



MEDIZINISCHE
FAKULTÄT
HEIDELBERG

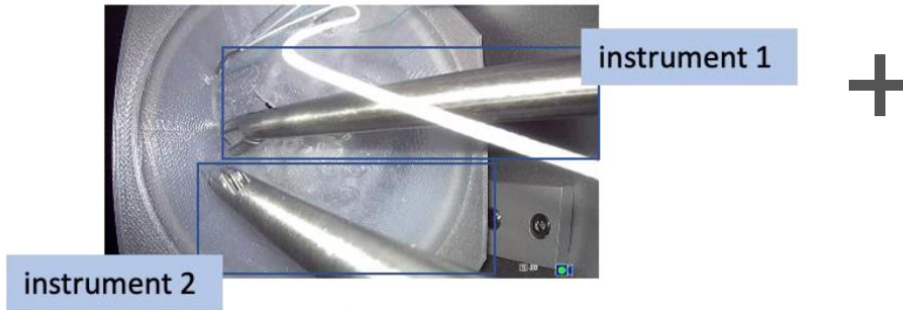
Multi-instrument detection for stereo-endoscopy in heart surgeries

Sommersemester 2022

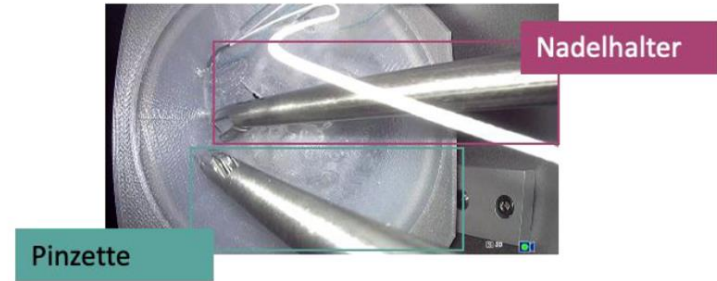
Bastian Westerheide, Lisa Wimmer, Nina Stegmayer, Moritz Bednorz und Nils-Christian Iseke

Aufgabenstellung / Ziele

Multi-instance detection
Wieviele Instrumente ?



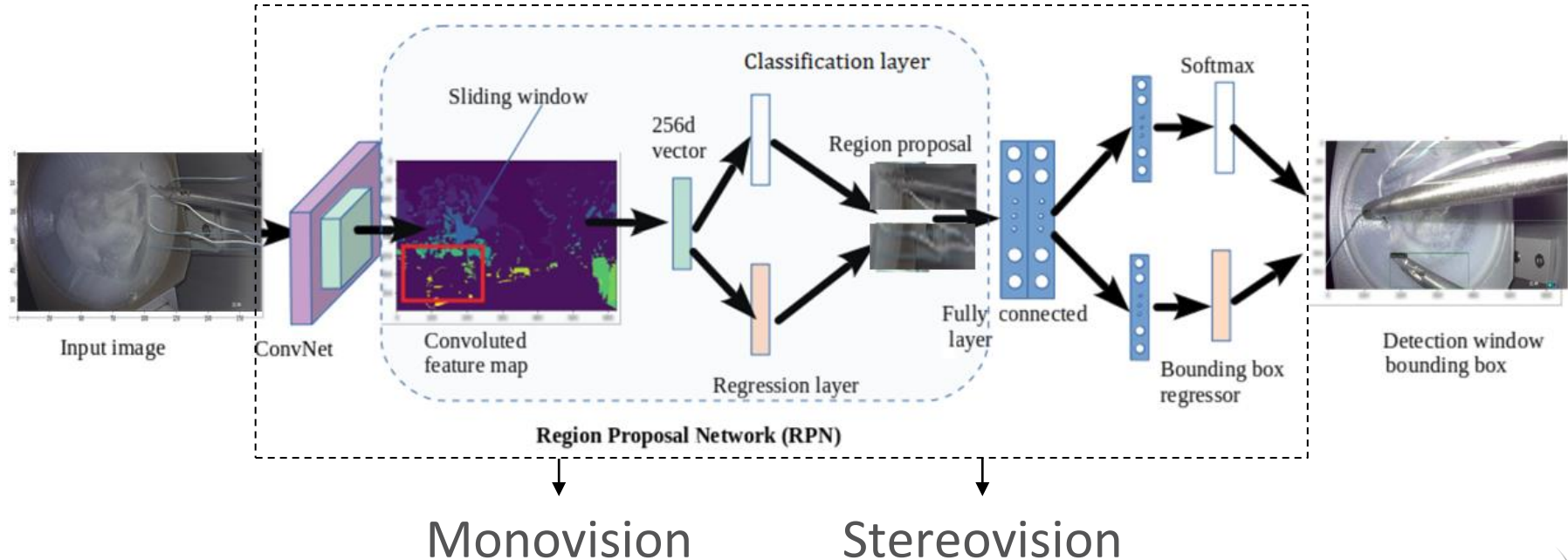
Source: wft_ss2022_session1_overview.pptx | Lalith Sharan



Multi-class detection
Welche Instrumente ?

Aufgabenstellung / Ziele

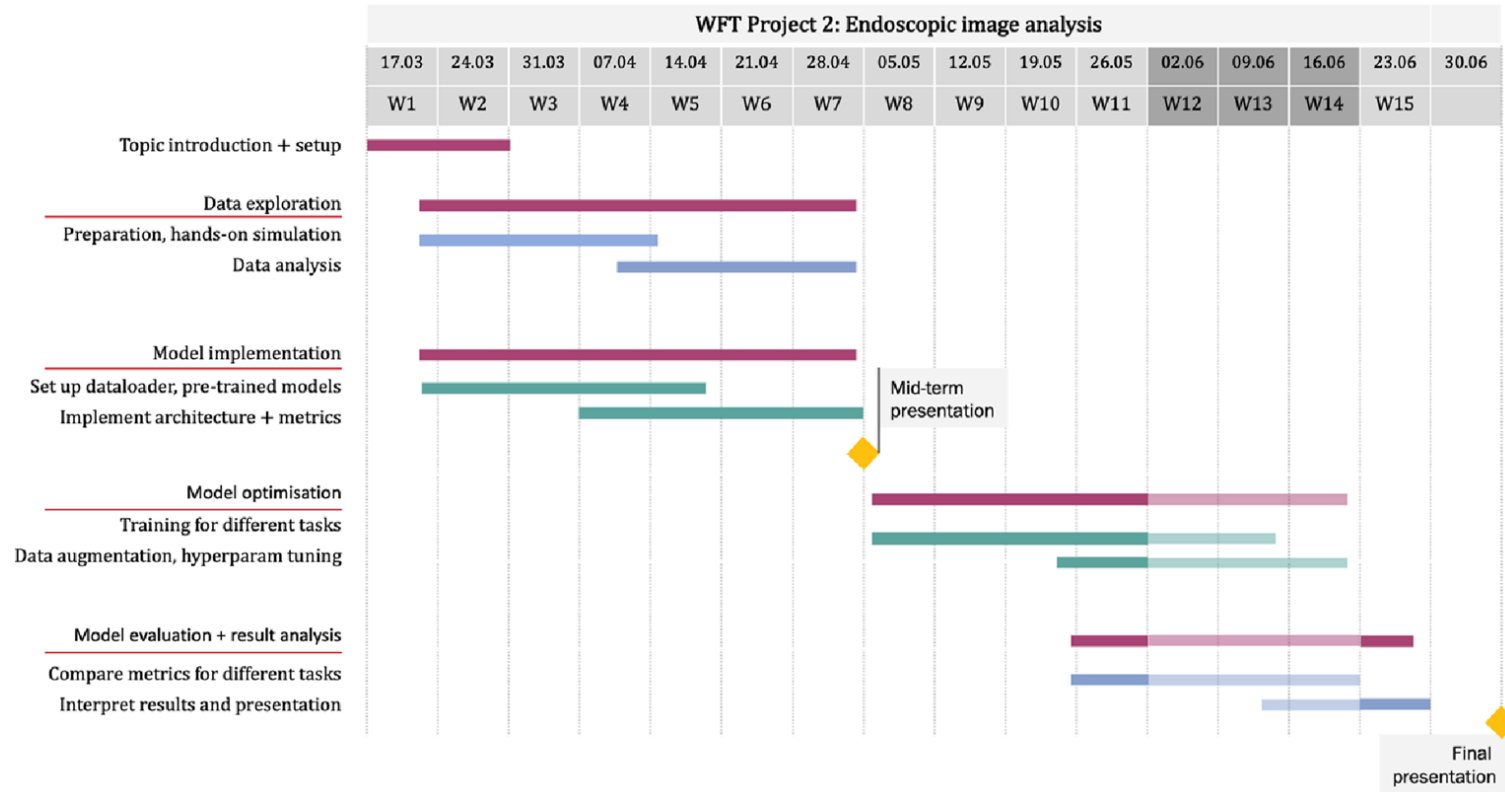
Faster R-CNN



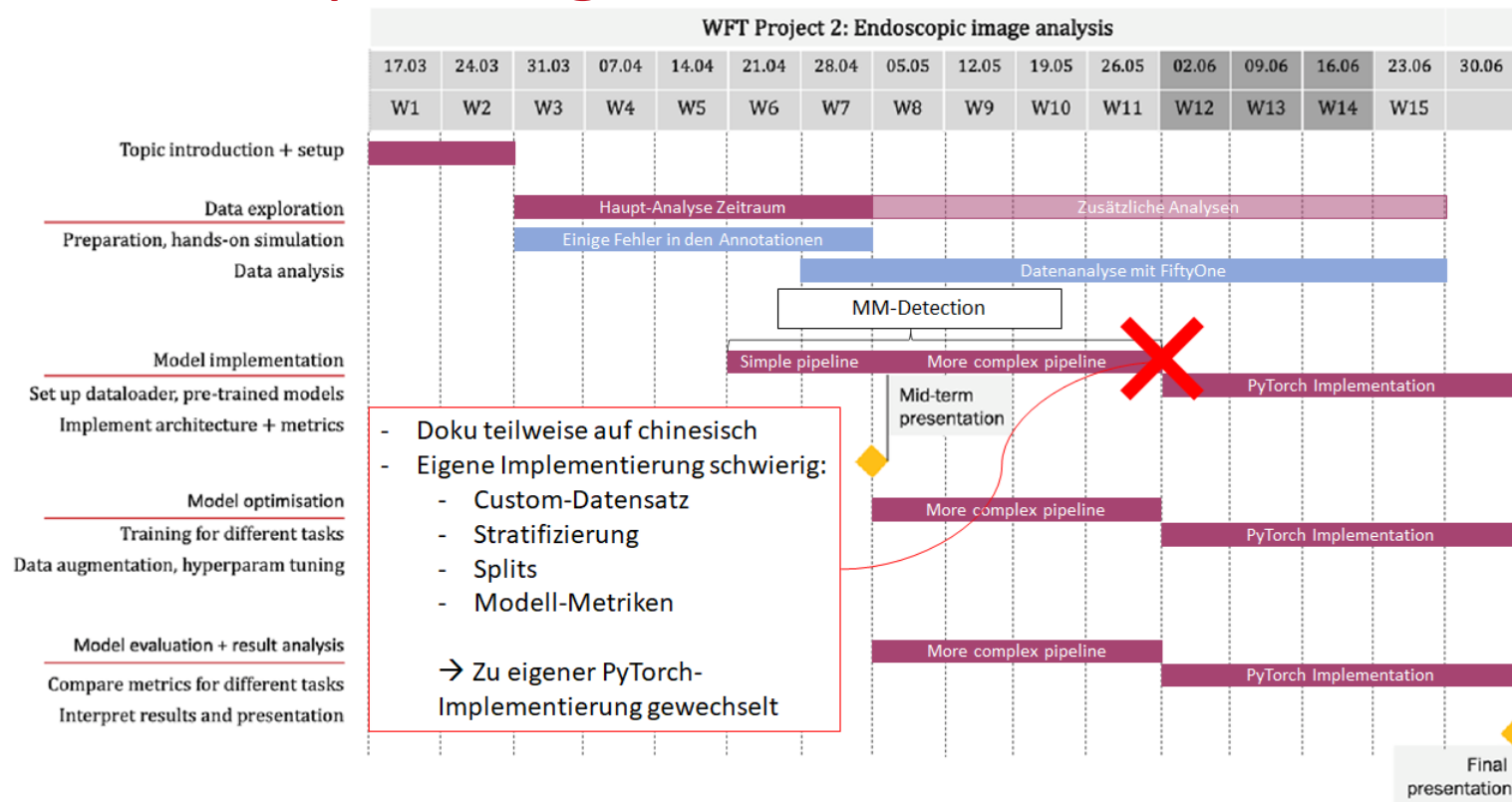
Source: <https://subscription.packtpub.com/book/data/9781838827069/7/ch07lvl1sec34/overview-of-faster-r-cnn>

Medizinische Fakultät Heidelberg | Juni 2022 | Jun. Prof. Dr. Sandy Engelhardt & Lalith Sharan

Meilensteinplanung



Meilensteinplanung

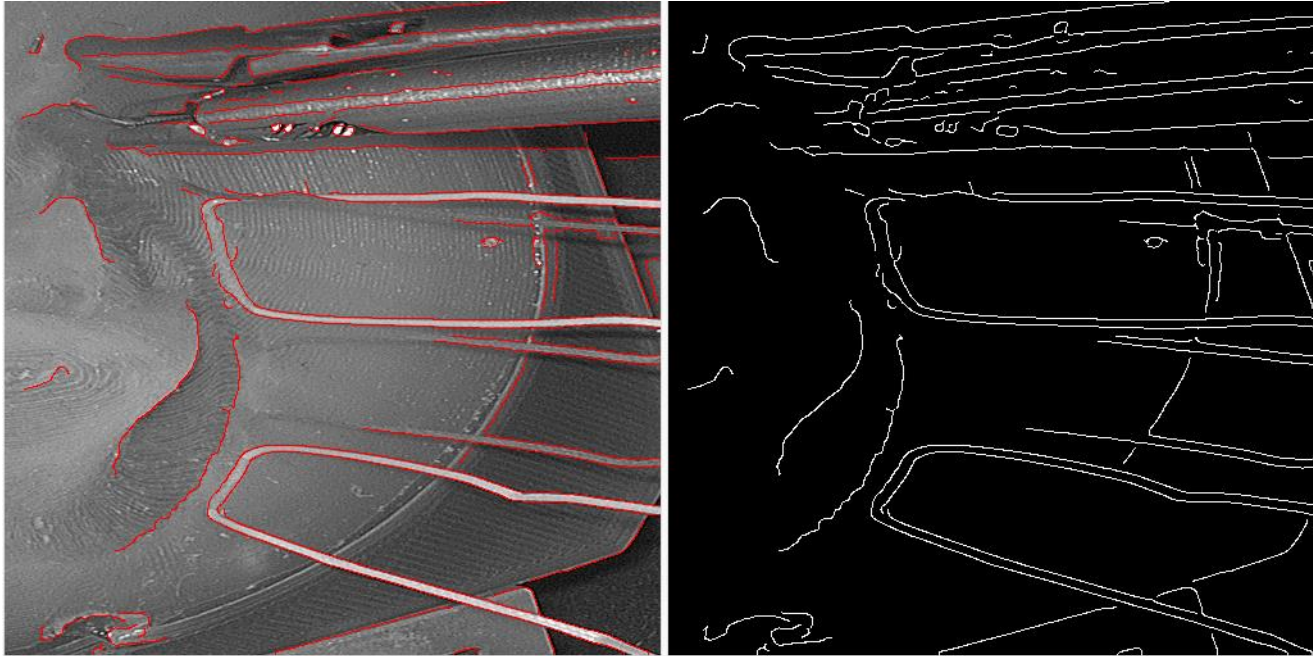


Aufgabenverteilung

- | | | | | | |
|---|---|---|---|---|---|
| 1. Annotation der Bilder | ● | ● | ● | ● | ● |
| a. Überprüfung der Annotierungen | | | | ● | ● |
| b. Annotierungsprozess (CVAT ,Fiftyone) | | | | ● | ● |
| 2. Datenanalyse und Statistik | | | ● | ● | |
| 3. Monovision Pipeline mit MMDetection | ● | ● | ● | ● | ● |
| a. CustomDataset, Augmentation, Evaluierung | | | | | ● |
| 4. Dataloader Implementierung | ● | ● | ● | | |
| 5. Monovision Model Implementierung | ● | ● | ● | | |
| 6. Stereovision Model Implementierung | | | ● | | |
| 7. Evaluierung | ● | ● | ● | | |
| 8. Dokumentation | | | | ● | ● |

● Wimmer ● Stegmayer ● Westerheide ● Iseke ● Bednorz

Motivation - Einsatz von Deep Learning



Annotierung



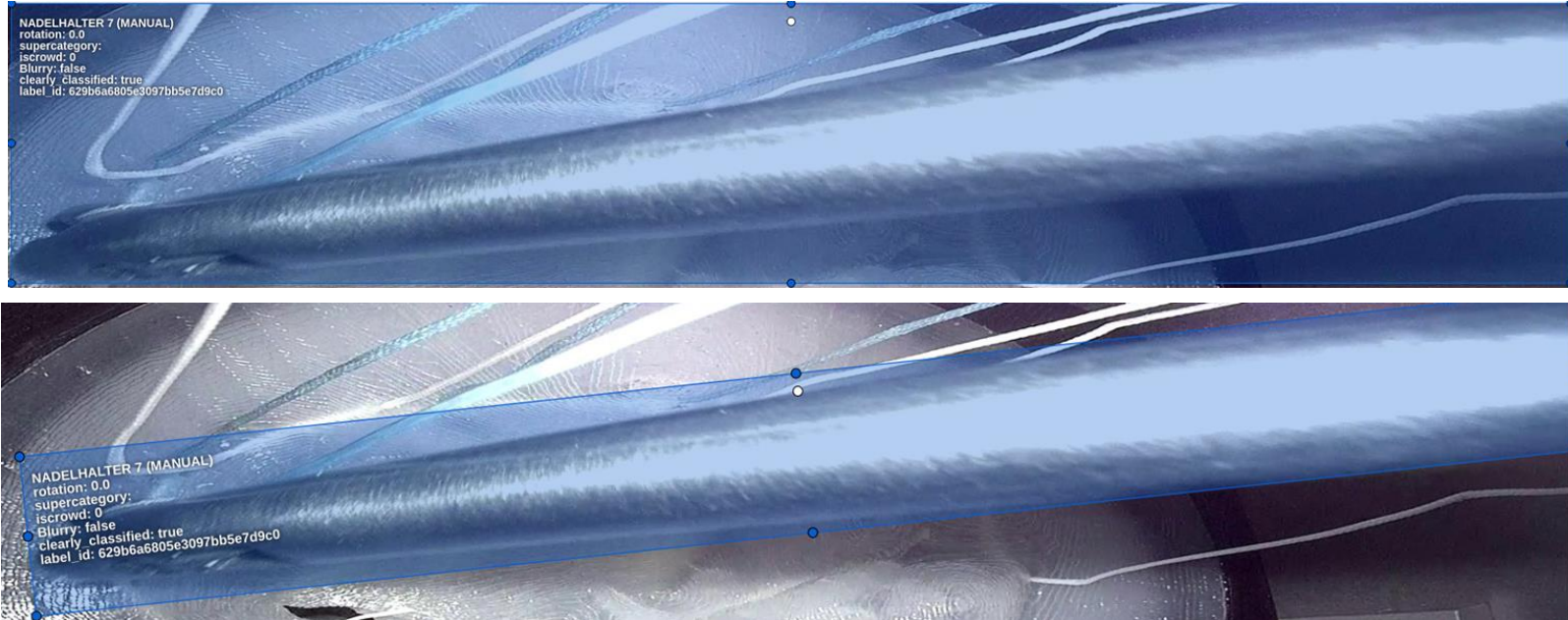
- Schlechte Annotierung → Schlechtes Ergebnis
- CVAT (Computer Vision Annotation Tool), lokal gehostet
- Wie sorgen wir für eine einheitliche Annotation?
 - Festlegung von Annotierungsstandards
 - “6-Augen-Prinzip”

Annotierung

- Welche Informationen stecken in diesem Bild?

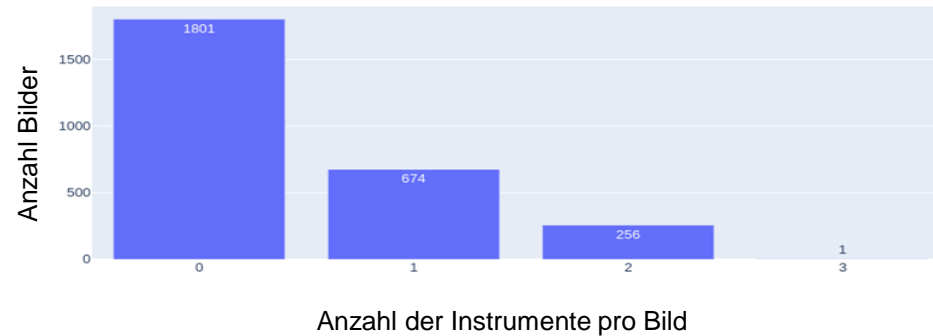
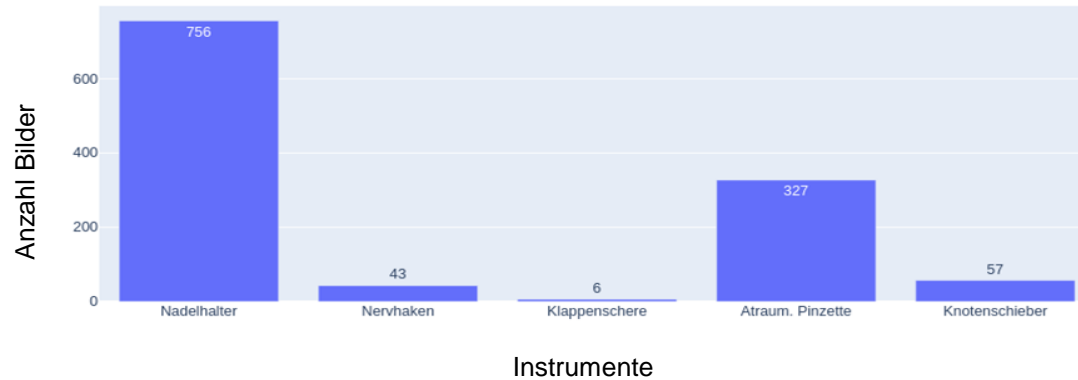


Annotierung - Rotierte Bounding Boxes

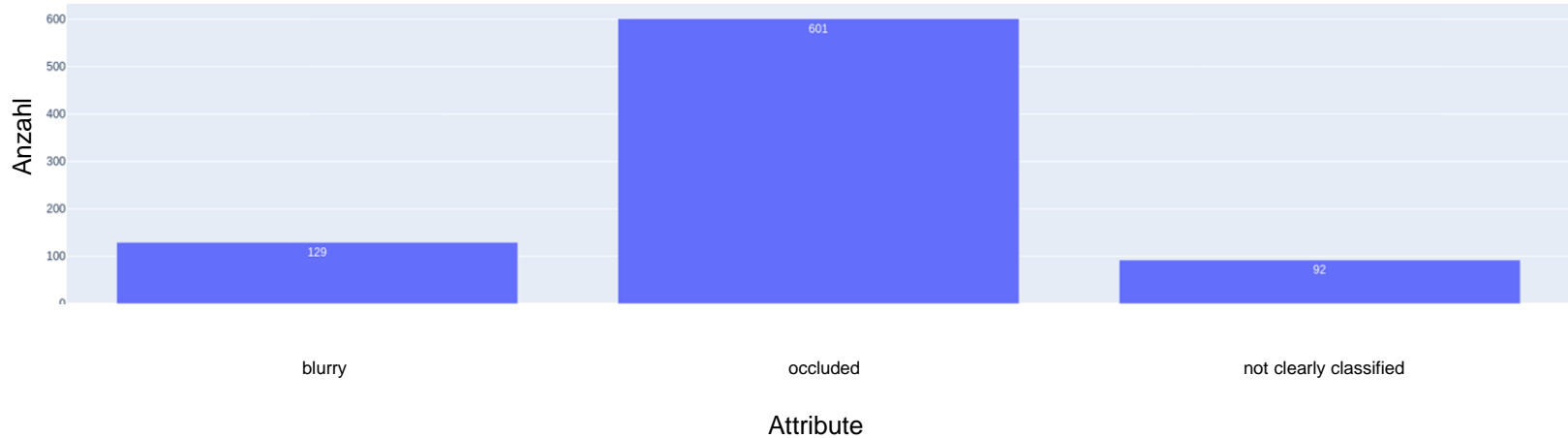


Toy-Datensatz im SDS: toy_rotated_bb

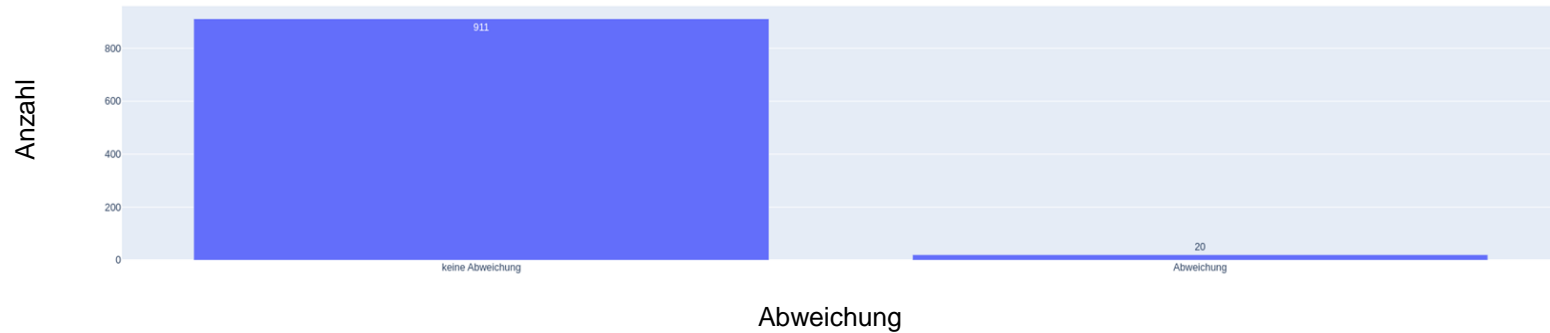
Datenauswertung



Datenauswertung Attribute



Unterschiede zwischen Stereodaten



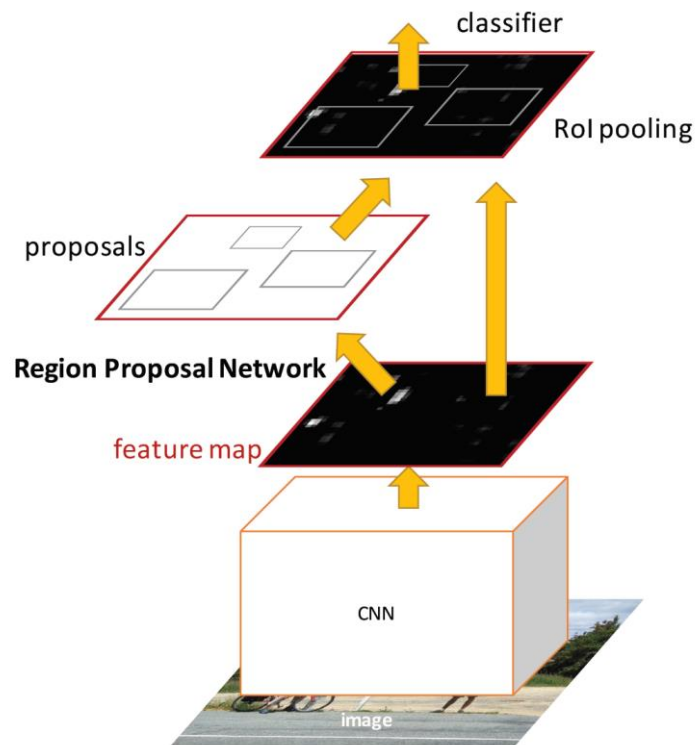
Umsetzung - Data Pipeline

- Augmentierung der Trainingsbilder
 - Formatieren der Bounding Boxes
 - Umgang mit Bildern ohne Bounding Boxes
-
- Für Stereo: Zusammenlegen der Bilder beider Kameraperspektiven



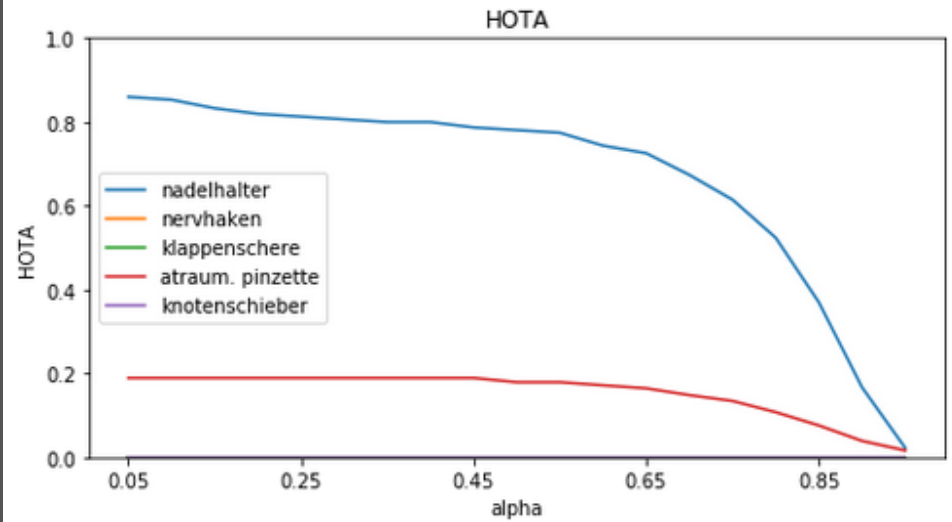
Umsetzung - Model Implementation

- Pretrained Faster-RCNN Modell
- ResNet50 Backbone
 - Mono: 3 Channel
 - Stereo: 6 Channel
- Output:
 - Bounding Boxen
 - Labels
 - Confidence score
- Non-Maximum-Supression

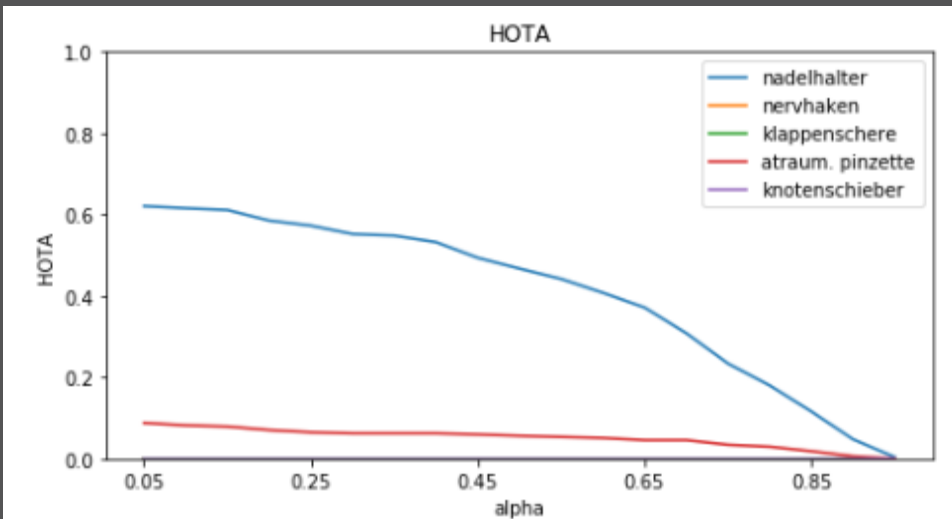


Resultate - HOTA metric

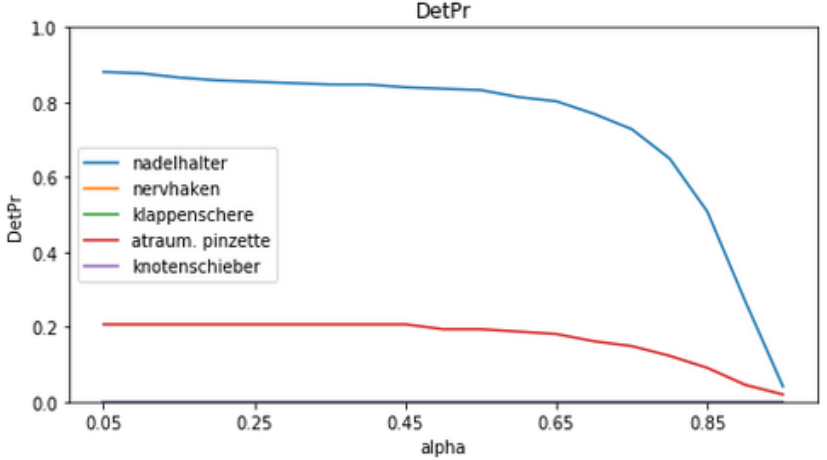
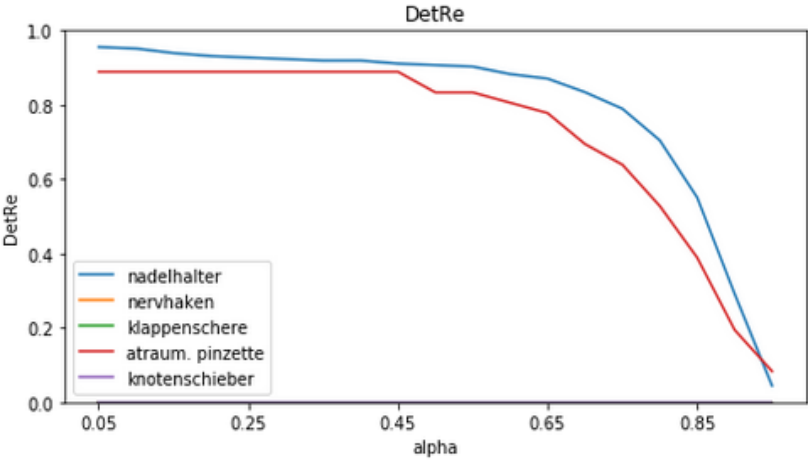
Mono Modell 100 Epochen



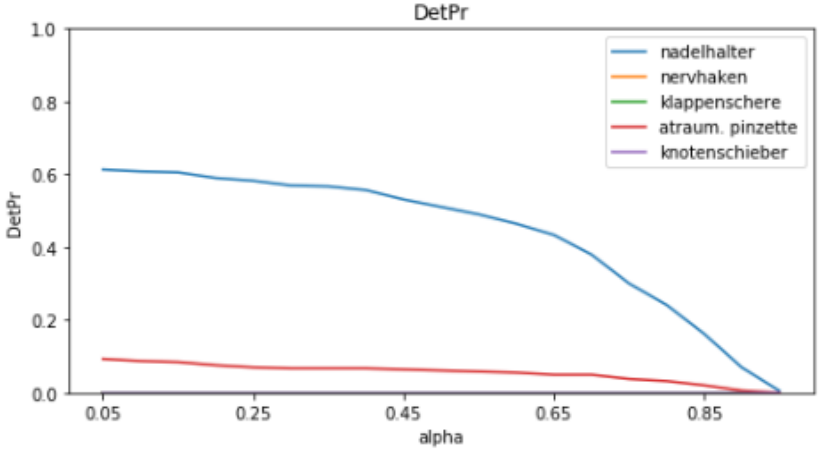
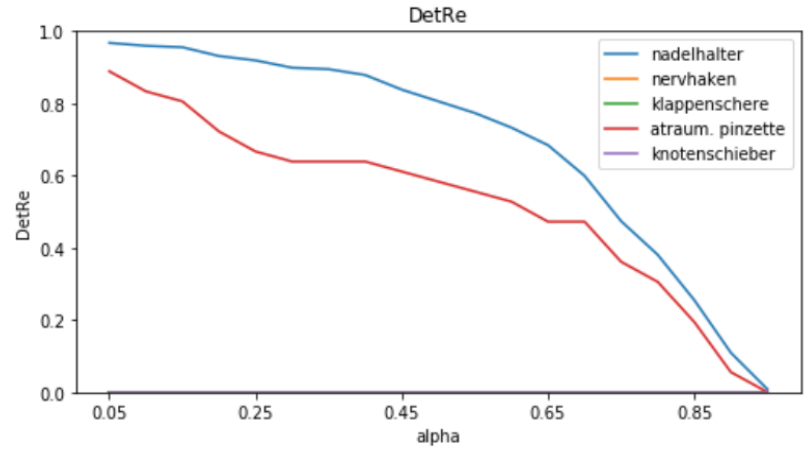
Stereo Modell 100 Epochen



mono-
Modell



stereo-
Modell

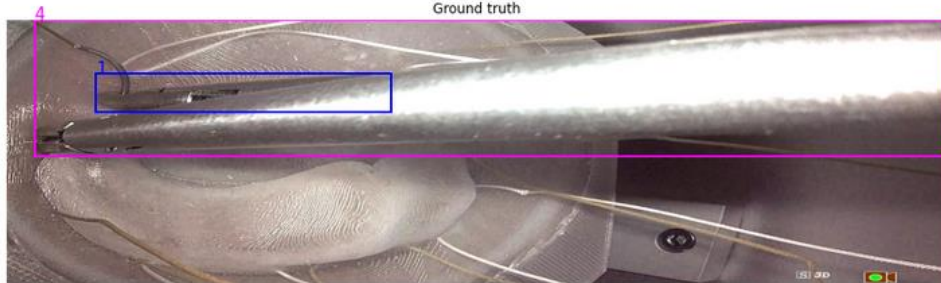


Herausforderungen

- Zwei Instrumente nebeneinander

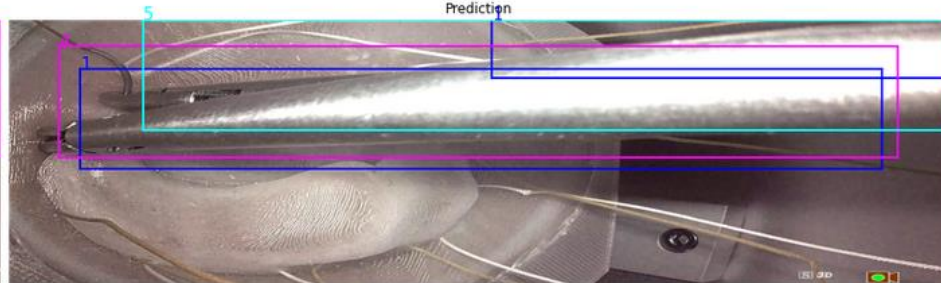
06_03_000131

Ground truth



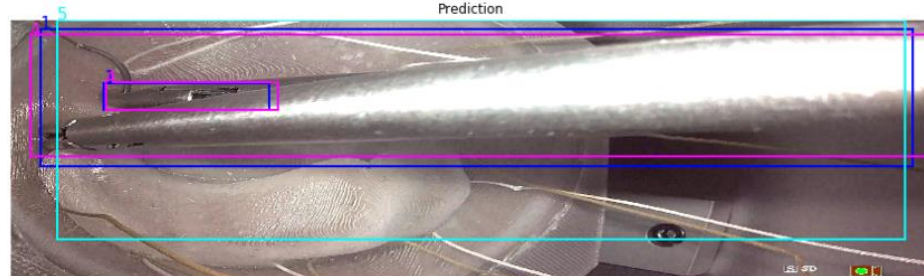
mono-Modell

Prediction



stereo-Modell

Prediction

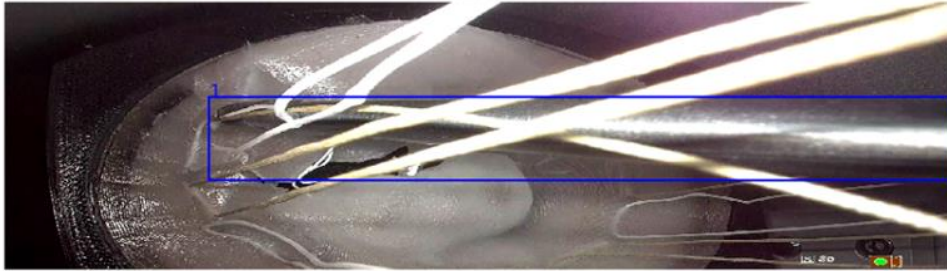


Herausforderungen

- Viele Fäden

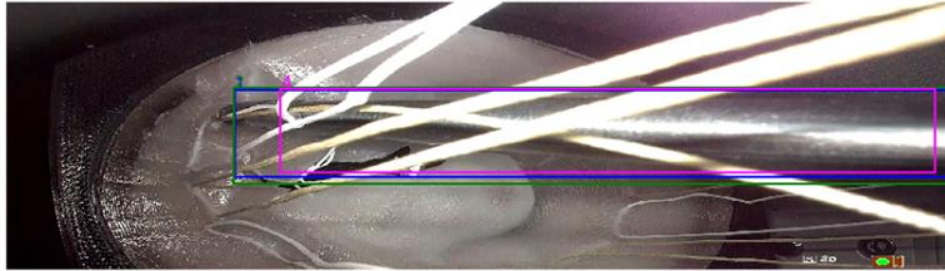
09_02_000044

Ground truth



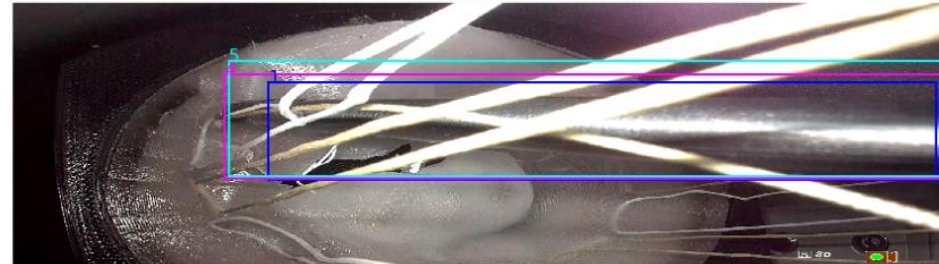
mono-Modell

Prediction



stereo-Modell

Prediction

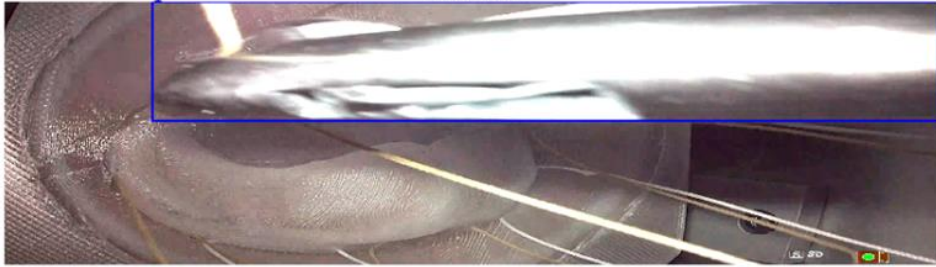


Herausforderungen

- Verschwommene Instrumente

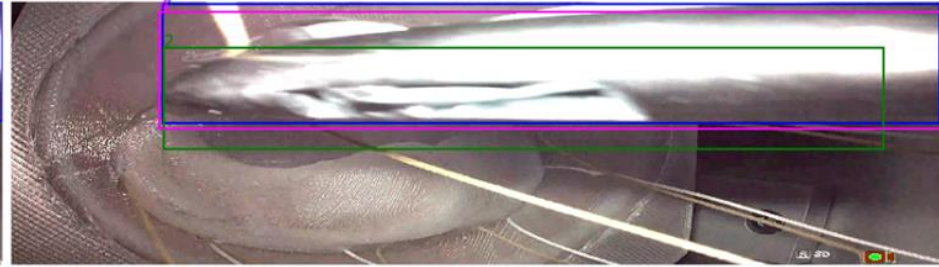
06_02_000149

Ground truth



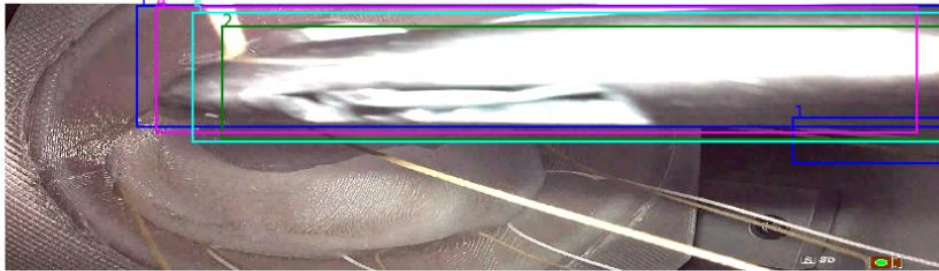
mono-Modell

Prediction



stereo-Modell

Prediction



Resultate - Dokumentation



- prozessierte Daten: SDS der Uni Heidelberg
- trainierte Modelle: SDS der Uni Heidelberg
- Pfad zum Repository: https://github.com/Cardio-AI/wft_ss22_multi_instrument_detection

Limitationen

- Stereomodell ist nicht wirklich spezialisiertes Stereomodell
 - kaum Paper gefunden
- Klassen sind sehr unbalanciert
- Bisher wenig Modelloptimierung

Herausforderungen

- Labeln der Daten
- labelMe
- bwVisu
- Zeitmangel durch Änderung des Lösungsansatzes
- viele leere Bilder

Verbesserungsmöglichkeiten

- stabilere Verbindung zu bwVisu
- Schreibrechte auf SDS
- evt. CVAT statt LabelMe
- Stereodaten → Kamerakonfiguration nutzen
- temporale Daten fürs Training nutzen
- rotierte Bounding Boxen nutzen?
- Modelle verbessern