ORIGINAL RESEARCH

Evaluation of Methods for Detection and Semantic Segmentation of the Anterior Capsulotomy in Cataract Surgery Video

Zixue Zeng[1],*, Binh Duong Giap [ID][2],*, Ethan Kahana[3], Jefferson Lustre [ID][4], Ossama Mahmoud[5], Shahzad I Mian[2], Bradford Tannen[2], Nambi Nallasamy [ID][2,6]

[1]School of Public Health, University of Michigan, Ann Arbor, MI, USA; [2]Kellogg Eye Center, Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI, USA; [3]Department of Computer Science, University of Michigan, Ann Arbor, MI, USA; [4]School of Medicine, University of Michigan, Ann Arbor, MI, USA; [5]School of Medicine, Wayne State University, Detroit, MI, USA; [6]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

*These authors contributed equally to this work

Correspondence: Nambi Nallasamy, Kellogg Eye Center, University of Michigan, 1000 Wall Street, Ann Arbor, MI, 48105, USA, Tel +1 (734) 763-5506, Fax +1 (734) 936-2340, Email nnallasa@med.umich.edu

Background: The capsulorhexis is one of the most important and challenging maneuvers in cataract surgery. Automated analysis of the anterior capsulotomy could aid surgical training through the provision of objective feedback and guidance to trainees.

Purpose: To develop and evaluate a deep learning-based system for the automated identification and semantic segmentation of the anterior capsulotomy in cataract surgery video.

Methods: In this study, we established a BigCat-Capsulotomy dataset comprising 1556 video frames extracted from 190 recorded cataract surgery videos for developing and validating the capsulotomy recognition system. The proposed system involves three primary stages: video preprocessing, capsulotomy video frame classification, and capsulotomy segmentation. To thoroughly evaluate its efficacy, we examined the performance of a total of eight deep learning-based classification models and eleven segmentation models, assessing both accuracy and time consumption. Furthermore, we delved into the factors influencing system performance by deploying it across various surgical phases.

Results: The ResNet-152 model employed in the classification step of the proposed capsulotomy recognition system attained strong performance with an overall Dice coefficient of 92.21%. Similarly, the UNet model with the DenseNet-169 backbone emerged as the most effective segmentation model among those investigated, achieving an overall Dice coefficient of 92.12%. Moreover, the time consumption of the system was low at 103.37 milliseconds per frame, facilitating its application in real-time scenarios. Phase-wise analysis indicated that the Phacoemulsification phase (nuclear disassembly) was the most challenging to segment (Dice coefficient of 86.02%).

Conclusion: The experimental results showed that the proposed system is highly effective in intraoperative capsulotomy recognition during cataract surgery and demonstrates both high accuracy and real-time capabilities. This system holds significant potential for applications in surgical performance analysis, education, and intraoperative guidance systems.

Keywords: cataract surgery, capsulotomy, deep learning, image segmentation, image classification

Introduction

The continuous curvilinear capsulorhexis (CCC) is an essential step in cataract surgery.[1] The skillful performance of the CCC allows for stability[2] of the capsular bag during nucleus disassembly and cortical removal. The continuity and appropriate sizing of the anterior capsulotomy are essential to achieving the appropriate positioning of the intraocular lens implant at the end of cataract surgery.[3] A capsulotomy that is too small can result in anterior capsular phimosis[4–6] and visual field constriction,[7] while a capsulotomy that is too large can lead to instability and tilting[8,9] of the intraocular lens (IOL) implant. The achievement of a smooth, continuous, round, appropriately sized anterior capsulotomy can be challenging for surgeons in training and providing

647

objective longitudinal feedback on CCC quality based on their actual surgical cases (rather than simply wet lab sessions or simulator-based training) may be valuable to the trainee surgeon.[10]

The traditional approach to providing feedback on surgical performance involves direct observation by expert surgeons. Providing feedback during surgery can distract focus from the patient, however, and can be disturbing to an awake patient. It also may not be possible for a trainee to be observed by the same surgical mentor for the entirety of their training, making longitudinal analysis difficult. An alternative to this approach is to provide automated objective feedback using computational analysis of surgical video.[11,12] This could improve objectivity by separating the mentor-mentee relationship from the analysis of performance, while allowing for longitudinal tracking of performance characteristics.[13] Since the quality of the CCC depends at least in part on the morphological characteristics of the anterior capsulotomy, an important step in developing an automated tool for analysis of capsulotomy performance (and cataract surgery performance) is the detection of the capsulotomy and segmentation of the completed anterior capsulotomy in surgical video. Therefore, we sought to create a system for identification and semantic segmentation of the completed anterior capsulotomy to aid in the analysis of its morphological characteristics as one portion of an overall framework for objective analysis of cataract surgery performance.

Leveraging these technical advances in semantic segmentation as well as the carefully curated BigCat database, we developed and analyzed deep learning models for the semantic segmentation of the anterior capsulotomy in surgical video frames. A capsulotomy segmentation model could aid in providing feedback for trainee surgeons and help improve surgeon control over the shape and size of the anterior capsulotomy. Furthermore, the detection and segmentation of a visible capsulotomy serves as a gateway to the morphologic analysis of the capsulotomy, encompassing critical parameters like its size, centration, circularity, and smoothness. A more nuanced understanding of these metrics may empower surgeons to achieve enhanced precision in their technique.

Methods

In this section, we present the proposed capsulotomy recognition system in detail. To develop and validate the system performance, we introduce two capsulotomy datasets dedicated to the separate components of the system. Furthermore, the system consisting of three primary components: video recording and preprocessing, capsulotomy video frame classification, and capsulotomy segmentation, as depicted in Figure 1, is also presented.

Data Collection

The dataset utilized for this study was derived from the BigCat database, which we have described previously.[14,15] In particular, the BigCat database consists of 190 surgical videos recordings from Kellogg Eye Center at the University of Michigan between 2020 and 2021. The study was approved (HUM00160950) by the Michigan Medicine IRB (IRBMED) in May of 2019 and was carried out in accordance with the tenets of the Declaration of Helsinki. Informed consent was not required for this study due to its use of anonymized data and its retrospective nature. Surgical imaging was performed using Zeiss high definition 1-chip imaging sensors integrated into ceiling-mounted Zeiss Lumera 700 operating microscopes (Carl Zeiss Meditec AG; Jena, Germany). Karl Storz AIDA recording devices were utilized to obtain high resolution (1920 × 1080) surgical video recordings (Karl Storz SE & Co. KG; Tuttlingen, Germany). Femtosecond laser cataract surgeries and complex cataract surgeries (those qualifying for Current Procedural Terminology code 66,982) were excluded from the dataset. The BigCat database contains over 4 million frames of surgical video in total.

In order to develop and validate the anterior capsulotomy recognition system in cataract surgery depicted in Figure 2, two image datasets were established based on the BigCat database: 1) the anterior capsulotomy classification dataset and 2) the anterior capsulotomy segmentation dataset. The anterior capsulotomy classification dataset, as shown in Figure 2, was created for training and validating the capsulotomy classification model. It consists of 1556 images belonging to two classes: Capsulotomy (V) and Non-Capsulotomy (NV), covering 13 surgical phases from 190 cataract surgeries. In this context, visibility refers to whether a capsulotomy was visible by human annotators or not. Nine hundred fifty images from the NV class were derived from five early surgical phases: Paracentesis, Medication Injection, Viscoelastic Insertion, Main Wound, and Capsulorrhexis Initiation. These phases occur prior to completion of the capsulotomy. The remaining 606 images in the V class were collected from eight later surgical phases, in which the
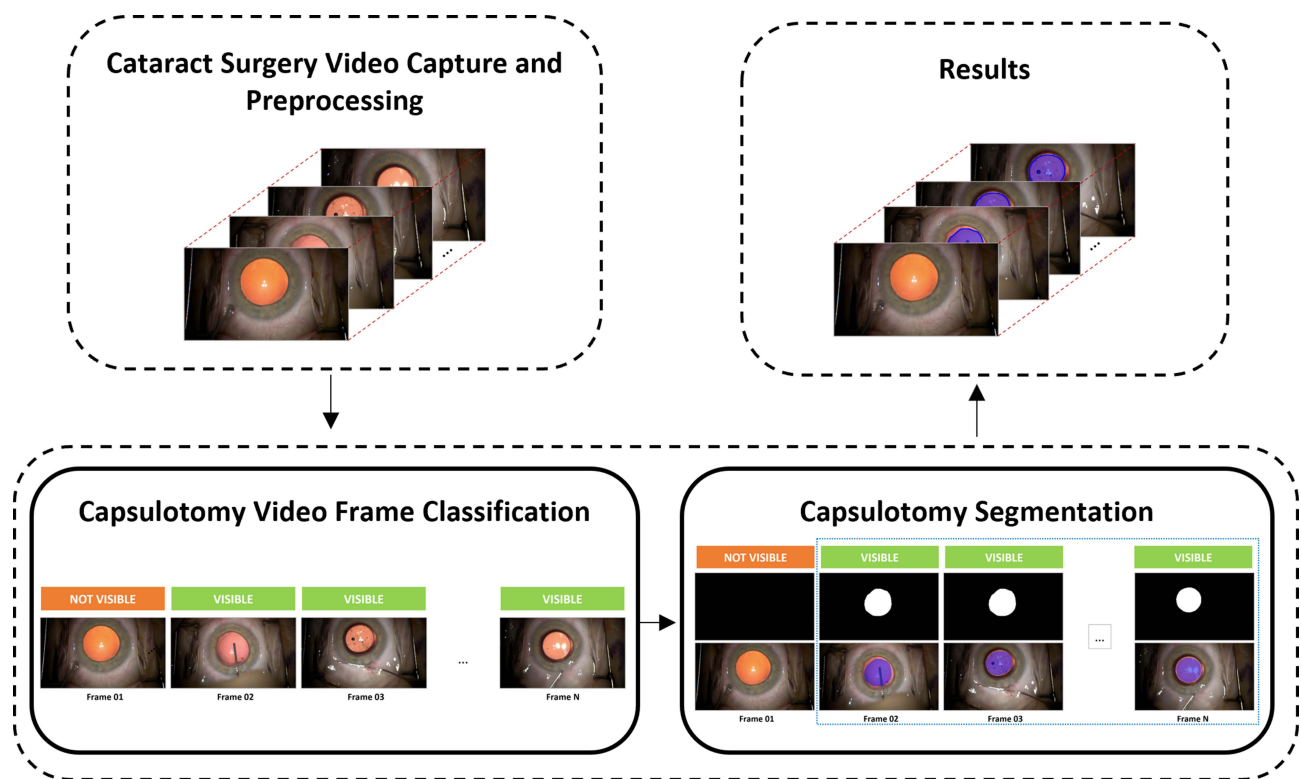
Figure I The propopsed two-stage capsulotomy recognition system (detection and segmentation) in cataract surgery.

anterior capsulotomy is visible for observation: Capsulorrhexis Formation, Hydrodissection, Phacoemulsification, Cortical Removal, Lens Insertion, Viscoelastic Removal, Wound Closure, and No Activity.

The semantic segmentation model was trained and validated using the capsulotomy segmentation dataset, which contains only the Capsulotomy-class images. Consequently, the anterior capsulotomy region in 594 capsulotomy-class images was manually labeled by two trained annotators using the Image Labeler applications in MATLAB (The MathWorks, Natick, MA, USA with a Wacom One drawing tablet (Wacom Co., Ltd.; Kazo, Saitama, Japan), as shown in Figure 3. The videos within each of the datasets were randomly divided into training, validation, and testing subsets consisting of 60% (114 videos), 20% (38 videos), and 20% (38 videos) of the datasets, respectively. In order to reduce model size as well as training and inference time, video frames were resized to 480×270 pixels.

Capsulotomy Detection and Segmentation Models

Automated capsulotomy recognition is a challenging task due to its subtle and intricate nature. The capsulotomy is a delicate, approximately circular opening within the crystalline lens capsule. The fine edge and translucency of the capsule, along with the presence of various anatomical structures and surgical instruments in the video frames make it a complex feature to detect accurately. Furthermore, the recognition of the capsulotomy in cataract surgery poses a challenge due to its appearance being limited to a few surgical phases. As such, using a segmentation model alone is inadequate to accurately identify the presence of a visible capsulotomy.

To address this complexity and ensure a high level of precision in capsulotomy recognition, we propose a two-step approach to detecting and segmenting the capsulotomy region: capsulotomy visibility classification and capsulotomy segmentation. First, we employ a deep learning-based classification model to identify whether a given frame contains a complete and visible capsulotomy. If present, the capsulotomy region is subsequently segmented using a deep learning-based semantic segmentation model. By separating the process in this manner, our approach aims to enhance the accuracy and reliability of capsulotomy recognition in cataract surgery.
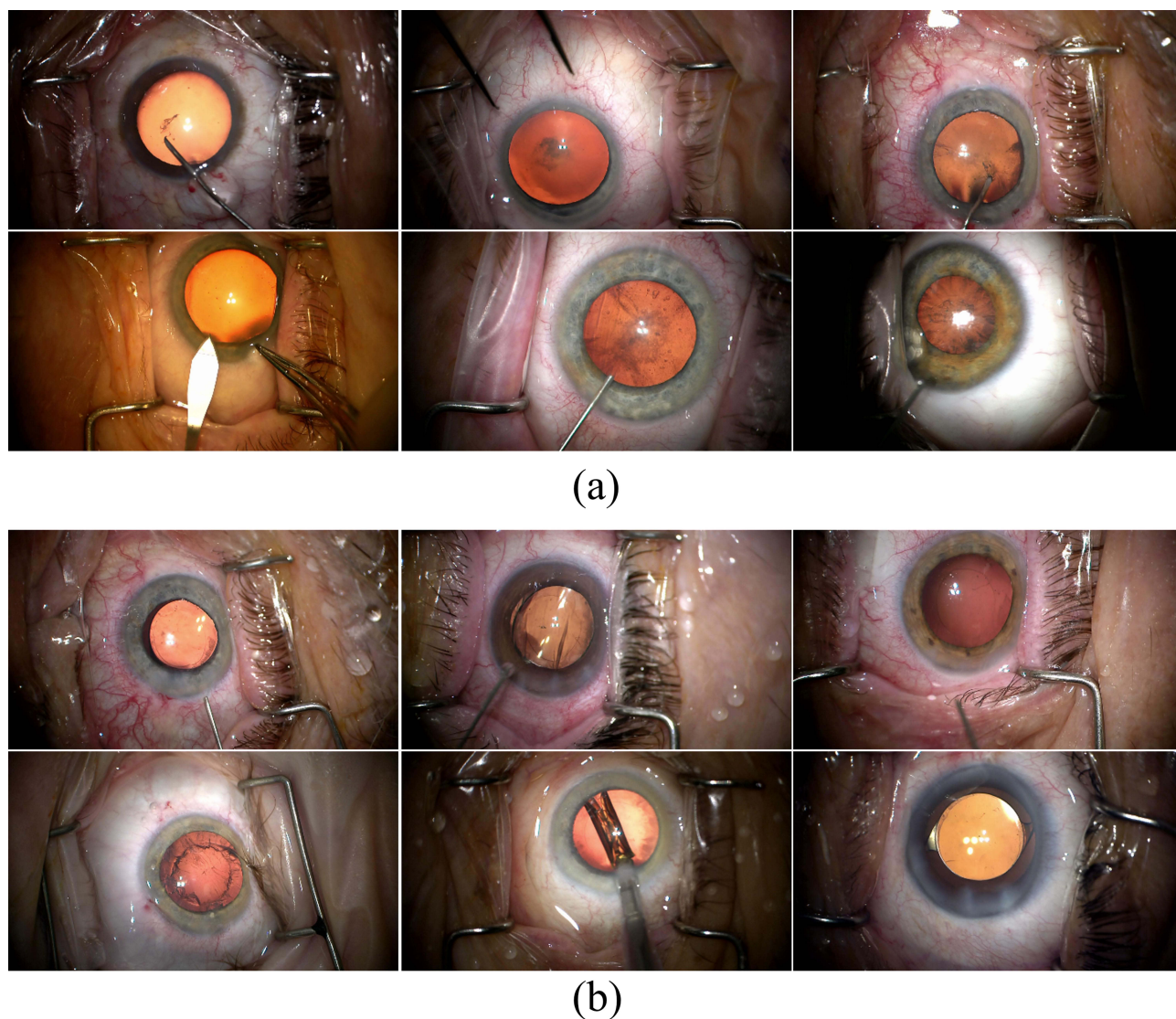
(a)

(b)

Figure 2 Samples in the anterior capsulotomy classification dataset. (**a**) The non-Capsulotomy images and (**b**) the Capsulotomy images.

In our investigation of optimal models for capsulotomy detection and segmentation tasks, we implemented an evaluation process to consider a range of deep learning classification and segmentation architectures with diverse backbone networks to conduct a comprehensive performance comparison. In particular, we conducted experiments employing state-of-The-art deep learning-based classification models from typical families including VGG, ResNet, and DenseNet.[16–18] Subsequently, we investigated the performance of five distinct deep learning architectures, namely FPN, LinkNet, UNet, HRNet, and DeepLabV3+, for the segmentation task.[19–23] The chosen deep learning models have been demonstrated previously to be effective to achieve a high performance in eye anatomy segmentation task.[24] Each investigated segmentation architecture was combined with multiple backbone networks (VGG16, ResNet50, and DenseNet-169). Our objective was to discern the model configuration that exhibited the highest level of performance and ensured the precision and robustness of our capsulotomy recognition system.

In our experiments, we alternatively trained all classification and segmentation models on the training sets of the anterior capsulotomy classification and segmentation datasets, respectively, for a maximum of 100 epochs. The batch size was constantly set at 32 for training models. All images were downsampled 224×224 pixels for the input to the CNN models. An initial learning rate of 10^{-4} was set for the Adam optimizer for all models. In this study, the learning rate was
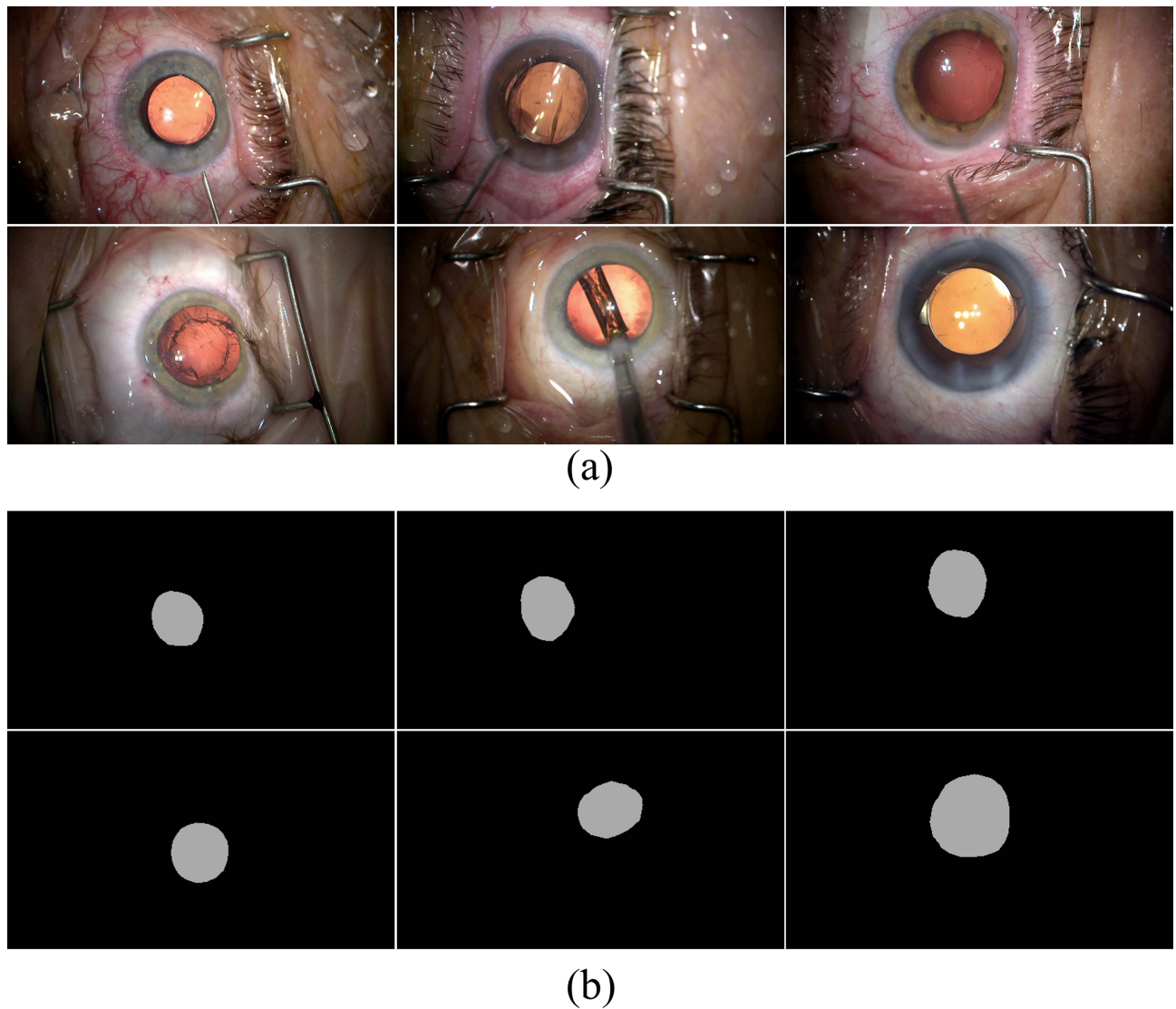
(a)

(b)

Figure 3 Samples in the anterior capsulotomy classification dataset. (**a**) The Capsulotomy images and (**b**) its corresponding capsulotomy masks.

then decreased by 0.1 when the validation loss stopped improving. The trained weights with the lowest validation loss during the training process were stored for validation. In order to reduce training time, all classification models and backbones of the segmentation models were pretrained on the ImageNet dataset.[25] Image shifting, rotating, flipping, zooming, and shearing transformations were applied for data augmentation in order to improve the generalizability of the models.

In this study, the performance of the classification and segmentation models was evaluated with precision, recall, and F1-Score (Dice coefficient) as the primary metrics, which are given below:

$$Precision = \frac{TP}{TP + FP}, \tag{1}$$

$$Recall = \frac{TP}{TP + FN}, \tag{2}$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}. \tag{3}$$

where *TP, FP, TN*, and *FN* denote the true positive, false positive, true negative, and false negative rates, respectively. In addition, the performance of the segmentation models was also evaluated by using the Intersection Over Union (*IoU*), determined as follows:

$$IoU = \frac{TP}{TP + FP + FN}. \tag{4}$$

Results

In this section, the performance of each component of the proposed capsulotomy detection and segmentation system is presented in detail. The experiments in this study were implemented on a workstation with a 24-Core Intel CPU, 128GB RAM, and four NVIDIA RTX 2080 Ti GPUs.

Capsulotomy Video Frame Classification Performance

The performance of eight deep learning-based classification models: VGG-16, VGG-19, DenseNet-121, DenseNet-169, DenseNet-201, ResNet-50, ResNet-101, and ResNet-152, was comprehensively assessed on the validation subset of the anterior capsulotomy classification dataset and detailed in Table 1. As seen, the evaluation of performance showed significant differences in the classification abilities of the examined models. Notably, the VGG architecture models, VGG-16 and VGG-19, demonstrated the lowest classification performance and achieved F1-scores of only 46.34%. In contrast, the DenseNet models exhibited considerably higher classification performance and achieved F1-scores of 91.94%, 88.43%, and 91.64% for the DenseNet-121, DenseNet-169, and DenseNet-201, respectively. While the ResNet-50 and ResNet-101 models demonstrated lower classification performance compared to all DenseNet models, achieving F1-scores of 87.11% and 88.43%, respectively, the ResNet152, the largest model within the ResNet family outperformed all other models. It attained an F1-score of 92.21% with a recall of 92.15% and precision of 92.28%.

We further assessed the selected classification models by conducting comparison experiments to evaluate the computational speed. As shown in Table 2, the inference time (milliseconds per frame) of ResNet-152 was faster than that of all the DenseNet models in our experiment. The inference time of ResNet-152 was slightly higher than that of VGG models, but this was outweighed by the substantial difference in classification performance. The ResNet-152 model was selected as the final architecture for the anterior capsulotomy classification stage in the proposed system due to its combination of top-level classification accuracy and computational speed suitable for real-world applications.

Table 1 The Anterior Capsulotomy Classification Performance of Deep Learning Models on the Validation Set of the Classification Dataset

Network Architecture	Precision (%)	Recall (%)	Dice (%)
VGG-16	37.32	61.09	46.34
VGG-19	37.32	61.09	46.34
DenseNet-121	91.71	91.96	91.94
DenseNet-169	88.69	88.20	88.43
DenseNet-201	91.21	91.21	91.64
ResNet-50	87.23	87.00	87.11
ResNet-101	88.23	88.10	88.14
ResNet-152	**92.28**	**92.15**	**92.21**

Note: Bold fonts indicate the better performance across models.

652

Clinical Ophthalmology 2024:18

Table 2 The Computational Speed Evaluation of the Capsulotomy Classification Models

Network Architecture	Size (M)	Speed (ms)
VGG-16	138	18.54
VGG-19	143	18.54
DenseNet-121	8	24.15
DenseNet-169	14.3	27.80
DenseNet-201	20	30.99
ResNet-50	28.5	18.88
ResNet-101	44.6	21.84
ResNet-152	60.2	22.11

Capsulotomy Segmentation Performance

In this subsection, we investigate the segmentation performance of state-of-The-art deep learning segmentation models with various backbone networks. We conducted an in-depth analysis of the chosen model performance across surgical phases to insight into the factors influencing segmentation task performance.

The performance of all eleven deep learning-based segmentation models studied is presented in Table 3. Notably, the HRNet model achieved the lowest performance, with a Dice coefficient of 83.58%. In contrast, the other models demonstrated a substantial improvement in segmentation performance and achieved Dice coefficients higher than 90%. The UNet model with the DenseNet-169 achieved the highest performance with an F1-score of 92.12%. The results of the segmentation by the UNet model with the DenseNet-169 backbone are illustrated in Figure 4 below.

To evaluate for differences among the validation Dice coefficients of the eleven models studied, we used the Friedman test, and found that there was a significant difference ($p<1e–15$). In addition, we performed post-hoc paired Wilcoxon signed-rank tests with Bonferroni correction to evaluate for statistical differences among the models. The UNet model with DenseNet-169 backbone had statistically superior performance ($p<1e–4$) compared to all other models studied.

An evaluation of computational efficiency of the segmentation models is shown in Table 4. The UNet model with a DenseNet-169 backbone had an inference time considerably lower than several of the investigated models (81.26

Table 3 The Anterior Capsulotomy Segmentation Performance of Deep Learning Models on the Validation Set of the Segmentation Dataset

Network Architecture	Backbone Network	Precision (%)	Recall (%)	IoU (%)	Dice (%)
HRNet	–	81.65	87.86	73.83	83.58
DeepLabV3+	–	91.02	93.28	85.73	91.31
FPN	VGG16	90.08	94.54	85.25	91.87
LinkNet		92.21	89.29	82.84	90.46
UNet		91.31	92.80	84.80	91.61
FPN	ResNet50	91.55	91.24	83.55	90.89
LinkNet		85.03	95.46	81.29	89.33
UNet		89.06	94.68	84.46	91.47
FPN	DenseNet-169	83.78	**96.92**	81.39	89.53
LinkNet		91.33	92.64	84.75	91.62
UNet		**92.42**	93.52	**87.59**	**92.12**

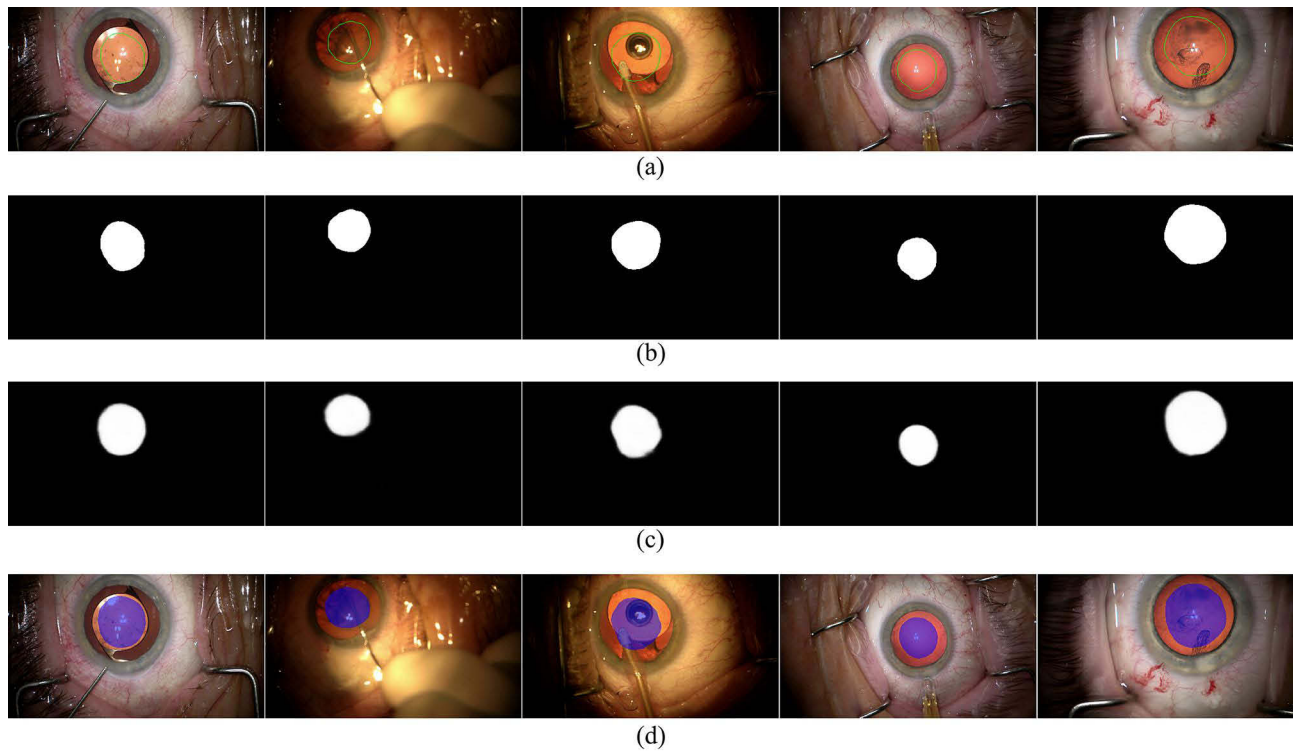**Note**: Bold fonts indicate the better performance across models.

Figure 4 The anterior capsulotomy segmentation results of the UNet (DenseNet-169) model. (**a**) The video frames in various surgical phases of cataract surgery (the edge of capsulotomy is marked by green lines); (**b**) the manually annotated capsulotomy masks of (**a**); (**c**) the masks generated by the UNet (DenseNet-169 model); and (**d**) the overlay images of (**a** and **c**).

milliseconds per frame), including HRNet and FPN. Notably, the UNet model managed to achieve a high segmentation accuracy in addition to its computational efficiency. Thus, the UNet model with DenseNet-169 backbone was chosen as the segmentation model in the proposed system.

The performance of the chosen UNet model with the DenseNet-169 backbone for each surgical phase is depicted in Figure 5. Notably, the highest performance (94.35%) was seen during the No Activity phase, in which the appearance of the complete anterior capsulotomy was unobstructed. The performance of the model in Phacoemulsification phase was

Table 4 The Computational Speed Evaluation of the Capsulotomy Segmentation Models

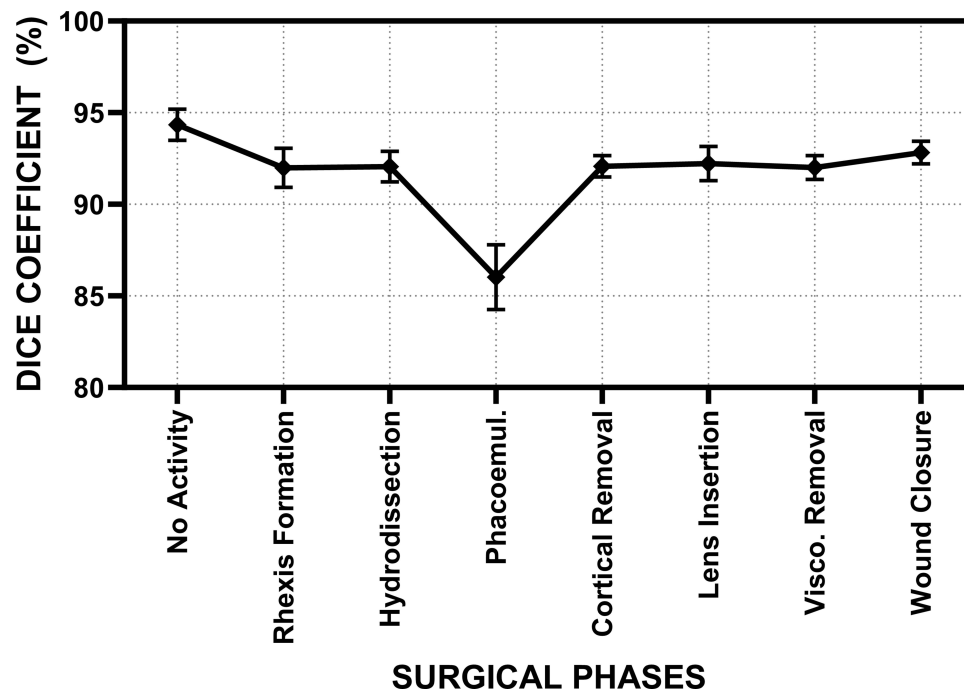| Network Architecture	Backbone Network	Size (M)	Speed (ms)
HRNet	–	28.65	149.29
DeepLabV3+	–	11.85	60.11
FPN	VGG16	17.57	40.81
LinkNet		20.32	41.66
UNet		23.75	40.44
FPN	ResNet50	15.71	106.55
LinkNet		28.78	58.15
UNet		32.56	57.24
FPN	DenseNet-169	15.71	104.95
LinkNet		15.62	103.47
UNet		19.51	81.26

Figure 5 Segmentation performance (Dice coefficient) of the selected UNet (DenseNet-169) model in the proposed system.

significantly lower than its performance in other phases with a Dice coefficient of 86.02%. The Dice coefficient of the model in other phases (capsulorhexis formation, hydrodissection, cortical removal, lens insertion, viscoelastic removal, and wound closure) was consistently higher than 92%.

Discussion

Accurate automated recognition of the anterior capsulotomy in cataract surgery presents several challenges. First, a complete anterior capsulotomy can only be identified in later surgical phases. Thus, performing segmentation tasks directly in the earlier phases, where the anterior capsulotomy does not yet exist, is an unnecessary task. This results in the unnecessary consumption of computing resources and may yield false positive results. Secondly, the anterior capsulotomy presents as an extremely thin contour, which can make it difficult for surgeons to accurately discern depending on the nature of the cataractous lens. Additionally, due to changes in the appearance of the crystalline lens due to cortical disruption and nuclear disassembly, capsulotomy recognition can be a highly challenging task.

In this study, we addressed the aforementioned challenges by proposing a novel deep learning-based system for precise recognition of the anterior capsulotomy during cataract surgery. Our approach involves dividing the recognition process into two tasks: capsulotomy video frame classification and capsulotomy segmentation. This separation enables accurate capsulotomy recognition while avoiding unnecessary segmentation tasks and the generation of false positive results. Specifically, recorded and preprocessed video frames are first classified by a deep learning-based model for the presence of a complete capsulotomy. The experimental results revealed significant variations in the performance of the investigated classification models. Despite the VGG-16 and VGG-19 models having large architectures, they achieved a relatively low performance with a Dice coefficient of 46.34%. This highlights the inherent difficulty in recognizing the capsulotomy in cataract surgery.

Among the models examined, the ResNet-152 model yielded the highest performance in capsulotomy classification with a Dice coefficient of 92.21%, which was significantly higher than the performance of the VGG models. The skip connections within the ResNet architecture play an important role enabling the model to learn residual functions. This characteristic facilitates easier training and the acquisition of more complex features, particularly in high-variation environments as in cataract surgery.

The video frames classified as "Visible" were subsequently employed for capsulotomy segmentation. Similar to the classification stage, we conducted a series of experiments to assess the performance of deep learning-based segmentation models using various backbones. The UNet model with the DenseNet-169 backbone demonstrated the highest segmentation performance and achieved a Dice coefficient of 92.12%.

In our experiments evaluating factors impacting segmentation performances, we observed that the performance of the UNet model with the DenseNet-169 backbone was affected during the Phacoemulsification phase. The Dice coefficient achieved in this phase was only 86.02%, which was significantly lower than its performance in other phases and the overall average (92.12%). The presence of surgical instruments and changes in the appearance of the crystalline lens due to nuclear disassembly emerged as primary factors contributing to the lower segmentation performance of the model during this phase of cataract surgery.

In order to enable real-time applications, we conducted experiments to assess the processing time of each component in the proposed system. The ResNet-152 model employed in the classification task consumed approximately 22.11 milliseconds per video frame. The segmentation model, UNet with DenseNet-169 backbone, required approximately 81.26 milliseconds to segment the capsulotomy in each "Visible" frame. In total, the system processed a video frame with complete capsulotomy in approximately 103.37 milliseconds. These findings indicate that the proposed system is well-suited for real-world applications to allow for high-speed capsulotomy recognition in cataract surgery, while maintaining high accuracy.

Conclusions

The capsulorhexis is a crucial and intricate step in cataract surgery. The automated high-speed analysis of the anterior capsulotomy can enable objective real-world feedback for trainee surgeons, as well as intraoperative guidance regarding capsulotomy position for more advanced surgeons in poor visibility situations. This study introduces an innovative system designed to automatically recognize and segment the capsulotomy during cataract surgery. The system utilizes the ResNet-152 model in the classification stage and the UNet model with DenseNet-169 backbone in the segmentation stage. Experimental results demonstrate impressive classification and segmentation performance, reaching 92.21% and 92.12%, respectively. Furthermore, the computational efficiency of the proposed system proves suitable for real-time applications including intraoperative analysis and guidance.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Masket S. Postoperative complications of capsulorhexis. *J Cataract Refract Surg*. 1993;19(6):721–724. doi:10.1016/S0886-3350(13)80340-9
2. Gimbel HV, Neuhann T. Development, advantages, and methods of the continuous circular capsulorhexis technique. *J Cataract Refract Surg*. 1990;16(1):31–37. doi:10.1016/S0886-3350(13)80870-X
3. Abell RG, Davies PE, Phelan D, Goemann K, McPherson ZE, Vote BJ. Anterior capsulotomy integrity after femtosecond laser-assisted cataract surgery. *Ophthalmology*. 2014;121(1):17–24. doi:10.1016/j.ophtha.2013.08.013
4. Narnaware SH, Bawankule PK. Anterior capsular phimosis. *Indian J Ophthalmol*. 2019;67(9):1476. doi:10.4103/ijo.IJO_392_19
5. Naik M, Sethi H, Mehta A. Capsular bag phimosis. *Am J Ophthalmol Case Rep*. 2020;20:100999. doi:10.1016/j.ajoc.2020.100999
6. Wong WK, Ing MR, Ling CJM. Complete anterior capsule phimosis following cataract surgery in a patient with a history of retinopathy of prematurity, nystagmus, and a narrow angle. *Case Rep Ophthalmol*. 2019;10(2):274–280. doi:10.1159/000502282
7. Mohr AM, Eckardt C. Diathermy capsulotomy to remove fibrotic anterior capsules in pseudophakic eyes. *J Cataract Refract Surg*. 1997;23(2):244–247. doi:10.1016/S0886-3350(97)80348-3
8. Kránitz K, Miháltz K, Sándor GL, Takacs A, Knorz MC, Nagy ZZ. Intraocular lens tilt and decentration measured by Scheimpflug camera following manual or femtosecond laser–created continuous circular capsulotomy. *J Refract Surg*. 2012;28(4):259–263. doi:10.3928/1081597X-20120309-01
9. Karahan E, Tuncer I, Zengin MO. The effect of ND: YAG laser posterior capsulotomy size on refraction, intraocular pressure, and macular thickness. *J Ophthalmol*. 2014;2014:846385. doi:10.1155/2014/846385
10. Sun R. Advantage and disadvantage of posterior continuous curvilinear capsulorhexis. *J Cataract Refract Surg*. 2007;33(12):2004–2005. doi:10.1016/j.jcrs.2007.10.001

11. Pissas T, Ravasio C, Cruz LD, Bergeles C. Effective semantic segmentation in Cataract Surgery: what matters most? Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference; 2021.

12. Mahmoud O, Zhang H, Matton N, Mian SI, Tannen B, Nallasamy N. CatStep: automated cataract surgical phase classification and boundary segmentation leveraging inflated 3D-convolutional neural network architecture and BigCat. *Ophthalmol Sci*. 2024;4(1):100405. doi:10.1016/j.xops.2023.100405

13. Pham DL, Xu C, Prince JL. A survey of current methods in medical image segmentation. *Annu Rev Biomed Eng*. 2000;2(3):315–337. doi:10.1146/annurev.bioeng.2.1.315

14. Matton N, Qalieh A, Zhang Y, et al. Analysis of cataract surgery instrument identification performance of convolutional and recurrent neural network ensembles leveraging BigCat. *Transl Vis Sci Tech*. 2022;11(4):1. doi:10.1167/tvst.11.4.1

15. Giap BD, Srinivasan K, Mahmoud O, Mian SI, Tannen BL, Nallasamy N. Adaptive tensor-based feature extraction for pupil segmentation in cataract Surgery. IEEE Journal of Biomedical and Health Informatics; 2023.

16. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2014.

17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.

18. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.

19. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017:936–944.

20. Chaurasia A, Culurciello E. LinkNet: exploiting encoder representations for efficient semantic segmentation. IEEE Visual Communications and Image Processing; 2017:1–4.

21. Ronneberger O, Fischer P, Brox R. U-Net: convolutional networks for biomedical image segmentation. Proc International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015:234–241.

22. Sun K, Xiao B, Liu D, Wang J. Deep High-resolution representation learning for human pose estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Long Beach, CA, USA; 2019:5686–5696.

23. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. European Conference on Computer Vision; 2018.

24. Giap BD, Srinivasan K, Mahmoud O, Mian SI, Tannen BL, Nallasamy N. Tensor-based feature extraction for pupil recognition in cataract surgery. Proc. the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2023.

25. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. Proc IEEE Conf Comput Vis Pattern Recognit; 2009:248–255.