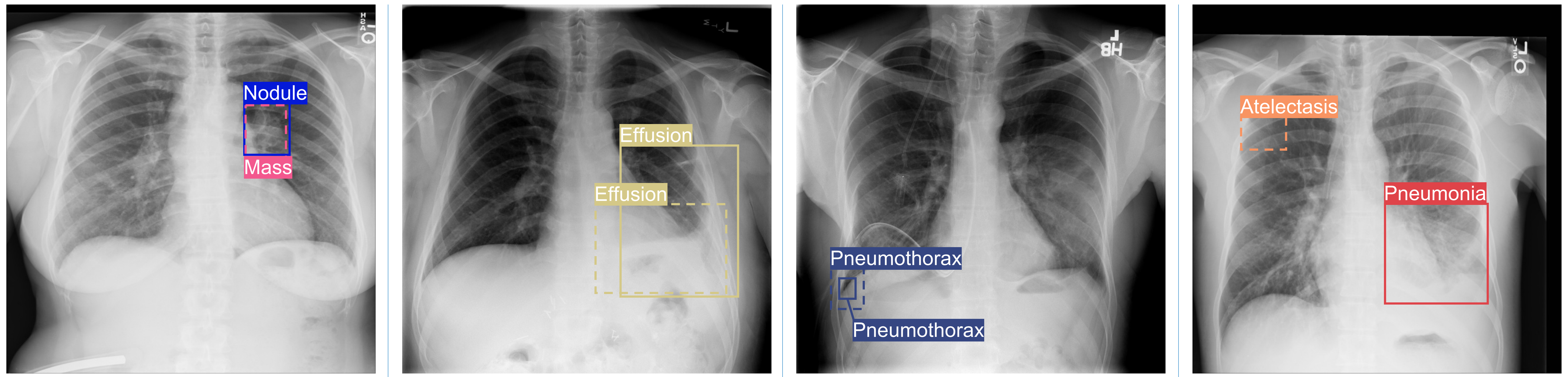
ROBUST DETECTION 🤠UTCOME

A METRIC FOR PATHOLOGY DETECTION IN MEDICAL IMAGES

FELIX MEISSEN, PHILIP MÜLLER, GEORGIOS KAISSIS, DANIEL RÜCKERT

WHY RODEO?

RoDeO reflects clinical needs better than current metrics



RoDeO total: **0.0%**	IoU: **81.5%**	RoDeO total: **74.4%**	IoU: **47.5%**	RoDeO total: **50.0%**	IoU: **25.0%**	RoDeO total: **0.0%**	IoU: **0.0%**
RoDeO cls: **0.0%**	acc@50: **75.0%**	RoDeO cls: **100.0%**	acc@50: **87.5%**	RoDeO cls: **100%**	acc@0.5: **87.5%**	RoDeO cls: **0.0%**	acc@0.5: **75.0%**
RoDeO shape: **89.2%**	AP@50: **0.0%**	RoDeO shape: **55.1%**	AP@50: **0.0%**	RoDeO shape: **25.0%**	AP@0.5: **0.0%**	RoDeO shape: **14.5**	AP@0.5: **0.0%**
RoDeO loc: **98.6%**		RoDeO loc: **82.2%**		RoDeO loc: **100%**		RoDeO loc: **0.0%**	

Problems with Existing Metrics

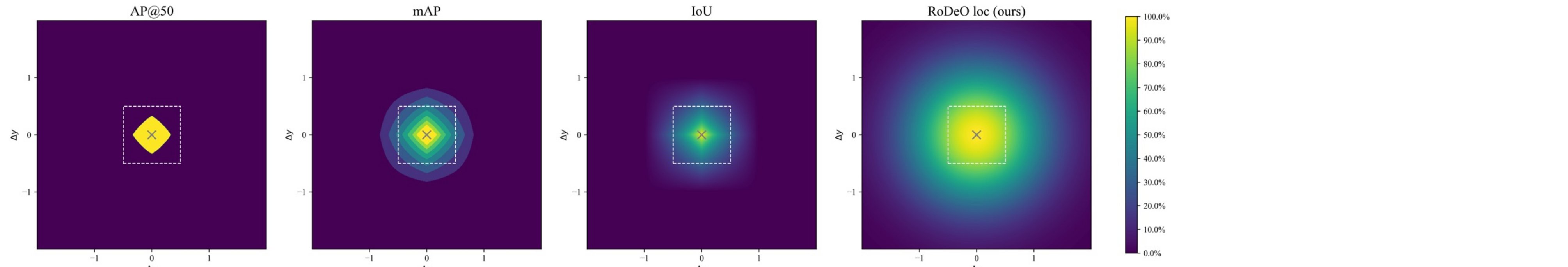
- AP@IoU declines sharply with small localization errors
- AP@IoU increases at high thresholds when predicting many boxes
- Acc@IoU is high when predicting nothing (many true negatives)
- Existing metrics are hard to interpret (why is the score low?)

What RoDeO does better

- Correct localization is valuable even under misclassification
- Uses a notion of distance that credits even coarse localization
- Can not be tricked by simple over- or under-prediction
- Easy to interpret through disentanglement of failure sources (localization, misclassification, shape mismatch)

HOW DOES RODEO WORK?

1. **Hungarian matching of predicted and target boxes**
   - Box overlap (gIoU)
   - Class matching, weighted based on classification performance (MCC) of model
2. **Sub-metric for Localization using center distance**
   - Based on (2D separable) normal distribution of (target-size normalized) center distance
   - Degrades slowly for small distances, then fast, until degrading slowly again for large distances
   - Degrades more smoothly and isotopically, further degrading even for non-overlapping boxes



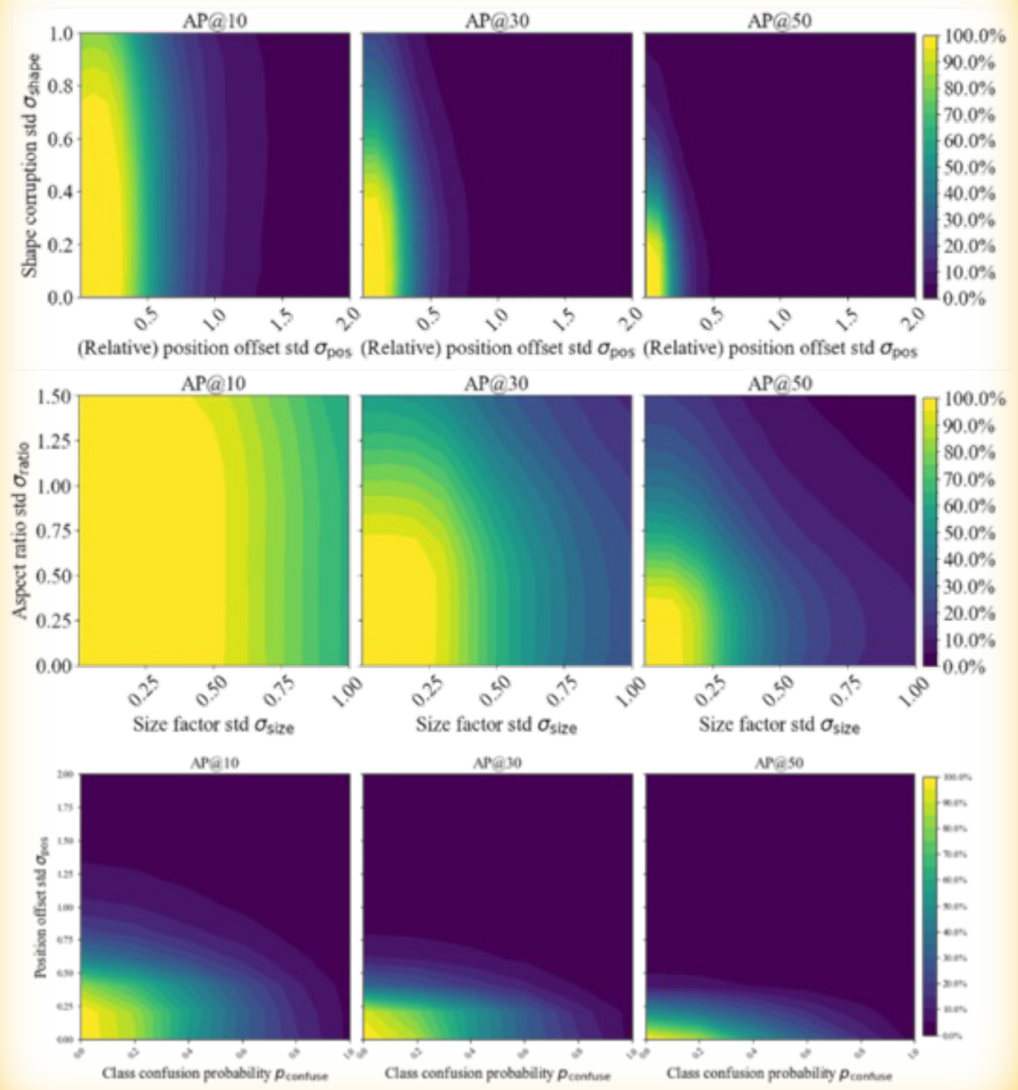
3. **Sub-metric for Shape similarity using centered IoU (cIoU)**
   - Measures box size and aspect ratio match
4. **Sub-metric for Classification using MCC**
   - MCC only gives positive scores for performance above random, regardless of class imbalances
   - Values < 0 are clipped
5. **Correction for Under- and Overprediction**
   - Penalize non-matched pred/target boxes using linear combination of matched sub-scores and zero
6. **Summary Metric using Harmonic Mean of corrected sub-metrics**

RoDeO degrades more smoothly and different error sources do not influence each other

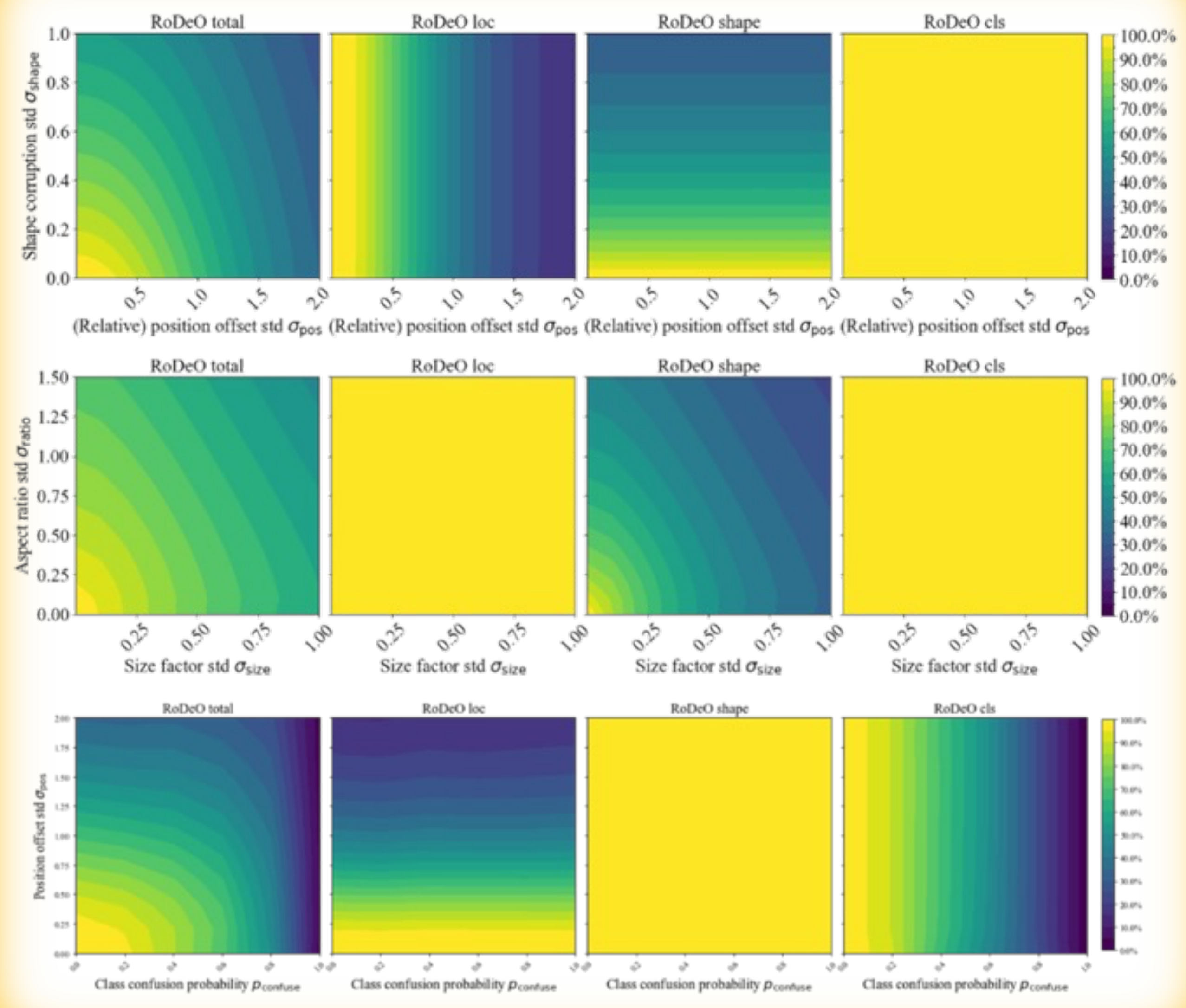
position/shape

size/aspect ratio

position/class

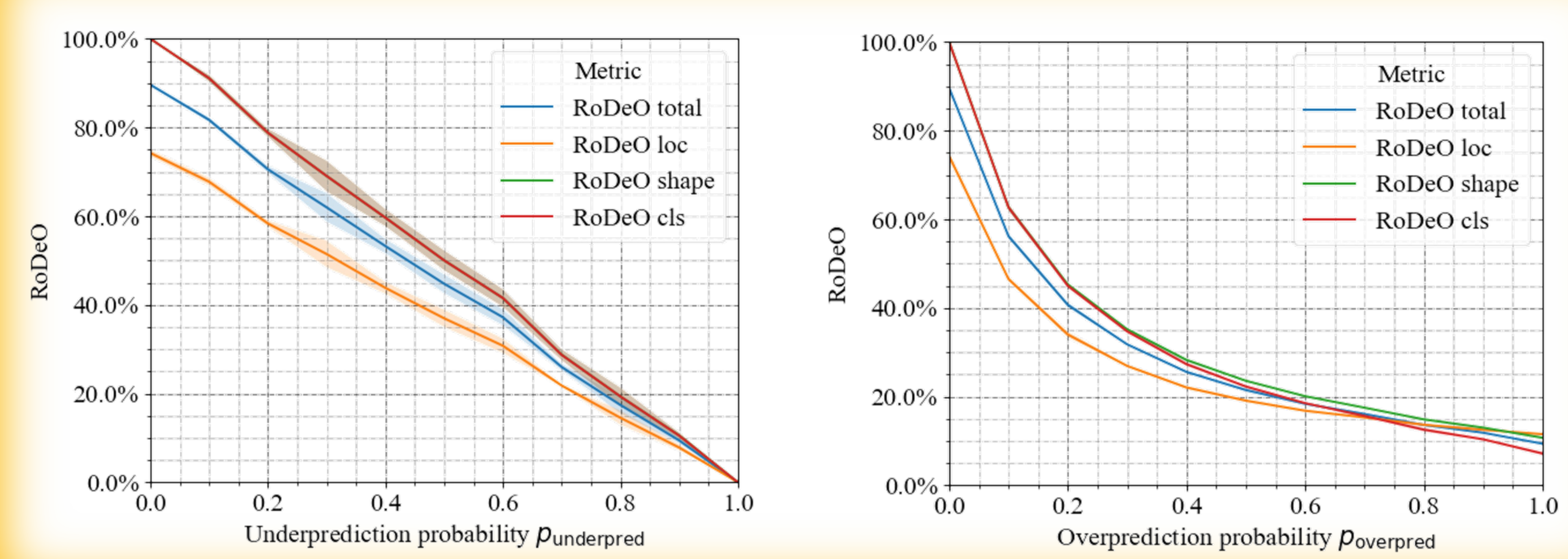


AP



RoDeO

RoDeO corrects for Under- and Overprediction



pip install rodeometric

twitter.com/FelixMeissen

felix.meissen@tum.de
philip.j.mueller@tum.de