BLI and LCI improve polyp detection rate and delineation accuracy for deep learning networks.

Tom Eelbode, Ingrid Demedts, Philip Roelandt, Cesare Hassan, Emmanuel Coron, Pradeep Bhandari, Helmut Neumann, Oliver Pech, Alessandro Repici, Frederik Maes, Raf Bisschops

Introduction Current state-of-the-art automated polyp detection and delineation techniques use white-light imaging as their base modality. Studies have however suggested that polyp detection rates can be improved by using other modalities such as linked color imaging (LCI) from Fujifilm. This might be true for human observers, but it has not yet been investigated how an artificial intelligence (AI) system is influenced by the choice of modality.

Aims The aim of this research is to investigate the influence of the chosen modality (WLI, BLI or LCI) on the performance of an AI system for polyp detection and delineation.

Methods Complete pull-through colonoscopy videos from 120 patients are included with a total of 280 polyps for training, validation and testing of the system ($n = 176, 27, 77$ respectively with no overlapping patients). Shorter video clips containing the first apparition of each polyp are extracted and for each clip, only a few frames are annotated by three individual experts. These 758 single-frame manual annotations are automatically propagated over the entire clip using dense object tracking. The resulting, much larger annotated dataset of 40887 images is then used to train a recurrent convolutional neural network (CNN) for polyp detection and delineation.

Frame-level sensitivity and specificity are reported for evaluation of the detection power of the network. For delineation accuracy, the Dice score is used which is a measure for the amount of overlap between a delineation map and its ground truth. The analysis is done for WLI, BLI (blue light imaging) and LCI.

Results The sensitivity is highest for BLI with also a significant increase for LCI (sens=0.81, 0.92, 0.85 for WLI, BLI and LCI respectively). Specificity shows the same trend (spec=0.76, 0.85, 0.82 respectively). This clearly indicates a higher performance for polyp detection with LCI and especially BLI when compared to using WLI. Similarly, for delineation, BLI has the highest accuracy (Dice score=0.69, 0.76, 0.63 respectively for WLI, BLI and LCI). For this task, LCI is inferior to WLI. Pairwise t-tests show that all differences are significant with a p value $<0,00001$ (significance level of 0,05).

Conclusion The choice of modality has a significant impact on the detection and delineation performance of an AI system. We show that our network performs best for both tasks on BLI and that LCI has a superior detection, but inferior delineation power compared to WLI.