

Video-assisted Dentistry with Deep Neural Networks

(Invited paper)

Daniel Węsierski

Multimedia Systems Department

Faculty of Electronics, Telecommunications, and Informatics

Gdańsk University of Technology

E-mail: daniel.wesierski@pg.edu.pl

Abstract—This study aims to develop a novel video-based technology to support the next generation of digital dentistry. By enabling enhanced and continuous intraoral visualization, the system has the potential to improve educational environments, clinical training, and real-time operational workflows in conservative dental procedures. The proposed approach is grounded in multi-task learning, where a unified neural network is trained to simultaneously perform multiple vision based tasks for enhanced human-system interaction. To facilitate this development, the study leverages existing and newly curated dental video datasets, including annotated sequences for training and evaluation.

Keywords—dentistry, multi-task learning, video

I. INTRODUCTION

This study aims to develop a novel video-based technology for next-generation digital dentistry. A miniature camera integrated into a dental handpiece could enable dentists to continuously monitor the treatment area during conservative dental procedures (Fig. 1). Such integration promises to enhance clinical outcomes, improve ergonomic efficiency for dental professionals, and elevate the quality of dental education and training. However, image acquisition in intraoral environments presents significant challenges. Miniaturized sensors and optics introduce visual distortions, while handpiece movements cause eye strain. Additionally, complex intraoral conditions—such as noise, blur, illumination changes, shadows, and dynamic fluids—degrade visual clarity, complicating continuous macro-visualization of the treatment field.

II. METHOD

To address the intraoral imaging challenges [5], we propose a vision-based system capable of jointly solving multiple interrelated video processing tasks. The proposed decoder-centric multi-task network [3], [2] enhances video quality under challenging conditions such as low light, noise, and camera shake. It integrates auxiliary tasks—teeth segmentation and teeth-based motion estimation—to support temporal alignment of encoder features, enabling more stable and accurate restoration. Perform robust long-range motion compensation [4] allows stabilizing the tooth's visual representation in the video, reducing eye fatigue even under fluid motion and fluctuating lighting.

III. CONCLUSIONS AND FUTURE WORK

Integrating video super-resolution techniques will further enhance image quality, offering visual feedback comparable to

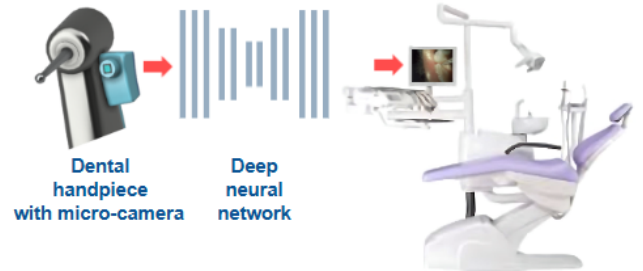


Fig. 1. Camera-assisted dental procedures offer significant potential to enhance training, education, and documentation, streamline clinical workflow and ergonomics, and ultimately improve patient care. However, the complexity of intraoral imaging—characterized by noise, motion blur, and varying lighting conditions—makes continuous macro-level visualization on custom displays particularly challenging. Effective video enhancement techniques are therefore essential to provide a clearer and more stable view of intraoral scenes throughout the entire course of tooth's treatment.

that provided by dental microscopes [1]. Real-time intraoral video enhancement could enable 3D reconstruction for augmented reality and immersive dental education, analogous to recent advances in minimally invasive surgery. Additionally, semantic segmentation of clinically relevant structures—such as caries, dentin, and pulp—can support young practitioners in performing more precise and less risky procedures, while enabling expert-guided assessment of their clinical performance. Through this comprehensive approach, we seek to lay the technological foundation for intelligent, vision-assisted dental procedures in modern digital dentistry.

REFERENCES

- [1] Efklidis Katsaros, Anna Jezierska, and Daniel Węsierski. Leveraging spatio-temporal features for joint deblurring and segmentation of instruments in dental video microscopy. In *2021 14th International Conference on Human System Interaction (HSI)*, pages 1–5. IEEE, 2021.
- [2] Efklidis Katsaros, Piotr Kopa Ostrowski, Anna Jezierska, Emilia Lewandowska, Jacek Rumiński, and Daniel Węsierski. Vident-lab: a dataset for multi-task video processing of phantom dental scenes, 2022.
- [3] Efklidis Katsaros, Piotr K Ostrowski, Krzysztof Włodarczyk, Emilia Lewandowska, Jacek Rumiński, Damian Siupka-Mróż, Łukasz Lassmann, Anna Jezierska, and Daniel Węsierski. Multi-task video enhancement for dental interventions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 177–187. Springer, 2022.
- [4] Emilia Lewandowska, Anna Jezierska, and Daniel Węsierski. Vident-synth: a synthetic intra-oral video dataset for optical flow estimation, 2024.
- [5] Daniel Węsierski, Anna Jezierska, Piotr Kopa Ostrowski, Emilia Lewandowska, Efklidis Katsaros, and Agata Żółtowska. Vident-real: an intra-oral video dataset for multi-task learning, 2024.