

Document Version

Final published version

Citation (APA)

Kore, A. A., te Nijenhuis, F. G., van der Sluijs, M., van Zwam, W., Majoie, C., Nijeholt, G. L. À., Ruijters, D., Vos, F., Cornelissen, S., Su, R., & van Walsum, T. (2026). OccluNet: Spatio-Temporal Deep Learning for Occlusion Detection on DSA. In R. Su, E. de la Rosa, L. Vorberg, L. Rist, J. Zhang, A. Hilbert, & T. van Walsum (Eds.), *Image Analysis in Stroke Diagnosis and Interventions - 5th International Workshop, SWITCH 2025, Held in Conjunction with MICCAI 2025, Proceedings* (pp. 22-31). (Lecture Notes in Computer Science; Vol. 16098 LNCS). Springer. https://doi.org/10.1007/978-3-032-07945-9_3

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

**Green Open Access added to [TU Delft Institutional Repository](#)
as part of the Taverne amendment.**

More information about this copyright law amendment
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:
the publisher is the copyright holder of this work and the
author uses the Dutch legislation to make this work public.



OccluNet: Spatio-Temporal Deep Learning for Occlusion Detection on DSA

Anushka A. Kore¹, Frank G. te Nijenhuis²✉, Matthijs van der Sluijs²,
Wim van Zwam³, Charles Majoie⁴, Geert Lycklama à Nijeholt⁵,
Danny Ruijters⁶, Frans Vos¹, Sandra Cornelissen², Ruisheng Su⁶,
and Theo van Walsum²

¹ TU Delft, Mekelweg 5, 2628 CD Delft, The Netherlands

`f.m.vos@tudelft.nl`

² Erasmus MC, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
`{f.tenijenhuis,p.m.vandersluijs,s.cornelissen,t.vanwalsum}@erasmusmc.nl`

³ Maastricht UMC+, P. Debyelaan 25, 6229 HX Maastricht, The Netherlands

`w.van.zwam@mumc.nl`

⁴ Amsterdam UMC, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

`c.b.majoie@amsterdamumc.nl`

⁵ Haaglanden MC, Lijnbaan 32, 2512 VA Den Haag, The Netherlands

`g.lycklama.a.nijeholt@haaglandenmc.nl`

⁶ TU Eindhoven, Groene Loper 5, 5612 AE Eindhoven, The Netherlands

`{d.ruijters,r.su}@tue.nl`

Abstract. Accurate detection of vascular occlusions during endovascular thrombectomy (EVT) is critical in acute ischemic stroke (AIS). Interpretation of digital subtraction angiography (DSA) sequences poses challenges due to anatomical complexity and time constraints. This work proposes OccluNet, a spatio-temporal deep learning model that integrates YOLOX, a single-stage object detector, with transformer-based temporal attention mechanisms to automate occlusion detection in DSA sequences. We compared OccluNet with a YOLOv11 baseline trained on either individual DSA frames or minimum intensity projections. Two spatio-temporal variants were explored for OccluNet: pure temporal attention and divided space-time attention. Evaluation on DSA images from the MR CLEAN Registry revealed the model's capability to capture temporally consistent features, achieving precision and recall of 89.02% and 74.87%, respectively. OccluNet significantly outperformed the baseline models, and both attention variants attained similar performance. Source code is available [here](#).

Keywords: Artificial Intelligence · Acute Ischemic Stroke · Large Vessel Occlusion · Object Detection

1 Introduction

Acute Ischemic Stroke (AIS) is a potentially life-threatening medical emergency caused by a sudden interruption of blood flow to a region of the brain due to

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

R. Su et al. (Eds.): SWITCH 2025, LNCS 16098, pp. 22–31, 2026.

https://doi.org/10.1007/978-3-032-07945-9_3

a vessel occlusion [2]. Endovascular thrombectomy (EVT) is a minimally invasive intervention to restore cerebral perfusion through mechanical removal of the occlusion. During EVT procedures, digital subtraction angiography (DSA) images are acquired to visualize the occlusion and to serve as a roadmap to navigate to the occlusion [6]. DSA is acquired in 2-4 frames per second to visualize the contrast agent flowing through the vasculature. It remains the gold standard for luminal imaging, delivering superior spatial and temporal resolution of the cerebral vasculature. Recent advances in deep learning (DL) have demonstrated potential in automating occlusion detection in various imaging modalities used for the diagnosis and management of stroke [4]. For DSA imaging in particular, recent studies to automate occlusion detection relied on DL methods applied to static frames. Karlsberg et al. [10] employed YOLOv3 (You Only Look Once), a single-stage object detector, to detect occlusions on single-frame datasets derived from DSA sequences. Khankari et al. [12] utilized a ResNet-50 convolutional neural network (CNN) to analyze patch-sampling-based regions of DSA minimum intensity projections (MinIPs) for carotid terminus detection and occlusion classification. Warman et al. [17] used an ensemble method, in which a DSA frame was evaluated by a ResNet-18-based detection model to predict if it had a large vessel occlusion (LVO), in which case the frame was then fed into a localization model based on a YOLOv5 backbone to localize the LVO with a bounding box. These approaches, however, treat DSA as a collection of independent images or collapse them into a single view using MinIP, ignoring the temporal information contained in the images. Leveraging the temporal nature of DSA imaging, Mittmann et al. [13] demonstrated that a thrombus yes-no-classification can be achieved by combining EfficientNet as a feature extractor with a bidirectionally configured long-short term memory (LSTM) or a gated recurrent unit (GRU) to establish spatio-temporal relations between the individual images of the DSA sequence. Kelly et al. [11] evaluated the performance of several DL models, including a 2D CNN Xception (using only single DSA frames), a stacked 2D CNN stacked-Xception (using multiple DSA frames), a 2D vision transformer (ViT), and a 3D-Inception CNN, the latter capturing full spatial and temporal resolution and achieving high sensitivity in occlusion detection using pre- and post-thrombectomy DSA video sequences. These studies confirmed that temporal context is critical for occlusion classification, particularly for distal occlusions where small thrombi may only become apparent through sequential frame analysis. Methods that explicitly localize the occlusion in the DSA are still under-explored. We hypothesized that occlusion detection on DSA is a task that can benefit from integration of spatio-temporal information, due to the inherently temporal aspect of occlusion visualization on DSA.

Building upon these advances in spatio-temporal DSA analysis, we develop a method that explicitly localizes the occlusion using a bounding box. We introduce OccluNet, a DL pipeline that combines YOLOX [7]— a high-performance single-stage object detector— with transformer-based temporal modules. Our proposed architecture compares two attention mechanisms for temporal learning: (1) a self-attention module that processes YOLO-extracted features across

frames to model long-range contrast flow dependencies, inspired by transformer-based temporal learning approaches [8, 15, 16, 18, 19]; and (2) a divided space-time attention mechanism adapted from the TimeSFormer architecture [3]. We compare both approaches with a baseline model trained on MinIPs, where all frames of the DSA sequence are collapsed along the time axis into a single image, demonstrating superior performance of the spatio-temporal approach.

2 Methods

YOLOv11 Baseline. To establish baseline results, we used the X variant of the YOLOv11 model with 56.9 million parameters, the latest iteration of the You Only Look Once (YOLO) series of real-time object detection models. The YOLOv11 model is commonly used across multiple vision tasks, including detection, segmentation, and pose estimation [9]. YOLOv11 predicts three outputs, bounding box locations, class probability ($P(\text{class}_i)$) and objectness probability ($P(\text{object})$). From these probabilities indicating whether a bounding box is an object, and which class that object is a member of, a confidence score can be calculated per bounding box, $\text{Confidence}_{\text{class}_i} = P(\text{object}) \times P(\text{class}_i)$.

OccluNet. Two custom architectures (OccluNet1 and OccluNet2) were developed, based on different temporal modules. For both models, the backbone is a pretrained CSPDarknet, the same backbone as used in the S variant of YOLOX with nine million parameters. To facilitate spatio-temporal processing, the current and neighboring frames ($t - 1, t, t + 1$) were passed as input channels. The backbone extracts multi-scale spatial features (C3, C4, C5) from each input frame stack. These frame-wise spatial features are then passed through a Path-Aggregation Feature Pyramid (PAFPN) neck module, as used by YOLOX, followed by a temporal module. There are two temporal approaches:

- OccluNet1 uses Temporal Attention, a Transformer-based attention across frames at each spatial location, using learned positional encoding.
- OccluNet2 uses Divided Space-Time Attention, which alternates between spatial and temporal attention layers, separating spatial (“where”) and temporal (“when”) dynamics for occlusion detection.

After temporal processing for the current frame group, we are left with spatio-temporal features at three levels (P3, P4, P5). These features are then processed by the YOLOX head module, containing the same three output models as the baseline. To enforce temporal consistency, we applied a trajectory-based optimization [15], linking detections across frames if their centers lay within a 15px radius. Each trajectory’s score was the sum of its confidence scores, multiplied by duration (number of frames) to reward temporally persistent predictions. The highest-scoring trajectory per sequence was selected, suppressing inconsistent false positives. See Fig. 1 for an overview of OccluNet.

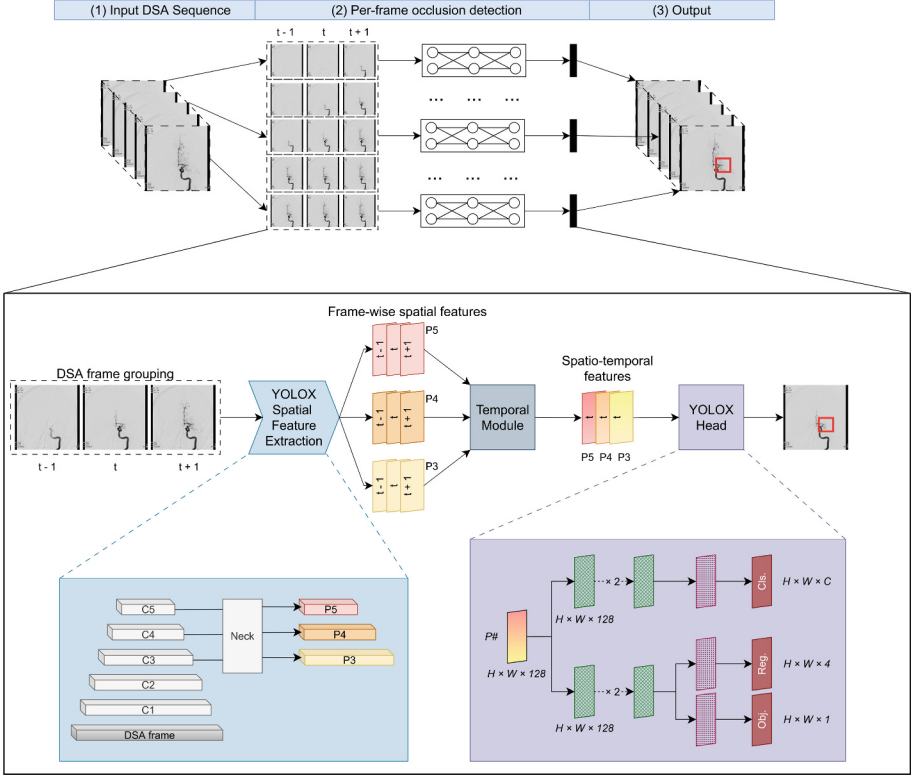


Fig. 1. Overview of OccluNet. Given an input DSA sequence (1), occlusions are first detected on each current frame (timepoint t) by taking neighbouring frames ($t \pm 1$) as context (2), and these per-frame detections form the final output (3). The model consists of a PAFPN feature extraction module, a temporal module and a YOLOX head. For end-to-end per-frame occlusion detection, first, frame-wise spatial features are extracted using the YOLOX spatial feature extraction module. The temporal module applies (temporal or divided space-time) attention to each frame-wise spatial feature level. The output of the temporal module constitutes enhanced spatio-temporal features at each feature level for the current frame, with incorporated context from the neighbouring frames. Finally, the YOLOX decoupled head performs the occlusion detection for each level of spatio-temporal features.

3 Experiments and Results

3.1 Data

The MR CLEAN Registry is a Dutch multi-center registry of acute ischemic stroke patients who underwent EVT between March 2014 and December 2018. The DSA sequences included in this dataset were acquired using various imaging systems. We included one lateral and one anteroposterior pre-EVT DSA per patient, ensuring visibility of the occlusion. These sequences are of size $1024 \times$

1024 pixels, containing between 3 and 76 frames. The temporal resolution varied from 1 to 4 frames per second, while the spatial resolution was approximately $0.4\text{mm} \times 0.4\text{mm}$ per pixel. From the MR CLEAN Registry, $n = 1061$ pre-EVT DSA DICOM sequences were selected from 809 patients, containing 19144 DSA frames in total. Six different occlusion location categories were considered, M1 ($n = 450$), M2 ($n = 238$), Intracranial ICA ($n = 192$), Extracranial ICA ($n = 28$), Unknown ($n = 61$), and No occlusion ($n = 92$).

Data Preprocessing and Annotation. DSA sequences were normalized and resized to target dimensions of 640×640 pixels using bicubic interpolation. Resizing was done due to computational constraints. During training, random left-right flipping was implemented at the sequence level, with a probability of 50%. The dataset was split into 80% training and 20% validation sets, with the same DSA sequences allocated to the train/val sets across all experiments to ensure fair comparison between models. A custom annotation tool was developed using MeVisLab, a platform for medical image processing and visualization (<https://www.mevislab.de/>). Using this tool, we annotated occlusion bounding box ($40\text{px} \times 40\text{px}$, for precise localization of the occlusion site rather than the entire affected vessel) coordinates, occluded vessel type (Extracranial ICA, Intracranial ICA, M1, M2, Unknown, and No occlusion), and the frame interval when the occlusion was visible, if present. All cases were annotated by AK and reviewed by MvdS and FtN. Challenging cases, such as those with ambiguous occlusions or multiple vessel involvement, were discussed with an experienced neurointerventionalist (SC) for verification.

3.2 Evaluation Metrics

Given the small size of the resampled occlusion bounding boxes ($25\text{px} \times 25\text{px}$) relative to the resampled DSA frame dimensions (640×640 pixels), we adopted lenient evaluation criteria. From an interventional radiology perspective, identifying the approximate occlusion location holds greater clinical value than precise bounding box overlap, as treatment decisions depend on recognizing the affected vascular territory rather than exact pixel-level localization, justifying our relatively lenient criteria. Predictions were assessed at the DSA sequence level. Initial per-frame outputs included bounding boxes, confidence scores, and class labels. A detection was considered true positive (TP) if the distance between centers of the highest-scoring trajectory and the ground truth bounding box for the DSA sequence is at most 25px ($\approx 6.25\text{mm}$, given the approximate spatial resolution of $0.4\text{mm} \times 0.4\text{mm}$ per pixel), with a minimum confidence score of 0.01, and a correct class label. False positives (FP) occurred when the distance exceeded 25px or the class was incorrect, while false negatives (FN) represent undetected ground truth occlusions.

Table 1. Performance comparison of baseline (YOLOv11) and OccluNet models across different datasets and occlusion classes. The MinIP_M and MinIP_S experiments represent multi-class and single-class detection on MinIP images of complete DSA sequences, while OccluNet1 and OccluNet2 were evaluated on full DSA sequences. Metrics are reported for both aggregate performance (“all”) and individual occlusion classes. Exp.: Experiment, P: Precision, R: Recall.

Exp.	Dataset	Model Type	Class	Samples	Instances	P (%)	R (%)
1	MinIP _M	YOLOv11	all	214	195	13.4	18.4
			M1	87	87	30.0	39.1
			M2	50	50	7.56	16.0
			Intracranial ICA	38	38	29.6	36.8
			Extracranial ICA	4	4	0.0	0.0
			Unknown	16	16	0.0	0.0
2	MinIP _S	YOLOv11	all	214	195	44.5	22.6
3	DSA	OccluNet1	all	214	195	89.02	74.87
			M1	87	87	84.81	77.01
			M2	50	50	52.38	44.0
			Intracranial ICA	38	38	93.75	78.94
			Extracranial ICA	4	4	100.0	100.0
			Unknown	16	16	45.45	31.25
4	DSA	OccluNet2	all	214	195	87.97	71.28
			M1	87	87	84.61	75.86
			M2	50	50	53.65	44.0
			Intracranial ICA	38	38	90.0	71.05
			Extracranial ICA	4	4	100.0	75.0
			Unknown	16	16	33.33	18.75

3.3 Implementation

The YOLOv11 baseline model was trained for 200 epochs with a batch size of 16. The remaining arguments used the default values recommended for training YOLOv11 [1]. OccluNet was implemented in Python using PyTorch [14] and the MMDetection library [5]. The models were trained on an NVIDIA A40 with 46 GB of memory. OccluNet variants were trained for 20 epochs. The batch size was set to four and one for training and validation, respectively. The input sequence length was set to 3. We employed standard YOLOX training protocols, using stochastic gradient descent (SGD) with Nesterov momentum (0.9) as the optimizer. The base learning rate was set to 0.005, with a two-phase learning schedule- a quadratic warm-up over the first two epochs for stable initialization, followed by cosine annealing that gradually reduces the learning rate to 0.1% of its initial value over the remaining 18 epochs. The final two epochs used a fixed low learning rate for fine-tuning. Weight decay was set to $5e^{-4}$.

3.4 Results

In total, we conducted four experiments. Experiments 1-2 assessed the YOLOv11 baseline model performance on MinIPs. For experiment 1, we trained the baseline model on MinIP images to detect individual occlusion classes. Experiment 2 assesses baseline performance on MinIPs when detecting a single ‘‘Occlusion’’ class, i.e., treating all occlusions as one class. Experiments 3-4 investigate OccluNet configurations on full DSA sequences, with Experiment 3 assessing the performance of the temporal attention variant (OccluNet1) in detecting single-class occlusions on DSA, whereas Experiment 4 assesses the divided space-time attention variant (OccluNet2). The YOLOv11 baseline performance is detailed in Table 1. When processing MinIPs (Exp. 1 and 2), multi-class detection (Exp. 1) achieved an overall precision of 13.4%, with varying performance on the subclasses ranging from 30% for the M1 occlusions to 0.0% for Extracranial and Unknown occlusions. Recall ranged from 39.1% to 0.0%. Single-class occlusion detection (Exp. 2) shows a precision of 44.5% and a recall of 22.6%.

OccluNet demonstrated distinct performance characteristics, as detailed in Table 1. On clinical DSA data (Exp. 3 and 4), both temporal attention variants achieved strong performance, with the temporal attention module (Exp. 3- OccluNet1) showing an overall precision and recall of 89.02% and 74.87%, respectively, with a particularly strong performance on Intracranial ICA (93.75% precision, 78.94% recall) and perfect detection of Extracranial ICA occlusions (100% precision, 100% recall). The divided space-time variant (Exp. 4- OccluNet2) reached similar values of overall precision and recall of 87.97% and 71.28%, respectively. McNemar’s test confirmed these improvements were statistically significant when comparing both OccluNet variants to the YOLOv11 baseline (MinIP_S vs OccluNet1: $p < 0.001$, MinIP_S vs OccluNet2: $p < 0.001$), demonstrating the importance of temporal information. McNemar’s test showed no significant differences between OccluNet1 and OccluNet2 performance ($p = 0.60$). Example of predictions made by both OccluNet configurations on clinical DSA after implementing temporal consistency can be seen in Fig. 2.

4 Discussion

We presented and assessed OccluNet, a novel spatio-temporal deep learning framework that integrates a YOLOX-based object detector with temporal attention mechanisms to exploit the dynamic flow information inherent in DSA sequences. We compared the performance of OccluNet to a YOLOv11 baseline model, showing a significant ($p < 0.001$) performance improvement when incorporating spatio-temporal information. The YOLOv11 baseline model is unable to incorporate temporal features, as it is trained on MinIP images (Exp. 1 and 2). OccluNet, by contrast, is able to assess temporal information by processing an entire DSA sequence during a single inference, either attending across time at each spatial location (OccluNet1) or by sequentially attending to temporal and spatial information (OccluNet2). The fact that both OccluNet variants significantly outperform the baseline models indicates that the integration of temporal

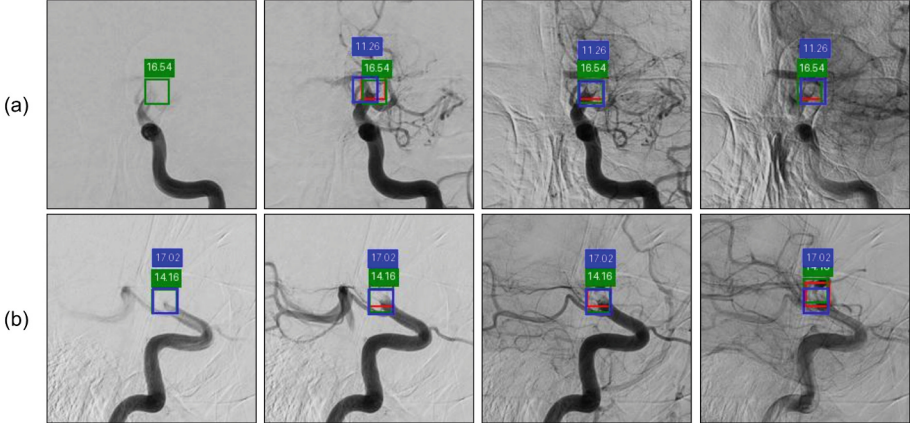


Fig. 2. Visualization of occlusion detection results after trajectory-based optimization to enforce temporal consistency. (a) represents an anteroposterior DSA sequence, and (b) a lateral DSA sequence from the same patient. The red bounding boxes indicate the ground truth annotations, while the green and blue bounding boxes represent predicted occlusions by OccluNet1 and OccluNet2, respectively. The numbers indicate the detection confidence scores. (Color figure online)

information is beneficial. These findings are in line with human behavior, as neurointerventionalists often “scroll through” the images to detect distinctive contrast flow patterns indicative of occlusions. OccluNet was trained for 20 epochs due to computational constraints, whereas the baseline was less demanding, allowing for a training over 200 epochs. Despite this imbalance, OccluNet outperformed the baseline models, highlighting the importance of temporal dynamics in DSA processing. The YOLOv11 baseline models showed increased performance when trained to detect any occlusions (single-class detection) compared to the models trained to detect multiple occlusions (multi-class detection) and differentiate between them. This behavior is expected, as it is easier for models to learn to detect a single class than to further differentiate between them. When comparing OccluNet1 and OccluNet2, we did not see statistically significant performance differences between the models (p -value= 0.6), indicating that both attention types are able to achieve similar performance. From this, we conclude that it is not relevant how exactly temporal information is integrated in an occlusion detection pipeline, as long as it is integrated in some form. Notably, the current OccluNet architecture processes a temporal window of three consecutive frames ($t - 1, t, t + 1$). Extending this window to five or seven frames might better exploit hemodynamic transitions important for detecting smaller or late-appearing occlusions, but such experiments were restricted by GPU memory constraints. Methodologically, OccluNet uses the YOLOX-S variant as its backbone. This is a relatively simple backbone which could be replaced with more complex variants, such as YOLOX-L, to make the model more powerful, potentially improving performance. Due to computational restrictions, we were

unable to evaluate this. Integrating anatomical priors (e.g., vessel segmentation masks) or multimodal inputs (e.g., adding CTA) may reduce false positives by constraining detections to plausible regions. Finally, more diverse datasets with standardized annotations for distal, rare occlusion types and multi-occlusion cases will be important for improving generalization.

OccluNet outperforms previous DSA occlusion detection models based on reported metrics. Future work should assess the applicability of OccluNet in a clinical setting. The detection of distal occlusions is the most clinically relevant, as these occlusions are easier to miss on DSA. As such, clinical evaluation of this model should focus on the added benefit in the detection of distal occlusions (M3, M4) in pre-EVT DSA as well as fragmented thrombi in post-EVT assessment. The relatively low prevalence of these distal occlusion types in DSA datasets presents a challenge.

5 Conclusion

This work presents OccluNet, a spatio-temporal deep learning framework for automated occlusion detection in DSA sequences, with the potential to assist radiologists during EVT procedures for AIS. OccluNet significantly outperforms baseline spatial models, suggesting a role for spatio-temporal processing in automated occlusion detection on DSA. This work lays the foundation for tools that improve clinical workflows during endovascular stroke interventions.

Acknowledgement. We thank the [ICAI Stroke Lab](#) collaborators for their contributions. This publication is part of the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), Philips Medical Systems Nederland B.V., and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Model Training with Ultralytics YOLO - Ultralytics YOLO Docs. <https://docs.ultralytics.com/modes/train/#train-settings>
2. Ischemic Stroke: Background, Anatomy, Pathophysiology (2025). <https://emedicine.medscape.com/article/1916852-overview>
3. Bertasius, G., Wang, H., Torresani, L.: Is space-time attention all you need for video understanding? <https://doi.org/10.48550/arXiv.2102.05095>
4. Chavva, I.R., et al.: Deep learning applications for acute stroke management. *Ann. Neurol.* **92**(4), 574–587 (2022). <https://doi.org/10.1002/ANA.26435>
5. Chen, K., et al.: MMDetection: open mmlab detection toolbox and benchmark. arXiv preprint [arXiv:1906.07155](https://arxiv.org/abs/1906.07155) (2019)

6. Czap, A.L., Sheth, S.A.: Overview of imaging modalities in stroke. *Neurology* **97**(20 Supplement 2), S42–S51 (2021). <https://doi.org/10.1212/WNL.0000000000012794>. <https://www.neurology.org/doi/abs/10.1212/WNL.0000000000012794>
7. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: exceeding YOLO series in (2021). <https://doi.org/10.48550/arXiv.2107.08430>
8. Han, T., et al.: Coronary artery stenosis detection via proposal-shifted spatial-temporal transformer in X-ray angiography **153**, 106546. <https://doi.org/10.1016/j.compbimed.2023.106546>. <https://www.sciencedirect.com/science/article/pii/S0010482523000112>
9. Jocher, G.: Ultralytics YOLO11 - Ultralytics YOLO Docs. <https://docs.ultralytics.com/models/yolo11/>
10. Karlsberg, A., Kim, S., Zaghari, N., Li, L.: Predictive deep learning model for neural vessel occlusion. <https://doi.org/10.1101/2023.09.27.559850>. <https://www.biorxiv.org/content/10.1101/2023.09.27.559850v1>
11. Kelly, B., et al.: DEEP MOVEMENT: deep learning of movie files for management of endovascular thrombectomy **33**(8), 5728–5739. <https://doi.org/10.1007/s00330-023-09478-3>
12. Khankari, J., Yu, Y., Ouyang, J., Hussein, R., Do, H.M., Heit, J.J., Zaharchuk, G.: Automated detection of arterial landmarks and vascular occlusions in patients with acute stroke receiving digital subtraction angiography using deep learning **15**(6), 521–525. <https://doi.org/10.1136/neurintsurg-2021-018638>. <https://jnis.bmj.com/content/15/6/521>
13. Mittmann, B.J., et al.: Deep learning-based classification of DSA image sequences of patients with acute ischemic stroke **17**(9), 1633–1644. <https://doi.org/10.1007/S11548-022-02654-8/FIGURES/5>. <https://link.springer.com/article/10.1007/s11548-022-02654-8>
14. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. <https://doi.org/10.48550/arXiv.1912.01703>
15. Su, R., et al.: Spatio-temporal deep learning for automatic detection of intracranial vessel perforation in digital subtraction angiography during endovascular thrombectomy. *Med. Image Anal.* **77**, 102377 (2022). <https://doi.org/10.1016/j.media.2022.102377>
16. Su, R., et al.: Cave: Cerebral artery-vein segmentation in digital subtraction angiography. *Comput. Med. Imaging Graph.* **115**, 102392 (2024)
17. Warman, R., et al.: A deep learning method to identify and localize large-vessel occlusions from cerebral digital subtraction angiography. **34**(3), 366–375. <https://doi.org/10.1111/jon.13193>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/jon.13193>
18. Xie, Q., et al.: DSCA: a digital subtraction angiography sequence dataset and spatio-temporal model for cerebral artery segmentation. <https://doi.org/10.48550/arXiv.2406.00341>
19. Xie, Q., et al.: DSNet: a spatio-temporal consistency network for cerebrovascular segmentation in digital subtraction angiography sequences. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MIC-CAI 2024*. pp. 199–208. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72111-3_19