmer, B.J., van Es, A.C., Boiten, J., Lycklama, G.J., Keizer, K., Lonneke, S., Lingsma, H.F., et al., 2021. Mr clean-no iv: intravenous treatment followed by endovascular treatment versus direct endovascular treatment for acute ischemic stroke caused by a proximal intracranial occlusion-study protocol for a randomized clinical trial. Trials 22 (1), 1–15.

WHO, G., 2018. Global health estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016.
Wollmann, T., Rohr, K., 2021. Deep consensus network: aggregating predictions to improve object detection in microscopy images. Med Image Anal 70, 102019.
Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.
Yan, J., Mu, L., Wang, L., Ranjan, R., Zomaya, A.Y., 2020. Temporal convolutional networks for the advance prediction of enso. Sci Rep 10 (1), 1–15.
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35.
Zhao, S., Wu, X., Chen, B., Li, S., 2021. Automatic vertebrae recognition from arbitrary spine mri images by a category-consistent self-calibration detection framework. Med Image Anal 67, 101826.
Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D., 2020. Distance-iou loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, pp. 12993–13000.
Zhou, D., Tian, F., Tian, X., Sun, L., Huang, X., Zhao, F., Zhou, N., Chen, Z., Zhang, Q., Yang, M., et al., 2020. Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. Nat Commun 11 (1), 1–9.