DeepPyramid: Enabling Pyramid View and Deformable Pyramid Reception for Semantic Segmentation in Cataract Surgery Videos

Negin Ghamsarian[1], **Mario Taschwer**[2], **Raphael Sznitman**[1], **Klaus Schoeffmann**[2]

[1]ARTORG Center for Biomedical Engineering Research, Artificial Intelligence in Medical Imaging Laboratory, Universität Bern, Switzerland
[2]Institute of Information Technology, Universität Klagenfurt, Klagenfurt, Austria
negin.ghamsarian@unibe.ch

Abstract

- The varying issues in segmenting the different relevant structures in cataract surgery videos make the designation of a unique network quite challenging. Specifically, a semantic segmentation network is required to simultaneously deal with: 1) a transparent artificial lens that undergoes deformations, 2) color, shape, size, and texture variations in the pupil, 3) unclear edges of the cornea, and 4) severe motion blur, reflection distortion, and scale variations in instruments (Figure 1-a).
- We propose a semantic segmentation network, termed DeepPyramid, that can deal with these challenges using three novelties: (1) a Pyramid View Fusion module which provides a varying-angle global view of the surrounding region centering at each pixel position in the input convolutional feature map; (2) a Deformable Pyramid Reception module which enables a wide deformable receptive field that can adapt to geometric transformations in the object of interest; and (3) a dedicated Pyramid Loss that adaptively supervises multi-scale semantic feature maps (Figure 1-b).

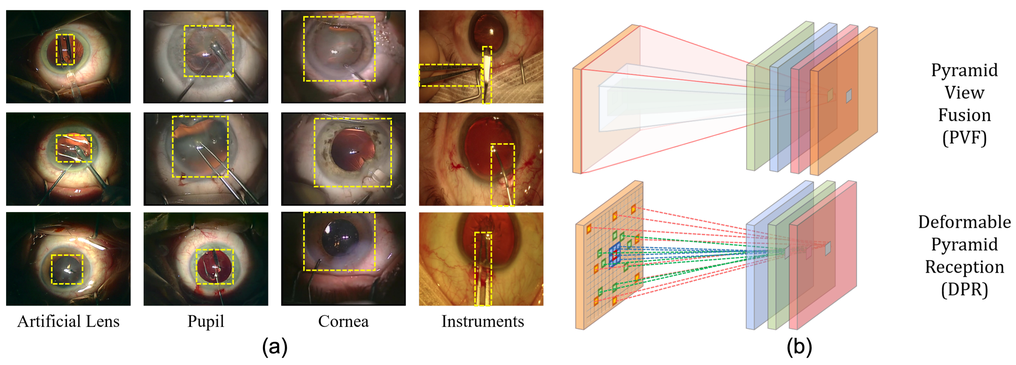

Figure 1: (a) Challenges in semantic segmentation of relevant objects in cataract surgery. (b) Proposed Pyramid View Fusion and Deformable Pyramid Reception modules.

Methodology

- Our proposed segmentation strategy aims to explicitly model deformations and context within its architecture. Using a U-Net-based architecture, our proposed model is illustrated in Figure 2. At its core, the encoder network remains that of a standard VGG16 network. Our approach is to provide useful decoder modules to help alleviate segmentation concerning relevant objects' features in cataract surgery.

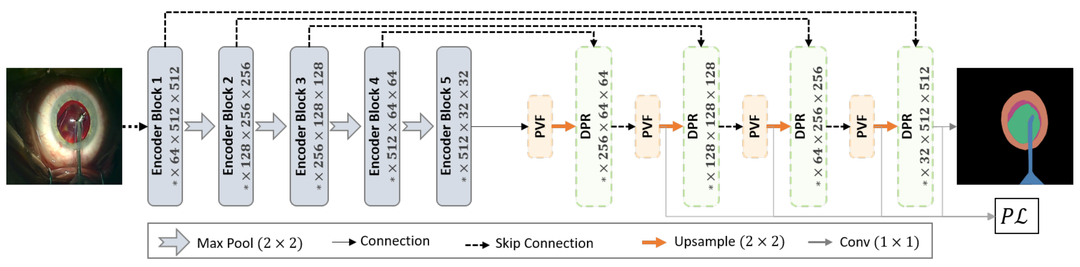

Figure 2: Overall architecture of the proposed DeepPyramid network. It contains Pyramid View Fusion (PVF), Deformable Pyramid Reception (DPR), and Pyramid Loss ($P\mathcal{L}$).

- Conceptually, the PVF module is inspired by the human visual system and aims to recognize semantic information found in images considering not only the internal object's content but also the relative information between the object and its surrounding area. Thus the role of the PVF is to reinforce the observation of relative information at every distinct pixel position. Specifically, we use average pooling to fuse the multi-angle local information for this novel attention mechanism (Figure 3).
- Our DPR module hinges on a novel deformable block based on dilated convolutions that can help recognize each pixel position's semantic label based on its cross-dependencies with varying-distance surrounding pixels without imposing additional trainable parameters. Due to the inflexible rectangle shape of the receptive field in regular convolutional layers, feature extraction procedures cannot be adapted to complex deformable shapes. Our proposed dilated deformable convolutional layers attempt to remedy this explicitly in terms of both scale and shape.

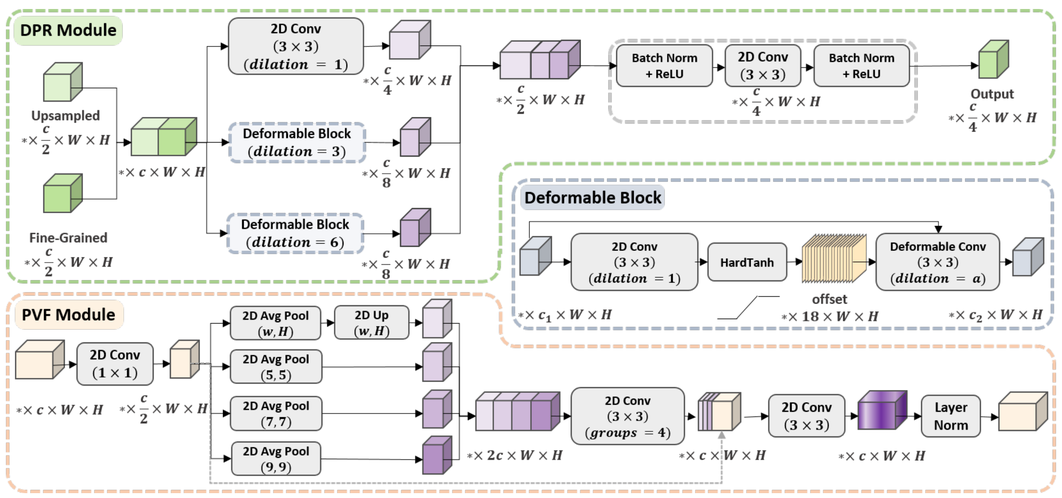

Figure 3: Detailed architecture of the Deformable Pyramid Reception (DPR) and Pyramid View Fusion (PVF) modules.

Experiments and results

- DeepPyramid has achieved more than 4% improvement in lens segmentation (85.61% vs. 81.32%) and more than 4% improvement in instrument segmentation (74.40% vs. 70.11%) compared to UNet++ as the second-best approach.
- Figure 4 further affirms the effectiveness of DeepPyramid in enhancing the segmentation results.

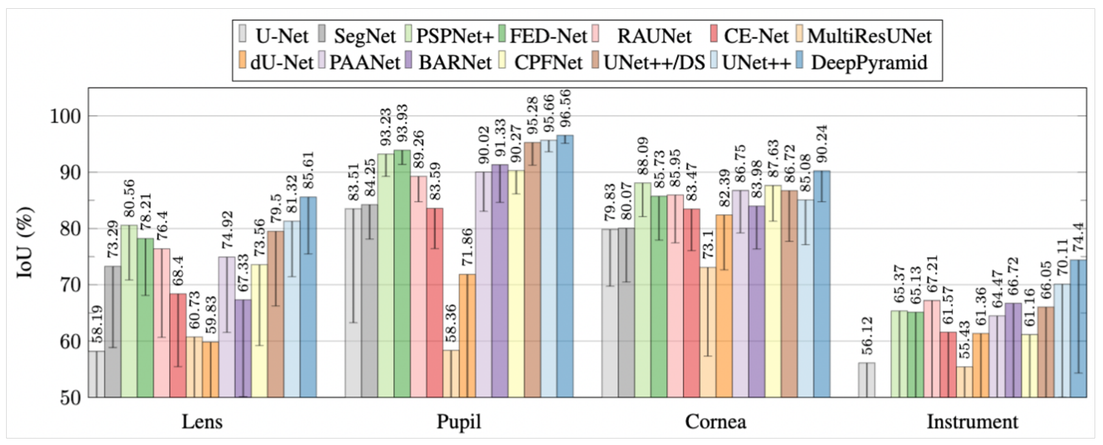

Figure 4: Quantitative comparisons among DeepPyramid and rival approaches based on average and standard deviation of IoU.

- we see from Figure 5 that DeepPyramid shows much less distortion in the region of edges (especially in the case of the cornea), and shows much better precision and recall in the narrow regions (for instruments and other relevant objects in the case of occlusion by the instruments).

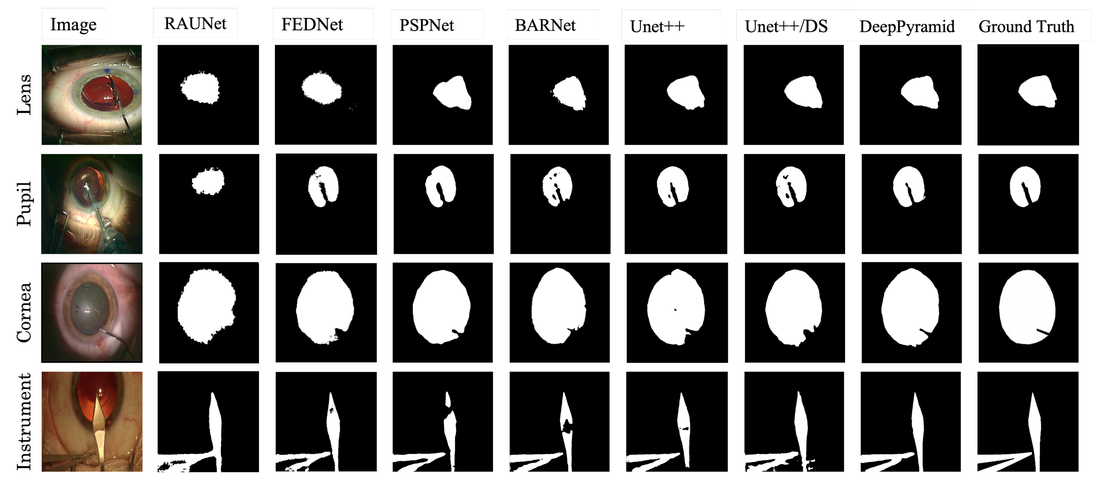

Figure 5: Qualitative comparisons among DeepPyramid and the rival approaches.

- Figure 6 visually compares the outstanding performance of DeepPyramid's modules versus alternative modules.

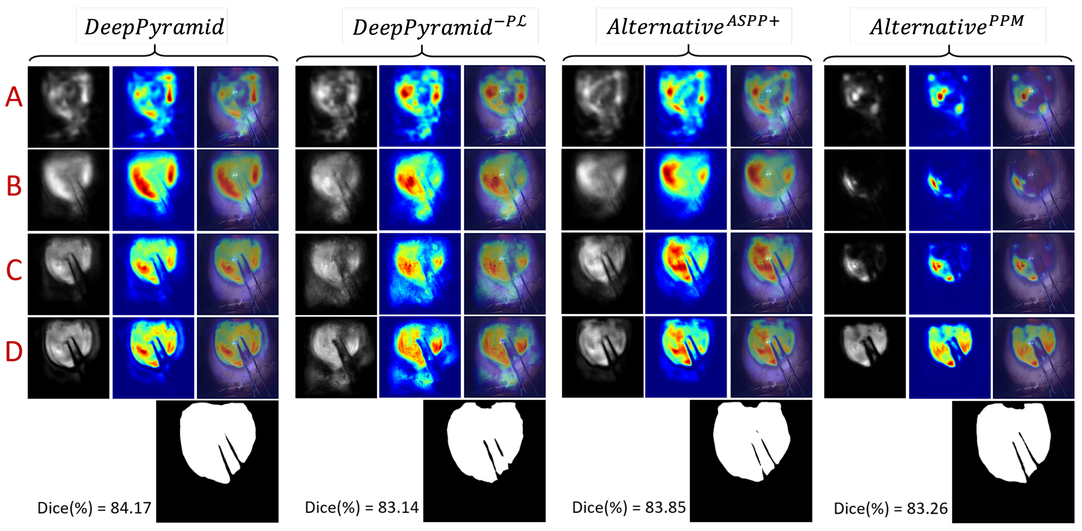

Figure 6: Qualitative comparisons among DeepPyramid and the rival approaches.

Conclusions

In this work, we have proposed a novel network architecture for semantic segmentation in cataract surgery videos. The proposed architecture takes advantage of two modules, namely "Pyramid View Fusion" and "Deformable Pyramid Reception", as well as a dedicated "Pyramid Loss", to simultaneously deal with (i) geometric transformations such as scale variation and deformability, (ii) blur degradation and blunt edges, and (iii) transparency, texture and color variation typically observed in cataract surgery images.

Funding: This work was funded by Haag-Streit Switzerland and the FWF Austrian Science Fund under grant P 31486-N31.