B54 (Fri-AM)

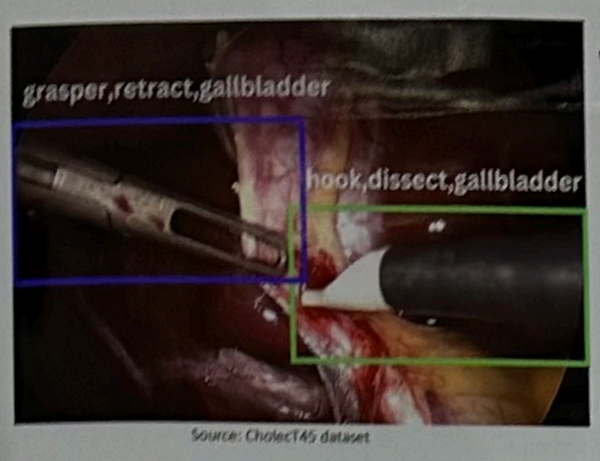
Smarter Self-Distillation: Optimizing the Teacher for Surgical Video Applications

Amine Yamlahi[1,2], Piotr Kalinowski[1,3,6], Patrick Godau[1,2,3,6], Rayan Younis[4], Martin Wagner[4], Beat Müller[5], Lena Maier-Hein[1,2,3,6,7,8]

[1]German Cancer Research Center (DKFZ) Heidelberg, Division of Intelligent Medical Systems, Germany, [2]National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and university medical center Heidelberg, Germany, [3]HIDSS4Health - Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany, [4]Department of Visceral, Thoracic and Vascular Surgery, University Hospital and Faculty of Medicine Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany, [5]University Digestive Healthcare Center Basel, Basel, Switzerland, [6]Faculty of Mathematics and Computer Science, Heidelberg University, Germany, [7]Medical Faculty, Heidelberg University, Germany, [8]Heidelberg University Hospital, Surgical Clinic, Surgical AI Research Group, Heidelberg, Germany

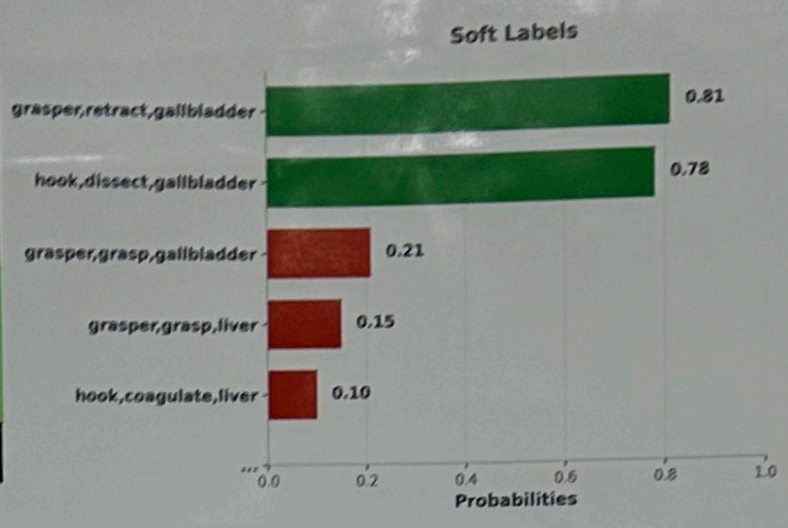
dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Abstract



Source: CholecT45 dataset

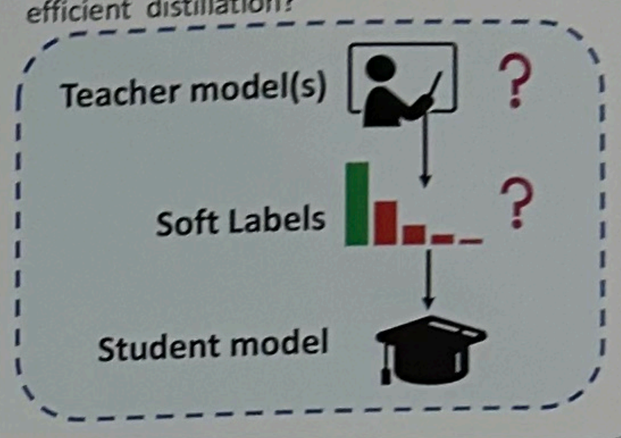
Surgical workflow analysis is challenging due to complex imaging, ambiguous annotations, and large number of classes.



Training with soft labels effectively tackles uncertainty and label ambiguity by capturing the teacher's knowledge and inter-class information.

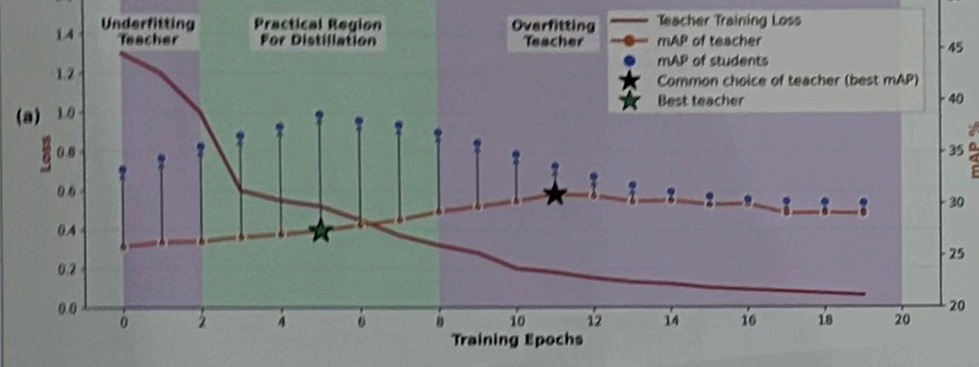
Contribution:

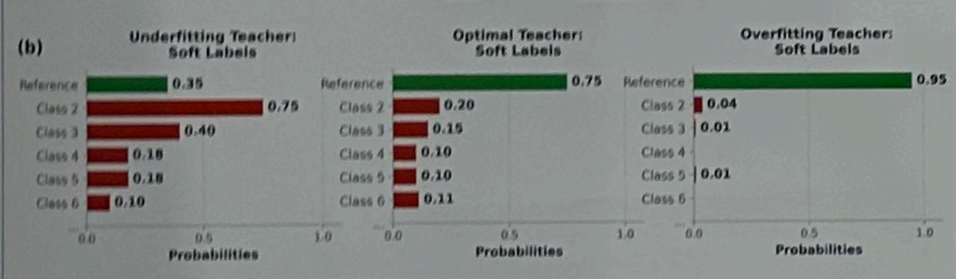
1. How to find the **optimal teacher(s)** for distillation?
2. How to generate **optimal soft labels** for a more efficient distillation?



1 How to find the **optimal teacher(s)** for distillation?

The strongest model is not always the best teacher!

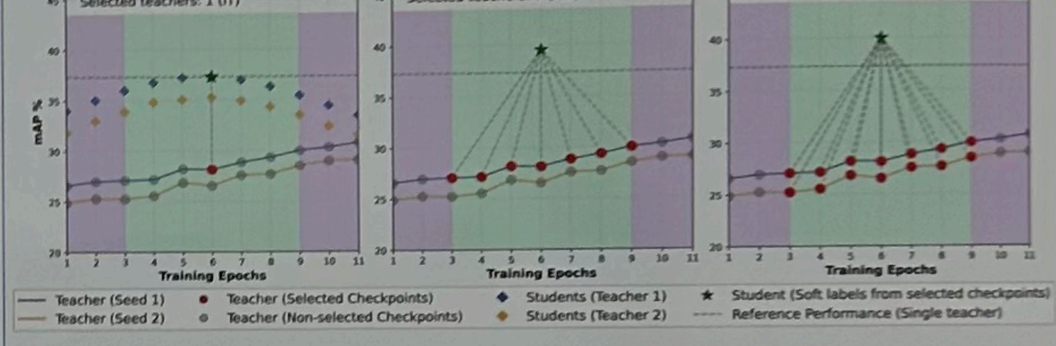




- **Teacher selection is critical for effective self-distillation**
- **Best validation teacher ≠ optimal distillation teacher**
- **Sweet spot exists**: Teachers that are neither too uncertain nor overconfident produce superior soft labels
- **Key insight**: Intermediate confidence levels maximize knowledge transfer quality

2 How to generate **optimal soft labels** for a more efficient distillation?

Optimized teachers ensemble across epochs/seeds



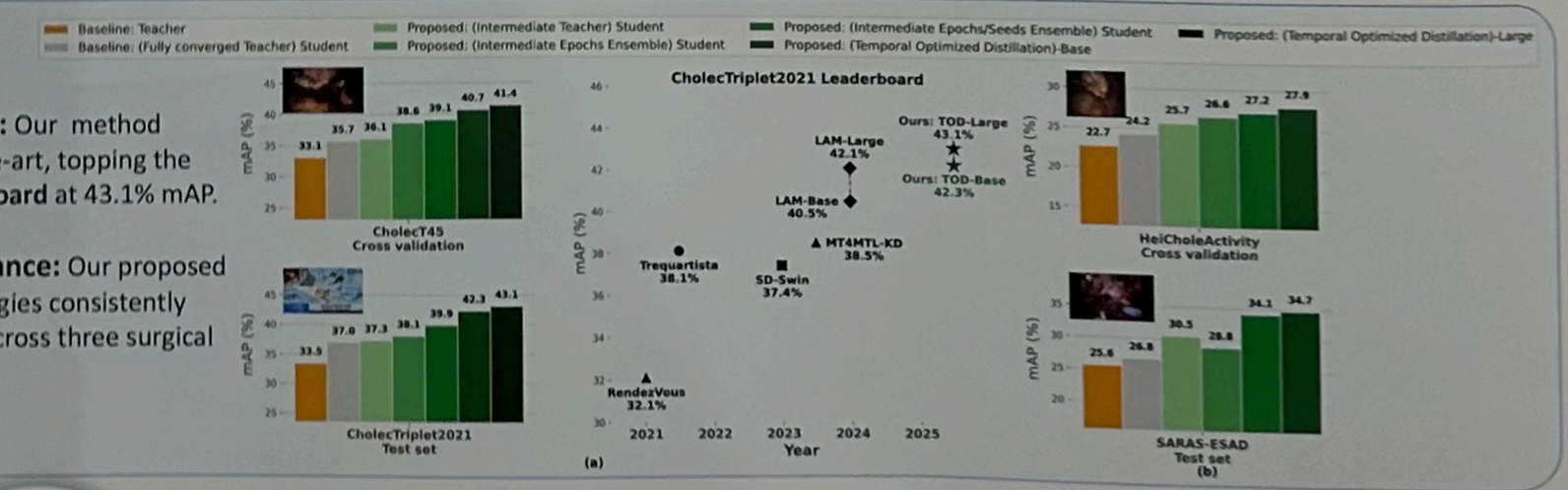
Optimized Teacher Selection for self-distillation:

1. **Intermediate Teacher (IT)**: Best Single intermediate checkpoint
2. **Intermediate Epochs Ensemble (IEE)**: Selected epoch checkpoints from one teacher capture intra-teacher temporal dynamics during training
3. **Intermediate Epochs/Seeds Ensemble (IESE)**: Multi-seed ensemble captures inter-teacher initialization diversity across different starting points

Strategic optimization of teacher selection consistently outperforms single teacher checkpoints

Results

- **State-of-the-Art Results:** Our method achieves a new state-of-the-art, topping the **CholecTriplet2021 leaderboard** at 43.1% mAP.
- **Consistent Outperformance:** Our proposed teacher optimization strategies consistently **outperform all baselines** across three surgical workflow analysis datasets.



Imaging and Radiooncology

IMSY Intelligent Medical Systems
m.elyamlahi@dkfz-heidelberg.de
@DKFZ_IMSY_lab

DECLARATION

Funding was provided by the Helmholtz Association under the joint research school "HIDSS4Health – Helmholtz Information and Data Science School for Health" and the Surgical AI Hub Germany and the National Center for Tumor Diseases (NCT), Heidelberg's Surgical Oncology Program.