RESEARCH ARTICLE

Predicting Postoperative Intraocular Lens Dislocation in Cataract Surgery via Deep Learning

NEGIN GHAMSARIAN[1], **DORIS PUTZGRUBER-ADAMITSCH**[2], **STEPHANIE SARNY**[2], **RAPHAEL SZNITMAN**[1], **KLAUS SCHOEFFMANN**[3], (Member, IEEE), **AND YOSUF EL-SHABRAWI**[2,4]

[1]ARTORG Center for Biomedical Engineering Research, University of Bern, 3012 Bern, Switzerland
[2]Department of Ophthalmology, Klinikum Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria
[3]Department of Information Technology, University of Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria
[4]Department of Ophthalmology, Medical University of Graz, 8010 Graz, Austria

Corresponding author: Klaus Schoeffmann (ks@itec.aau.at)

ABSTRACT A critical yet unpredictable complication following cataract surgery is intraocular lens dislocation. Postoperative stability is imperative, as even a tiny decentration of multifocal lenses or inadequate alignment of the torus in toric lenses due to postoperative rotation can lead to a significant drop in visual acuity. Investigating possible intraoperative indicators that can predict post-surgical instabilities of intraocular lenses can help prevent this complication. In this paper, we develop and evaluate the first fully automatic framework for the computation of lens unfolding delay, rotation, and instability during surgery. Adopting a combination of three types of CNNs, namely recurrent, region-based, and pixel-based, the proposed framework is employed to assess the possibility of predicting postoperative lens dislocation during cataract surgery. This is achieved via performing a large-scale study on the statistical differences between the behavior of different brands of intraocular lenses and aligning the results with expert surgeons' hypotheses and observations about the lenses. We exploit a large-scale dataset of cataract surgery videos featuring four intraocular lens brands. Experimental results confirm the reliability of the proposed framework in evaluating the lens' statistics during the surgery. The Pearson correlation and t-test results reveal significant correlations between lens unfolding delay and lens rotation and significant differences between the intra-operative rotations stability of four groups of lenses. These results suggest that the proposed framework can help surgeons select the lenses based on the patient's eye conditions and predict post-surgical lens dislocation.

INDEX TERMS Cataract surgery, semantic segmentation, phase recognition, action recognition, computer-assisted intervention, irregularity detection, intraocular lens complication.

I. INTRODUCTION

Cataract refers to the cloudiness of the eye's natural lens, usually due to aging, resulting in vision blur, dimness, distortion, double vision, and degraded color perception. Cataracts are the major cause of blindness worldwide [1]. Due to the aging population and longer life expectancies, the World Health Organization (WHO) predicts that he incidence of cataract-related blindness will rise to 40 million by 2025 [2]. This common disease can be remedied by replacing

the natural lens with an artificial lens termed intraocular lens (IOL) during cataract surgery [3], [4]. Cataract surgery is the most frequent eye surgery and one of the most frequently performed surgeries worldwide. With the continuous introduction of technological advancements, cataract surgery techniques are constantly evolving. The progression has witnessed significant shifts, starting from intracapsular cataract extraction (ICCE) in the 1960s and 1970s, to extracapsular cataract extraction (ECCE) in the 1980s and 1990s, and currently, the widely adopted technique is sutureless small-incision phacoemulsification surgery with injectable intraocular lens (IOL) implantation.[1] These advancements in surgical methods demonstrated tangible enhancements in visual outcomes and safety [2], [5]. Although not exceeding 10% with mostly transient effects, the intra-operative and post-operative complications in cataract surgery may lead to visual impairment and severe patient discontent [6], [7], [8], [9]. Due to the prevalence of cataract surgery and its considerable impact on the patient's quality of life, predicting and avoiding its post-operative complications is of prime concern for the surgical community.

Intraocular lens dislocation is a major post-operative complication following cataract surgery [10], [11], [12]. During the procedure, the eye's natural lens is removed and a folded artificial lens (IOL) is inserted into the eye's capsular bag. The lens then unfolds and possibly rotates and dislocates until completely unfolded. Despite being aligned and centralized at the end of the surgery, in some cases, the IOL rotates or dislocates following the surgery. Even minor misalignments of the torus in toric IOLs and decentration and tilting of multifocal IOLs can lead to significant vision distortion and dissatisfied patients. Follow-up surgery is currently the only way to address this post-operative complication, entailing additional costs, surgical risks, and patient discomfort. There is an unmet clinical demand to identify intra-operative indicators to predict and avoid this post-operative complication during the surgery.

It is argued that early rotation in toric IOLs during cataract surgery is the leading cause of post-operative misalignments [13]. Since the unfolding delay differs between various IOL brands, it is hypothesized that there is a direct correlation between the lens' behavior during unfolding and its post-operative stability. Besides, an incomplete unfolding of the IOLs at the end of surgery may lead to an inadequate pressure of the haptics against the capsular bag, thus resulting in post-operative rotation, decentration, or tilting of the IOL. In recent years, extensive research has been conducted to compare and predict the rotation stability of different IOLs [14], [15], [16], [17], [18]. However, a reliable study for evaluating the behavior of IOL during its unfolding or other risk factors during the surgery requires large-scale comparisons, necessitating an automatic lens' feature extraction method from surgical microscope video feeds.

In this paper, we aim to investigate the possibility of automating the statistical analysis for different intraocular lens (IOL) behaviors during surgery to predict post-operative lens rotational stability. The main contributions of this paper are as follows.

1) We introduce the first deep-learning-based framework for automatic analysis and comparison of four brands of IOLs based on (i) lens unfolding delay, (ii) lens instability, and (iii) lens rotation during the surgery. To achieve this, three deep-learning-based architectures are employed to tackle different problems in surgical scene understanding: (a) a recurrent convolutional neural network for precise implantation phase detection, (b) a U-Net-based network for lens and pupil segmentation after the implantation phase, and (c) a region-based network for lens' hook detection.

2) The proposed framework is evaluated using a large-scale dataset of cataract surgery videos.

3) Using the proposed framework with trained models, a large-scale study is conducted based on the statistics of four groups of intraocular lenses.

4) The fully-automated statistical comparisons among these four brands of intraocular lenses for the first time suggest significant correlations between lens unfolding delay and rotation and significant differences among the rotation degrees of different lenses.

The efficacy of each stage in the proposed framework is evaluated using relevant metrics, including (I) precision, recall, f1-score, and accuracy for phase recognition, (II) Jaccard metric and dice coefficient for semantic segmentation, and (III) mean average precision for object detection and pose estimation. The phase recognition network achieved 100% accuracy in detecting the implantation phase, which is the starting time to compute lens unfolding delay and rotation. The segmentation network showed outstanding performance in lens and pupil segmentation (a dice coefficient equal to 92.62% for lens segmentation and 97.98% for pupil segmentation). Ultimately, the proposed lens orientation calculation method demonstrates a mean error as small as 3.707 degrees, confirming the detections' high reliability. Our statistical evaluation results align with the surgeons' hypotheses regarding the correlation between lens behaviors during and after surgery. By demonstrating the possibility of predicting and subsequently reducing post-operative complications of intraocular lenses through lens behavior evaluation during surgery, our results provide evidence for potential improvements in patient outcomes.

The rest of this paper is organized as follows. In Section II, we position our work in the literature by reviewing state-of-the-art methods on artificial-intelligence-assisted analysis of cataract surgery videos. Section III details our proposed framework for computing the IOL statistics during the surgery. We explain the experimental setup in Section IV and present the experimental results in Section V. Finally, Section VI discusses the achievements of our work and concludes the paper.
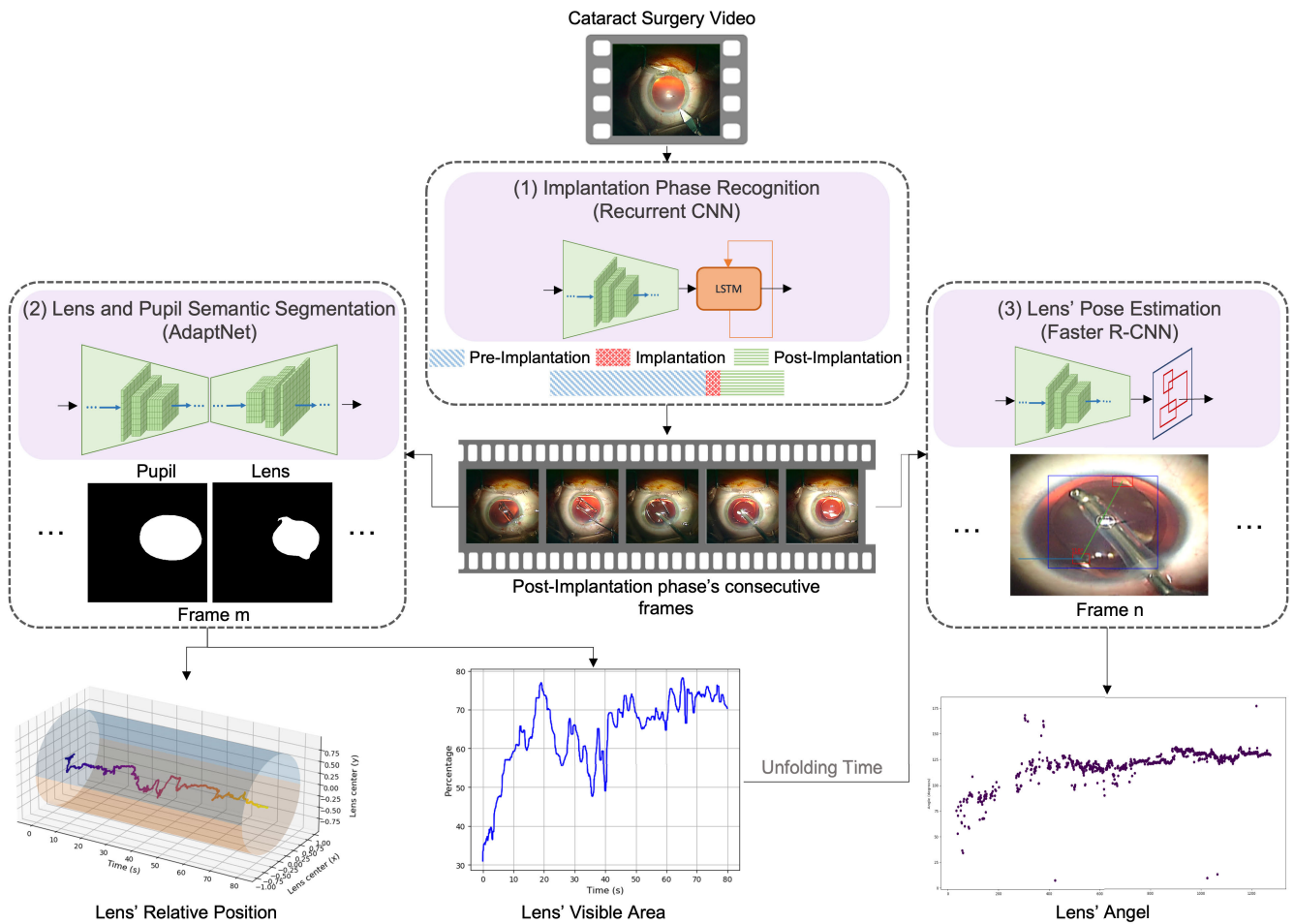
---

[1] In this paper, the term "Cataract Surgery" refers to "Phacoemulsification Cataract Surgery."

FIGURE 1. The overall framework of the proposed intraocular lens qualification method. In the first stage, a recurrent CNN detects the implantation phase to label the last frame of this phase as the starting point for intraocular lens statistics computation. Afterward, a semantic segmentation network outputs the masks of the intraocular lens and pupil to be used for lens unfolding delay and instability computation. Finally, a region-based CNN is employed to compute absolute lens rotations after full unfolding until the end of the video.

II. RELATED WORK

The field of cataract surgery has witnessed the integration of artificial intelligence (AI) to address a spectrum of demands spanning pre-operative, intra-operative, and post-operative applications. Regarding pre-operative requisites, AI has been instrumental in supporting surgical diagnosis and decision-making, including cataract detection and grading, as evidenced by numerous studies [19], [20], [21], [22]. Classic AI-based methods for the intra-operative and post-operative applications focused on instrument tracking [23], surgical process modeling [24], surgical training [25], [26], [27], [28], robot-assisted surgery [29], [30], and surgical time prediction [31]. Furthermore, AI has demonstrated its effectiveness in predicting outcomes related to cataract surgery, notably in the calculation of intraocular lens power [32].

In recent years, convolutional neural networks (CNN) have become the predominant driving engine in computerized surgical workflow analysis. Recent studies have showcased the capabilities of CNN-based frameworks in pre-operative cataract diagnosis [33], [34], [35], and cataract type and severity classification [36]. The intra-operative deep-learning-based methods can be categorized into two primary areas (i) operation room planning and (ii) intra-operative surgical guidance. Operation room planning encompasses tasks like predicting remaining surgery duration [37], [38] and surgical site confirmation [39]. Real-time guidance in particular phases [40] and pupil reaction detection [41] are examples of the latter group of methods. Post-operative cataract surgery analysis methods primarily focus on surgical training and prognosis. Workflow analysis methods are integral in this context and include but are not limited to CNN architectures for phase classification [42], [43], [44], joint phase segmentation-classification [45], instrument tracking [46], and deblurring cataract surgery videos [47]. Furthermore, several CNN-RNN-based frameworks have been proposed to perform relevance-based compression [48], and surgical training expedition [49], [50]. Recent studies
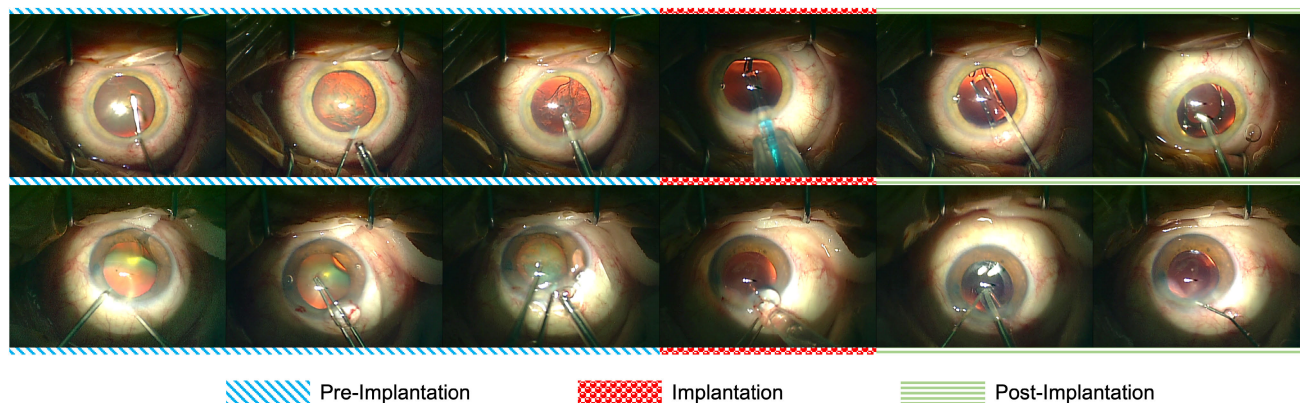
Pre-Implantation   **Implantation**   **Post-Implantation**

FIGURE 2. Sample frames from the pre-implantation, implantation, and post-implantation phases from two representative videos.

have also explored automated technical skill assessment in robotic surgeries [51]. Given the fundamental role of semantic segmentation in various surgical workflow analysis applications, significant efforts have been invested in improving semantic segmentation performance in cataract surgery [52], [53], [54]. In recent years, some efforts have been made to enable surgical prognosis, such as post-surgical visual acuity prediction [55]. These developments collectively underscore the transformative role of CNNs in advancing the field of cataract surgery and its associated analysis techniques.

Regarding research about intraocular lenses, substantial attention has been dedicated to various pre-operative methods, including IOL power calculation [56], [57], [58], IOL parameter verification [59], and IoL segmentation [60]. However, when it comes to the prognosis of cataract surgery, the role of artificial intelligence remains relatively underexplored, with limited work on predicting posterior capsule opacification [61].

This study aims to leverage the CNNs' power to predict lens dislocation as a significant postoperative complication in cataract surgery. Our primary objective is to develop a fully automatic framework capable of computing critical parameters such as lens unfolding delay, rotation, and instability from cataract surgery videos. By achieving this, we aim to facilitate an in-depth analysis that compares intra-operative statistics across various brands of IOLs. Moreover, this research serves as a foundational step toward predicting and preventing post-operative lens dislocation, a significant concern in cataract surgery.

III. METHODOLOGY

Figure 1 demonstrates the pipeline of the proposed method for automatic lens statistic computation during cataract surgery. The pipeline mainly consists of three modules: (1) implantation phase recognition, (2) lens and pupil semantic segmentation, and (3) lens' pose estimation. For the two first modules, we use our proposed neural network architectures [54]. It should be mentioned that all modules within the proposed framework are interdependent and

contribute indispensably to the overall functionality. Each module serves a distinct purpose, and the presence of all modules is imperative for the coherent operation of the framework.. In this section, we detail the functionality of each module in the proposed framework. We then explain the lens statistic computation and correlation analysis in III-D and III-E, respectively. The pseudocode of our proposed framework for IOL evaluation is present in **Algorithm** 1.

A. IMPLANTATION-PHASE RECOGNITION

As the first step toward lens evaluation, we set the starting point for lens statistics computation to be the post-implantation phase, where the folded IOL is inserted inside the eye using a cartridge. Figure 2 illustrates randomly sampled frames from pre-implantation, implantation, and post-implantation phases for two representative videos. We utilize a recurrent CNN with a many-to-many architecture to detect the implantation phase accurately. Recurrent convolutional neural networks can detect the label associated with a sequence of input frames considering the intertwined spatiotemporal features. Moreover, by exploiting features from the neighboring frames, recurrent CNNs can mitigate degraded frame quality typical in cataract surgery videos, such as harsh motion blur and defocus blur.

To enhance the model's robustness, we consider variations in surgeons' speed and intra-operative irregularities, which can alter the order and duration of surgical phases. Consequently, we develop a phase recognition technique that is independent of both order and duration, coupled with a random sampling strategy. This strategy avoids network overfitting to the speed and skill level of the surgeons and improves the network's generalization performance.[2] We address sequence length or temporal context variations during training by employing a systematic approach. Initially, we extracted 12 three-second video clips with overlapping

---

[2]Some other variations, such as lighting conditions, remain relatively consistent within a single hospital. Adapting the model to different domains, such as varying camera settings or hospitals, involves "domain adaptation" techniques [62], [63], [64], which are beyond the scope of this paper.
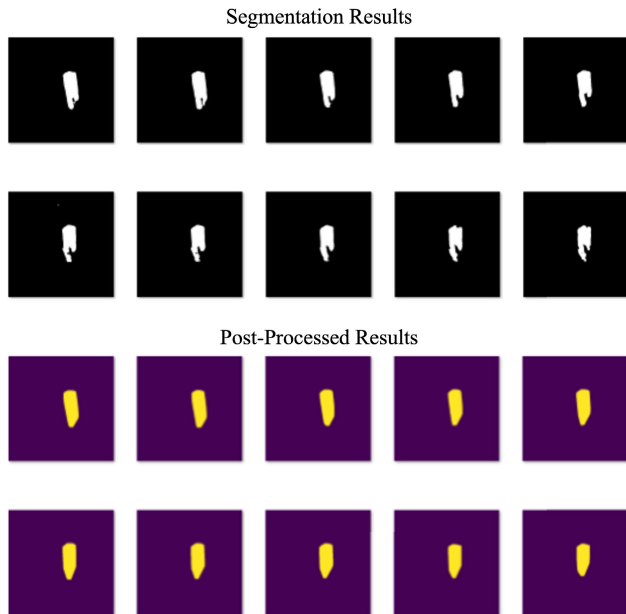
Segmentation Results

Post-Processed Results

FIGURE 3. The IOL segmentation results for ten consecutive frames containing dents in the regions of instrument occlusion, and their corresponding refined versions where occluded segments are recovered.

frames from the implantation phase of each cataract surgery video. To handle variable-length sequences, we further divide video segments before and after the implantation phase into eight and four clips, respectively, resulting in a balanced dataset. During training, each example is generated by selecting a three-second clip with a stochastic variable. More specifically, the network detects the associated phase label to each three-second clip as follows: (1) the three-second sequence with the rate of 25 frames per second is split into five subsequences, each containing 15 consecutive frames; (2) a frame is randomly sampled as the keyframe from each subsequence; (3) the five sampled frames are fed to the network, and for every frame, the network outputs the probability of belonging to the implantation phase and (4) the output probabilities are averaged to obtain the predicted label for the three-second input sequence.

During inference, We utilize overlapping three-second video clips with a two-minute overlap. The associated label for each second in the surgery is determined using a three-second clip centered at that time interval. We perform uniform sampling to provide better diversity in the input frames. Following the detection of the labels for all consecutive three-second clips, we have the time-slot range of the pre-implantation, implantation, and post-implantation phases. We use the post-implantation phase for computing the IOL statistics. Importantly, while the training phase maximizes diversity by randomly selecting frames, the inference phase uniformly samples frames to optimize information feeding into the network.

B. LENS AND PUPIL SEMANTIC SEGMENTATION
To compute the lens unfolding time and instability, we track changes in the size of the IOL over time. Accordingly,

we require a semantic segmentation mask of the pupil and IOL. We employ AdaptNet, which is particularly tailored to the challenges in segmenting the artificial lens and pupil in cataract surgery videos [54].

Upon implantation, the transparent folded artificial lens inherits the pupil color. The IOL unfolds rapidly, introducing unpredictable shapes until reaching complete expansion. Factors like occlusion, defocus blur, and motion blur further complicate segmentation and tracking. Similarly, the diversity in color and texture poses challenges in the segmentation of the pupil. AdaptNet, our chosen solution, employs a U-Net-based encoder-decoder architecture comprising three key components: the encoder, utilizing the VGG16 network; the cascade pooling fusion (CPF) module, which enriches feature representation through pyramid features; and the shape/scale-adaptive feature fusion (SSF) module, introducing adaptability through a strategic combination of convolutional and deformable convolutional layers. This integrated approach ensures a nuanced handling of the challenges posed by the artificial lens and pupil segmentation in cataract surgery videos.

These modules can effectively deal with various difficulties in segmenting the IOL due to its transparency, unpredictable formation during unfolding, occlusion by the instruments, defocus blur, and motion blur.[3]

The lens and pupil segmentation results are then post-processed. Since there are several phases after implantation, the lens and pupil are usually occluded with the instruments, and the segmentation networks cannot detect the occluded regions. However, these regions should be included in the area of the lens and pupil. We adopt the domain-specific knowledge related to these objects to retrieve the occluded regions in their semantic segmentation masks. Specifically, since the IOL and pupil are often convex objects, we draw a convex polygon around each detected object to retrieve the occluded regions. Figure 3 compares the segmentation results and refined masks for some consecutive frames. After post-processing, the visible area of the IOL is computed by counting the pixels belonging to its mask. To compute the lens instabilities, we propose to track the relative location of the lens segment inside the pupil. Hence, the pupil that is unstable due to the unconscious eye movements and surgical operations will be calibrated, and the relative position of the IOL is calculated by computing the distance between the centers of the IOL's mask and the pupil's mask, as shown in Figure 4.

---

[3]It is noteworthy that incorporating deformable convolutions into the neural network architecture may extend inference times, requiring more robust hardware infrastructure to achieve real-time performance. However, as detailed in Section IV-D, our algorithm offers flexibility for substituting neural networks, dealing with tasks like phase recognition, semantic segmentation, and object localization, with lightweight architectures in scenarios with lower hardware configurations. This adjustment comes with the trade-off of slightly reduced accuracy expectations.
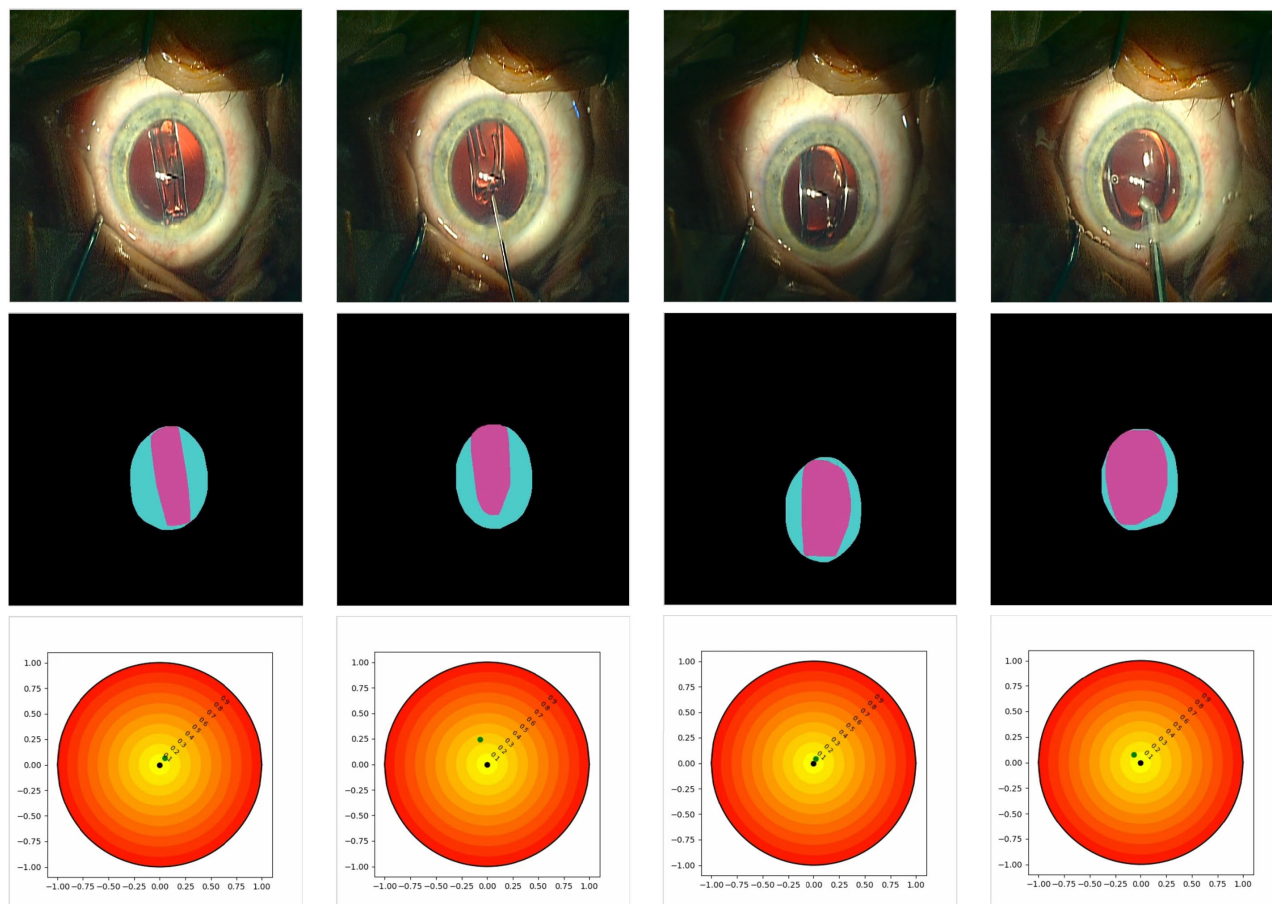
FIGURE 4. Visualization of lens unfolding, rotation, and instability for a representative video. The frames from left to right show the IOL unfolding procedure. Top: sample frames from the post-implantation phase; Middle: the corresponding refined IOL and pupil masks; Bottom: the lens center relative to the pupil center corresponding to each frame.

C. LENS POSE ESTIMATION

To compute the lens rotation amount, we propose estimating the lens orientation per frame based on the hooks' location. We use the Faster R-CNN [65] framework for lens and hook localization. The Faster R-CNN network is a region-based CNN consisting of a backbone network and a region proposal network (RPN) followed by two branches: (1) a localization branch trying to output the most-fitted bounding box to each object and (2) a classification branch that detects the label associated with each detected object.[4] We adopt several strategies to utilize the detection results in the inference stage. First, we only consider the detected bounding boxes with more than 60% detection confidence. Considering that the IOL has only two hooks that are not always visible, we calculate the hooks' location using the detection results based on three scenarios:

1) If the network detects up to one hook fulfilling the determined threshold, the detected bounding box does not undergo any post-processing step.
2) In case two hooks are detected by the network, we consider the position of the detected bounding boxes relative to the center of the IOL. If the angle between the two detected hooks is around 180 degrees, both detections are kept. Otherwise, only the detection with the higher confidence is considered a hook.
3) In the condition that more than two bounding boxes fulfill the confidence threshold for the hook label, we perform hierarchical clustering with two clusters using the centers of the detected bounding boxes. Afterward, the bounding box with the highest detection confidence is selected as the best bounding box among all detections within each cluster. Having two bounding boxes, we further check their relative positions as described in scenario 2.

To compute lens orientation, we calculate the angle of the line connecting the centers of the two hooks' bounding boxes relative to the x-axis (in case of having two final hooks) or the angle of the line connecting the lens center and the visible

---

[4]Region-based CNNs provide an efficient solution for localizing specific objects within the surgical scene. In contrast, segmentation networks face challenges in accurately segmenting intricate structures like hooks and require additional post-processing techniques for location computation. The region-based approach is favored for its simplicity and effectiveness in addressing the specific characteristics of the surgical setting.

Algorithm 1 IOL Evaluation Pseudocode

Input : The trained *phase recognition* network
(CNN-RNN);
The trained *Semantic Segmentation* network
(AdaptNet);
The trained *Pose Estimation* network
(Faster R-CNN [65]).

Output: Lens correlation results.

for *lens brand in {Technis, AvanSee, NC1, XC1}* **do**
  for *video in lens group* **do**
    Feed the video to the phase recognition network to
    detect the implantation phase;
    Feed the frames of the post-implantation phase to the
    segmentation network to achieve the lens and pupil
    segmentation results;
    Refine the segmentation results by drawing a convex
    polygon around the detected masks;
    Compute lens-unfolding time based on the lens area
    (Eq. (4));
    Compute lens instability based on the relative position
    of the lens inside the pupil (Eq. (5));
    Feed the frames starting from the computed unfolding
    time until the end of surgery to the pose estimation
    network and compute lens orientation;
    Compute lens rotation based on lens orientation
    results (Eq. (6));
  end
  Compute the correlation between lens unfolding delay
  and lens rotation (Eq. (7));
end
for *Pair of lens brands* **do**
  Perform three T-tests to evaluate the statistical differences
  between unfolding delay, instability, and rotation of the
  two groups (Eq. (8)).
end

hook's bounding box relative to the x-axis (in case of having only one final hook). In the last case, we also consider the location of the detected hook relative to the lens center in calculating the lens orientation.

D. LENS STATISTIC COMPUTATION

The statistics of the IOL are calculated based on the lens' pose, visible area, and relative position. Supposing that we have the refined masks of the pupil ($\mathcal{M}^p = \{\mathcal{M}^p_1, \ldots, \mathcal{M}^p_n\}$) and IOL ($\mathcal{M}^l = \{\mathcal{M}^l_1, \ldots, \mathcal{M}^l_n\}$) starting from the post-implantation phase until the end of surgery with the rate of 25 *fps* ($n = 25 \times (t_{surgery} - t_{post-implantation})$), the masks' centers ($\mathcal{C}$) and areas ($\mathcal{A}$) can be denoted as:

$$\text{Pupil} \begin{cases} \mathcal{C}^p = [\mathcal{C}^p_1, \mathcal{C}^p_2, \ldots, \mathcal{C}^p_n] \\ \mathcal{A}^p = [\mathcal{A}^p_1, \mathcal{A}^p_2, \ldots, \mathcal{A}^p_n] \end{cases} \tag{1}$$

$$\text{IOL} \begin{cases} \mathcal{C}^l = [\mathcal{C}^l_1, \mathcal{C}^l_2, \ldots, \mathcal{C}^l_n] \\ \mathcal{A}^l = [\mathcal{A}^l_1, \mathcal{A}^l_2, \ldots, \mathcal{A}^l_n] \end{cases} \tag{2}$$

where $\mathcal{A}_1$ and $\mathcal{C}_1 = [\mathcal{C}_{1x}, \mathcal{C}_{1y}]$ correspond to the area and center of the masks in the first frame of the post-implantation phase, respectively. To mitigate the effect of the lens mask's prediction error on lens unfolding time prediction, we pass the lens area vector ($\mathcal{A}^l$) through a mean filter with a window size of 15 frames:

$$\tilde{\mathcal{A}}^l = \begin{cases} \frac{1}{15} \sum_{i-7}^{i+7} \mathcal{A}^l_i & 8 < i < n-7 \\ \mathcal{A}^l_i & else \end{cases} \tag{3}$$

For lens unfolding time (t_U), we compute the difference between the starting time of the post-implantation phase and when the frame-averaged visible lens' area is maximum for the first time:

$$t_U = \underset{t}{\operatorname{argmax}}(\tilde{\mathcal{A}}^l) \tag{4}$$

Lens instability (*Ins*) is computed based on the sum of the lens' absolute relative movements inside the pupil as follows:

$$\text{Ins} = \sum_{i=1}^{n-1} ||\mathcal{C}^l_{i+1} - \mathcal{C}^p_{i+1}| - |\mathcal{C}^l_i - \mathcal{C}^p_i||. \tag{5}$$

Besides, considering the vector of lens orientations as $\mathcal{O}^l = [\mathcal{O}^l_1, \ldots, \mathcal{O}^l_n]$, we compute lens rotation based on the sum of absolute relative lens orientation changes, starting from the time when the lens is unfolded based on lens unfolding results:

$$\mathcal{R}^l = \sum_{i=t_U}^{n-1} |\mathcal{O}^l_{i+1} - \mathcal{O}^l_i| \tag{6}$$

E. CORRELATION ANALYSIS

We adopt the Pearson correlation coefficient to evaluate the correlations between lens unfolding delay and rotation. Assuming two subsets $x \subset \mathcal{X}$ and $y \subset \mathcal{Y}$ to be the representatives of unfolding delays and rotations of a particular brand of IOLs ($x = t_U$ and $y = R^l$), the Pearson correlation coefficient can be calculated as the covariance of x and y divided by the multiplication of standard deviation of these two sets:

$$\text{Pearson}_{x,y} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \times \sum_{i=1}^m (y_i - \bar{y})^2}} \tag{7}$$

To evaluate the significant differences between the rotations of different IOL brands, we employ a t-test as follows:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{m} \times \sum_{i=1}^m (x_i - \bar{x})^2 \times \sum_{i=1}^m (y_i - \bar{y})^2}} \tag{8}$$

In the last two equations, m is the number of samples in each set, being equal to 94 in our experiments. In our evaluations, we set the significance level as 0.05.

IV. EXPERIMENTAL SETUP

In this section, we first describe the datasets prepared and utilized for training the three neural network architectures as well as our large-scale IOL evaluations using the trained networks and the proposed framework in IV-A. Afterward, we explain the network training settings in IV-B and introduce the evaluation metrics in IV-C. The inference configurations are explained in IV-D.

A. VIDEO AND IMAGE DATASETS

This study adhered to the tenets of the Declaration of Helsinki with the approval of the ethics committee (EK 28/17). All patients provided written informed consent before the study. The studies follow the reporting guidelines of the Standards for Quality Improvement Reporting Excellence (SQUIRE) and the Standards for Reporting of Diagnostic Accuracy (STARD). No patient received compensation or was offered any incentive for participating in this study.

The dataset used in this study stems from the Cataract-1k video dataset [66], containing 1000 videos of cataract surgeries recorded at the Klinikum Klagenfurt in 2020-2021. Since the task of intraocular lens qualification involves three independent trained networks, namely *phase recognition*, *pose estimation*, and *semantic segmentation*, we prepared three training/testing datasets. It should be noted that all training and testing images and videos are split patient-wise, meaning no images in the training and testing sets are sampled from the same video. This separation is intrinsic to building models working in real-world conditions. In addition to the training dataset, we use a large-scale dataset for validation, which we denote *inference dataset*. The four mentioned datasets are explained as follows[5]:

1) **Phase recognition dataset**: Contains annotations of the implantation phase versus the rest of the phases using the first and last implantation phase frames for 100 cataract surgery videos.

2) **Pose estimation dataset**: Includes bounding box annotations of the IOL and its hooks. Overall, 532 frames from 45 videos with the condition that at least one of the two lenses' hooks is visible were manually selected from the post-implantation phase. Next, 532 bounding boxes of the lenses and 821 bounding boxes of the lens' hooks were manually annotated. From these frames, 409 frames are used for training, and the remaining frames are used for testing.

3) **Semantic segmentation dataset**: This dataset includes the segmentation masks of the IOL and pupil. The lens dataset contains 401 frames from 27 videos, and the pupil dataset contains 189 frames from 16 videos. From these annotations, we use 13 videos containing 141 frames with pupil annotation and 21 videos containing 292 frames with lens annotation for training. The remaining frames are utilized for testing.

4) **Inference dataset**: Includes 376 other videos of cataract surgery and is used for validating trained models against different lenses' statistics. This dataset contains lenses from four brands: Technis, AvanSee, NC1, and XC1. From each brand, 94 videos were included.

B. TRAINING SETTINGS

For the phase recognition stage, all networks are trained for 20 epochs, with initial learning rates of 0.0002 for the VGG19 backbone and 0.0004 for the Resnet50 backbone. The learning rates are halved after ten epochs.

For the semantic segmentation task, all networks are trained for 30 epochs. To account for differences in segmentation networks used for evaluations, three different initial learning rates ($lr_0 \in 0.0005, 0.001, 0.002$) are used, with a learning rate decrease of 0.8 every other epoch, and the results with the highest Dice coefficient are reported for each network.

For the pose estimation task, we use the ResNet101 backbone and set the initial learning rate to 0.001. This network is trained for 50 epochs.

The backbones of all networks evaluated for phase recognition, lens/pupil semantic segmentation, and lens/hook localization are initialized with ImageNet [67] weights. The input image size for all networks is set to $512 \times 512 \times 3$. Data augmentation techniques, including motion blur, Gaussian blur, random contrast, random brightness, shift, scale, and rotation, are applied to prevent overfitting and improve generalization performance. For the phase recognition and object localization tasks, we use binary cross entropy and the region-proposal-network (RPN) loss, respectively. For the semantic segmentation task, we adopt the cross-entropy-log-dice loss function consisting of categorical cross entropy and the logarithm of the soft Dice coefficient as follows:

$$\mathcal{L} = \lambda \times \text{CE}(\mathcal{X}_{\text{Pred}}, \mathcal{X}_{\text{True}})$$
$$-(1 - \lambda) \times \log_2 \text{Dice}(\mathcal{X}_{\text{Pred}}, \mathcal{X}_{\text{True}}) \quad (9)$$

where CE stands for *Cross Entropy*, and $\mathcal{X}_{\text{Pred}}$ and $\mathcal{X}_{\text{True}}$ are the predicted masks and ground-truth segmentations, respectively. Besides, λ is set to 0.8 in our experiments.

C. EVALUATION METRICS

We evaluated the performance of each module separately. Phase recognition performance is evaluated using the standard classification metrics, namely Precision, Recall, F1-Score, and Accuracy. We use the Dice coefficient and intersection over union (IoU) to evaluate the semantic segmentation performance. For pose estimation evaluation, we adopt two schemes: (1) using the mean average precision (mAP) metric to measure the performance in object detection and localization, and (2) using the orientation detection error.

D. INFERENCE

Our proposed framework, leveraging the existing neural network architectures, is capable of real-time performance

---

[5]The annotated datasets will be publicly released in https://ftp.itec.aau.at/datasets/ovid/lens-dislocation/

with the support of four RTX3090 GPUs. It's crucial to underscore the adaptability of our framework, recognizing the diversities in hardware infrastructures. In the case of lower hardware configurations, our algorithm allows for a seamless substitution of neural networks—such as those handling phase recognition, semantic segmentation, and object localization—with lightweight architectures. This ensures faster inference times without seriously compromising performance.

V. EXPERIMENTAL RESULTS

In this section, we first assess the effectiveness of each distinct module within the proposed framework based on the results reported in Table 1. Next, we use the proposed framework with trained networks for the statistical evaluations of the four mentioned IOL brands to assess the possibility of automating IOL evaluation and post-operative IOL irregularity prediction.

A. SURGICAL PHASE CLASSIFICATION RESULTS

Table 1-A presents the implantation phase recognition results of the utilized recurrent CNN considering two different backbone networks and four different recurrent layers. It can be perceived from the table that the utilized architecture can effectively capture the joint spatio-temporal features associated with the implantation phase disregarding the backbone network's model and the recurrent layer. Surprisingly, the network with the bidirectional LSTM layer and the VGG19 backbone could retrieve 100% of the three-second clips belonging to the implantation phase. Moreover, this network was 100% precise in discriminating the implantation phase versus the rest of the phases. These results confirm the effectiveness of the proposed approach in detecting the starting point for lens statistics computation.

B. SEMANTIC SEGMENTATION RESULTS

To highlight the superiority of the utilized semantic segmentation network in segmenting the artificial lens and pupil, we have compared its results with several state-of-the-art networks. Table 1-B lists the semantic segmentation results of the proposed approach (AdaptNet) and rival approaches based on the mean and standard deviation of the IoU and Dice coefficient. According to the IoU results, AdaptNet has achieved the best performance in segmenting both IOL and pupil. It gains at least 4% relative improvement in IoU and 2.4% relative improvement in Dice coefficient compared to the best alternative network (UNet++) in lens segmentation.

C. LENS' POSE ESTIMATION RESULTS

Table 1-C lists the lens and hooks localization results based on mean average precision (mAP) and mean average precision at 50% intersection over union (mAP@0.5IoU). The results indicate that the network is 82% accurate in localizing the bounding boxes with at least 50% intersection over union.

According to Table 1-D, the proposed lens orientation computation method shows a mean error equal to

3.707 degrees, with its standard deviation being equivalent to 7.499 degrees. Moreover, the model shows less than 1.5 degrees error in orientation computation for at least 50% of samples in the test set. These results confirm the reliability of the proposed method for lens orientation computation.

D. IOL EVALUATIONS AND STATISTICAL COMPARISONS

In order to statistically compare the behavior of different groups of lenses, we have computed lens unfolding time, instability, and unfolded-lens rotation (rotation for short) for four brands of intraocular lenses. Figure 5 demonstrates the boxplots of the three mentioned measurements for each group of lenses containing 94 cataract surgery videos. Regarding intra-operative rotation after unfolding, the four brands of lenses show significantly different behaviors based on the middle quartile and the interquartile range and the amount of skewness. According to the unfolding plots, XC1 shows the smallest interquartile range (IQR) and the smallest overall spread ($[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$), suggesting that XC1 lenses have less dispersed unfolding time. On the other hand, the Tecnis and NC1 boxplots have a substantially higher upper whisker and positive skew, meaning more dispersed unfolding time and non-normally distributed data. The distant outliers in the two latter lenses can also imply more irregularities. The boxplots of instability for all four groups of lenses have a relatively close length of interquartile range. We can conclude that intra-operative instability of the lens during unfolding cannot be used as an indicator of postoperative rotation.

We infer from the boxplots that there is a higher statistical difference between the unfolding delay and rotation of the four types of IOLs compared to their instabilities. Hence, we have further computed the Pearson correlations between the unfolding delay and rotation of these lenses. As listed in Table 2-A, the p-values for correlations between lens unfolding time and rotation of the AvanSee (0.0202), NC1 (0.0003), and XC1 (0.0101) are less than 0.05. This suggests that there is a significant correlation between the unfolding time and rotation of each mentioned group of IOLs. Besides, we have computed the p-value based on a t-test between the rotation of the four groups of IOLs (Table 2-B). The t-test results confirm the statistically significant differences between the rotations of Tecnis vs. Avansee (0.0322), Tecnis vs. NC1 (5.01×10^{-6}), Tecnis vs. XC1 (8.26×10^{-18}), AvanSee vs. NC1 (0.0019), AvanSee vs. XC1 (1.48×10^{-14}), and NC1 vs. XC1 (0.0363). The results of the t-test between the unfolding time of different IOLs in Table 2-C also indicate a nearly significant difference between the unfolding delay of the XC1 and Tecnis lenses (0.054).

VI. DISCUSSION

We formulated a hypothesis suggesting that variations in the unfolding delay of intraocular lenses play a significant role in their susceptibility to post-operative rotational instability. To substantiate this idea, we conducted a comparative analysis of the unfolding time in our studies and cross-referenced

TABLE 1. Stepwise evaluation of the proposed lens irregularity detection framework.

(A) Phase recognition results of the end-to-end recurrent convolutional neural networks								
	Backbone:VGG16				Backbone:ResNet50			
RNN	Precision	Recall	F1–Score	Accuracy	Precision	Recall	F1–Score	Accuracy
GRU [68]	0.97	0.96	0.96	0.96	0.90	0.94	0.94	0.94
LSTM [69]	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.96
BiGRU	0.97	0.96	0.96	0.96	0.95	0.95	0.95	0.95
BiLSTM	**1.00**	**1.00**	**1.00**	**1.00**	0.98	0.98	0.98	0.98

(B) Lens and pupil segmentation results based on intersection over union (IoU) and Dice								
Object	Metric	U-Net [70]	PSPNet [71]	CE-Net [72]	CPFNet [73]	UNet++/DS [74]	UNet++ [74]	AdaptNet [54]
Lens	IoU	61.89	71.40	70.56	75.38	82.32	83.61	**87.09**
	Dice	73.86	81.53	82.22	85.26	89.95	90.44	**92.62**
Pupil	IoU	83.51	89.55	87.66	92.33	95.28	96.02	**96.06**
	Dice	89.36	94.18	93.32	95.99	97.53	97.96	**97.98**

(C) Lens and hook localization results of Faster R-CNN [65]		
Backbone	mAP	mAP@0.5IoU
ResNet50	0.547	0.828

(D) Lens orientation computation results				
Mean Error	Std of Error	Top 75% Error	Top 50% Error	Top 25% Error
3.707	7.499	3.455	1.31	0.439

TABLE 2. Statistical analysis of the behavior of different intraocular lens brands. The statistically significant results are bold.

(A) Pearson correlation and p-values between lens unfolding delay and lens rotation				
Lens Brand	Tecnis	AvanSee	NC1	XC1
Pearson Correlation	0.0238	0.2392	0.3592	0.2639
P-Value	0.8194	**0.0202**	**0.0003**	**0.0101**

(B) P-values resulting from the t-test between the rotations of different IOLs				
	Tecnis	AvanSee	NC1	XC1
Tecnis	N/A	**0.0322**	$\mathbf{5.01 \times 10^{-6}}$	$\mathbf{8.26 \times 10^{-18}}$
AvanSee	**0.0322**	N/A	**0.0019**	$\mathbf{1.48 \times 10^{-14}}$
NC1	$\mathbf{5.01 \times 10^{-6}}$	0.0019	N /A	0.0363
XC1	$\mathbf{8.26 \times 10^{-18}}$	$\mathbf{1.48 \times 10^{-14}}$	0.0363	N /A

(C) P-values resulting from the t-test between the unfolding of different IOLs				
	Tecnis	AvanSee	NC1	XC1
Tecnis	N/A	0.143	0.798	0.054
AvanSee	0.143	N /A	0.204	0.471
NC1	0.798	0.204	N /A	0.116
XC1	0.054	0.471	0.116	N/A

it with data published on post-operative IOL rotation, as indicated by [15], [17], and [75].

In the study that compared the post-operative rotational stability of the Hoya Vivinex lens (XC1) and the Tecnis IOL, conducted by Osawa et al. [75], it was revealed that the Hoya Vivinex lens displayed less early post-operative rotation than the Tecnis IOL. Our research identified distinct differences in the unfolding behavior of these two lenses, with the Tecnis IOL exhibiting a less predictable unfolding time and more irregularities when compared to XC1. Furthermore, a recent cohort study involving 647 implanted Tecnis IOLs highlighted a noteworthy post-operative rotation issue, with 8.1% of cases experiencing rotations exceeding 5 degrees and 3.1% requiring secondary interventions for re-positioning [15]. This pattern of high post-operative absolute rotation in Tecnis lenses has also been corroborated by other studies, as cited in [17].

Our proposed framework has effectively generated features that confirm statistically significant differences in the intra-operative rotational stability of various IOL brands. The direct correlation between the intraoperative rotation of a pair of lenses (Tecnis vs. XC1) and their subsequent post-operative rotational stability reaffirms the predictability of this complication during surgery. These findings can provide valuable guidance to surgeons in their IOL selection, particularly in cases where patient-specific factors, such as myopia (where lens dislocation is more likely and lens rotation can significantly impact vision, as discussed in [10]), need to be considered. Ultimately, our research contributes to the prevention of post-operative
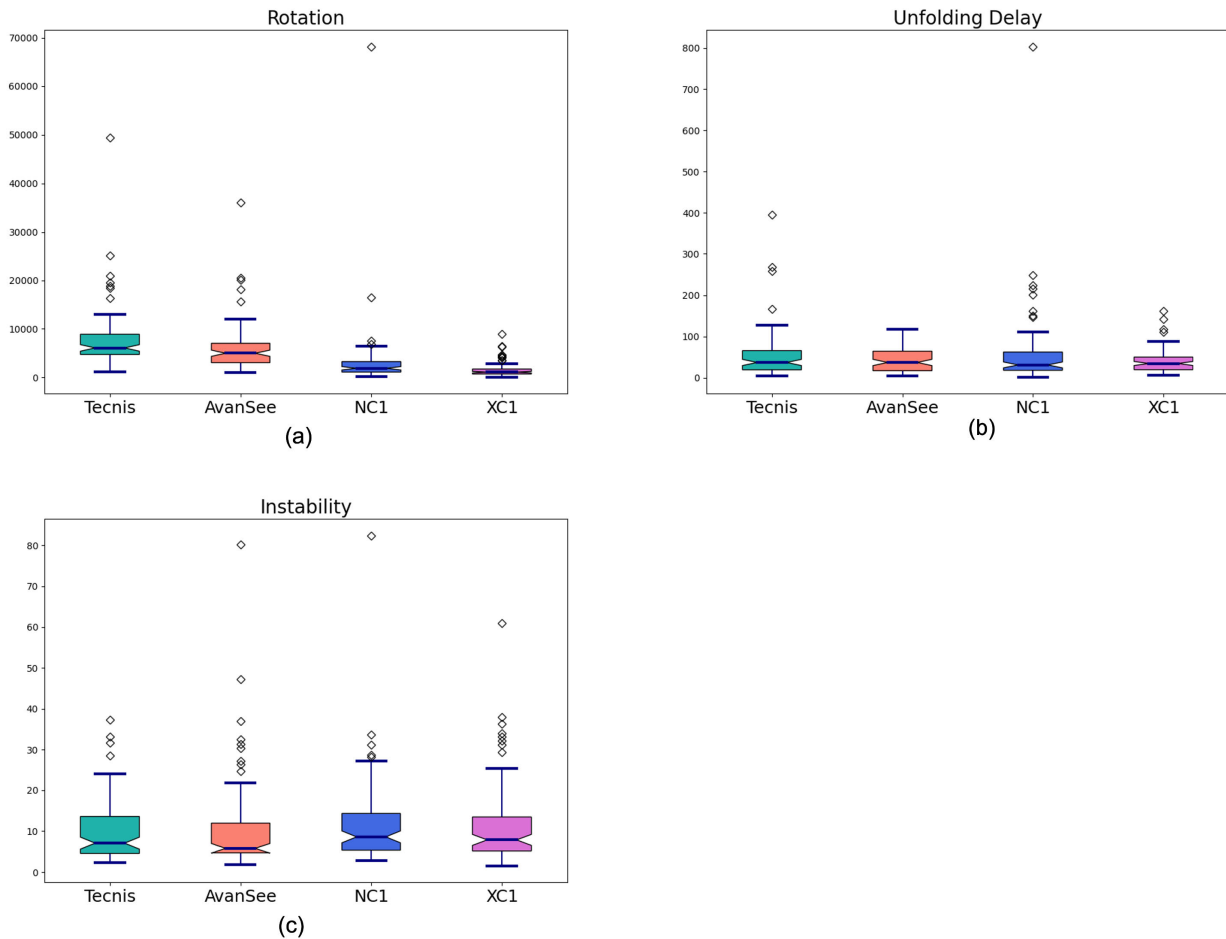
FIGURE 5. The lens statistics for one representative cataract surgery video.

IOL complications, enhancing patient outcomes and satisfaction.

A. LIMITATIONS OF THE PRESENT STUDY

Evaluating the intra-operative behavior of IOLs can empower the predictability of the post-operative IOL's dislocation. However, in addition to the lens characteristics, intrinsic factors such as physicochemical and surface properties of acrylic IOL, modifiable factors such as the temperature of the IOL at the time of implantation, and the viscoelastic used can play a role in the unfolding of an IOL during cataract surgery. The proposed framework enables surgeons to evaluate the intra-operative behavior of IOLs and correlate them with IOLs' early postoperative stability. These results help the surgeons measure each factor's influence on IOL post-operative complications.

VII. CONCLUSION

Intraocular lens dislocation stands as a pivotal postoperative complication in cataract surgery, warranting considerable attention due to its implications on patient outcomes and healthcare costs. This paper presents a pioneering framework,

representing the first endeavor to automate the extraction of critical lens statistics including intraocular lens unfolding delay, instability, and rotation during cataract surgery. We have proposed, evaluated, and utilized a CNN-RNN-based framework for a large-scale evaluation of four brands of IOLs. These results have enabled us to measure statistical correlations between different features in each IOL and differences in the behavior of four commonly used IOL brands during the surgery. The proposed framework not only helps enhance our understanding of IOL behavior during cataract surgery but also offers a major step toward predicting and ultimately preventing such a crucial irregularity. By improving the predictability of complications such as lens dislocation, we can substantially reduce the economic burden for patients and healthcare systems alike. Beyond the financial aspect, preventing such complications translates to heightened patient satisfaction, improved quality of care, and overall enhanced surgical experiences.

Expanding upon our current work, future investigations should focus on an exhaustive examination of the multifaceted factors influencing the intraoperative behavior of intraocular lenses (IOLs) during cataract surgery. This

could encompass a comprehensive study of the relationships between lens characteristics, the inherent properties of acrylic IOLs, the temperature during implantation, and the type of viscoelastic used. In addition to this multifaceted exploration, the practical applicability of our framework for real-time analysis of IOL behavior during surgical procedures is reliant on the existing hardware infrastructure within operating rooms. Accordingly, a key consideration for enhancing the usability of such frameworks is the reduction of dependency on powerful GPUs by optimizing network parameters. This optimization will not only improve the efficiency of our system but also promote its wider adoption across diverse surgical environments.

REFERENCES

[1] M. J. Burton, J. Ramke, A. P. Marques, R. R. Bourne, N. Congdon, I. Jones, B. A. A. Tong, S. Arunga, D. Bachani, and C. Bascaran, "The lancet global health commission on global eye health: Vision beyond 2020," *IHOPE J. Ophthalmol.*, vol. 1, pp. 16–18, Jan. 2022.

[2] S. Srinivasan, "In-the-bag intraocular lens dislocation: A ticking time bomb," *J. Cataract Refractive Surg.*, vol. 45, no. 3, pp. 259–260, 2019.

[3] T. Ianchulev, D. Litoff, D. Ellinger, K. Stiverson, and M. Packer, "Office-based cataract surgery: Population health outcomes study of more than 21 000 cases in the United States," *Ophthalmology*, vol. 123, no. 4, pp. 723–728, 2016.

[4] V. L. Qin, F. F. Conti, and R. P. Singh, "Measuring outcomes in cataract surgery," *Current Opinion Ophthalmol.*, vol. 29, no. 1, pp. 100–104, 2018.

[5] D. C. Minassian, "Extracapsular cataract extraction compared with small incision surgery by phacoemulsification: A randomised trial," *Brit. J. Ophthalmol.*, vol. 85, no. 7, pp. 822–829, Jul. 2001.

[6] M. Lanza, R. Koprowski, R. Boccia, K. Krysik, S. Sbordone, A. Tartaglione, A. Ruggiero, and F. Simonelli, "Application of artificial intelligence in the analysis of features affecting cataract surgery complications in a teaching hospital," *Frontiers Med.*, vol. 7, Dec. 2020, Art. no. 607870.

[7] J. V. Han, D. V. Patel, K. Liu, B. Z. Kim, T. Sherwin, and C. N. J. McGhee, "Auckland cataract study IV: Practical application of NZCRS cataract risk stratification to reduce phacoemulsification complications," *Clin. Experim. Ophthalmol.*, vol. 48, no. 3, pp. 311–318, Apr. 2020.

[8] A. C. Day, P. H. J. Donachie, J. M. Sparrow, and R. L. Johnston, "The royal college of ophthalmologists' national ophthalmology database study of cataract surgery: Report 1, visual outcomes and complications," *Eye*, vol. 29, no. 4, pp. 552–560, Apr. 2015.

[9] I. I. K. Ahmed and A. S. Crandall, "In-the-bag intraocular lens dislocation," *Amer. J. Ophthalmol.*, vol. 139, no. 5, pp. 952–953, May 2005.

[10] G.-I. Lee, D. H. Lim, S. A. Chi, S. W. Kim, D. W. Shin, and T.-Y. Chung, "Risk factors for intraocular lens dislocation after phacoemulsification: A nationwide population-based cohort study," *Amer. J. Ophthalmol.*, vol. 214, pp. 86–96, Jun. 2020.

[11] S. Subasi, N. Yuksel, V. L. Karabas, and B. Y. Tugan, "Late in-the-bag spontaneous iol dislocation: Risk factors and surgical outcomes," *Int. J. Ophthalmol.*, vol. 12, no. 6, p. 954, 2019.

[12] O. Kristianslund, M. Raen, A. E. Østern, and L. Drolsum, "Late in-the-bag intraocular lens dislocation," *Ophthalmology*, vol. 124, no. 2, pp. 151–159, Feb. 2017.

[13] T. Oshika, K. Negishi, T. Noda, H. Arai, M. Inamura, Y. Inoue, T. Miyoshi, Y. Fujita, K. Miyata, and Y. Hasegawa, "Prospective assessment of plate-haptic rotationally asymmetric multifocal toric intraocular lens with near addition of +1.5 diopters," *BMC Ophthalmol.*, vol. 20, no. 1, p. 454, Nov. 2020.

[14] B. S. Lee and D. F. Chang, "Comparison of the rotational stability of two toric intraocular lenses in 1273 consecutive eyes," *Ophthalmology*, vol. 125, no. 9, pp. 1325–1331, Sep. 2018.

[15] D. Schartmüller, L. Schwarzenbacher, E. L. Meyer, S. Schriefl, C. Leydolt, and R. Menapace, "Comparison of long-term rotational stability of three commonly implanted intraocular lenses," *Amer. J. Ophthalmol.*, vol. 220, pp. 72–81, Dec. 2020.

[16] Y. Inoue, H. Takehara, and T. Oshika, "Axis misalignment of toric intraocular lens: Placement error and postoperative rotation," *Ophthalmology*, vol. 124, no. 9, pp. 1424–1425, Sep. 2017.

[17] D. Schartmüller, S. Schriefl, L. Schwarzenbacher, C. Leydolt, and R. Menapace, "True rotational stability of a single-piece hydrophobic intraocular lens," *Brit. J. Ophthalmol.*, vol. 103, no. 2, pp. 186–190, Feb. 2019.

[18] C. F. Mayer-Xanthaki, G. Pregartner, N. Hirnschall, T. Falb, M. Sommer, O. Findl, and A. Wedrich, "Impact of intraocular lens characteristics on intraocular lens dislocation after cataract surgery," *Brit. J. Ophthalmol.*, vol. 105, no. 11, pp. 1510–1514, Nov. 2021.

[19] H. Li, J. H. Lim, J. Liu, D. W. K. Wong, N. M. Tan, S. Lu, Z. Zhang, and T. Y. Wong, "An automatic diagnosis system of nuclear cataract using slit-lamp images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 3693–3696.

[20] A. A. Siddiqui, J. G. Ladas, and J. K. Lee, "Artificial intelligence in cornea, refractive, and cataract surgery," *Current Opinion Ophthalmol.*, vol. 31, no. 4, pp. 253–260, 2020.

[21] D. J. Lindegger, J. Wawrzynski, and G. M. Saleh, "Evolution and applications of artificial intelligence to cataract surgery," *Ophthalmol. Sci.*, vol. 2, no. 3, Sep. 2022, Art. no. 100164.

[22] Y. Xu, X. Gao, S. Lin, D. W. K. Wong, J. Liu, D. Xu, C. Y. Cheng, C. Y. Cheung, and T. Y. Wong, "Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Berlin, Germany: Springer, 2013, pp. 468–475.

[23] P. Smith, L. Tang, V. Balntas, K. Young, Y. Athanasiadis, P. Sullivan, B. Hussain, and G. M. Saleh, "'PhacoTracking': An evolving paradigm in ophthalmic surgical training," *JAMA Ophthalmol.*, vol. 131, no. 5, pp. 659–661, 2013.

[24] T. Neumuth, "Surgical process modeling," *Innov. Surgical Sci.*, vol. 2, no. 3, pp. 123–137, May 2017.

[25] S. Mamtora, R. Jones, A. Rabiolo, G. M. Saleh, and J. D. Ferris, "Remote supervision for simulated cataract surgery," *Eye*, vol. 36, no. 6, pp. 1333–1334, Jun. 2022.

[26] M. F. Jacobsen, L. Konge, D. Bach-Holm, M. L. Cour, L. Holm, K. Hojgaard-Olsen, H. Kjærbo, G. Saleh, and A. S. Thomsen, "Correlation of virtual reality performance with real-life cataract surgery performance," *J. Cataract Refractive Surg.*, vol. 45, no. 9, pp. 1246–1251, 2019.

[27] A. S. S. Thomsen, D. Bach-Holm, H. Kjærbo, K. Højgaard-Olsen, Y. Subhi, G. M. Saleh, Y. S. Park, M. la Cour, and L. Konge, "Operating room performance improves after proficiency-based virtual reality cataract surgery training," *Ophthalmology*, vol. 124, no. 4, pp. 524–531, Apr. 2017.

[28] G. M. Saleh, J. R. Wawrzynski, K. Saha, P. Smith, D. Flanagan, M. Hingorani, C. John, and P. Sullivan, "Feasibility of human factors immersive simulation training in ophthalmology: The London pilot," *JAMA Ophthalmol.*, vol. 134, no. 8, pp. 905–911, 2016.

[29] T. Bourcier, J. Chammas, P.-H. Becmeur, A. Sauer, D. Gaucher, P. Liverneaux, J. Marescaux, and D. Mutter, "Robot-assisted simulated cataract surgery," *J. Cataract Refractive Surg.*, vol. 43, no. 4, pp. 552–557, 2017.

[30] C. Chen, Y. Lee, M. J. Gerber, H. Cheng, Y. Yang, A. Govetto, A. A. Francone, S. Soatto, W. S. Grundfest, J. Hubschman, and T. Tsao, "Intraocular robotic interventional surgical system (IRISS): Semi-automated OCT-guided cataract removal," *Int. J. Med. Robot. Comput. Assist. Surgery*, vol. 14, no. 6, p. e1949, Dec. 2018.

[31] S. P. Devi, K. S. Rao, and S. S. Sangeetha, "Prediction of surgery times and scheduling of operation theaters in optholmology department," *J. Med. Syst.*, vol. 36, no. 2, pp. 415–430, Apr. 2012.

[32] D. Carmona González and C. Palomino Bautista, "Accuracy of a new intraocular lens power calculation method based on artificial intelligence," *Eye*, vol. 35, no. 2, pp. 517–522, Feb. 2021.

[33] Y.-C. Tham, J. H. L. Goh, A. Anees, X. Lei, T. H. Rim, M.-L. Chee, Y. X. Wang, J. B. Jonas, S. Thakur, and Z. L. Teo, "Detecting visually significant cataract using retinal photograph-based deep learning," *Nature Aging*, vol. 2, no. 3, pp. 264–271, Feb. 2022.

[34] P. Zéboulon, C. Panthier, H. Rouger, J. Bijon, W. Ghazal, and D. Gatinel, "Development and validation of a pixel wise deep learning model to detect cataract on swept-source optical coherence tomography images," *J. Optometry*, vol. 15, pp. S43–S49, 2022.

[35] P. Tripathi, Y. Akhter, M. Khurshid, A. Lakra, R. Keshari, M. Vatsa, and R. Singh, "MTCD: Cataract detection via near infrared eye images," *Comput. Vis. Image Understand.*, vol. 214, Jan. 2022, Art. no. 103303.

[36] T. D. Keenan, Q. Chen, E. Agron, Y.-C. Tham, J. H. L. Goh, X. Lei, Y. P. Ng, Y. Liu, X. Xu, C.-Y. Cheng, and M. M. Bikbov, "DeepLensNet: Deep learning automated diagnosis and quantitative classification of cataract type and severity," *Ophthalmology*, vol. 129, no. 5, pp. 571–584, May 2022.

[37] A. Marafioti, M. Hayoz, M. Gallardo, P. Márquez Neila, S. Wolf, M. Zinkernagel, and R. Sznitman, "CataNet: Predicting remaining cataract surgery duration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham, Switzerland: Springer, 2021, pp. 426–435.

[38] B. Wang, L. Li, Y. Nakashima, R. Kawasaki, and H. Nagahara, "Real-time estimation of the remaining surgery duration for cataract surgery using deep convolutional neural networks and long short-term memory," *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, p. 80, May 2023.

[39] T. K. Yoo, E. Oh, H. K. Kim, I. H. Ryu, I. S. Lee, J. S. Kim, and J. K. Kim, "Deep learning-based smart speaker to confirm surgical sites for cataract surgeries: A pilot study," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, Art. no. e0231322.

[40] R. Garcia Nespolo, D. Yi, E. Cole, N. Valikodath, C. Luciano, and Y. I. Leiderman, "Evaluation of artificial intelligence-based intraoperative guidance tools for phacoemulsification cataract surgery," *JAMA Ophthalmol.*, vol. 140, pp. 170–177, Feb. 2022.

[41] N. Sokolova, K. Schoeffmann, M. Taschwer, S. Sarny, D. Putzgruber-Adamitsch, and Y. El-Shabrawi, "Automatic detection of pupil reactions in cataract surgery videos," *PLoS ONE*, vol. 16, no. 10, Oct. 2021, Art. no. e0258390.

[42] F. Yu, G. Silva Croso, T. S. Kim, Z. Song, F. Parker, G. D. Hager, A. Reiter, S. S. Vedula, H. Ali, and S. Sikder, "Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques," *JAMA Netw. Open*, vol. 2, Apr. 2019, Art. no. e191860.

[43] T. Wang, J. Xia, R. Li, R. Wang, N. Stanojcic, J.-P. O. Li, E. Long, J. Wang, X. Zhang, and J. Li, "Intelligent cataract surgery supervision and evaluation via deep learning," *Int. J. Surg.*, vol. 104, Aug. 2022, Art. no. 106740.

[44] S. Touma, F. Antaki, and R. Duval, "Development of a code-free machine learning model for the classification of cataract surgery phases," *Sci. Rep.*, vol. 12, no. 1, p. 2398, Feb. 2022.

[45] N. Ghamsarian, M. Taschwer, D. Putzgruber-Adamitsch, S. Sarny, and K. Schoeffmann, "Relevance detection in cataract surgery videos by spatio-temporal action localization," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 10720–10727.

[46] H.-H. Yeh, A. M. Jain, O. Fox, K. Sebov, and S. Y. Wang, "PhacoTrainer: Deep learning for cataract surgical videos to track surgical tools," *Translational Vis. Sci. Technol.*, vol. 12, no. 3, p. 23, Mar. 2023.

[47] N. Ghamsarian, M. Taschwer, and K. Schoeffmann, "Deblurring cataract surgery videos using a multi-scale deconvolutional neural network," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 872–876.

[48] N. Ghamsarian, H. Amirpourazarian, C. Timmerer, M. Taschwer, and K. Schöffmann, "Relevance-based compression of cataract surgery videos using convolutional neural networks," in *Proc. 28th ACM Int. Conf. Multimedia*, NY, NY, USA, Oct. 2020, pp. 3577–3585.

[49] N. Ghamsarian, "Enabling relevance-based exploration of cataract videos," in *Proc. Int. Conf. Multimedia Retr.*, New York, NY, USA, Jun. 2020, pp. 378–382.

[50] J. Whitten, J. McKelvie, and M. Mayo, "Clinically-relevant summarisation of cataract surgery videos using deep learning," in *Recent Challenges in Intelligent Information and Database Systems*. Ho Chi Minh City, Vietnam: Springer, Nov. 2022, pp. 711–723.

[51] J. Ruzicki, M. Holden, S. Cheon, T. Ungi, R. Egan, and C. Law, "Use of machine learning to assess cataract surgery skill level with tool detection," *Ophthalmol. Sci.*, vol. 3, no. 1, Mar. 2023, Art. no. 100235.

[52] N. Ghamsarian, M. Taschwer, R. Sznitman, and K. Schoeffmann, "DeepPyramid: Enabling pyramid view and deformable pyramid reception for semantic segmentation in cataract surgery videos," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 276–286.

[53] N. Ghamsarian, M. Taschwer, D. Putzgruber-Adamitsch, S. Sarny, Y. El-Shabrawi, and K. Schóffmann, "Recal-Net: Joint region-channel-wise calibrated network for semantic segmentation in cataract surgery videos," in *Neural Information Processing*, T. Mantoro, M. Lee, M. A. Ayu, K. W. Wong, and A. N. Hidayanto, Eds. Cham, Switzerland: Springer, 2021, pp. 391–402.

[54] N. Ghamsarian, M. Taschwer, D. Putzgruber-Adamitsch, S. Sarny, Y. El-Shabrawi, and K. Schoeffmann, "LensID: A CNN-RNN-based framework towards lens irregularity detection in cataract surgery videos," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham, Switzerland: Springer, 2021, pp. 76–86.

[55] L. Wei, W. He, J. Wang, K. Zhang, Y. Du, J. Qi, J. Meng, X. Qiu, L. Cai, Q. Fan, Z. Zhao, Y. Tang, S. Ni, H. Guo, Y. Song, X. He, D. Ding, Y. Lu, and X. Zhu, "An optical coherence tomography-based deep learning algorithm for visual acuity prediction of highly myopic eyes after cataract surgery," *Frontiers Cell Develop. Biol.*, vol. 9, May 2021, Art. no. 652848.

[56] J. G. Ladas, A. A. Siddiqui, U. Devgan, and A. S. Jun, "A 3-D 'super surface' combining modern intraocular lens formulas to generate a 'super formula' and maximize accuracy," *JAMA Ophthalmol.*, vol. 133, pp. 1431–1436, Dec. 2015.

[57] R. B. Melles, J. X. Kane, T. Olsen, and W. J. Chang, "Update on intraocular lens calculation formulas," *Ophthalmology*, vol. 126, no. 9, pp. 1334–1335, Sep. 2019.

[58] A. R. Brant, J. Hinkle, S. Shi, O. Hess, T. Zubair, S. Pershing, and G. C. Tabin, "Artificial intelligence in global ophthalmology: Using machine learning to improve cataract surgery outcomes at Ethiopian outreaches," *J. Cataract Refractive Surg.*, vol. 47, no. 1, pp. 6–10, 2021.

[59] G. Kiuchi, M. Tanabe, K. Nagata, N. Ishitobi, H. Tabuchi, and T. Oshika, "Deep learning-based system for preoperative safety management in cataract surgery," *J. Clin. Med.*, vol. 11, no. 18, p. 5397, Sep. 2022.

[60] L. Schwarzenbacher, P. Seeböck, D. Schartmüller, C. Leydolt, R. Menapace, and U. Schmidt-Erfurth, "Automatic segmentation of intraocular lens, the retrolental space and Berger's space using deep learning," *Acta Ophthalmologica*, vol. 100, no. 8, pp. e1611–e1616, Dec. 2022.

[61] S.-F. Mohammadi, M. Sabbaghi, H. Z. Mehrjardi, H. Hashemi, S. Alizadeh, M. Majdi, and F. Taee, "Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification," *J. Cataract Refractive Surg.*, vol. 38, no. 3, pp. 403–408, 2012.

[62] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.

[63] N. Ghamsarian, J. G. Tejero, P. Márquez-Neila, S. Wolf, M. Zinkernagel, K. Schoeffmann, and R. Sznitman, "Domain adaptation for medical image segmentation using transformation-invariant self-training," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2023, pp. 331–341.

[64] R. Wen, H. Yuan, D. Ni, W. Xiao, and Y. Wu, "From denoising training to test-time adaptation: Enhancing domain generalization for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jan. 2024, pp. 464–474.

[65] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[66] N. Ghamsarian, Y. El-Shabrawi, S. Nasirihaghighi, D. Putzgruber-Adamitsch, M. Zinkernagel, S. Wolf, K. Schoeffmann, and R. Sznitman, "Cataract-1K: Cataract surgery dataset for scene segmentation, phase recognition, and irregularity detection," 2023, *arXiv:2312.06295*.

[67] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. CVPR*, 2009, pp. 248–255.

[68] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax Semantics Struct. Stat. Transl. (SSST)*, Oct. 2014, pp. 103–111.

[69] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[71] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[72] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[73] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.

[74] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[75] R. Osawa, T. Oshika, M. Sano, T. Yuguchi, and T. Kaiya, "Rotational stability of modified toric intraocular lens," *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0247844.

STEPHANIE SARNY received the Ph.D. degree in medical science. She was a Senior Ophthalmologist with the Department of Ophthalmology, Klinikum Klagenfurt. She has authored multiple research articles in peer-reviewed journals. Her research interests include various topics in epidemiologic medical studies and different aspects of ophthalmic surgical video analysis.

RAPHAEL SZNITMAN received the degree in cognitive systems from The University of British Columbia, Canada, in 2007, and the Ph.D. degree in computer science from Johns Hopkins University, USA, in 2011. He was then a Postdoctoral Fellow with École Polytechnique Fédérale de Lausanne, Switzerland, where he involved in the Computer Vision Laboratory. In 2015, he joined the Faculty of the ARTORG Center, University of Bern, Switzerland, where he is currently a Full Professor of AI for medical imaging and the Director of the ARTORG Center for Biomedical Engineering Research. His primary research interests include computational vision, probabilistic methods, and statistical learning, applied to applications in medical imaging.

NEGIN GHAMSARIAN received the Ph.D. degree in computer science from the University of Klagenfurt, Austria, in 2021. She was a Research Assistant on the project "Deep-Learning-Assisted Analysis of Surgical Videos," in 2019. She is currently a Postdoctoral Researcher with the AI for Medical Imaging Laboratory, Department of Medicine, University of Bern, Switzerland. Her research interests revolve around the application of deep learning techniques, including supervised learning, semi-supervised learning, and unsupervised domain adaptation, for action recognition in surgical videos and semantic segmentation in medical images.

KLAUS SCHOEFFMANN (Member, IEEE) is currently an Associate Professor with the Institute of Information Technology (ITEC), University of Klagenfurt, Austria. He has coauthored more than 100 publications on various topics in multimedia, inclusive of many works on different aspects of medical video analysis. He has co-organized several international conferences, workshops, and special sessions in the field of multimedia. His research interests include video content understanding (in particular medical/surgery videos), multimedia retrieval, interactive multimedia, and applied deep learning. He is a member of ACM and a regular reviewer of international conferences and journals in the field of multimedia.

DORIS PUTZGRUBER-ADAMITSCH is currently a board-certified Senior Ophthalmologist with the Department of Ophthalmology, Klinikum Klagenfurt. Besides extensive clinical surgical expertise, her main research interests include the use of digitalized surgical videos and their use in teaching and evaluation of adverse events.

YOSUF EL-SHABRAWI is currently the Head of the Department of Ophthalmology, Klinikum Klagenfurt. He is an Associate Professor of Ophthalmology with the Department of Ophthalmology, Medical University of Graz, and a Lecturer with the Medical University of Vienna. He authored more than 100 articles published in peer-reviewed journals. His main research interests include the various aspects of immunologic disorders in the eye and the use of machine learning in evaluating ophthalmic surgical videos.

● ● ●