

P1: Benchmarking



Evaluated on

- private
- public



#Datasets used

- 1
- 2
- 3

P2: Baselines

On MML data	MRE
U-Net [1]	2.7
U-Net [2]	2.2
U-Net [3]	1.9

→ no meaningful conclusions possible

P3: Usability



Public code

- no
- yes

→ Exhaustive parameter tuning and risk for reimplementation errors

Requirements for a novel landmark detection framework

- (1) Provide easy access to public benchmarking datasets (P1)
- (2) Propose a state-of-the-art baseline (P2)
- (3) Generalize out-of-the-box to new datasets (P3)
- (4) Serve as a standardized environment for evaluation of new methods (P3)