

Explainable AI with Domain Adapted FastCAM for Endoscopy Images

Jan Stodt¹[0000-0001-9115-7668], Christoph Reich¹[0000-0001-9831-2181], and Nathan Clarke²[0000-0002-3595-3800]

¹ Institute for Data Science Cloud Computing and IT-Security (IDACUS)
Furtwangen University for Applied Science, 78120 Furtwangen im Schwarzwald,
Germany {Jan.Stodt, Christoph.Reich}@hs-furtwangen.de

² Centre for Security, Communications, and Networks Research (CSCAN)
Plymouth University, Portland Square, Plymouth PL4 8AA, UK
N.Clarke@plymouth.ac.uk

Abstract. Enormous potential of artificial intelligence (AI) exists in numerous products and services, especially in healthcare and medical technology. Explainability is a central prerequisite for certification procedures around the world and the fulfilment of transparency obligations. Explainability tools increase the comprehensibility of object recognition in images using Convolutional Neural Networks, but lack precision. This paper adapts FastCAM for the domain of detection of medical instruments in endoscopy images. The results show that the Domain Adapted (DA)-FastCAM provides better results for the focus of the model than standard FastCAM weights.

Keywords: XAI · FastCAM · CNN · Healthcare · Endoscopy

1 Introduction

Explainable Artificial Intelligence (XAI) is essential for artificial intelligence products and services in healthcare and medical technology and is a central prerequisite for certification procedures around the world [7, 2, 5]. How black box models such as neural networks arrive at their results cannot be understood due to their complex processes. However, explainability tools can be used to increase comprehensibility at least for local explanation of individual decisions. In image processing with neural networks, one already speaks of an explanation of a decision when the areas in the input image that have led to the classification of an object are highlighted [2]. In this case, the inner workings of the model are not explained, only the data that is most significant for the decision-making process is highlighted, for example the focus of the model. The motivation of the work to optimize the weights of FastCAM for the endoscopy domain (DA-FastCAM), to achieve better results as the standard FastCAM. The paper works on endoscopy instrument recognition to perform plausibility tests to achieve trustable Convolutional Neural Networks (CNNs) based object detection. The aim is to support the certification of AI-based applications in medicine based on plausibility tests.

2 Related Work

In SHAP (SHapley Additive exPlanations) [6] each input feature is weighted regarding model output. All combinations of features are considered to determine the importance (positive and negative) of a single feature. Integrated Gradients [10] visualizes the importance of the input features of a CNN that contribute to the output of the model via heat maps. GradCAM (Gradient-weighted Class Activation Mapping) [9] allows generating visual explanations that highlight the most important regions of an image that predict the feature.

3 Domain Adapted (DA)-FastCAM

The goal is to optimize the weights of the XAI framework FastCAM [8] for the domain of endoscopy instrument recognition. FastCAM generates a focus area, which is an explanation for the results of object classification. The optimized FastCAM weights more accurately reflect the reality of the focus for the target domain compared to the original weights which are an average of the weights for the ImageNet, CSAIL Places and COWC datasets [8]. Areas that are important for the recognition are the focus of the model (Fig. 1 - Focus area). Focus areas that are not connected to main focus area are distraction area (Fig. 1 - Distraction area). Unimportant (out-of-focus) areas are masked (Fig. 1 - Masked area). Masked areas that occlude a part of the instrument are the occlusion areas (Fig. 1 - Occlusion area). Tests with the original FastCAM weights (Table 2) showed high occlusion, although the model recognized them correctly and a significant occurrence of distraction areas.

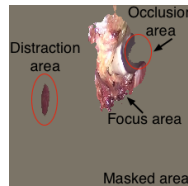


Fig. 1: Areas within an explanation frame

4 Optimization Approach for DA-FastCAM

The dataset used is CholecSeg8k [4] which provides segmentation for 8080 images of laparoscopic cholecystectomy. Each image annotated at the pixel level for thirteen classes common in laparoscopic cholecystectomy. The two tool classes grasper and L hook electrocautery (from here on referred to as hook) are the focus of this paper. The AI model utilized in this paper is the AlexNet based model by Ranem et al.³ which has an average test accuracy of 67% for the Cholec80 [11] dataset. The AI model was used to classify frames of CholecSeg8k dataset.

³ <https://github.com/amrane99/CAI-Classification>

4.1 Optimization Framework and Algorithms

The Optuna framework⁴ was used to optimize the FastCAM mask weights of the five 2-D convolution layers of AlexNet. The metric used to evaluate the optimization is the Root-Mean-Square Difference (RMS). The RMS is calculated for a segmented frame and the explanation frame computed by FastCAM (for the input frame); see Fig. 2b, 2c and 2a. All RMS values are summed, and the average is calculated. The goal of the optimization is to minimize this average RMS value. The algorithms CMA-ES [3] and TPE [1] were used for optimization.

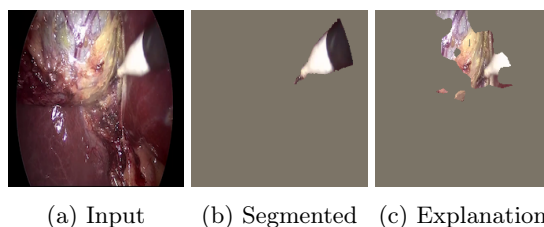


Fig. 2: Example frames of the RMS calculation

5 Evaluation of DA-FastCAM

A series of experiments were conducted for the optimization of the weights: a) selected frames of grasper and hook together, b) selected frames of grasper, c) selected frames of hook, d) grasper of frame 312 and e) hook of frame 28926. Table 1 provides an visual overview of the most important results. Numerical experiment results can be seen in Table 2, the best and worst RMS for the original and optimized weights can be seen in Table 3. 10 frames for the grasper and 10 frames for the hook were selected for the optimization. These 20 frames represent the typical views of the instruments and their position in the frame with different recognition rates by the model.

5.1 Optimization of Weights for Grasper and Hook Together

The weights are optimized for the selected frames of the grasper and hook.

Grasper & Hook CMA-ES "312 - Grasper - Best": the focus includes the entire instrument (A) and does not cut off the upper half of the jaw (B). The distraction area changes its shape (C). "613 - Grasper - Worst": the focus includes the transition between the jaws and the outer tube (A) and a sharper edge to the masked area (B). However, a new occlusion area is created that covers the fenestrated opening in the jaws (C), considered a small disadvantage. "2850 - Hook - Best": the focus includes the transition between the white body and the outer tube (A). The distraction area changes its shape (B). "28911 - Hook - Worst", the focus is reduced so much that only a small section of to the white body of the instrument is visible (A). **Result:** Except for "28911 - Hook - Worst" a clearer focus was created.

⁴ <https://optuna.org/>

Grasper & Hook TPE The results are identical to Section Grasper & Hook CMA-ES; see Section 5.1 for discussion and visual presentation of the results.

5.2 Optimization of Weights for Grasper

The weights are optimized for the selected frames of the grasper. As seen in Table 3, the use of CMA-ES results in better RMS values compared to TPE. Therefore, for the following experiments the TPE algorithm is omitted.

Grasper CMA-ES Despite the optimization for graspers, the results are visually identical to the results of Grasper & Hook CMA-ES, Grasper & Hook TPE and Grasper CMA-ES, therefore see Section 5.1 for discussion and visual presentation of the results.

5.3 Optimization of Weights for Hook

The weights for the selected frames are optimized for the Hook.

Hook CMA-ES "293 - Grasper - Best" includes the transition between the jaws and the outer tube (A). The distraction area (B, C) changed their shape and the distraction area (C) is divided into two parts. "561 - Grasper - Worst": no occlusion area on the grasper mechanism (A) but grasper is more occluded (B). The distraction area changed its shape (D). An additional distraction area has been created (C). "28605 - Hook - Best" the focus includes more area of the outer tube (A). "2850 - Hook - Worst" the hook is completely occluded (A) and the focus area widens (B). **Result:** Except for "2850 - Hook - Worst", a clearer focus was created after optimization. Even if only hooks was optimized, the grasper focus was also optimized.

5.4 Optimization of Weights for Grasper of Frame 312

The weights are optimized for the frame 312. This frame was selected because it was the worst occluded explanation frame with the original weights for all selected graspers. The aim to see if it is possible to improve the masking of one grasper without degrading the results of the other graspers.

Grasper - 312 CMA-ES "312 - Grasper - Optimized for", the focus includes both parts of the jaws (A) and the transition between the jaws and the outer tube (B). The two distraction areas (C, D) change their geometry and the distraction area (D) is divided into two parts. **Result:** Even if optimization was done on a specific grasper, for all graspers the focus after optimization is more precise on the respective instruments (also hook, albeit with limitations).

5.5 Optimization of Weights for Hook of Frame 28926

The weights are optimized for frame 28926. This frame was selected because it was the worst occluded explanation frame with the original weights for all selected hooks. The aim is to see if it is possible to improve the masking of one hook without degrading the results of the other hooks.

Hook - 28926 CMA-ES "28926 - Hook - Optimized For" the focus contains the white body of the instrument (A). But also contains more of the black background (B). The distraction area is smaller but now divided into two parts (C). **Result:** Even if a specific hook is optimized, for all hooks (and grasper) except "2850 - Hook - Worst", the focus after optimization is more precise.

5.6 Optimization Result Overview

Table 2 shows the average RMS, the weights and at which epoch the optimum was achieved. Table 3 shows the best and worst RMS values per frame for the original weights and the optimized weights for grasper and hook. Interesting outcomes are: First, for grasper & hook the algorithm CMA-ES and TPE have the same RMS value and layer weights. Second, experiment grasper CMA-ES have the same RMS value as grasper & hook (CMA-ES, TPE). Third, the CMA-ES finds the optimal weights faster than TPE.

6 Conclusion

Experiments showed that the DA-FastCAM archives a general improvement of the original FastCAM weights via an automated process, validated by reduced RMS values after optimization compared to the RMS values for the original FastCAM weights. Through this optimization, the explanation frames had in average a smaller distraction area and a smaller occlusion area, and a more precise focus area. For a high accuracy of object recognition (e.g. graspers) the RMS values of the DA-FastCAM decreased significantly compared to a lower accuracy of object recognition (e.g. hooks). It is worth mentioning, that through the optimization of FastCAM, a bad CNN model gives bad XAI images. Optimization for all frames of the grasper and hook instruments was shown to be the best approach for optimization. It should be noted that even when optimizing only for a specific frame of grasper, optimization can be performed on average for all frames, whether grasper or hook, albeit with exceptions for some frames of hook. An expected result is that in general hooks are less recognized by the CNN model. For the choice of algorithm, the CMA-ES is recommended, as it not only finds the best weights, but also has the best performance compared to TPE. This is particularly important when numerous images (videos) have to be the basis of the optimization.

Acknowledgment

The authors would like to acknowledge the financial support from the German Federal Ministry of Research and Education (Bundesministerium für Bildung und Forschung) under grant CoHMed/PersonaMed A for this research. Thanks to the DigNest project the results will be disseminated at seminars and workshops.

References

1. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. *Advances in neural information processing systems* **24** (2011)

Table 1: Results of Grasper & Hook CMA-ES, Grasper & Hook TPE, Grasper CMA-ES, Grasper - 312 CMA-ES, Hook CMA-ES and Hook - 28926 CMA-ES


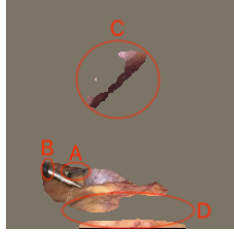

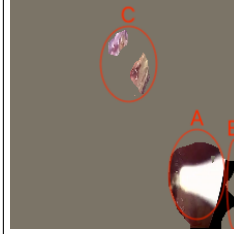
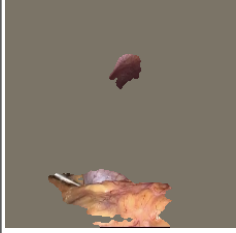


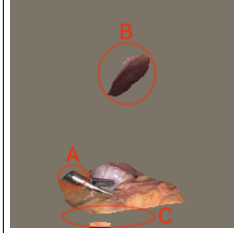



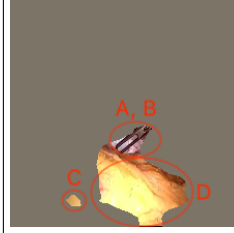

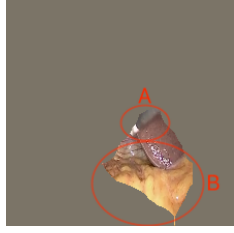
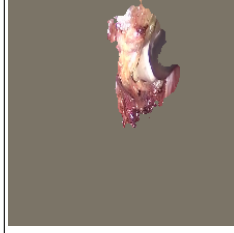
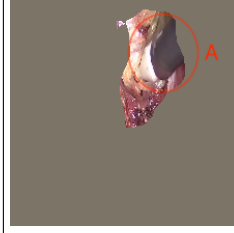
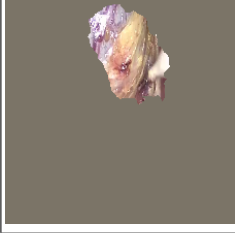

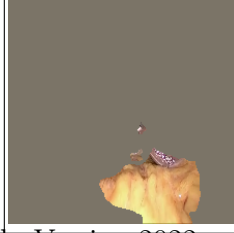
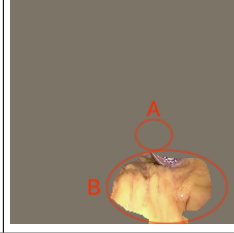
Original Weights	Optimized Weights	Original Weights	Optimized Weights
Frame: 312 - Tool: Grasper Type: Optimized For - RMS: 42.1729		Frame: 28926 - Tool: Hook Type: Optimized For - RMS: 46.0587	
			
Frame: 312 - Tool: Grasper Type: Best - RMS: 41.0978		Frame: 293 - Tool: Grasper Type: Best - RMS: 42.2980	
			
Frame: 613 - Tool: Grasper Type: Worst - RMS: 50.9357		Frame: 561 - Tool: Grasper Type: Worst - RMS: 60.3581	
			
Frame: 2850 - Tool: Hook Type: Best - RMS: 43.0347		Frame: 28605 - Tool: Hook Type: Best - RMS: 43.9202	
			
Frame: 28911 - Tool: Hook Type: Worst - RMS: 60.7620		Frame: 2850 - Tool: Hook Type: Worst - RMS: 60.7620	
			

Table 2: Optimization result overview

Index	Average RMS	Weight Layer A	Weight Layer B	Weight Layer C	Weight Layer D	Weight Layer E	Epoch of best result
Original FastCAM weights	51.18	0.18	0.15	0.37	0.40	0.72	N/A
Grasper & Hook CMA-ES	48.62 ^a	0.00	0.00	0.00	0.00	0.87	471
Grasper & Hook TPE	48.62	0.00	0.00	0.00	0.00	0.87	855
Grasper CMA-ES	48.62 ^b	0.00	0.00	0.00	0.00	0.89	355
Hook CMA-ES	49.28	0.00	0.16	0.00	0.61	0.12	550
Grasper - 312 CMA-ES	49.81	0.01	0.95	0.01	0.06	0.34	396
Hook - 28926 CMA-ES	49.28	0.00	0.22	0.00	0.86	0.17	693

^a Optimum reached earlier than TPE ^b Same optimum RMS as for Grasper & Hook CMA-ES

Table 3: Overview of best and worst RMS for the original and optimized weights

Index	Original Weights RMS					Optimized Weights RMS				
	Average	Grasper		Hook		Average	Grasper		Hook	
		Best	Worst	Best	Worst		Best	Worst	Best	Worst
Grasper & Hook CMA-ES	51.14	44.20	59.42	47.19	59.09	48.62	41.09	53.56	43.03	60.76
Grasper & Hook TPE						48.62	41.09	53.56	43.03	60.76
Grasper CMA-ES	↓	↓	↓	↓	↓	48.62	41.09	53.56	43.03	60.76
Hook CMA-ES						49.28	42.29	60.35	43.92	52.17
Grasper - 312 CMA-ES	↓	↓	↓	↓	↓	49.81	42.15	58.54	41.31	60.76
Hook - 28926 CMA-ES						49.28	42.32	60.34	43.81	52.14

- DIN: DIN SPEC 13288, Guideline for the development of deep learning image recognition systems in medicine. Tech. rep., <https://dx.doi.org/10.31030/3235648>
- Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation* **9**(2), 159–195 (2001)
- Hong, W.Y., Kao, C.L., Kuo, Y.H., Wang, J.R., Chang, W.L., Shih, C.S.: Cholec-Seg8k: A Semantic Segmentation Dataset for Laparoscopic Cholecystectomy Based on Cholec80. arXiv:2012.12453 [cs] (Dec 2020), <http://arxiv.org/abs/2012.12453>
- ISO/IEC AWI TS 6254: Objectives and approaches for explainability of ml models and ai systems. Standard, ISO, Geneva, CH
- Lundberg, S., Lee, S.I.: A unified approach to interpreting model predictions (2017)
- Maack, S., Bertovic, M., Radtke, M.: Deutsche Normungsroadmap KI (2020)
- Mundhenk, T.N., Chen, B.Y., Friedland, G.: Efficient saliency maps for explainable ai (2019)
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision* (2017)
- Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: *International Conference on Machine Learning*. pp. 3319–3328. PMLR (2017)
- Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., Padoy, N.: Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging* **36**(1), 86–97 (2016)