International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-023-02909-y

ORIGINAL ARTICLE

Unsupervised out-of-distribution detection for safer robotically guided retinal microsurgery

Alain Jungo[1] · Lars Doorenbos[1] · Tommaso Da Col[2] · Maarten Beelen[2] · Martin Zinkernagel[3] · Pablo Márquez-Neila[1] · Raphael Sznitman[1]

Received: 8 February 2023 / Accepted: 3 April 2023

Abstract

Purpose A fundamental problem in designing safe machine learning systems is identifying when samples presented to a deployed model differ from those observed at training time. Detecting so-called out-of-distribution (OoD) samples is crucial in safety-critical applications such as robotically guided retinal microsurgery, where distances between the instrument and the retina are derived from sequences of 1D images that are acquired by an instrument-integrated optical coherence tomography (iiOCT) probe.

Methods This work investigates the feasibility of using an OoD detector to identify when images from the iiOCT probe are inappropriate for subsequent machine learning-based distance estimation. We show how a simple OoD detector based on the Mahalanobis distance can successfully reject corrupted samples coming from real-world ex vivo porcine eyes.

Results Our results demonstrate that the proposed approach can successfully detect OoD samples and help maintain the performance of the downstream task within reasonable levels. MahaAD outperformed a supervised approach trained on the same kind of corruptions and achieved the best performance in detecting OoD cases from a collection of iiOCT samples with real-world corruptions.

Conclusion The results indicate that detecting corrupted iiOCT data through OoD detection is feasible and does not need prior knowledge of possible corruptions. Consequently, MahaAD could aid in ensuring patient safety during robotically guided microsurgery by preventing deployed prediction models from estimating distances that put the patient at risk.

Keywords Out-of-distribution detection · Instrument-integrated OCT · Medical robotics · Retinal microsurgery

Introduction

Ensuring safe machine learning models is one of the key challenges for real-world medical systems. While the need for reliable models is highly important for image-based diagnostics with human-in-the-loop users, it is mission-critical when combined with medical robotic systems that tightly couple image-based sensing for augmented visualizations or automation.

In this context, one of the fundamental problems in designing safe machine learning is identifying when samples presented to a deployed model differ from those observed at training time. This problem, commonly known as *out-of-distribution* (OoD) detection [1], aims to alleviate the risks of evaluating OoD samples, as performances on these are known to be erratic and typically produce wrong answers with high confidences, whereby making them potentially dangerous. As machine learning has become increasingly prevalent in mission-critical systems, the problem of OoD detection has gathered significant attention both in general computer vision research [1], and in applied medical imaging systems [2–8].

OoD detection for robotically assisted surgery is particularly relevant as erratic machine learning predictions can have extremely serious consequences for the patient. For example, a misprediction in the distance estimation between an instrument and its targeted tissue could lead to important inadvertent trauma. Surprisingly, the topic of OoD detection for robotically assisted surgery has received little attention to date, despite its necessity and advantages. More broadly,

✉ Alain Jungo
alain.jungo@unibe.ch

[1] ARTORG Center, University of Bern, Bern, Switzerland

[2] Preceyes B.V., Eindhoven, The Netherlands

[3] Department of Ophthalmology and Department of Clinical Research, Bern University Hospital, Bern, Switzerland

Published online: 03 May 2023

the potential benefits of OoD detection in this context remain largely unexplored. This work aims to close this gap by analyzing the implications of integrating an OoD detector in a relevant robot-assisted surgery use case.

Specifically, we consider the setting of retinal microsurgery, where a machine learning model is needed to infer the distance between a robotically manipulated instrument and the retina of the eye (see Fig. 1). As with most of the recently proposed robotic systems for retinal microsurgery [9–13], the goal is to assist an operating surgeon when manipulating micron-sized retinal structures using an *optical coherence tomography* (OCT) imaging probe which yields 1D OCT measures over time, also known as *M-scans*. When using such a probe to help guide the robot to an intra-retinal injection site, automatic estimation between the instrument and the retinal surface is key. Yet, for any robot-tissue interacting system, a critical necessity is to ensure that the inferred distances derived from the imaging probe are safe for the robotic system to use.

To this end, this work investigates the feasibility of using OoD detection to identify when images from an *intra-operative instrument-integrated OCT* (iiOCT) probe are inappropriate for subsequent machine learning-based distance estimation (see Fig. 2). We show how data from this probe, in combination with the simple MahaAD OoD [14] detector, can be rejected from further evaluation when the data are corrupted. We demonstrate the implications of our approach on the downstream task of distance estimation using simulated corruptions and report OoD detection performance on ex vivo porcine eyes with real-world corruptions.

Methods

Problem setting

Our retinal microsurgical setup is equipped with a robot that manipulates an injection needle with an iiOCT sensor. The sensor captures the retinal tissue in front of the instrument in a M-scan, which is a sequence of one-dimensional depth signals, denoted *A-scans*. Specifically, M-scans contain useful information about the layers of the retina and the sensor's distance to the different layers (see Fig. 2). However, extracting distance information from M-scans is challenging due to the large appearance variability and noise observed in these signals.

To this end, machine learning and deep learning models are a natural approach to do so consistently and reliably. We thus train a deep learning model $r : \mathbb{R}^P \to [0, 1]^P$ to estimate the location of the internal limiting membrane (ILM) of the retina. Given an M-scan $\mathbf{x}$, the retinal detection model r receives individual A-scans as one-dimensional vectors $\mathbf{x}_j$ and produces one-dimensional heatmaps $\hat{\mathbf{y}}_j = r(\mathbf{x}_j)$ indicating the probability that the ILM is located at each location of the input A-scan. The location of maximum probability determines the ML-based distance as shown in Fig. 1. Similar to [15], the model r is trained by minimizing the standard L2 loss over a training dataset $\mathcal{T} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$ of M-scans and their corresponding ground-truth retinal maps.

At inference time, the retinal detection model r is robust to the types of A-scan variability learned from the training set $\mathcal{T}$, but not to others never seen in this dataset. This poses a risk to the safety of the surgical system in practice, as we cannot ensure that the range of potential perturbations that can occur during surgery are present in the training dataset. The range is simply too large to build a representative dataset that covers all cases.

Unsupervised OoD detection

We augment our system with an unsupervised out-of-distribution detection method to tackle the abovementioned limitation. Our approach is unsupervised in the sense that we do not have examples of OoD cases from which we can train a supervised model to perform OoD. Instead, we have only the dataset from which the distance estimation model, r, is trained. In this context, we leverage the MahaAD method

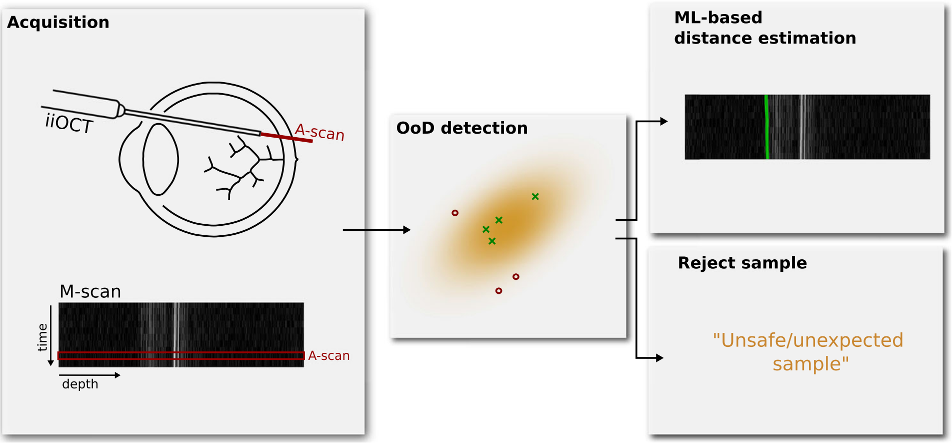

Fig. 1 Out-of-distribution detection of an inappropriate sequence of 1D images, or *M-scan*, acquired by an iiOCT probe. These should be rejected rather than processed by a subsequent machine learning-based distance estimation method

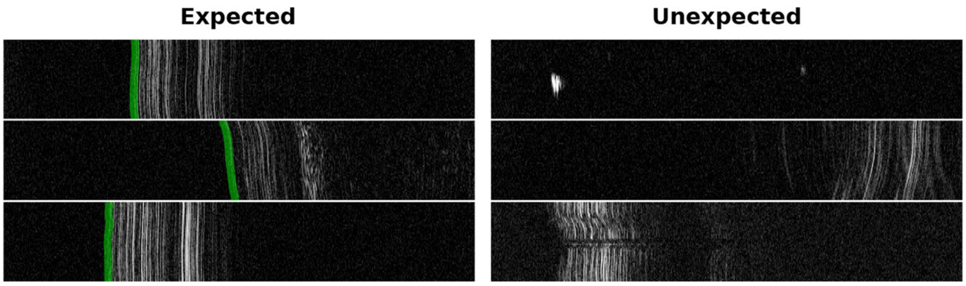

Fig. 2 Six M-scans acquired from a 1D OCT image probe from which distance estimates to the ILM of the retina (shown in green) need to be computed. Evaluating unexpected images (right column) can lead to incorrect estimates and endanger the intervention. Images were resized for improved visualization

proposed by Rippel et al. [14] to learn the appearance of M-scans in the training dataset and detect when novel M-scans are too far from the training distribution to be safely processed by r. We select this model as it has been shown to be highly effective in a large number of cases while being interpretable and computationally lean [16].

At training time, MahaAD learns the training distribution by fitting multiple multivariate Gaussians to latent representations of the training data at different scales. More specifically, we build a training dataset $\mathcal{T}' = \{\mathbf{x}^{(i)}\}_{i=1}^{M}$, where each sample $\mathbf{x}^{(i)} \in \mathbb{R}^{10\times P}$ is a M-scan of 10 consecutive A-scans. M-scans in $\mathcal{T}'$ may come from the training data $\mathcal{T}$ used to train r or, given the unsupervised nature of MahaAD, from any other dataset of M-scans without annotations. Given a pre-trained network f with K layers for feature extraction, MahaAD first describes each training sample $\mathbf{x}^{(i)}$ as a collection of K feature vectors $\{\mathbf{f}_{i,k} = f_k(\mathbf{x}^{(i)})\}_{k=1}^{K}$, where each vector $f_k(\mathbf{x}^{(i)})$ is the spatial average of the k-th feature map for the input $\mathbf{x}^{(i)}$. The collection of the feature vectors for all training samples is then used to fit K multivariate Gaussians, one per layer k, with parameters,

$$\boldsymbol{\mu}_k = \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}_{i,k} \quad \text{and } \boldsymbol{\Sigma}_k = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{f}_{i,k} - \boldsymbol{\mu}_k)(\mathbf{f}_{i,k} - \boldsymbol{\mu}_k)^T$$
$$\forall k \in \{1,\ldots,K\}.$$

At test time, MahaAD computes K Mahalanobis distances between an M-scan $\mathbf{x}$ and the means $\boldsymbol{\mu}_k$ of the learned Gaussians as shown in Fig. 3,

$$d_k(\mathbf{x}) \; = d(\mathbf{x}, \boldsymbol{\mu}_k) = \sqrt{(\mathbf{f}_k - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma_k}^{-1}(\mathbf{f}_k - \boldsymbol{\mu}_k)},$$
$$\forall k \in \{1,\ldots,K\}.$$

The final OoD score for a test-time sample $\mathbf{x}$ is then the sum over all distances,

$$s(\mathbf{x}) = \sum_{k=1}^{K} d_k(\mathbf{x}).$$

The M-scan is then considered OoD if its score $s(\mathbf{x})$ is larger than a threshold τ, which is the only hyperparameter of the method. When an M-scan is considered OoD, we treat all of its individual A-scan components $\mathbf{x}_j$ as OoD and assume they are not suitable for safe estimation with the subsequent retina detection model r. We experimentally found that applying MahaAD on M-scans produced more reliable results than on individual A-scans.

Experimental setting

Data

Our data consist of four recordings from ex vivo trials on four different pig eyes, with each recording containing approximately 900’000 A-scans. The iiOCT device produced temporal A-scans at a frequency of approximately 700Hz with a resolution of 3.7μm/pixel and a scan depth of $P = 674$ pixels (2.49mm).Of the four pig recordings, three were used for training (and validation), and the fourth recording was held out for evaluation. From the training recordings, a collection of 334 in-distribution M-scans consisting of 10 A-scans was used to train the OoD detector. We manually selected samples with identifiable retina to ensure that they are in-distribution samples.

Implementation details

To measure the impact of OoD samples on the retinal model r, we compared six OoD detection strategies and one reference baseline.

MahaAD The method proposed in Sect. 2.2. As proposed in [14], our feature extractor f is an EfficientNet-B0 [17] with $K = 9$ blocks pre-trained on ImageNet. The input to f are M-scans $\mathbf{x}$ resized from 10 × 674 to 64 × 224 with bicubic interpolation to increase the importance of the temporal information and to make the input size closer to the training sizes of the EfficientNet-B0. We applied z-score normalization with the mean and standard deviation from ImageNet to all the input sequences.

Supervised An OoD detector implemented as a binary classifier and trained in a supervised fashion with both in-distribution and OoD samples. Given that OoD samples are not available in large amounts, we synthetically

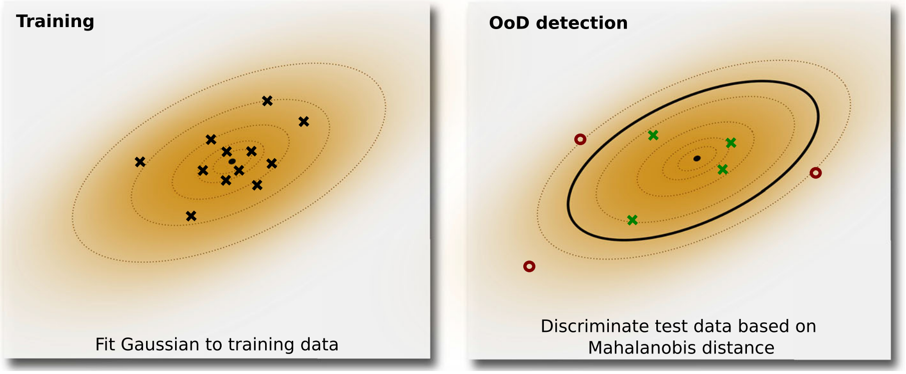
Training
Fit Gaussian to training data
OoD detection
Discriminate test data based on Mahalanobis distance

Fig. 3 Example of MahaAD with a single bivariate Gaussian (*i.e.*, a single 2D latent representation). A multivariate Gaussian is fit to the latent representations of the training samples and is used to determine the Mahalanobis distance of the test samples' latent representations. Based on the distance, samples are considered in- or out-of-distribution

generated them by perturbing 50% of the training data using four types of perturbations: noise, smoothing, shifts and intensity (see Fig. 4). The OoD detector uses an ImageNet-pre-trained EfficientNet-B0 as backbone with a classification head adapted to the binary OoD detection task. We fine-tuned all layers with Adam optimizer and learning rate 10^{-5}.

Glow A generative flow-based model [18] used as OoD detector. We use the model's negative likelihood output as the OoD score (*i.e.*, the lower the likelihood, the less probable a sample is in-distribution). The employed architecture has three blocks of 32 layers and was trained with the public implementation of [19].

Uncertainty OoD samples tend to produce estimations with lower maximum softmax probabilities, (*i.e.*, higher uncertainty [20]). We take the maximum probability of the estimated heatmap $\hat{\mathbf{y}}_j = r(\mathbf{x}_j)$ and use its entropy as the OoD score.

Raw-MahaAD Similar to **MahaAD** but, instead of the feature vectors $\mathbf{f}_{i,k}$, we use the raw signal to fit a single $(K = 1)$ multivariate Gaussian. This can be seen as an ablation of the deep features.

SNR A simple measure of scan quality directly used as OoD score. We measure the signal-to-noise ratio (SNR) as $\mu_{\mathbf{x}}/\sigma_{\mathbf{x}}$.

No-rejection Reference baseline that considers all samples as inliers (*i.e.*, no OoD detection is applied).

In all cases, we used a retinal model r that was implemented as a one-dimensional U-Net-like architecture [21] with four down-pooling/upsampling steps and one convolutional layer per step. We used Adam with a learning rate of 10^{-4} for optimization and performed early stopping according to the performance on the validation split. To train and validate r, the location of the ILM was manually annotated for a random collection of 14'700 M-scans from the original pig recordings.

Experiments

OoD detection for distance estimation

The first experiment measures the impact of our approach in a simulated scenario of retinal surgery where the retinal model r only receives the samples considered safe for estimation by the OoD detector. For this purpose, we employed a test set of 2'000 M-scans with annotated ILM locations. To account for the lack of real OoD samples, OoD samples were synthetically generated by perturbing a fraction p of elements from the test data with eight types of corruptions:

Noise Additive Gaussian noise with μ=0 and σ=50.

Smoothing Gaussian filtering with σ=5.

Contrast Contrast increase/decrease by a factor uniformly sampled from $\{0.1, 0.2, 0.3, 2, 3, 4\}$.

Intensity Equally probable positive/negative shift of the intensity uniformly sampled from the set $\mathcal{U}([-50, -25] \cup [25, 50])$.

Stripes Randomly replacing one or two A-scans in a sequence with a constant intensity sampled from $\mathcal{U}(100, 200)$.

Rectangle Randomly placing a rectangle with size (*i.e.*, M-scan stretch×depth) sampled from $\mathcal{U}([6, 10] \times [15, 30])$ pixels and a constant intensity sampled from $\mathcal{U}(100, 200)$.

Shift Roll in depth for a random split of A-scans in the sequence with positive/negative shift sampled from $\mathcal{U}(25, 100)$ pixels.

Zoom Zoom each A-scan in a sequence by a factor sampled from $\mathcal{U}(1.5, 1.75)$.

All perturbations were applied with equal probability to samples with intensities rescaled to the range [0, 255]. Figure 4 shows examples of produced synthetic corruptions.

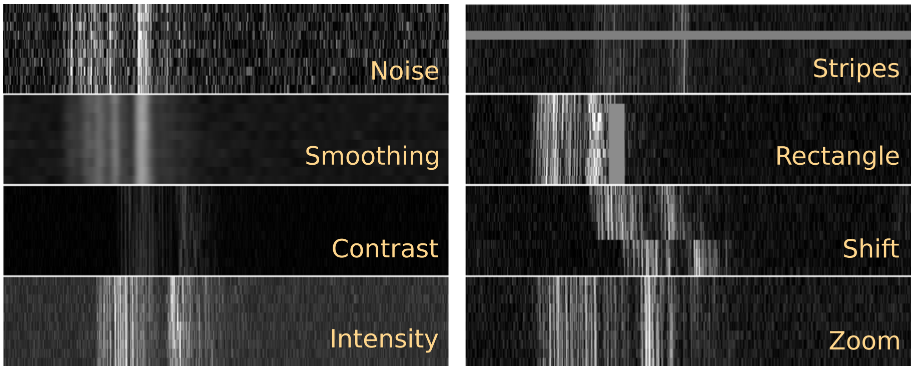

Fig. 4 Examples of the eight types of perturbations applied to simulate OoD samples. Each sample is an M-scan with a depth of 674 pixels and 10 consecutive A-scans. Images were resized for improved visualization

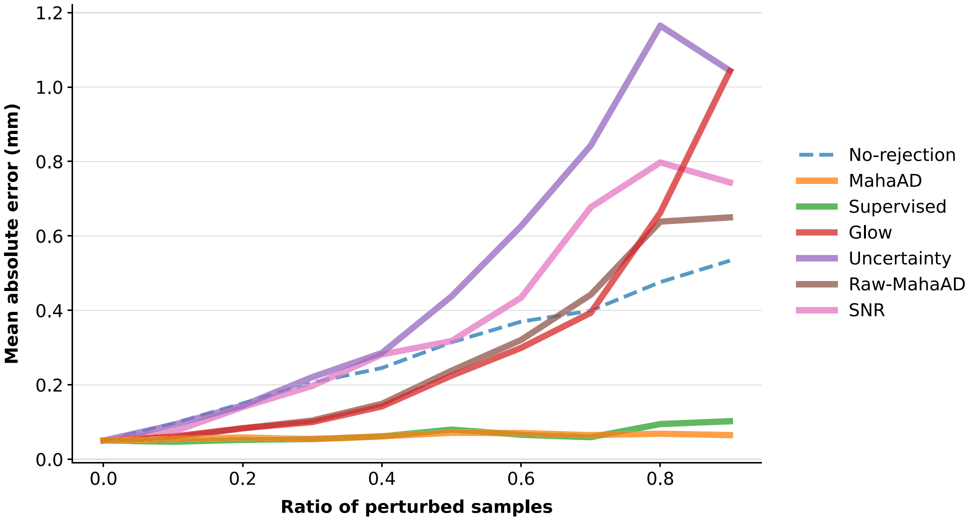

Fig. 5 Effect of different OoD methods on the retinal surgery pipeline. Mean absolute distance error (MAE) is shown for different perturbation ratios p. For each baseline, a proportion of p M-scans were considered OoD and rejected from MAE computation

Real OoD samples in ex vivo data

In a second experiment, we explore the behavior of the methods when presented with real OoD M-scans that were manually identified in our data. For this purpose, we built a test dataset with 258 real OoD M-scans and 258 in-distribution M-scans, where each M-scan consists of 10 A-scans. Figure 8 includes a few examples. As these samples are real OoD cases, it is impossible to label the location of the ILM and thus prevents us from using the above experimental protocol. Instead, we compared the performance of the baselines in the task of detecting OoD samples in this small dataset, omitting the retinal network r.

Results

OoD detection for distance estimation

We measured the performance of r in terms of the mean absolute error (MAE) between the estimated and the real distances for a progressively increasing ratio of corruptions p, which ranged from 0 (*i.e.*, no corruptions) to 0.9. To quantify the impact of each OoD detection approach, M-scans detected as OoD were discarded from MAE computation. For proper comparison, an M-scan was considered OoD if it was among the top-p highest OoD-scoring M-scans. Hence, a perfect OoD detector will discard all the corrupted M-scans, keeping the MAE low.

MahaAD outperformed all baselines, with its MAE staying almost constant for all the perturbation ratios (Fig. 5). **Raw-MahaAD**, **Glow**, **Uncertainty**, and **SNR** underperformed compared to **No-rejection**, suggesting that they flag a large proportion of correct samples as OoD while allowing corrupted A-scans to be processed by the subsequent retinal network. The poor behavior of **Uncertainty** and **SNR** is noticeable for perturbation ratios as low as 0.2, which makes them unsuitable for OoD detection in the present setting. Finally, **Supervised** matched the performance of **MahaAD**, but given that it was trained with the same kind of synthetic perturbations used in the evaluation, this is most likely an overoptimistic result.

Additionally, we compared **MahaAD** and **Supervised** based on their isolated OoD detection performance for individual corruption types at a proportion p of 0.5. To investigate our presumption of **Supervised**'s overoptimistic performance due to known perturbations, we analyzed the

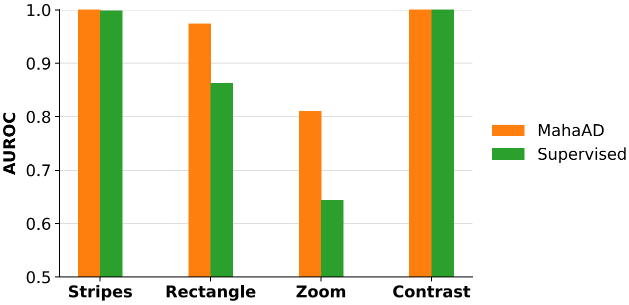

Fig. 6 Comparison between **MahaAD** and **Supervised** on the OoD detection performance in terms of area under the receiver operating characteristic curve (AUROC) for corruptions not used for training **Supervised**

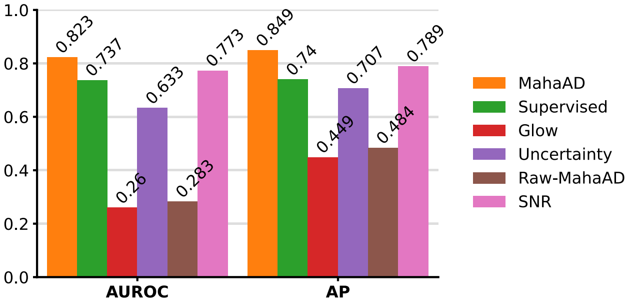

Fig. 7 Area under the receiver operating characteristic curve (AUROC) and average precision (AP) performance for the detection task on real OoD samples

corruptions that **Supervised** has not seen during training (*i.e.*, *stripes*, *rectangle*, *zoom*, *contrast*). Figure 6 shows that **MahaAD** is outperforming **Supervised** in terms of OoD detection on the unseen corruptions. Specifically, the difference is notable for *zoom* and *rectangle*, which seem to be the most difficult perturbations to detect. This result indicates that **MahaAD** is a better OoD detector when the type of expected perturbations is unknown and for which we cannot train.

Real OoD samples in ex vivo data

Figure 7 reports the results for the second experiment on the selection of real OoD samples. As previously found, **MahaAD** outperformed the other baselines, demonstrating its ability to generalize beyond simulated perturbations in a more realistic scenario. Furthermore, **Supervised** performs significantly worse than **MahaAD** on real data, suggesting that the results of Fig. 5 were indeed overoptimistic and that **Supervised** is not suitable as an OoD detector in a realistic scenario. In contrast, **SNR**'s performance improved on real data, likely due to a selection bias facilitating discrimination of OoD samples through low-order statistics. Surprisingly, **Glow** and **Raw-MahaAD** seem to produce OoD scores that better describe in-distribution samples than OoD samples.

Figure 8 shows visual examples of correctly and incorrectly classified in- and out-of-distribution samples for the **MahaAD** approach. The examples confirm that **MahaAD** typically classifies obvious in-distribution or OoD samples correctly but can misclassify borderline samples. For instance, some false negatives may be considered in-distribution based on their retinal-like structure, while false positives often exhibit a hyperreflective retinal pigment epithelium layer, which might lead to their OoD classification.

Discussion and conclusion

In this work, we showed how corrupted data from an iiOCT probe in the context of retinal microsurgery can be rejected from further evaluation by using unsupervised OoD detection. The simple MahaAD approach was able to maintain good performance of distance estimation by reliably detecting and rejecting simulated corruptions, and showed promising results on OoD cases from an ex vivo porcine trial.

The experiments revealed that the benefits of MahaAD observed for a variety of scenarios on 2D images [16] translate well to temporal iiOCT scans with high levels of noise and limited lateral view. Another benefit is its computational efficiency, allowing it to cope with high-frequency A-scan acquisition with minimal latency. Additionally, the experiments point to the challenges of supervised OoD detection when not all unknowns (*i.e.*, possible corruptions) are known and why unsupervised OoD detection might be suitable for improved generalization. In conclusion, we showed that detecting corrupted iiOCT data through unsupervised OoD detection is feasible and that MahaAD could potentially be used to improve safety in retinal microsurgery.

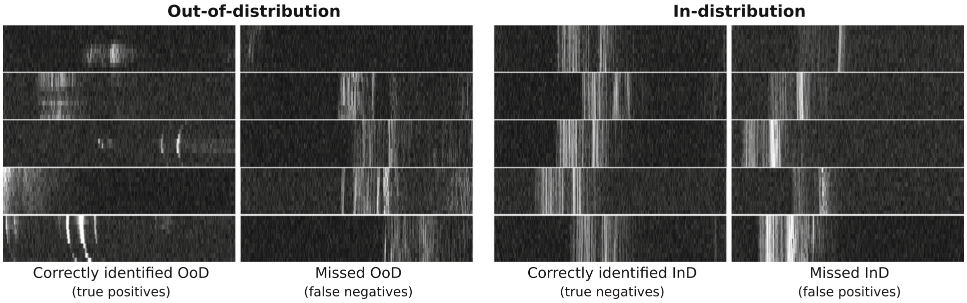

Fig. 8 Examples of correctly detected and missed OoD and in-distribution samples with **MahaAD**. Images have been resized for improved visualization

However, one limitation of this work is that the temporal component of the iiOCT is largely ignored as individual samples were considered for the distance estimation without any knowledge of the past. In the future, we plan to take this temporal information into account by combining the MahaAD OoD detection with dedicated techniques such as Bayesian filters to further improve performances.

Funding Open access funding provided by University of Bern This work was supported by EUREKA Eurostars (project #114442) and H2020 project GEYEDANCE.

Data availability Availability of data, materials, and code: Data are private, code and models are available at https://github.com/alainjungo/ipcai23-iioct-ood.

Declarations

Conflict of interest The authors have no conflict of interest.

Ethical approval Not applicable.

Consent Not applicable.

References

1. Yang J, Zhou K, Li Y, Liu Z (2021) Generalized out-of-distribution detection: a survey. arXiv preprint arXiv:2110.11334
2. Márquez-Neila P, Sznitman R (2019) Image data validation for medical systems. MICCAI 2019:329–337. https://doi.org/10.1007/978-3-030-32251-9_36
3. Zimmerer D, Isensee F, Petersen J, Kohl S, Maier-Hein K (2019) Unsupervised anomaly localization using variational auto-encoders. MICCAI 2019:289–297. https://doi.org/10.1007/978-3-030-32251-9_32
4. Jungo A, Meier R, Ermis E, Herrmann E, Reyes M (2018) Uncertainty-driven sanity check: application to postoperative brain tumor cavity segmentation. MIDL 2018
5. Zimmerer D, Full PM, Isensee F, Jäger P, Adler T, Petersen J, Köhler G, Ross T, Reinke A, Kascenas A, Jensen BS, O'Neil AQ, Tan J, Hou B, Batten J, Qiu H, Kainz B, Shvetsova N, Fedulova I, Dylov DV, Yu B, Zhai J, Hu J, Si R, Zhou S, Wang S, Li X, Chen X, Zhao Y, Marimont SN, Tarroni G, Saase V, Maier-Hein L, Maier-Hein K (2022) Mood 2020: a public benchmark for out-of-distribution detection and localization on medical images. IEEE Trans Med Imag 41(10):2728–2738. https://doi.org/10.1109/TMI.2022.3170077
6. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. MICCAI 2017:146–157. https://doi.org/10.1007/978-3-319-59050-9_12
7. Berger C, Paschali M, Glocker B, Kamnitsas K (2021) Confidence-based out-of-distribution detection: a comparative study and analysis. UNSURE 2021:122–132. https://doi.org/10.1007/978-3-030-87735-4_12
8. González C, Gotkowski K, Fuchs M, Bucher A, Dadras A, Fischbach R, Kaltenborn IJ, Mukhopadhyay A (2022) Distance-based detection of out-of-distribution silent failures for covid-19 lung lesion segmentation. Med Image Anal 82:102596. https://doi.org/10.1016/j.media.2022.102596
9. Balicki M, Han J-H, Iordachita I, Gehlbach P, Handa J, Taylor R, Kang J (2009) Single fiber optical coherence tomography microsurgical instruments for computer and robot-assisted retinal surgery. MICCAI 2009:108–115. https://doi.org/10.1007/978-3-642-04268-3_14
10. Üneri A, Balicki MA, Handa J, Gehlbach P, Taylor RH, Iordachita I (2010) New steady-hand eye robot with micro-force sensing for vitreoretinal surgery. BioRob 2010:814–819. https://doi.org/10.1109/BIOROB.2010.5625991
11. Vander Poorten E, Riviere CN, Abbott JJ, Bergeles C, Nasseri MA, Kang JU, Sznitman R, Faridpooya K, Iordachita I (2020) Robotic retinal surgery. In: Handbook of robotic and image-guided surgery, pp. 627–672 https://doi.org/10.1016/B978-0-12-814245-5.00036-0
12. Cereda MG, Parrulli S, Douven YGM, Faridpooya K, van Romunde S, Hüttmann G, Eixmann T, Schulz-Hildebrandt H, Kronreif G, Beelen M, de Smet MD (2021) Clinical evaluation of an instrument-integrated oct-based distance sensor for robotic vitreoretinal surgery. Ophthalmol Sci 1(4):100085. https://doi.org/10.1016/j.xops.2021.100085
13. Weiss J, Rieke N, Nasseri MA, Maier M, Eslami A, Navab N (2018) Fast 5dof needle tracking in ioct. IJCARS 13(6):787–796. https://doi.org/10.1007/s11548-018-1751-5
14. Rippel O, Mertens P, Merhof D (2021) Modeling the distribution of normal data in pre-trained deep features for anomaly detection. ICPR 2020:6726–6733. https://doi.org/10.1109/ICPR48806.2021.9412109
15. Lee S, Kang JU (2021) CNN-based CP-OCT sensor integrated with a subretinal injector for retinal boundary tracking and injection guidance. J Biomed Optics 26(6):1–14. https://doi.org/10.1117/1.JBO.26.6.068001
16. Doorenbos L, Sznitman R, Márquez-Neila P (2022) Data invariants to understand unsupervised out-of-distribution detection. ECCV 2022:133–150. https://doi.org/10.1007/978-3-031-19821-2_8
17. Tan M, Le Q (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: ICML 2019, vol. 97, pp. 6105–6114
18. Kingma DP, Dhariwal P (2018) Glow: Generative flow with invertible 1x1 convolutions. In: NeurIPS 2018, vol. 31
19. Amersfoort Jv (2022) Glow https://github.com/y0ast/Glow-PyTorch Accessed 08 Nov 2022
20. Hendrycks D, Gimpel K (2017) A baseline for detecting misclassified and out-of-distribution examples in neural networks. ICLR 2017
21. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. MICCAI 2015:234–241. https://doi.org/10.1007/978-3-319-24574-4_28