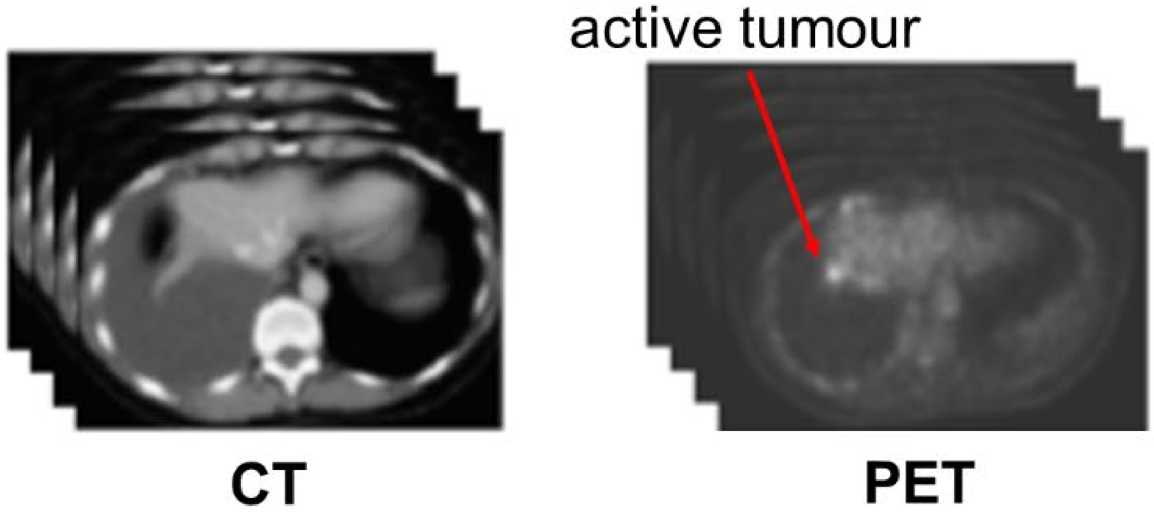
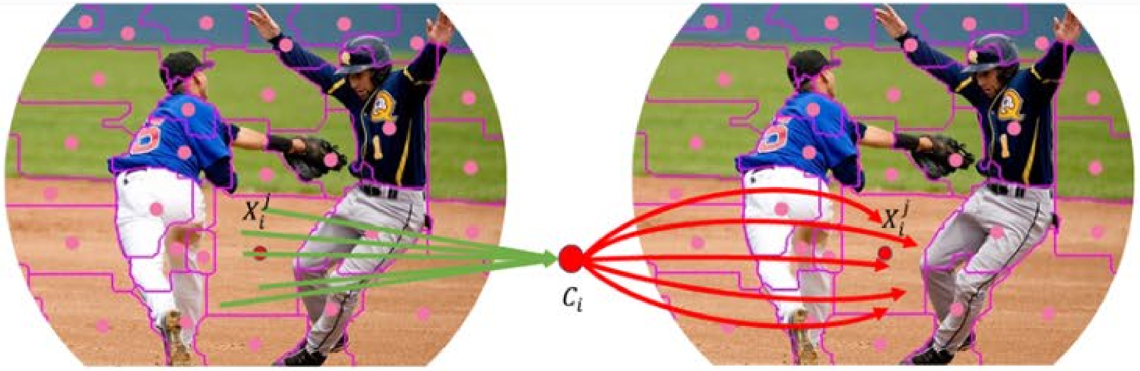
Practical Course: Computer Vision for Human-Computer Interaction

SS 2023
M.Sc. Zdravko Marinov
Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT)

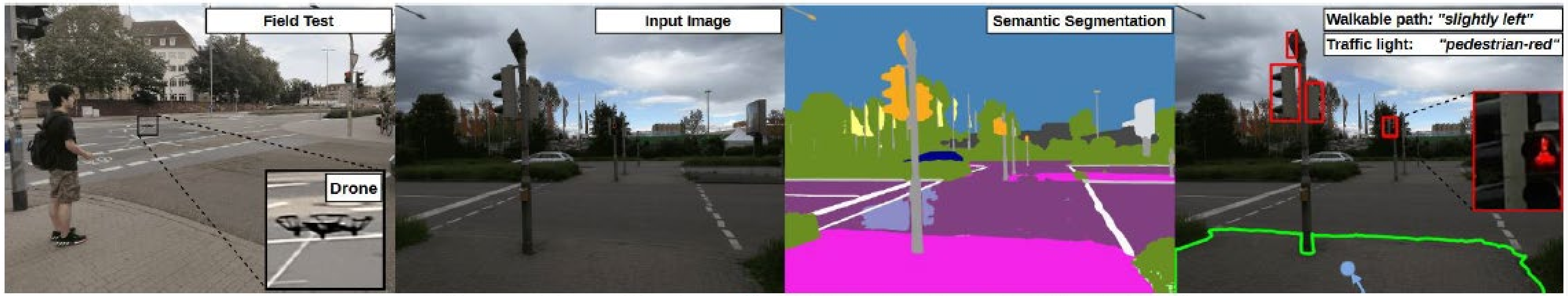




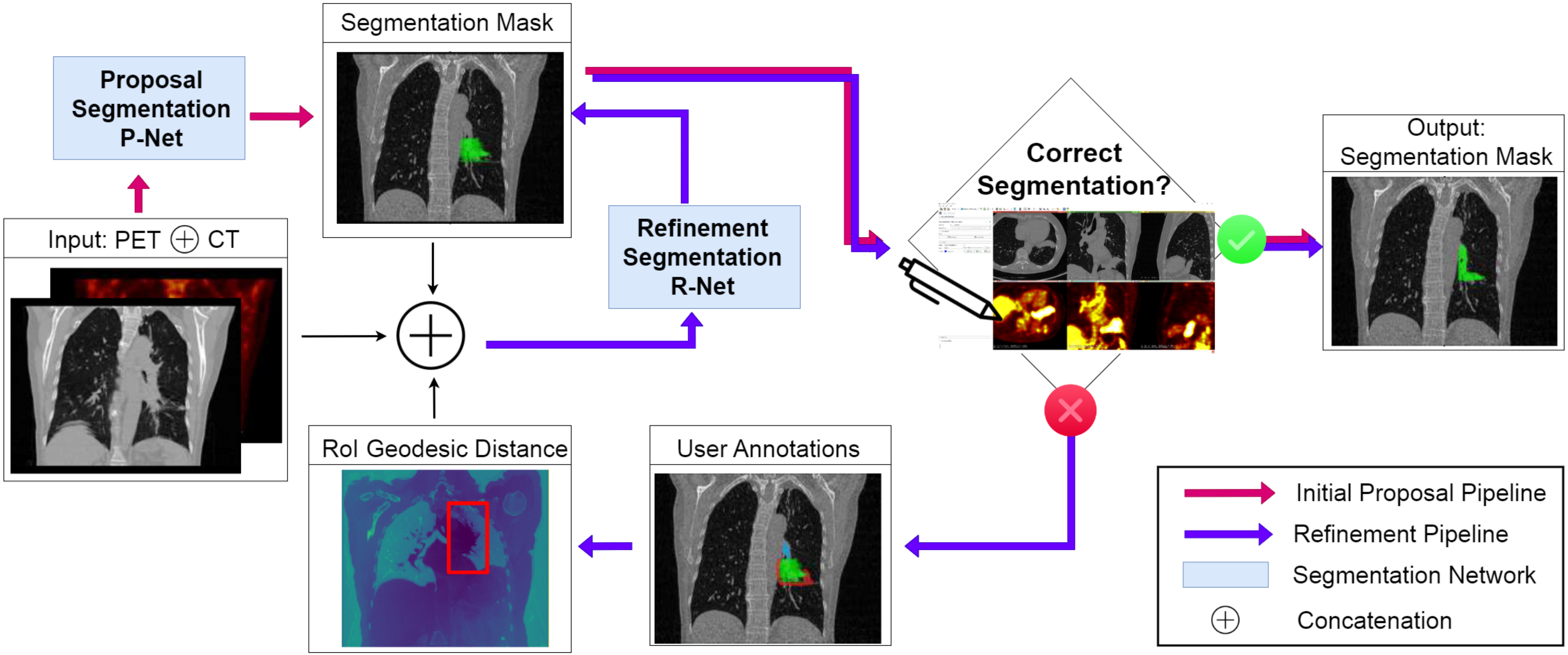
What will you learn?

- Apply algorithms from lectures and papers
- Hands-on experience
- Get comfortable with machine learning tools
- Learn about current problems and applications in machine learning and vision
- Find solutions to difficult problems

Examples from previous semesters: SS21 – Flying Guide Dog, ROBIO 2021



Examples from previous semesters:
SS22 – Interactive PET/CT annotation, ISBI 2023



General Information

Weekly meeting (MS Teams)

- Compulsory Attendance
- Talk about intermediate results & problems
- Ask for help and guidance
- Weekly goal: stay on “track”

3 Students per Team

- Use version control (e.g. git)
- Internal git repos provided via the SCC’s GitLab (https:/git.scc.kit.edu/)
- Divide work into separate tasks and distribute withing group

At the end of the Practical Course…

- Final presentation of each group (1/3 grading)
  - 15 minute talk (5min/student)
  - The presentation should be about:
    - Goals and usefulness of your topic
    - Your proposed approach
    - Results

- Written report describing the topic/approach/results (1/3 grading)
  - 4-pages in standard paper format
    - Abstract/Introduction/Method/Results/Conclusion
    - References do not count in the 4-pages!
    - Written in a conference template

- Working implementations of your algorithms (1/3 grading)
  - A Readme-file describing how the code can be used to reproduce the results
    - If the team agrees → make code publicly available to the community

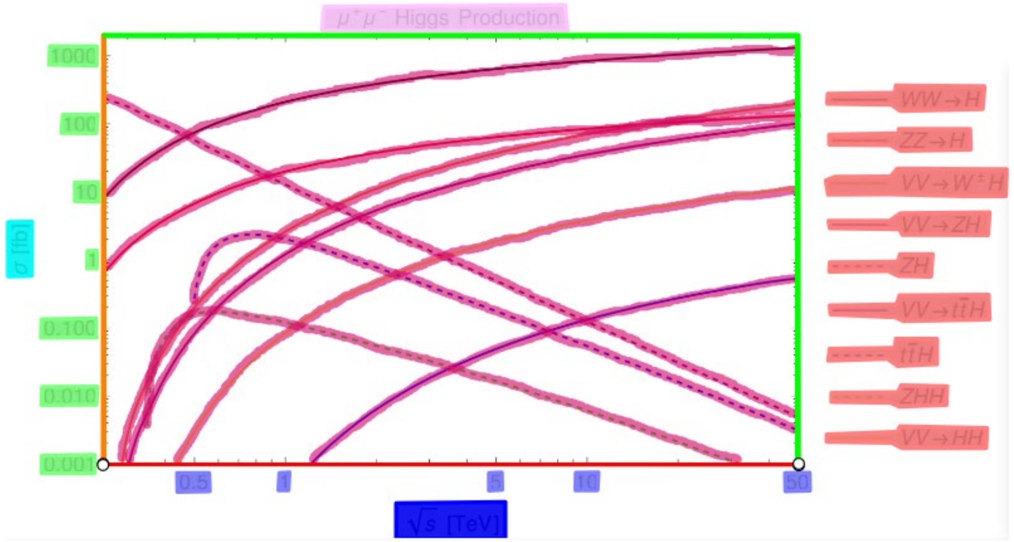
Topics SS 2023

- **A:** DeepTrace: A Transformer-based Line Tracing for Panoptic Segmentation of Line Charts
- **B:** Fall detection with Synthetic Data
- **C:** Action Recognition with Noisy Labels
- **D:** Anatomy-Guided Interactive Segmentation of PET/CT lesions
- **E:** Guided mobile HR-3D-Scanning
- **F:** Context Cluster for panoramic semantic segmentation

DeepTrace: A Transformer-based Line Tracing for Panoptic Segmentation of Line Charts

- By segmenting line plots, visually impaired individuals can more easily access and interpret the information presented in the plot.

- They can focus on individual sections of the plot, rather than interpret the entire plot as a whole.
  - Lines, Labels, Legend etc.



Example of line plot segmentation

DeepTrace: A Transformer-based Line Tracing for Panoptic Segmentation of Line Charts

Task

- Go through literature on current line tracing approaches
  - [4], [5], [6]
- Train Transformer models for Panoptic Segmentation of Line Charts
  - Mask2Former [1] and YOSO [2]

Criteria

- Panoptic Quality [3] and F1-scores [4]

Process

- Literature research
- Training and Evaluating on Datasets:
  - CHART2019-S [7], ExcelChart400K [8], FigureSeer [9], ICPR-2020-CHART-UB PMC [10]

DeepTrace: A Transformer-based Line Tracing for Panoptic Segmentation of Line Charts

ACCESS @KIT

KIT Karlsruher Institut für Technologie

Resources

[1] Mask2former for video instance segmentation. *arXiv 2021* [paper]

[2] You Only Segment Once: Towards Real-Time Panoptic Segmentation. *arXiv* 2023 [paper]

[3] Panoptic segmentation. CVPR 2019 [paper]

[4] Parsing line chart images using linear programming. WACV 2022 [paper]

[5] LineEX: Data Extraction from Scientific Line Charts. WACV 2023 [paper]

[6] Review of chart image detection and classification. IJDAR 2023 [paper]

[7] Competition on HArvesting Raw Tables (CHART) - CHART2019-S_1 dataset [link]

[8] Chartocr: Data extraction from charts images via a deep hybrid framework. WACV 2021 [paper]

[9] Figureseer: Parsing result-figures in research papers. ECCV 2016 [paper]

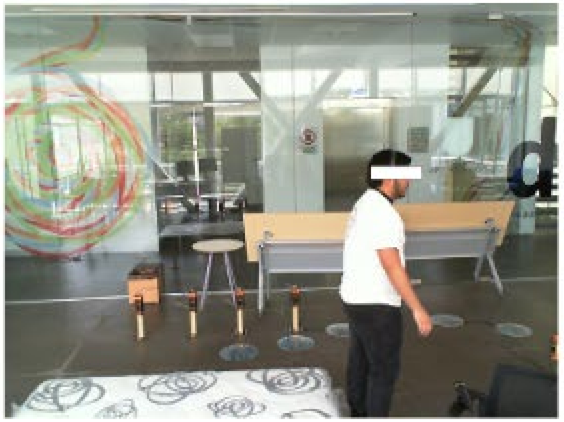
[10] Competition on HArvesting Raw Tables - ICPR-2020-CHARTUB PMC dataset [link]

Fall Detection with Synthetic Data

- Falls are common and especially dangerous for the elderly
- Many public fall detection datasets available
  - However all are very small compared to general activity recognition datasets
  - Transferability to real-world scenarios is questionable

Solution

- Synthesize fall detection datasets to increase the amount of training data and improve fall detection.



Example of a “Fall” video sample [1]

Fall Detection with Synthetic Data

Task

- Collect existing fall detection datasets into a superset
  - Datasets are listed in Table 2 in [1]
- Extract 3D poses from combined superset
- Use 3D poses as input to our provided synthetic data generation engine
  - Create a large synthetic superset
  - Synthetic data generator will be given to you in advance ☺
- Train a state-of-the-art network on fall detection
  - Real-world data only (real superset only)
  - Combination of real and synthetic supersets

Criteria

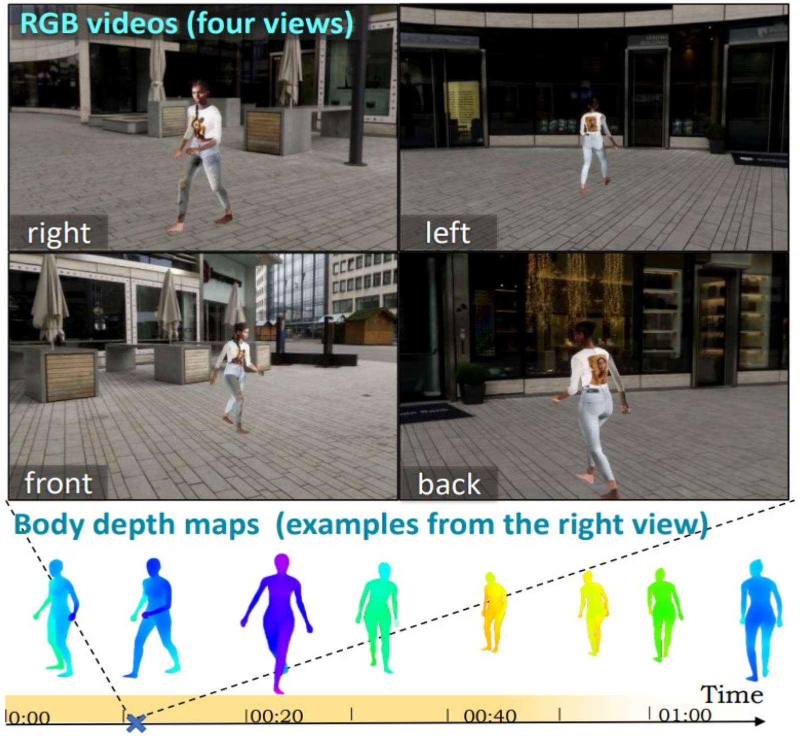
- Accuracy, Sensitivity, Specificity

Process

- Literature research on state-of-the-art fall detection models
- Collection of datasets from [1] into real-world superset
- 3D Pose extraction of real-world superset
- Synthetic superset generation from extracted poses
- Training on (real) and (real + synthetic) supersets

Resources:

- [1] UP-fall detection dataset: A multimodal approach – Sensors 2019 [paper]

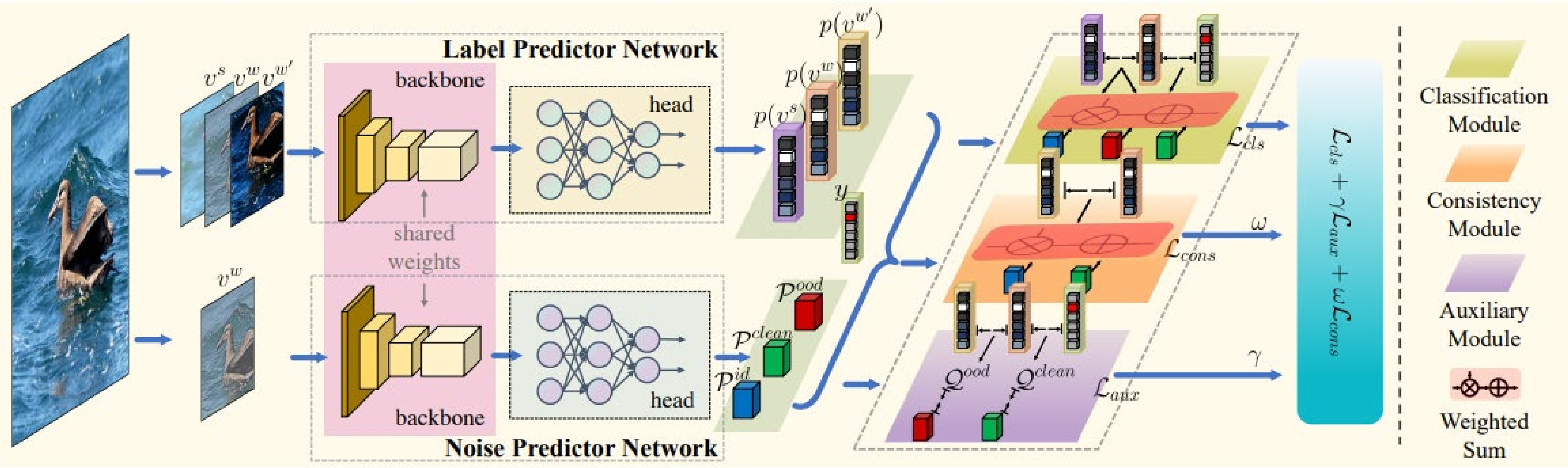


Action Recognition with Noisy Labels

- In action recognition, depth sensors like Kinect can produce noisy or incomplete depth maps due to occlusions, reflections, or environmental factors
- Using perfect depth labels can often lead to a domain shift in the model performance on real-world data

Solution

- Simulate noisy labels during training to make models more robust



PnP: Learning from Noisy Labels [2]

Action Recognition with Noisy Labels

Task

- **Baselines**: Train transformer- ([3], [10]), CNN- ([9], [4]) and skeleton-based approaches ([8], [5]) on the NTU60 dataset [11]
  - Add noise to a portion [**p**]=% of the training labels to simulate noise
  - 6 models in total
- **Qualitative:** Analyze the embeddings in the latent space via t-SNE [7] regarding different models trained with different noisy label ratio **p**
- **Quantitative:** Recognition performance
- Combine the baselines with PnP [2] and Label correction [6] to improve the representation quality and recognition performance

Criteria

- Representation quality (distinct clusters)
- Recognition performance (Accuracy, Sensitivity, Specificity)

Process

- Literature Research on Noisy Labels [2], [6]
- Model training on NTU60 [11]
- Model improvement using PnP [2] and Label correction [6]
- Qualitative and Quantitative evaluation

Action Recognition with Noisy Labels

Resources

[1] Learning from noisy labels with deep neural networks: A survey. IEEE Transactions on Neural Networks and Learning Systems, 2022 [paper]

[2] Pnp: Robust learning from noisy labels by probabilistic noise prediction. CVPR 2022 [paper]

[3] Mvitv2: Improved multiscale vision transformers for classification and detection. CVPR 2022 [paper]

[4] X3d: Expanding architectures for efficient video recognition. CVPR 2020 [paper]

[5] Hierarchically Decomposed Graph Convolutional Networks for Skeleton-Based Action Recognition. arXiv 2022 [paper]

[6] Adaptive early-learning correction for segmentation from noisy annotations. CVPR 2022 [paper]

[7] Visualizing data using t-SNE. Journal of machine learning research, 2008 [paper]

[8] Stacked spatio-temporal graph convolutional networks for action segmentation. WACV 2020 [paper]

[9] Quo vadis, action recognition? a new model and the kinetics dataset. CVPR 2017 [paper]

[10] Is space-time attention all you need for video understanding. ICML 2021 [paper]

[11] Ntu rgb+ d: A large scale dataset for 3d human activity analysis. CVPR 2016 [paper]

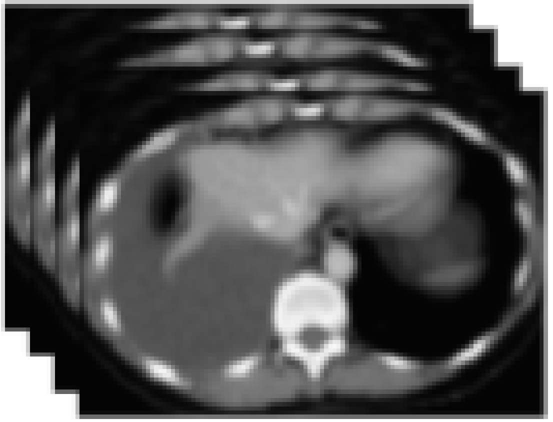
Anatomy-Guided Interactive Segmentation of PET/CT lesions

University Medicine Essen
Institute for Artificial Intelligence in Medicine

cv:hci @KIT

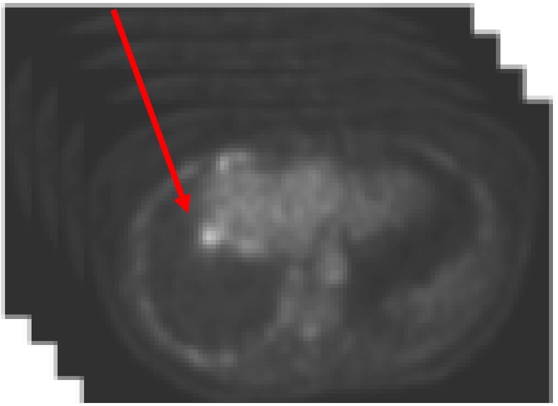
KIT
Karlsruher Institut für Technologie

- PET is an imaging technique in nuclear medicine
  - Often in combination with CT scans
- PET can visualize physiological activities like tumour metabolism
- CT utilizes X-rays from multiple angles to visualize the patient body composition, i.e., anatomy



CT

active tumour



PET

Anatomy-Guided Interactive Segmentation of PET/CT lesions

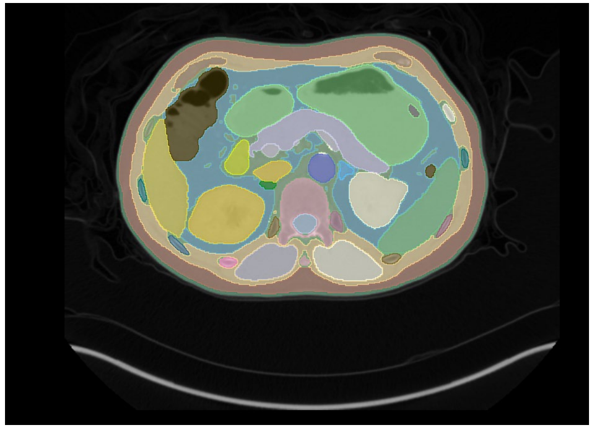
University Medicine Essen
Institute for Artificial Intelligence in Medicine

cv:hci @KIT

Karlsruher Institut für Technologie

Task

- Interactive segmentation using anatomical labels as additional input
- Train a DeepEdit [1] interactive model with clicks
  - 3 channel-input: PET, CT, and anatomical labels (organs and tissues)
- Visualize affected anatomy: size and location of lesion
  - Example: 3ml lesions in the left lung
- Evaluate with user study in collaboration with University Clinic Essen.



Anatomical labels: Different colours correspond to different anatomical structures (organs, tissues)

Criteria

- User annotation time, Dice score

Process

- Train DeepEdit [1] on the AutoPET dataset [2] utilizing anatomical labels
- Integrate model with MONAI Label [3] into 3D Slicer [4]
- Compare to other interactive approaches
  - GrabCut [5], DeepIGeoS [6] (Praktikum SS22, ISBI), DeepEdit [1] without anatomical labels

[3] MONAI+

[4] 3D Slicer

Anatomy-Guided Interactive Segmentation of PET/CT lesions

University Medicine Essen
Institute for Artificial Intelligence in Medicine

cv:hci @KIT

Karlsruher Institut für Technologie

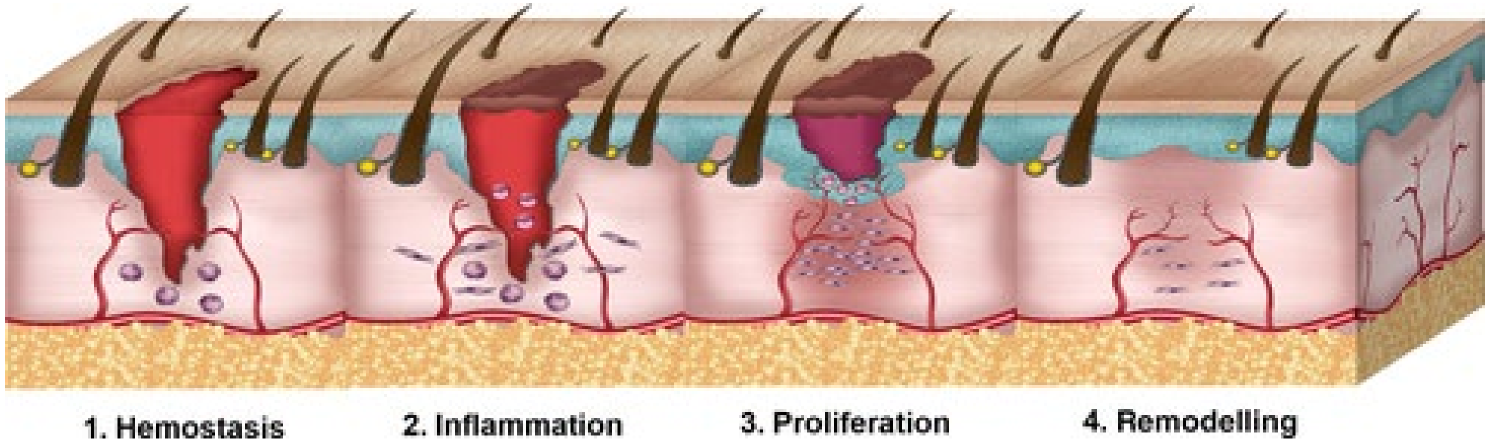
Resources

- [1] DeepEdit: Deep Editable Learning for Interactive Segmentation of 3D Medical Images. MICCAIw 2022 [paper]
- [2] A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. Nature Scientific Data 2022 [paper]
- [3] MONAI Label [link]
- [4] 3D Slicer [link]
- [5] GrabCut" interactive foreground extraction using iterated graph cuts. TOG ACM 2004 [paper]
- [6] DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. T-PAMI 2018 [paper]

- Demo videos for interactive segmentation: [video_1] [video_2]

Guided mobile HR-3D-Scans of Wounds

- Chronic open wounds occur due to diseases (diabetes) or trauma (burns, infection)
  - Take a long time to heal
  - Must be monitored for progress
- Wound surface structure is a characteristic metric for improvement
  - Can be obtained through depth information (e.g. LiDAR)



Four stages of wound healing [1]

[1] Emerging treatment strategies in wound care. International Wound Journal (2022)

Guided mobile HR-3D-Scans of Wounds

Task

- A cube (3D-box) is defined by a medical doctor to describe the size and orientation of the **[wound]**
- Guide a mobile device (iPhone 14 Pro) to an optimal position to obtain the **[wound]** surface
  - Issue commands like “Get closer” or “Move the phone to the left”
- Obtain a true-to-scale 3D model of the **[wound]**
  - Using the Apple Object Capture API: https://developer.apple.com/augmented-reality/object-capture/
- **[wound]** will be initially replaced by a proxy task
  - **[ear], [nose],** other body parts, or everyday objects **[coffee mug]**
  - Once the workflow is implemented, the user guidance can/will be tested on real patients with wounds

Criteria

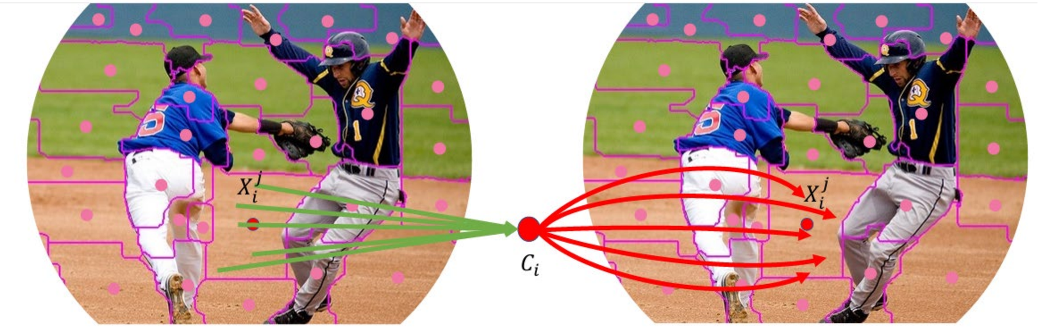
- Qualitative evaluation of the 3D model
- Possible user study on usability

Process

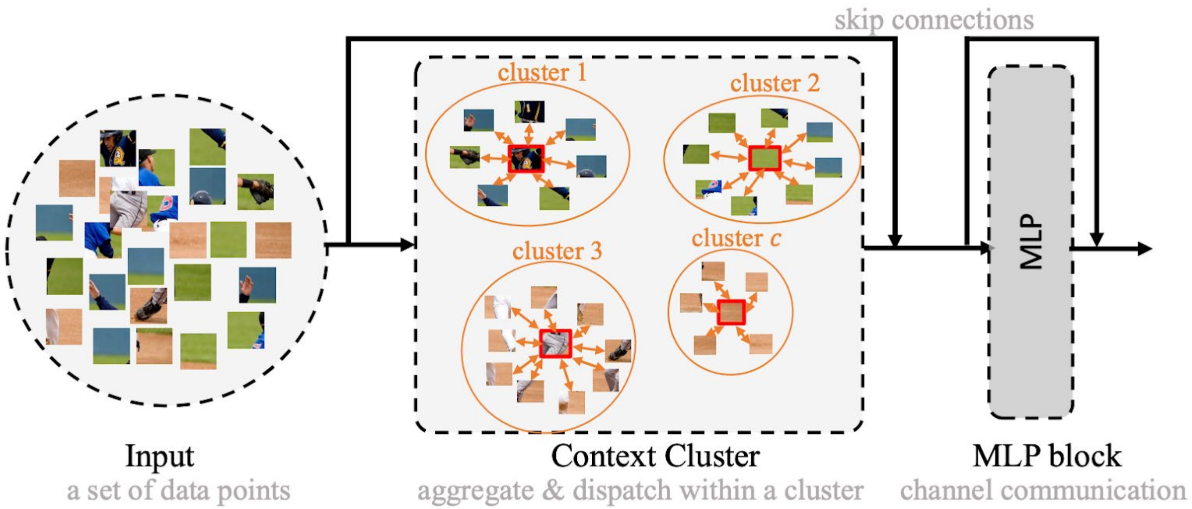
- Get to know the iPhone 14 Pro, Apple Object Capture API, and define a meaningful proxy task
  - Collaboration with MDI
- Implement the user guidance with the proxy task
- Real test on proxy objects and wounds

Context Cluster for panoramic semantic segmentation

- **Context clusters** (CoCs) view an image as a set of unorganized points
  - A point includes the raw feature (e.g., color) and positional information (e.g., coordinates)
  - A clustering algorithm is employed to group and extract deep features hierarchically



Context cluster points [1]



Context cluster block [1]

Context Cluster for panoramic semantic segmentation

Task

- Understand and implement the Context Cluster (CoC) block [1]
- Propose an architecture for panoramic semantic segmentation with CoC block, since CoC works well on irregular images
  - Panoramas, fisheye, etc.
- Train and Evaluate on Standfor2D3D [3] and DensePASS [4] datasets
- Visualize results and compare to previous work [2]

Criteria

- mIoU, IoU per class

Process

- Literature Research on Context Clusters (CoC) [1], [5], [6]
- Implementation and Training of CoC-based models on Standfor2D3D [3] and DensePASS [4]
- Comparison to previous work [2] and documentation of results

Context Cluster for panoramic semantic segmentation

Resources

[1] Image as Set of Points. arXiv 2023 [paper]

[2] Bending Reality: Distortion-aware Transformers for Adapting to Panoramic Semantic Segmentation. CVPR 2022 [paper]

[3] Joint 2D-3D-Semantic Data for Indoor Scene Understanding. arXiv 2017 [paper]

[4] DensePASS: Dense Panoramic Semantic Segmentation via Unsupervised Domain Adaptation with Attention-Augmented Context Exchange. ITSC 2021 [paper]

[5] Attention-based Point Cloud Edge Sampling. CVPR 2023 [paper]

[6] Parameter is Not All You Need: Starting from Non-Parametric Networks for 3D Point Cloud Analysis. arXiv 2023 [paper]

Topic Selection

Karlsruher Institut für Technologie

- Find a team of three people (i.e. through the MS-Teams chat)
- Each team sends us a ranking of the presented topics until 23rd 23:59 of April per Email at zdravko.marinov@kit.edu (1 – most preferred; 6 – least preferred)
  - Example: A2, B4, C1, D3, E5, F6

- If you cannot find a team, you can also send personal preferences

- Students will be assigned to the respective topics based on their preferences and the order of registration

Organization

- Meeting schedule (Potential process)
  - Week 0 [17.04.23]: Introduction and topic selection
  - Week 1: Read related work and present ideas on how to approach the problem
  - Week 2: Implementation
  - …
  - Week 15 [24.07.23] (Monday 14:00-15:30): Final Presentations
- Weekly meeting for discussion and status updates with corresponding supervisor
  - Set a consistent date for weekly meetings
- Register Projektpraktikum with KIT's Studienbüro (Modulhandbuch M-INFO-102966, Teilleistung T-INFO 105943)
- For these slides, other information, announcements and updates → check website [coursemember/321meins] and MS Teams