The RoDEM Benchmark: Evaluating the Robustness of Monocular Single-shot Depth Estimation Methods in Minimally-Invasive Surgery

Rasoul Sharifian[♁], *Navid Rabbani*[♁] *and Adrien Bartoli*

EnCoV, Institut Pascal, UMR 6602 CNRS/UCA, DIA2M, Clermont-Ferrand University Hospital, France, SURGAR, Surgical Augmented Reality, Clermont-Ferrand, France

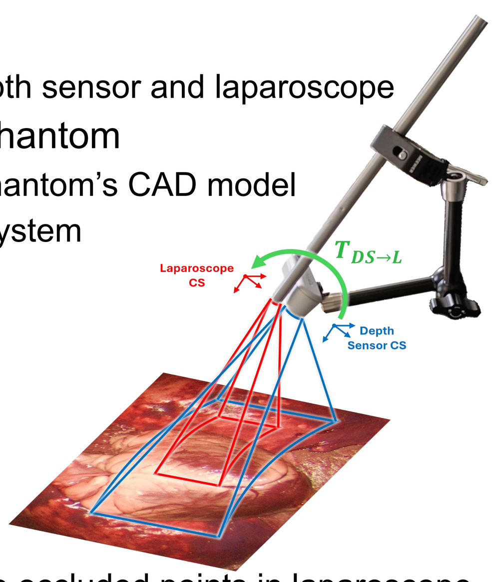
Overview

- **Monocular Single-shot Depth Estimation (MoSDE)**
  - MoSDE methods estimate depth from single RGB images, without relying on stereo pairs or temporal cues
  - Learning-based MoSDE methods have shown promising results for urban scenarios, motivating their adaptation to MIS
- **Effective Benchmark for MoSDE Methods**
  An effective benchmark should:
  - Evaluate the performance under clean MIS conditions
  - Assess robustness to perturbations: specific domain conditions which complicate depth estimation
- **RoDEM Benchmark:**
  RoDEM introduces three essential components for robust evaluation:
  - Relevant perturbation list: Blood, Lens dirtiness, Defocus, Bright light, Dark light, Smoke, Motion blur, Deformation and incision
  - Realistic dataset with depth labels covering all perturbations
  - Appropriate evaluation metrics

Dataset Acquisition

The RoDEM dataset acquisition process has **three** phases:

1. **Pre-acquisition phase**
   a. Synchronisation Between depth sensor and laparoscope streams
   b. Calibration
      i. Estimating camera parameters
      ii. Estimating extrinsics between the depth sensor and laparoscope
   c. Validating using a 3D-printed liver phantom
      i. 1.4 mm average error between the phantom's CAD model and the point-cloud obtained by our system
2. **Acquisition phase**
   a. Ex-vivo sheep organs setups
      1) liver, 2) kidney, and 3) heart-lung
   b. Inducing perturbations



3. **Post-acquisition phase**
   a. Z-buffering: Detecting and removing the occluded points in laparoscope coordinates
   b. Excluding Surgical tools
   c. Severity levels: categorised by visual inspection
   d. Final dataset: 29,803 images with depth GT, 44 video sequences, 9 perturbations

- **Comparison of existing MIS datasets with RoDEM**

Dataset	Data type	Anatomical zone	Image depth pairs	Depth acquisition	Image perturbations S	M	L	D	F	B	Sequences C	D	I
SCARED[2]	Ex-Vivo	Abdomen	45	Structured light	-	-	-	-	-	-	✓	-	-
SERV-CT	Ex-Vivo	Abdomen, Thorax, Pelvis	16	CT scan	-	-	-	-	-	-	-	-	-
Hamlyn	Ex-Vivo, In-Vivo, Phantom	Abdomen, Pelvis	92694	Stereo matching	-	-	-	-	-	-	✓	-	-
EndoAbs	Phantom	Abdomen	120	Laser scanner	✓	-	✓	-	-	-	-	-	-
RoDEM (proposed)	Ex-Vivo	Thorax, Abdomen	29803	Depth sensor	✓	✓	✓	✓	✓	✓	✓	✓	✓

Benchmarking

- **Prediction to GT Alignment**
  To accurately evaluate the models, it is necessary to first estimate the optimal scale and shift parameters that align the predicted disparity with the ground truth, after which the depth evaluation metrics can be computed.

No-alignment (metric depth)	Direct use of depth values; typically designed for metric MoSDE methods predicting absolute depth.
s-alignment (up to scale depth)	Rescaling of the predicted depth; typically designed for disparity prediction MoSDE methods.
s&t-alignment (affine disparity)	Rescaling and shifting in disparity space (a complex depth transformation); typically designed for relative depth prediction MoSDE methods.

- **Metrics**

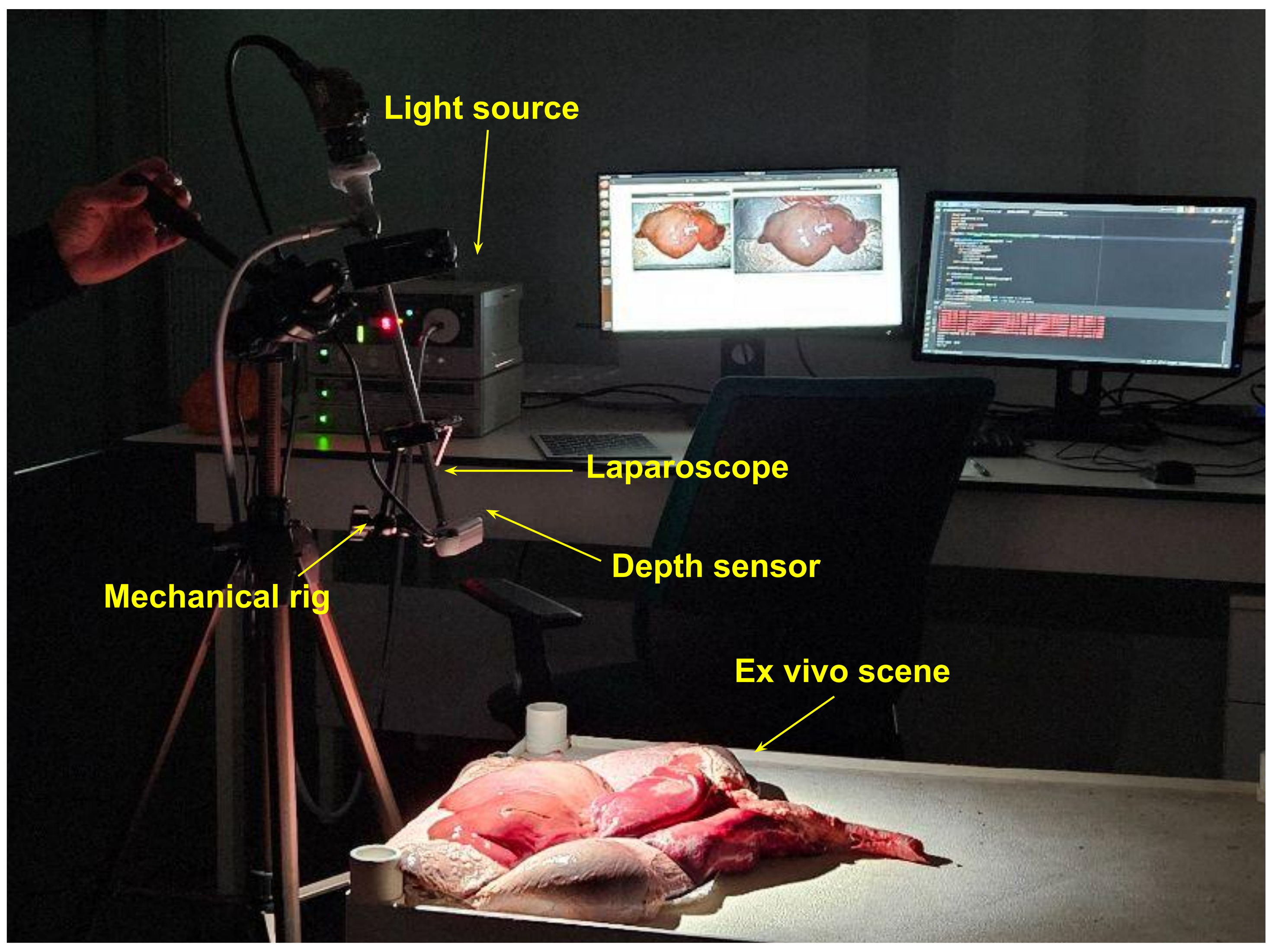
After aligning the prediction and ground-truth, the metrics for clean images and the robustness analysis can be computed.

Clean error	The error computed on the clean data (no perturbation)
mean Corruption Error (mCE) [1, 2]	Robustness comparison against perturbations, relative to a baseline method
mean Resilience Rate (mRR) [1]	Accuracy retainment under different perturbations

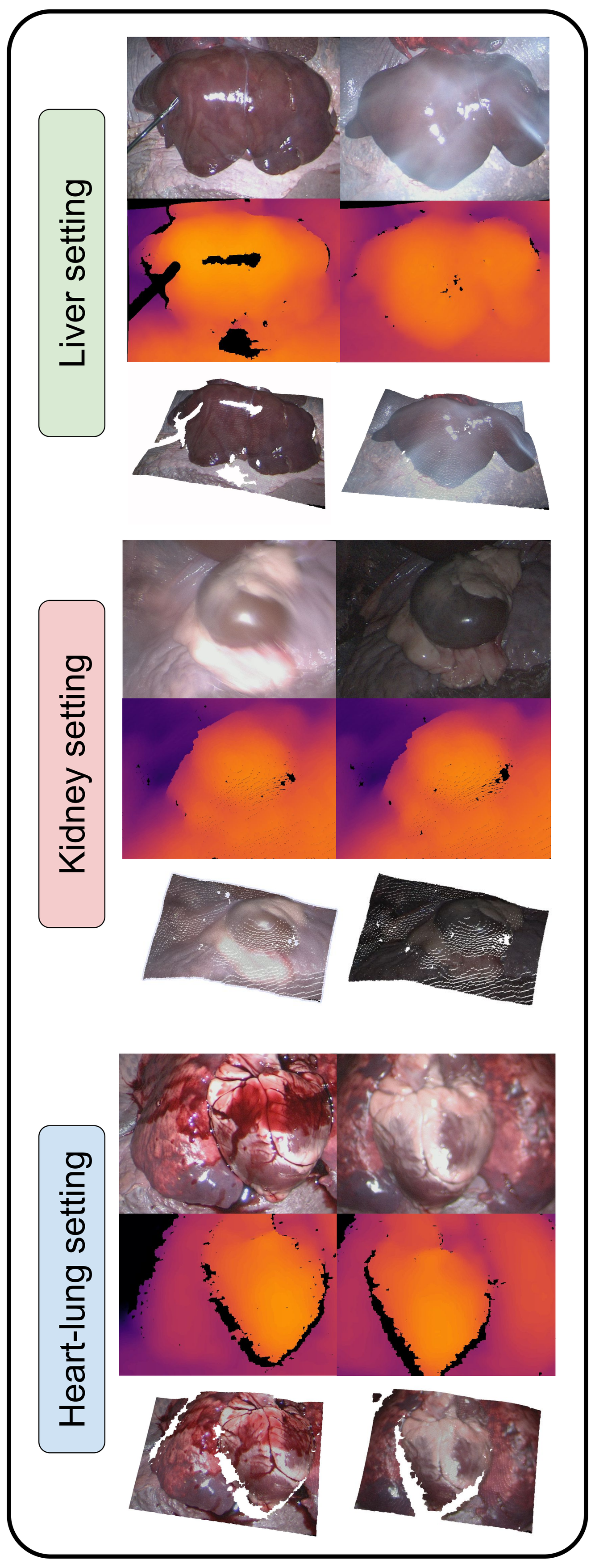
- **Benchmarked methods**
  - **MoSDE foundation models:** MiDaS V3 [1], Depth anything V1 [2], Depth anything V2 [3]
  - **MoSDE models for MIS images:** AF-SfM learner [4], EndoDAC [5], TRMDSV [6]
  - **Metric MoSDE models:** ZoeDepth [7], Depth Anything V2-metric [8], Unidepth [9]

RoDEM Characteristics

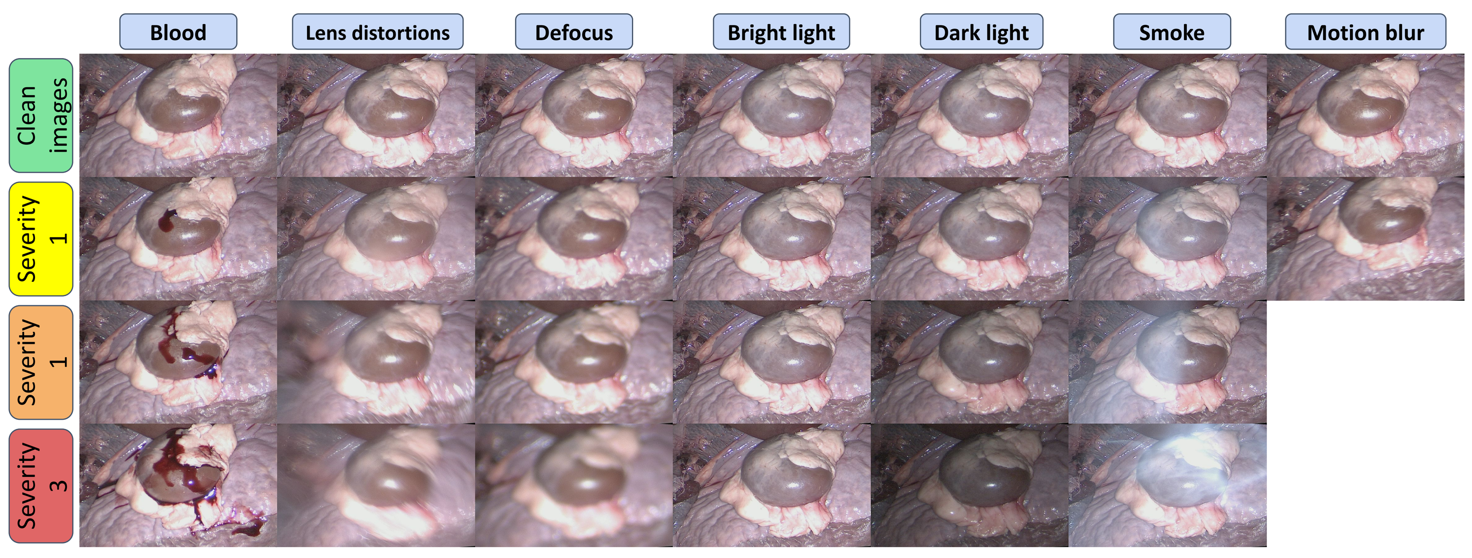
- **RoDEM aquestion setup**



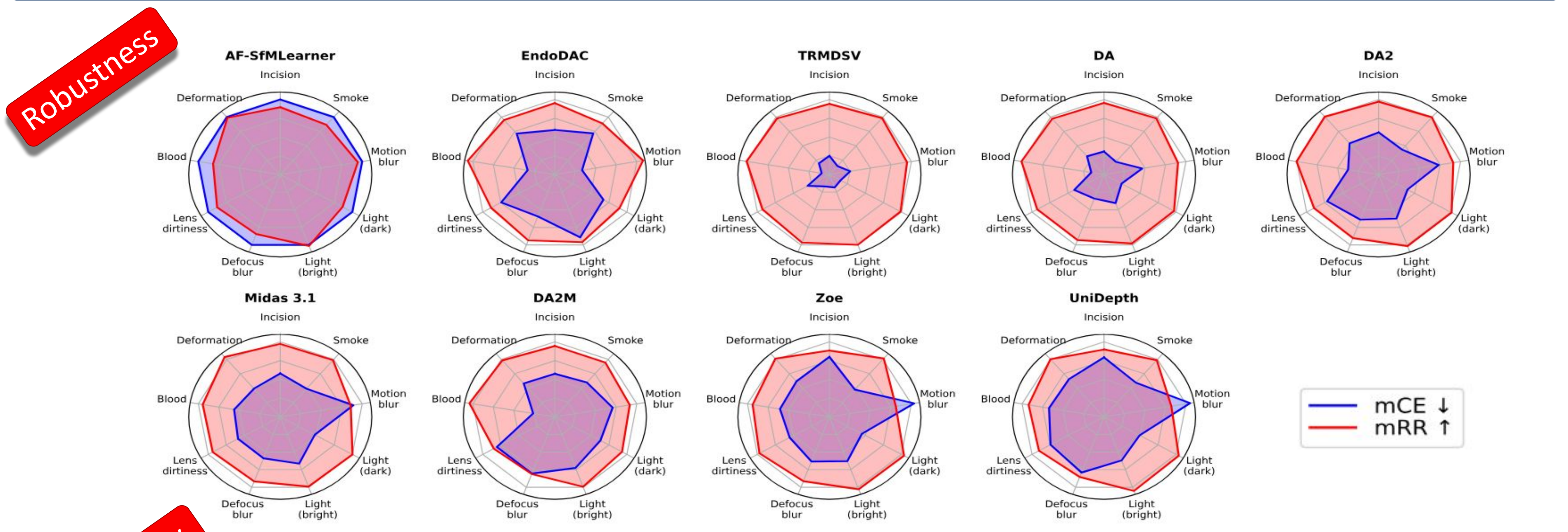
- **RoDEM dataset includes three scenes**

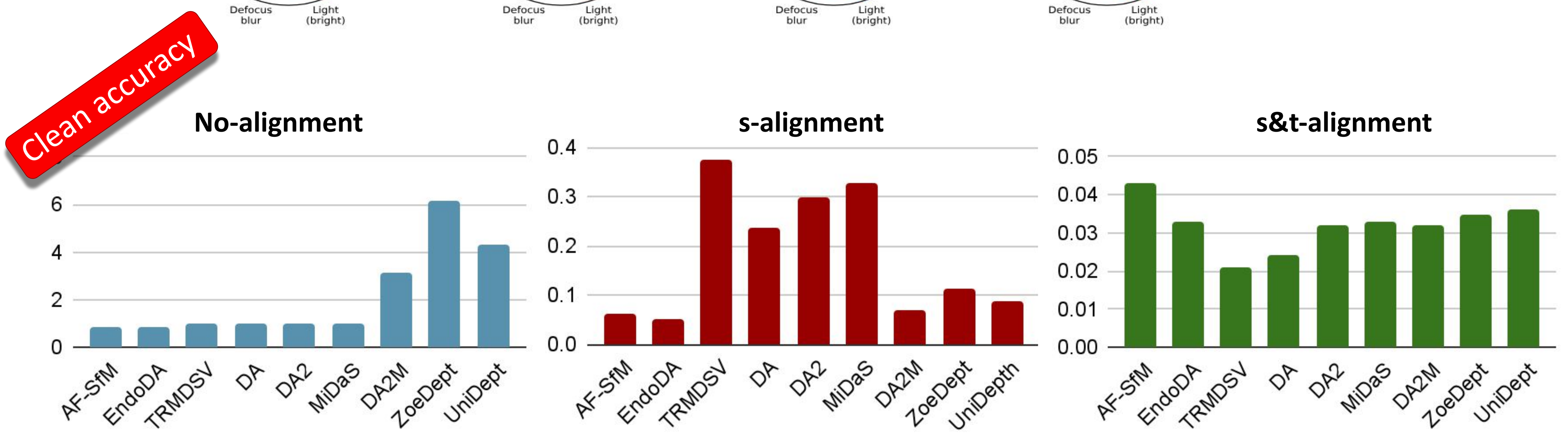


- **Various Surgical perturbation categorised in different severity levels**



MIS MoSDE Evaluations





Conclusions

- MoDSE foundation models outperform in both accuracy and robustness
- All methods were robust to motion blur and bright light. Methods trained on large datasets were robust against smoke, blood, and low light whereas the other methods exhibited reduced robustness.
- None of the methods coped with lens dirtiness and defocus blur.

Download RoDEM here!

References

[1] *Birkl, R., Wofk, D., M¨uller, M., Midas v3. 1–a model zoo for robust monocular relative depth estimation, arXiv preprint 2023*
[2] *Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., Depth anything: Unleashing the power of large-scale unlabeled data. CVPR 2024*
[3] Depth anything V2: *Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., Depth anything V2, Advances in Neural Information Processing Systems, 2024*
[4] AF-SfM learner: *Shao, S., Pei, Z., Chen, W., Zhu, W., Wu, X., Sun, D., Zhang, B., Self-supervised monocular depth and ego-motion estimation in endoscopy: Appearance flow to the rescue. MedIA, 2022*
[5] EndoDAC: *Cui, B., Islam, M., Bai, L., Wang, A., Ren, H., Efficient adapting foundation model for self-supervised depth estimation from any endoscopic camera, MICCAI 2024*
[6] TRMDSV: *Budd, C., Vercauteren, T., Transferring relative monocular depth to surgical vision with temporal consistency, MICCAI 2024*
[7] ZoeDepth: *Bhat, S.F., Birkl, R., Wofk, D., Wonka, P., Muller, M., Zero-shot transfer by combining relative and metric depth, arXiv preprint 2023*
[8] Depth Anything V2-metric: *Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., Depth anything V2, Advances in Neural Information Processing Systems, 2024*
[9] Unidepth: *Piccinelli, L., Yang, Y.-H., Sakaridis, C., Segu, M., Li, S., Van Gool, L., Yu, F., Universal monocular metric depth estimation, CVPR 2024*