The project goal: The project aims to develop new algorithms for video objects segmentation in the presence of noise. We will focus on objects whose properties change in time, particularly on objects with irregular shape. The meaning of noise is twofold. Firstly we consider noise/artifacts in the video data. Secondly we consider noise in labels, related to incertain expertise provided by the experts labeling the data.

Description of research: The research will focus on answering the following questions: (1) How to combine the general knowledge about the objects and their dynamic with automatic knowledge extraction of high-level futures from data? Will it be helpful for video segmentation of objects with properties changing in time? (2) How to use such a combined knowledge for video object segmentation when dealing with noisy data, noisy labels and objects with irregular shapes ? (2) How to evaluate such an approach/ compare with other approaches if there is no ground truth data? How to prepare a synthetic dataset to cover the complexity of the problem? (3) How to minimize the impact of label noise/ data noise on the segmentation results? By introducing hierarchy? By introducing a multi-task learning paradigm? By using temporal information?

Motivation: We are motivated to conduct this research by two practical problems. The first is the video instance segmentation of full sperm. The second is the video semantic segmentation of teeth structures. Both problems, while seem to be very different, are very similar from the project point of view. Common features of these two applications consist of: (1) changing object properties: The shape of teeth structure is unpredictable and changes during the dental treatment. Ideally some structures like cavities should disappear at the end of the treatment. Generally the shape and morphology of the sperm are predictable but not flagella beat rate/motion pattern. With small width and quick movements, the flagella are hardly visible under blurry and low-contrast imaging conditions, often confusing the human raters. (2) repetitive environment: there is limited number of objects which can be present in the observed scene, and hence we can possess domain knowledge, e.g. reconstruct 3D model of dental scene, motion model of flagella beating; (3) low quality of video data: in both cases we deal with the problem of low quality video data, subject to noise and blur; In case of dental data video is taken with miniaturized vision sensors and optics, the continuous camera vibrations, splashing water and saliva increase level of blur. In case of ICSI video the rapid moving sperm easily enters and leaves in-focus zone, which produces space-variant blur. There is a high level of noise. (4) low interrater agreement: the qualitative performance of deep learning based supervised approaches for video segmentation depends on the quality of labels. However, human experts labeling processes depend on expert's experience, bias and competence levels. Consequently in many practical problems there is no access to "true" labels. Reliability of an expert is unknown. (4) high social value of successful segmentation method: For standard ICSI imaging discovering locomotion patterns involves reconstructing all points along the centerline of the sperm tail. This requires some segmentation of the entire spermatozoon. Discovery of sperm locomotion patterns contributes to our understanding of how sperm navigate inside the female reproductive system. These behaviors can be used to design approaches for selecting sperm that are highly fertile. Infertility affects up to 15% of reproductive-aged couples worldwide. On the other hand, a close and continuous view of the operated tooth enables a more effective and safer dental bur maneuver within the tooth to remove caries and limit the risk of exposing pulp tissue to infection.

Substantial results expected: The new deep learning algorithms exploring combined knowledge composed of the general knowledge about the objects and their dynamic with automatic knowledge extraction of high-level futures from data will play a huge role in the video segmentation of irregular objects from difficult, noisy data. Range of application of video segmentation will be enlarged from objects with some specific properties to objects whose properties change in time. Current best video segmentation algorithms still are built on fully-supervised deep learning techniques, depending on a huge amount of annotated data. Our result can be further explored by the community for designing new annotation efficient techniques for video segmentation. This will have a profound impact well beyond computer vision (e.g. for robotics, biomedical engineering). The project results will be shared with a computer vision community to accelerate further research by publishing: (1) open-sourced implementations of algorithms, (2) datasets,(3) description of algorithms and obtained results in top computer vision journals and conferences.