lamar.ethz.ch

LaMAR Tutorial
3. Benchmarking

Viktor Larsson, Mihai Dusmanu

ETH zürich

ECCV TEL AVIV 2022

Microsoft

Outline

a) Approaches & baselines
b) Metrics & tracks
c) Limitations and open problems
d) Q&A

a) Approaches & baselines

Image Matching

Global features
- NetVLAD
- Fusion (NetVLAD + APGeM)

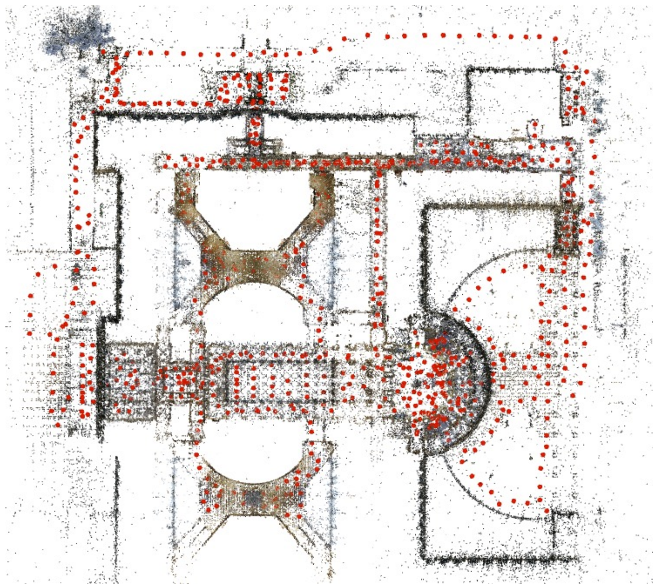
Local features
- SIFT + AdaLAM
- DoG + SOSNet + AdaLAM
- R2D2
- SuperPoint + SuperGlue

Dense matchers
- LoFTR

Mapping

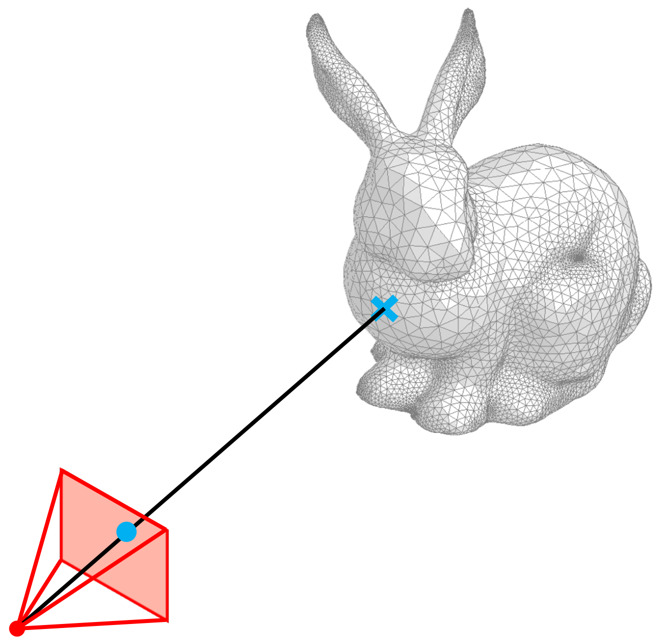
Triangulation with COLMAP

- Pairs to match
  - Frustum + pose distance filtering
  - Global descriptor similarity ordering
- GV using map poses



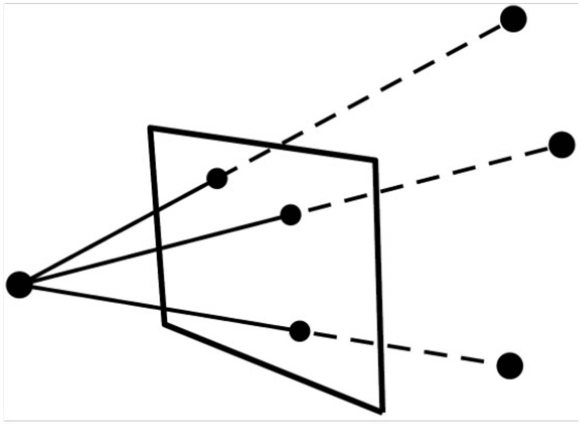
Lifting from mesh

- Requires dense depth (GT mesh)

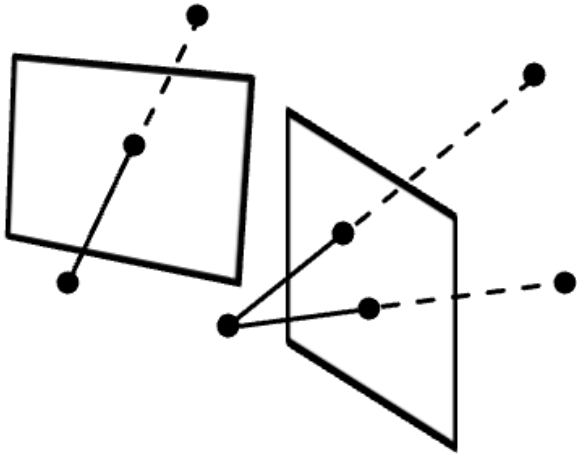


Pose Estimation

LO-RANSAC + P3P



LO-RANSAC + GP3P



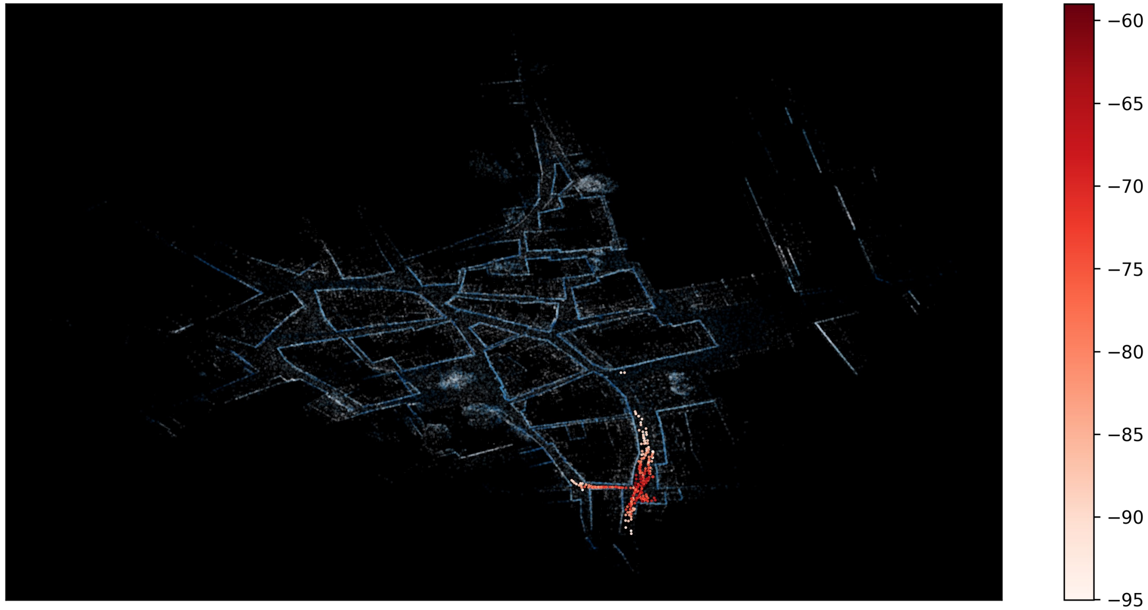
One can also use gravity constraints!
UP2P/UGP2P

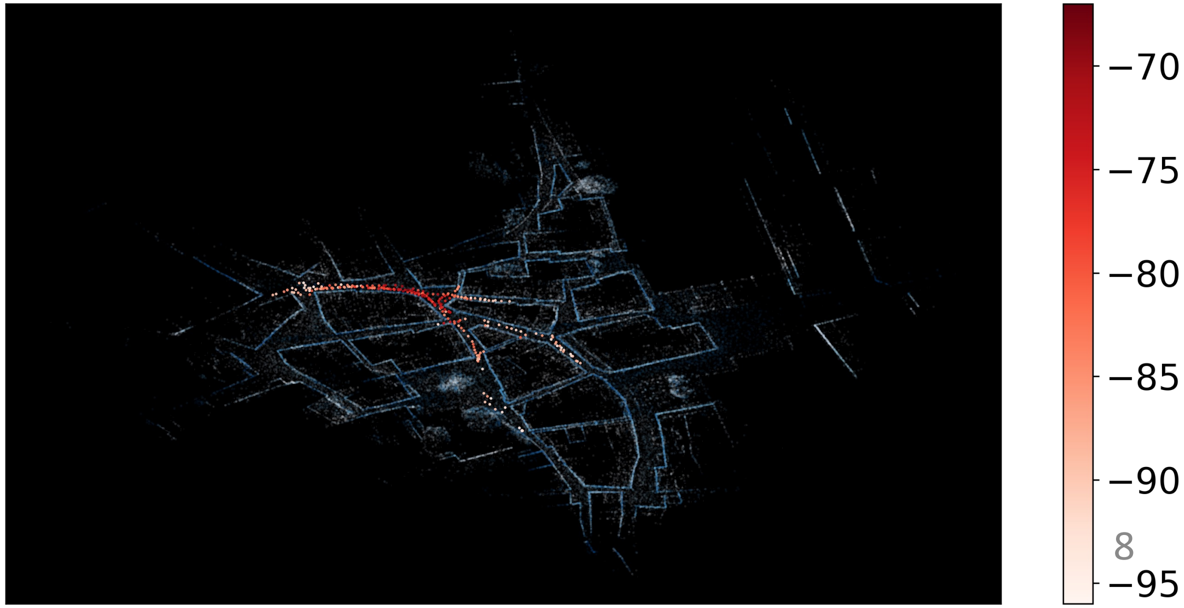
Radios Baseline

- Radio signal reception – main information
  - Timestamp (of receiving)
  - WiFi MAC Address / BT GUID
  - Signal strength – RSSI (Received Signal Strength Indicator)

Radios Baseline

- Radio signal reception – main information
  - Timestamp (of receiving)
  - WiFi MAC Address / BT GUID
  - Signal strength – RSSI (Received Signal Strength Indicator)
- Usefulness: reception is locally bound!



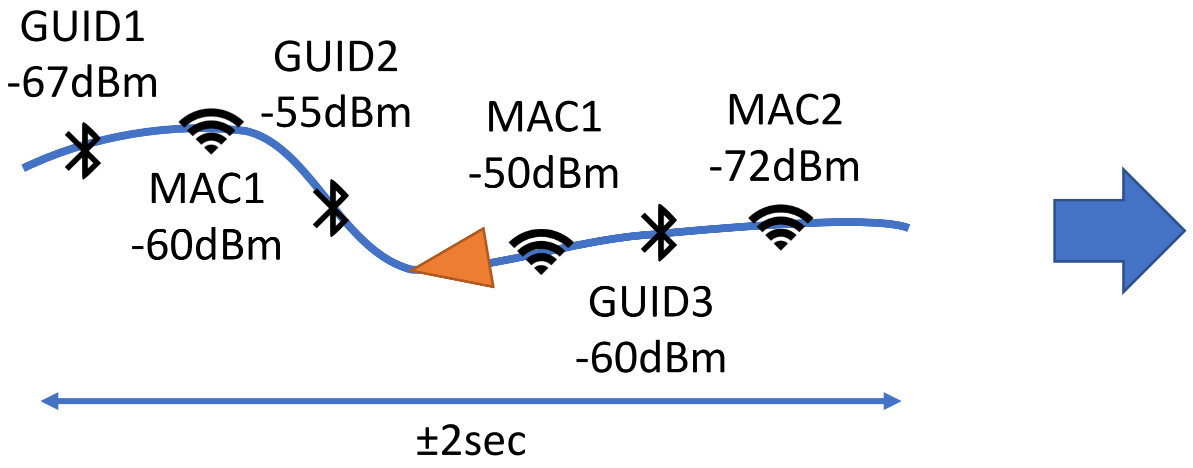


Radios Baseline for Image Retrieval

- Approach: radio descriptors compared by L2 distance

Radios Baseline for Image Retrieval

- Approach: radio descriptors compared by L2 distance
- Temporal window around keyframes (±2sec, only past for queries)

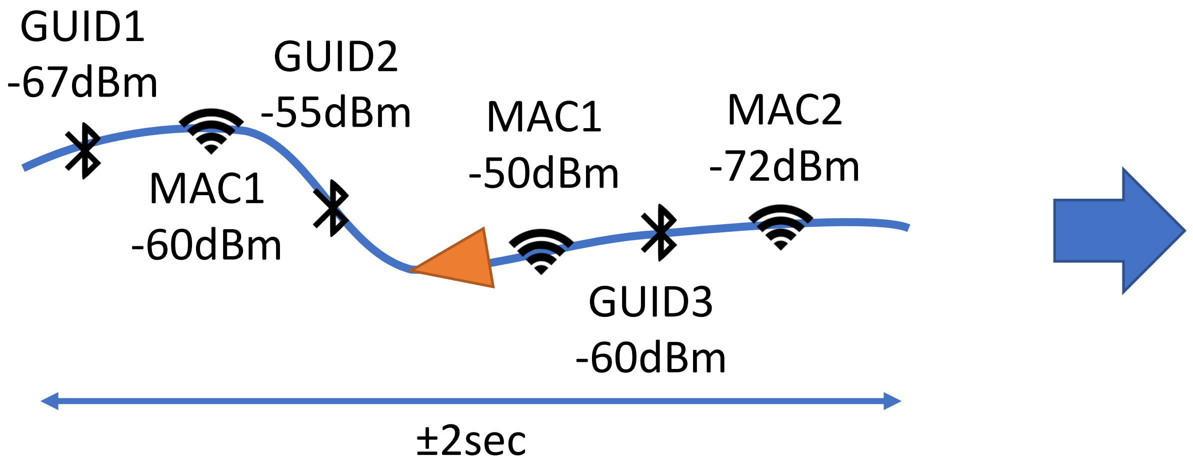


ID	RSSI (dBm)
MAC1	(-60 + -50) / 2 = -55
MAC2	-72
GUID1	-67
GUID2	-55
GUID3	-60

ECCV TEL AVIV 2022

Radios Baseline for Image Retrieval

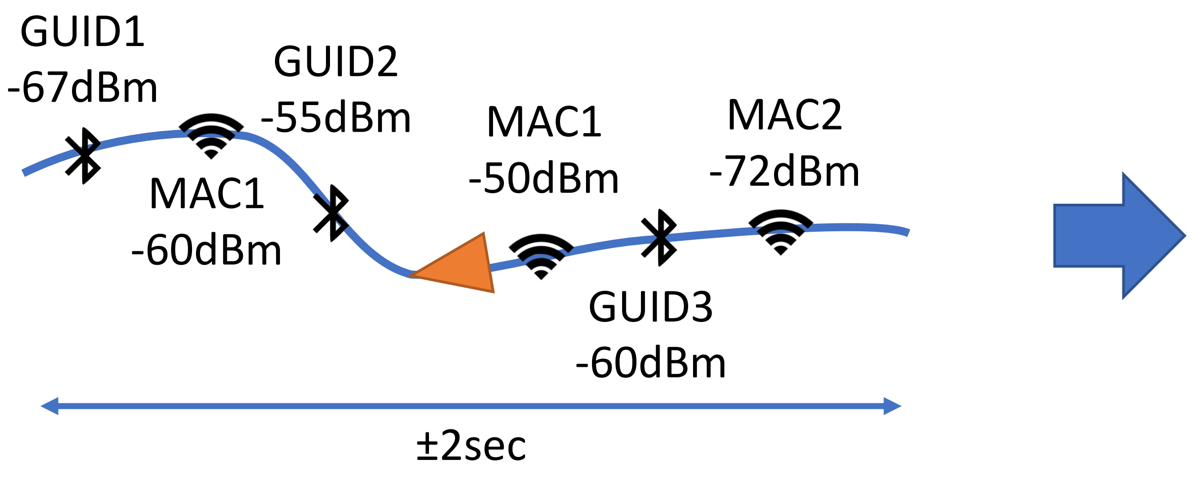
- Approach: radio descriptors compared by L2 distance
- Temporal window around keyframes (±2sec, only past for queries)
- Mapping: group keyframes in a voxel grid (1m)



ID	RSSI (dBm)
MAC1	(-60 + -50) / 2 = -55
MAC2	-72
GUID1	-67
GUID2	-55
GUID3	-60

Radios Baseline for Image Retrieval

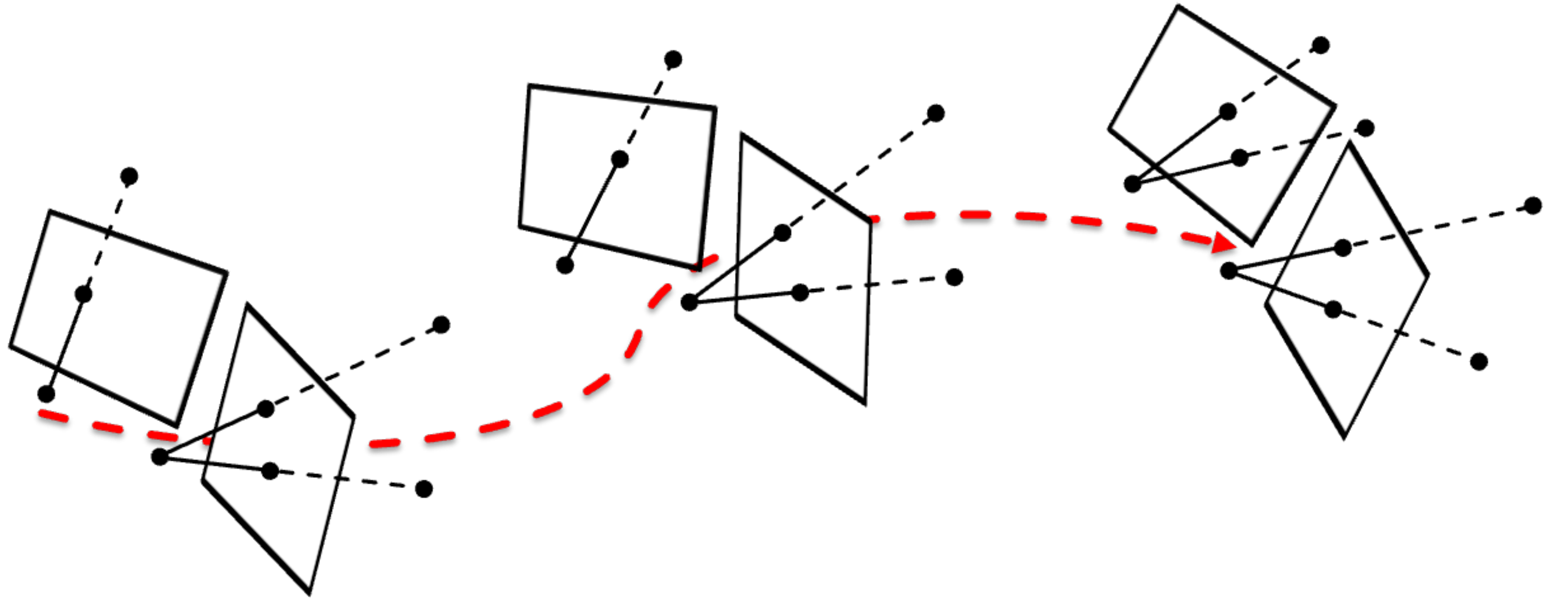
- Approach: radio descriptors compared by L2 distance
- Temporal window around keyframes (±2sec, only past for queries)
- Mapping: group keyframes in a voxel grid (1m)
- Consider only 2.5% of map visual retrieval



ID	RSSI (dBm)
MAC1	(-60 + -50) / 2 = -55
MAC2	-72
GUID1	-67
GUID2	-55
GUID3	-60

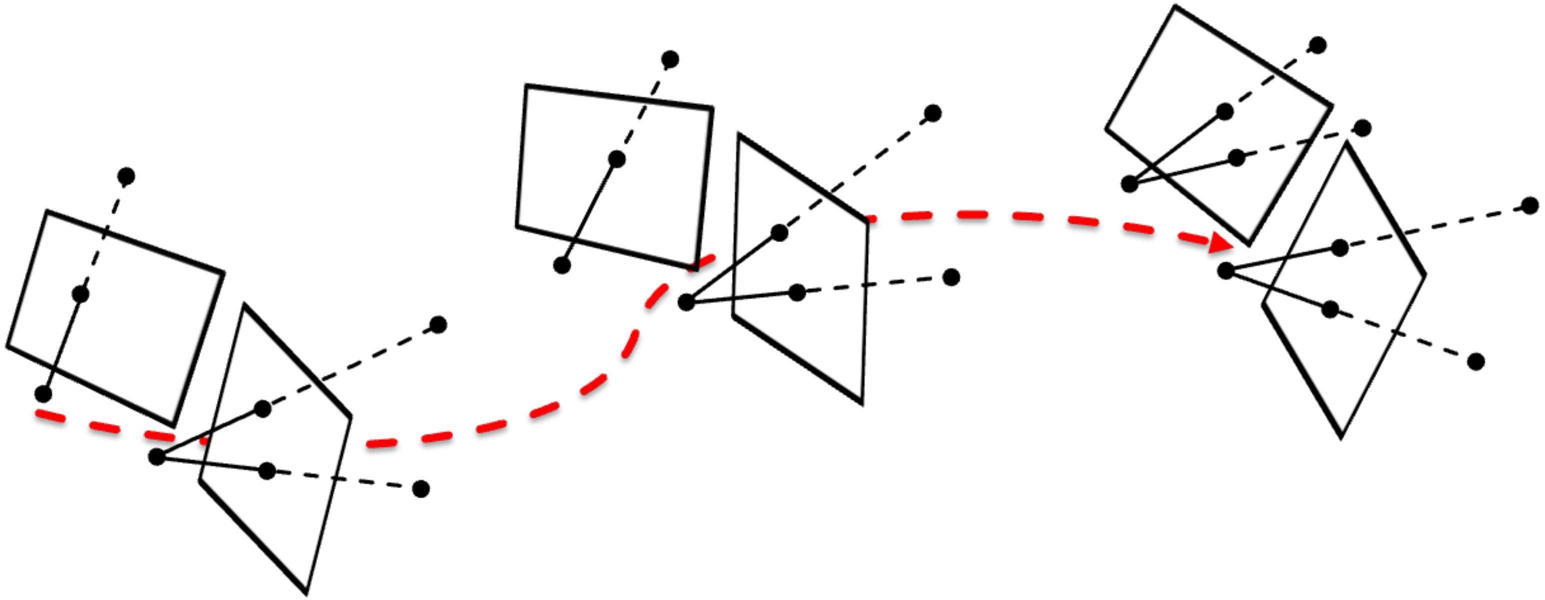
Sequence Localization

- Sequence as generalized camera (Sattler, 2018)
- Sequence as local reconstruction (Schneider, 2018)



Sequence Localization

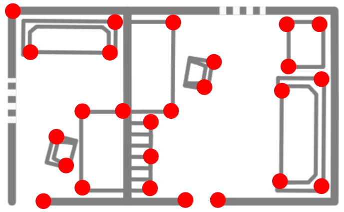
- Sequence as generalized camera (Sattler, 2018)
- Sequence as local reconstruction (Schneider, 2018)



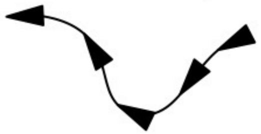
- Local poses from VIO are not perfect
- Degrade localization performance if not handled

Sequence Localization Baseline

Sparse Map

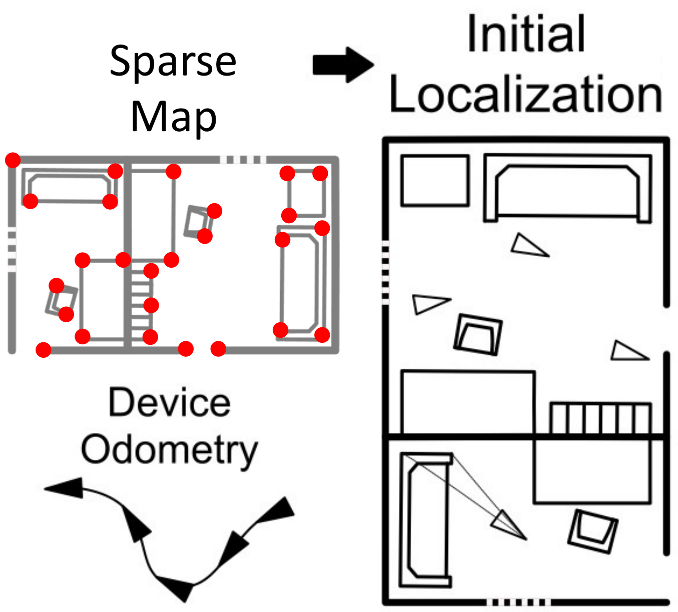


Device Odometry



ECCV TEL AVIV 2022

Sequence Localization Baseline

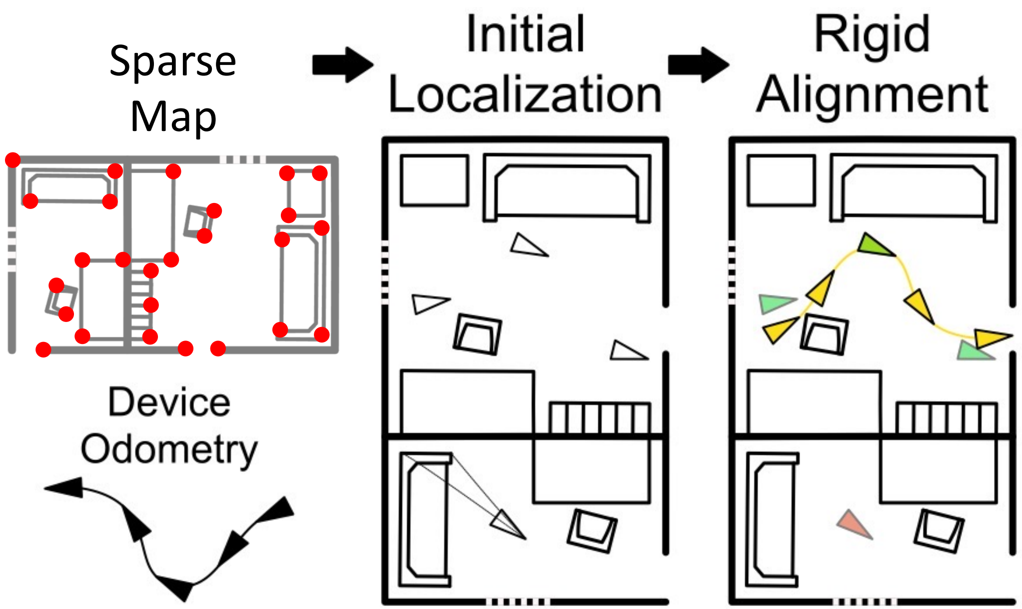


Fusion
(+ radio)

SP + SG

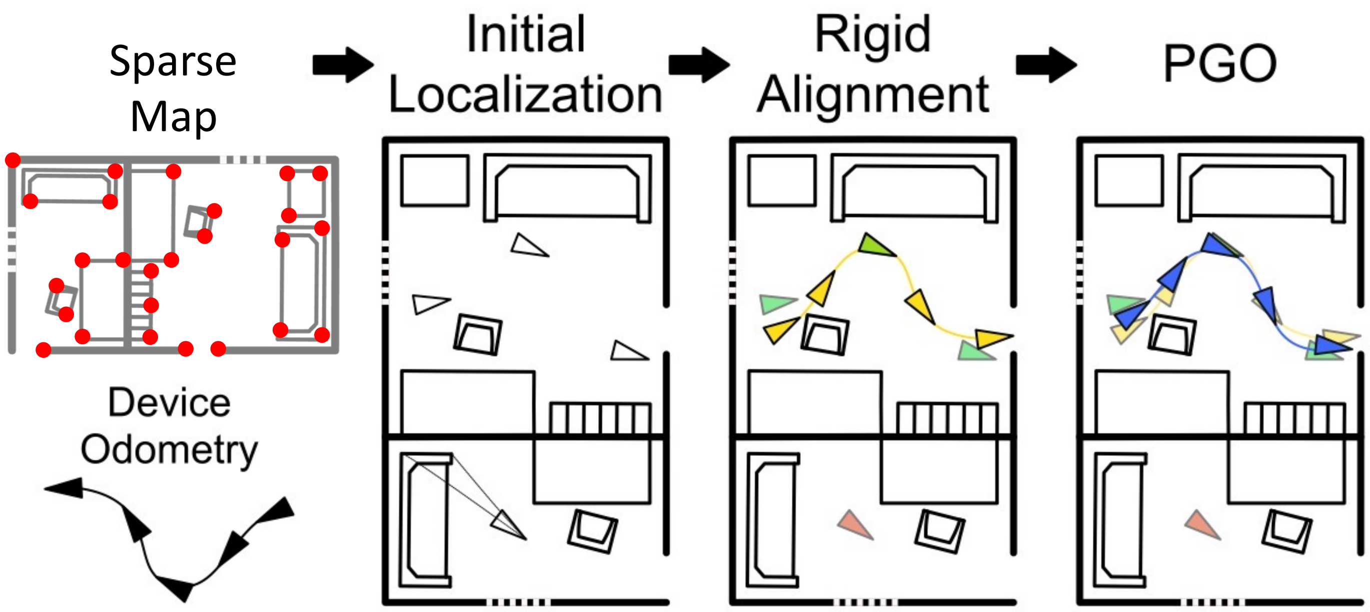
(G)P3P

Sequence Localization Baseline



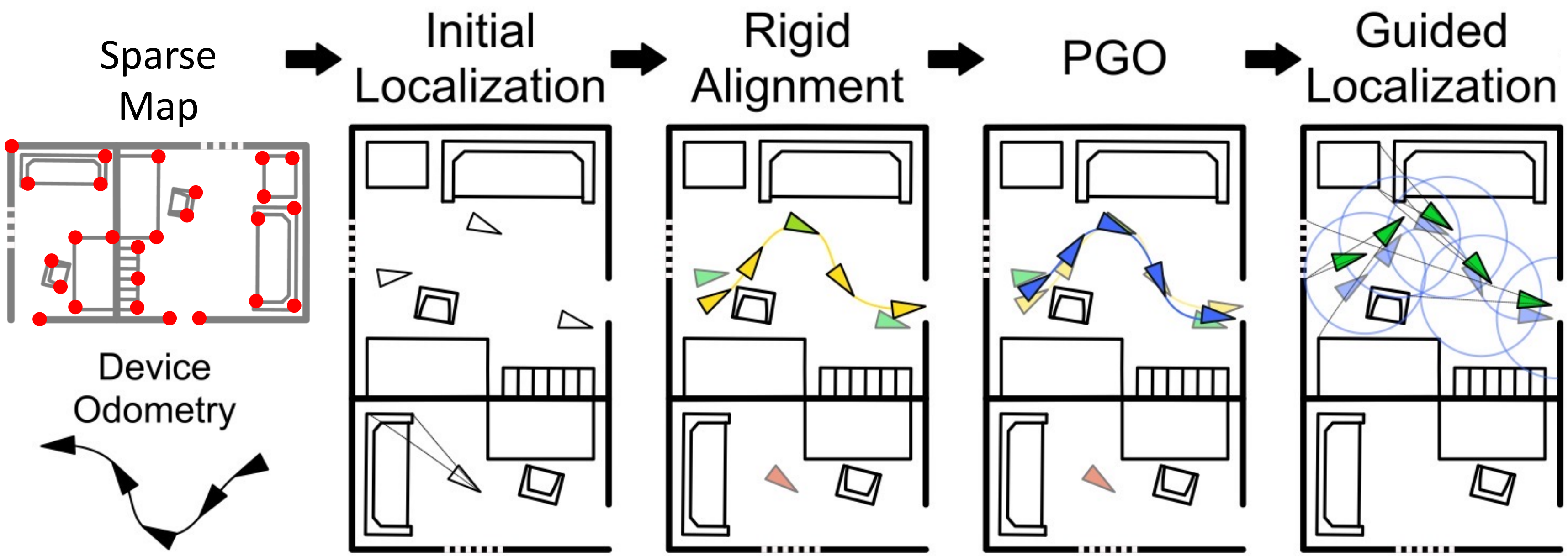
Voting

Sequence Localization Baseline



Absolute pose
measurements
+
Relative VIO
frame-to-frame
measurements

Sequence Localization Baseline

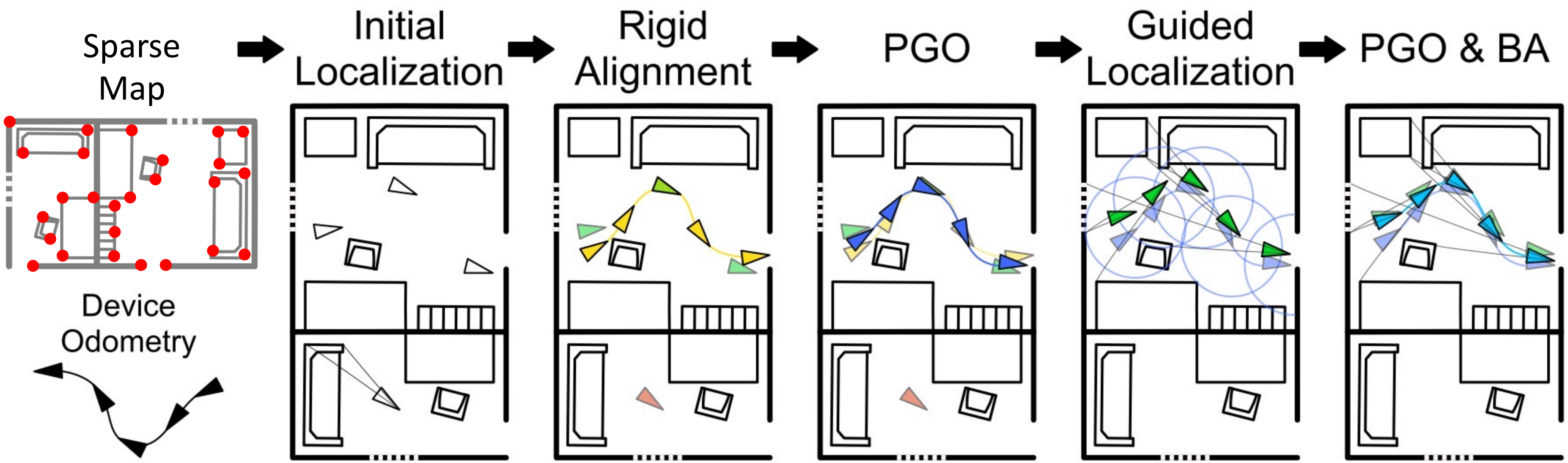


Fusion with frustum & pose filter

SP + SG

(G)P3P

Sequence Localization Baseline



Reprojection
errors
+
Relative VIO
frame-to-frame
measurements

b) Metrics and tracks

Single-frame Localization

Single-frame – Metrics

- Localization recall at (translation, rotation) threshold
  - 10cm, 1deg => tight threshold, good target for AR applications
  - 1m, 5deg => coarse threshold, fixable with better features / matching

Single-frame – Metrics

- Localization recall at (translation, rotation) threshold
  - 10cm, 1deg => tight threshold, good target for AR applications
  - 1m, 5deg => coarse threshold, fixable with better features / matching

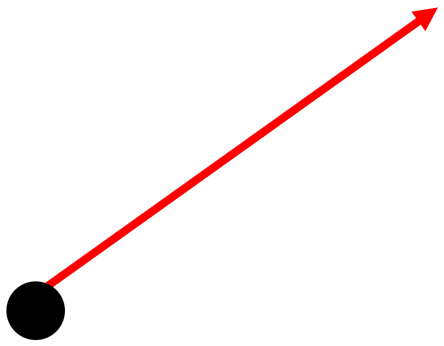
- Alternative metrics:
  - Perceptive: Aachen v1.1, RIO10, MapFreeReloc Niantic
  - Uncertainty based: Aachen v1.1
  - We didn't focus the evaluation on them, but we are considering adding them to the leaderboard

Single-frame – Tracks

•

Single-frame – Tracks

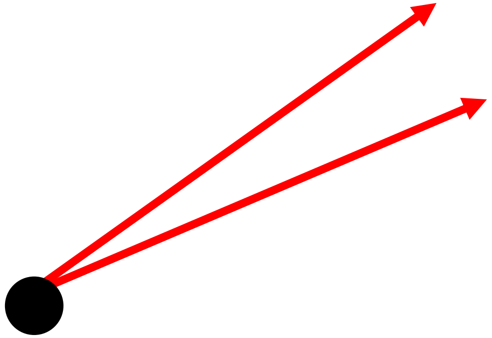
single-images, single-rigs



Single-frame – Tracks

single-images, single-rigs

+ radios

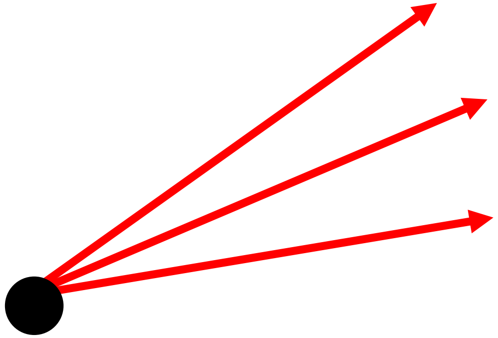


Single-frame – Tracks

single-images, single-rigs

+ radios

mapping modality SfM, mesh lifting



Single-frame – Tracks

single-images, single-rigs

+ radios

mapping modality SfM, mesh lifting

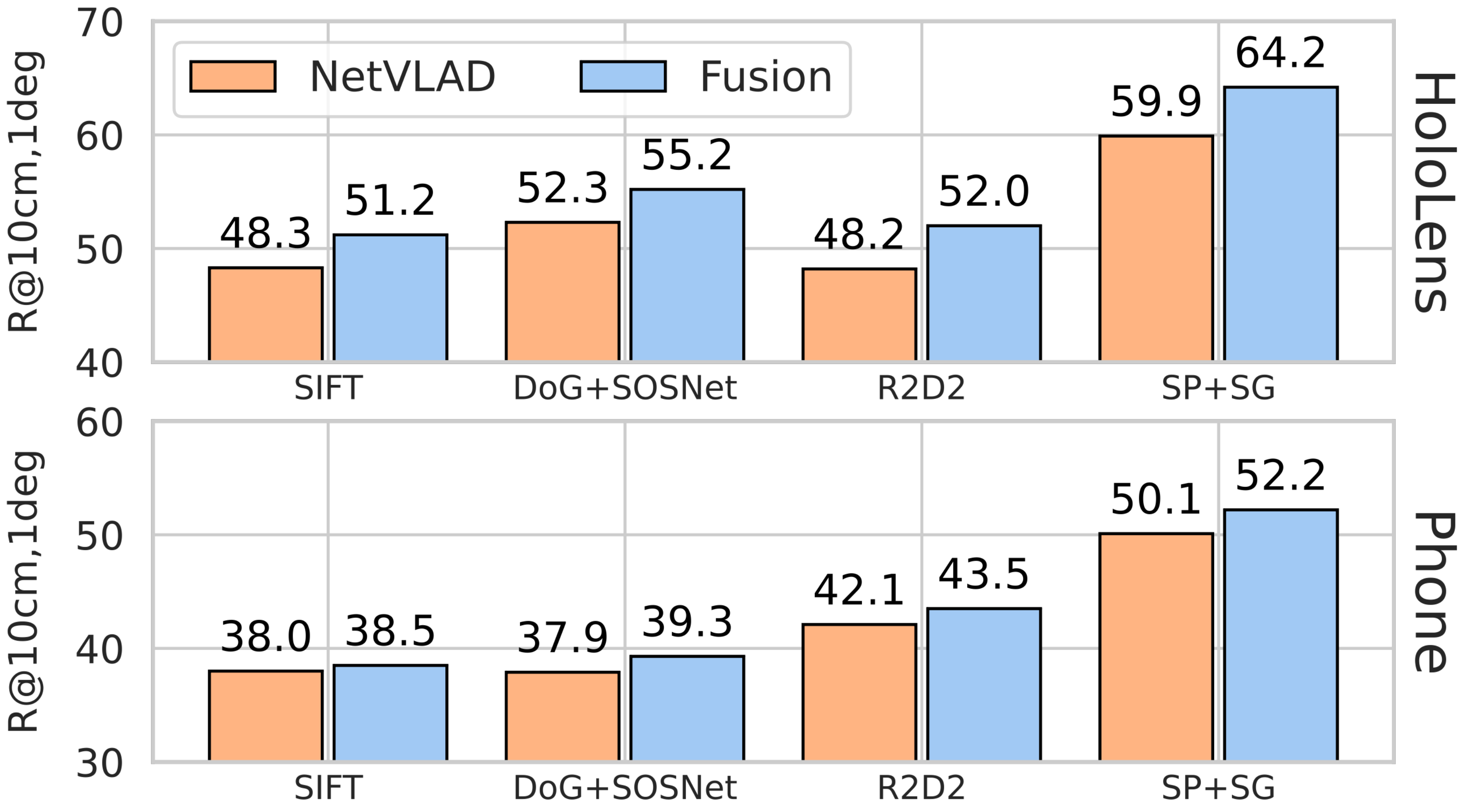
localization device HL, phone

per scene HGE, LIN, CAB

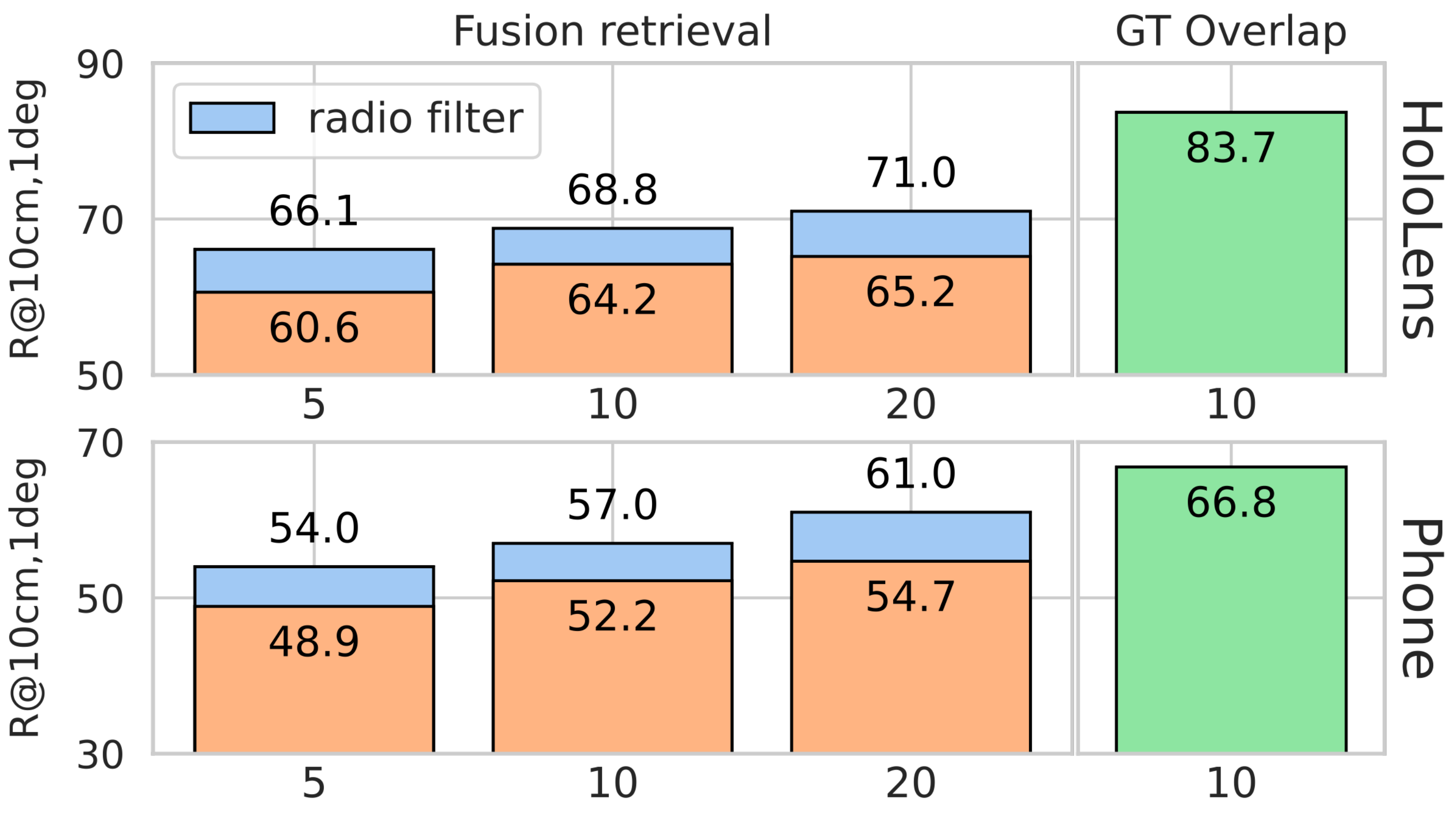
per type of scene indoor vs outdoor

per query condition day vs night

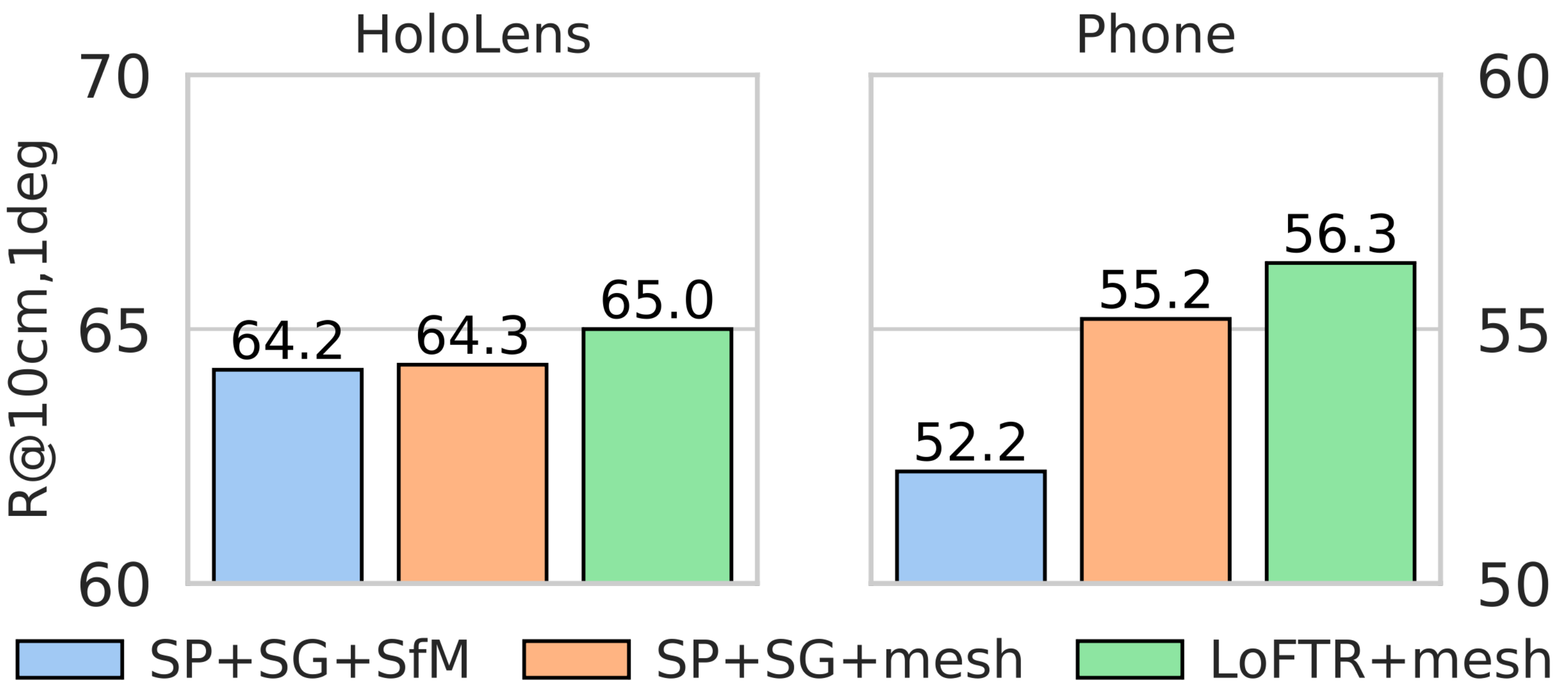
Comparing Global and Local Features



Radio retrieval



Mapping modality



Imaging conditions

Condition	CAB scene		HGE scene		LIN scene
	Indoor	Outdoor	Indoor	Outdoor	Outdoor
day	66.5 / 74.7	73.9 / 88.1	52.7 / 65.9	43.0 / 64.3	71.2 / 82.5
night	30.3 / 44.8	18.8 / 40.6	47.9 / 59.4	12.1 / 33.6	38.6 / 55.6

Imaging conditions

Condition	CAB scene		HGE scene		LIN scene
	Indoor	Outdoor	Indoor	Outdoor	Outdoor
day	66.5 / 74.7	73.9 / 88.1	52.7 / 65.9	43.0 / 64.3	71.2 / 82.5
night	30.3 / 44.8	18.8 / 40.6	47.9 / 59.4	12.1 / 33.6	38.6 / 55.6

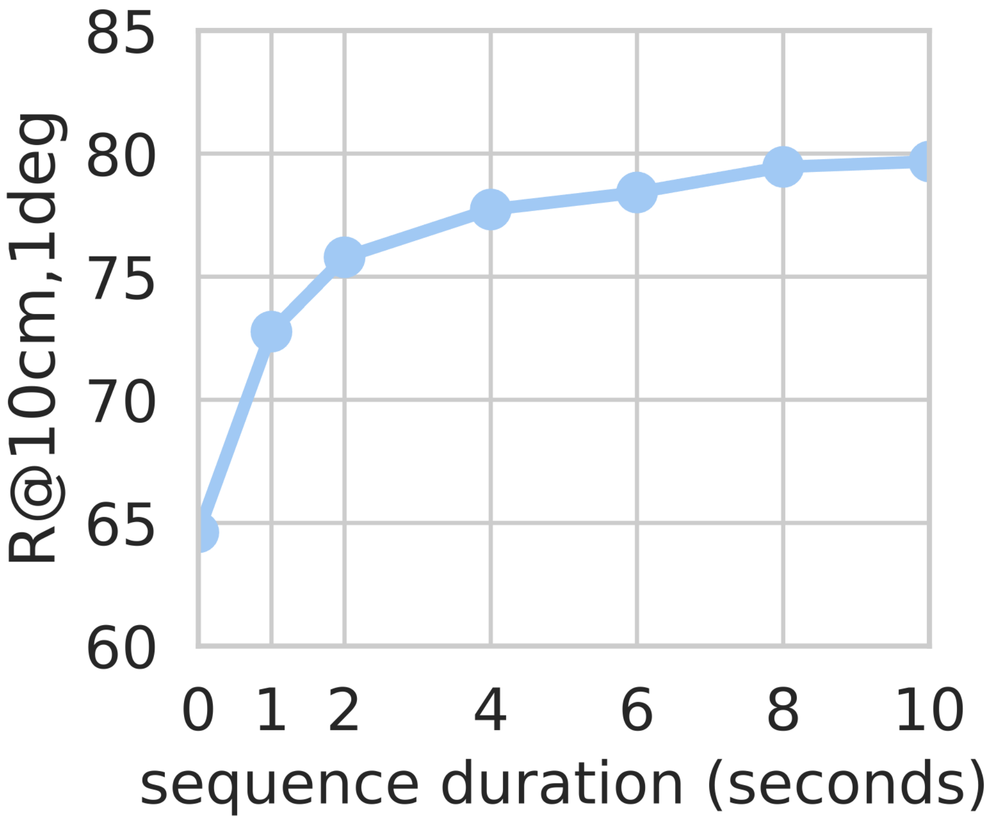
Imaging conditions

Condition	CAB scene		HGE scene		LIN scene
	Indoor	Outdoor	Indoor	Outdoor	Outdoor
day	66.5 / 74.7	73.9 / 88.1	52.7 / 65.9	43.0 / 64.3	71.2 / 82.5
night	30.3 / 44.8	18.8 / 40.6	47.9 / 59.4	12.1 / 33.6	38.6 / 55.6

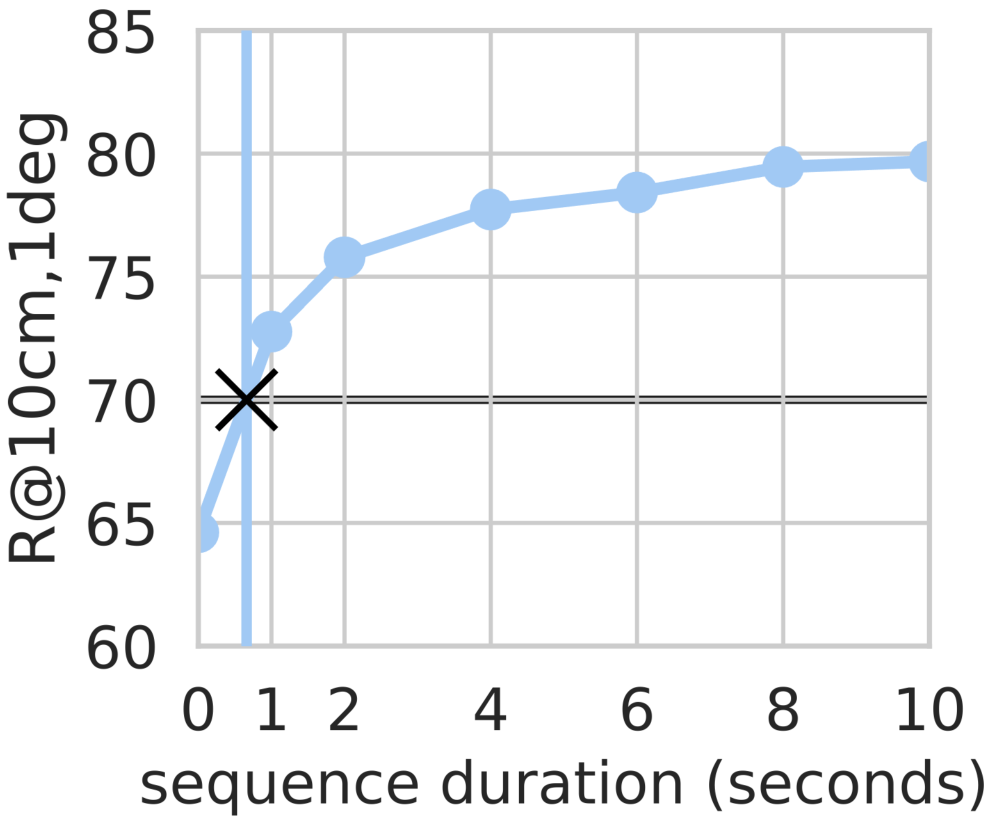
b) Metrics and tracks

Sequence Localization

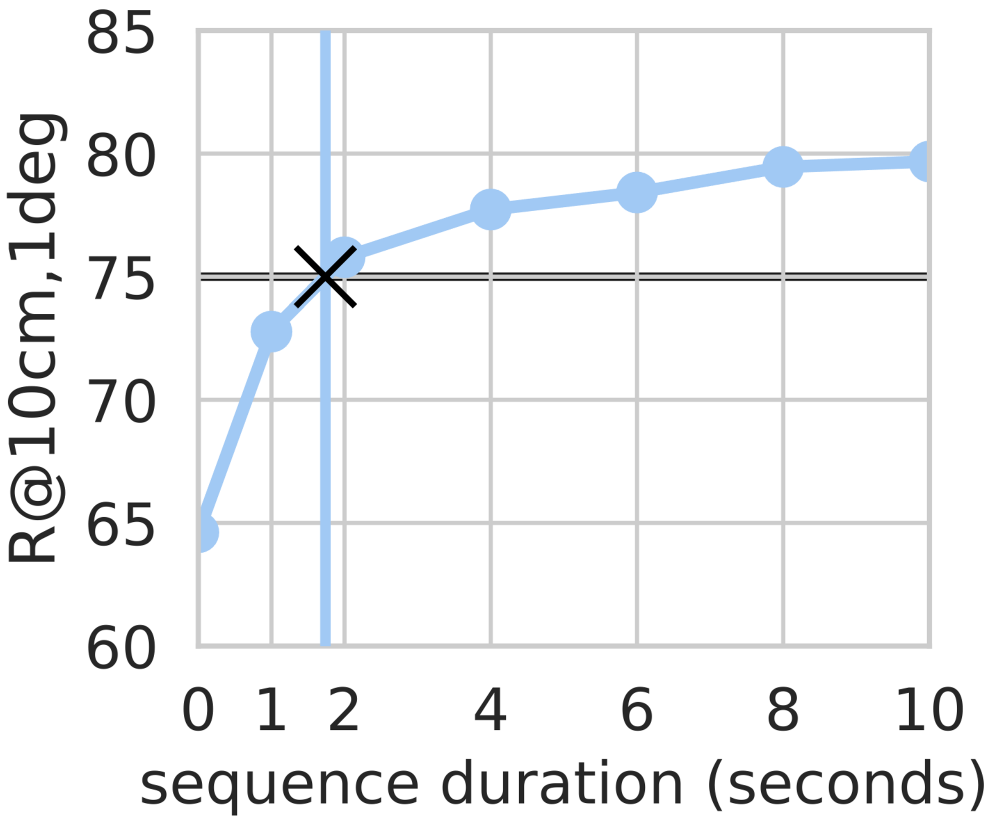
Sequence – Metrics



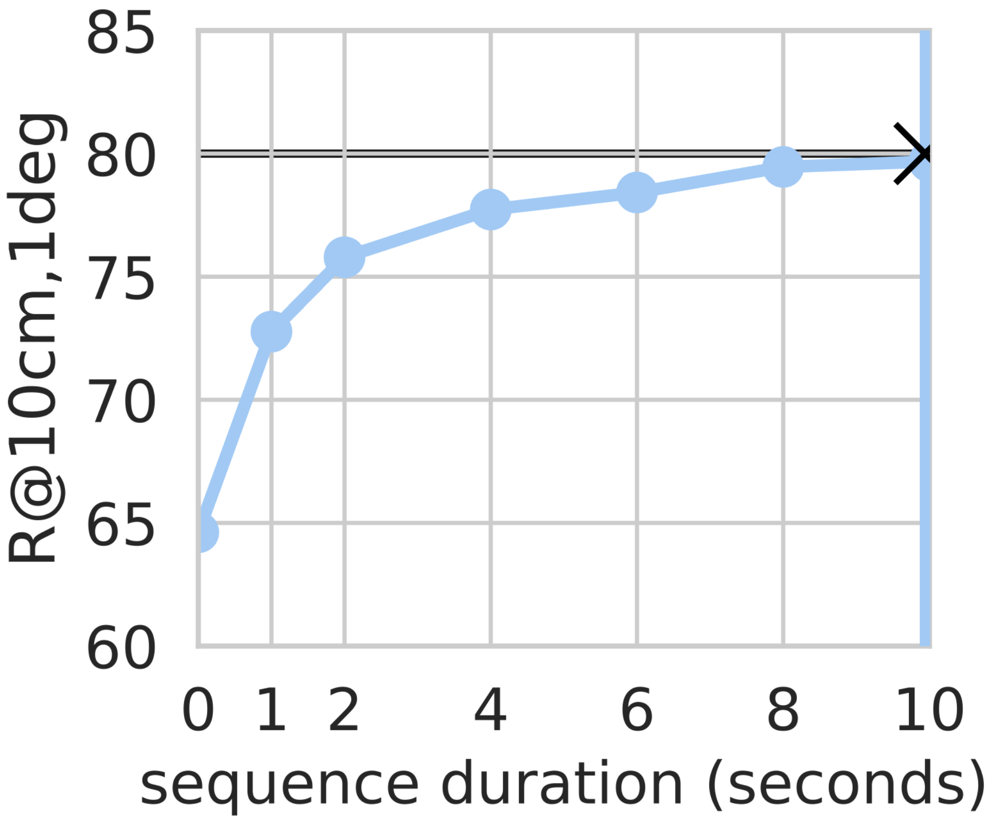
Sequence – Metrics



Sequence – Metrics



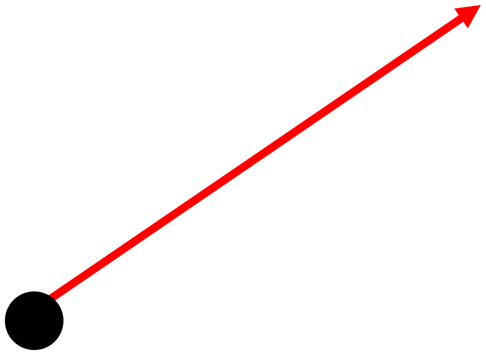
Sequence – Metrics



Sequence – Tracks

Sequence – Tracks

sequence duration

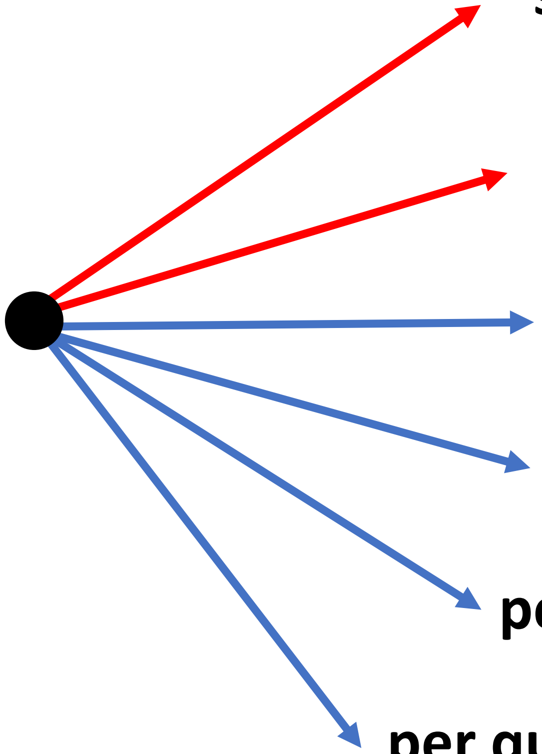


Sequence – Tracks

sequence duration

+ radios

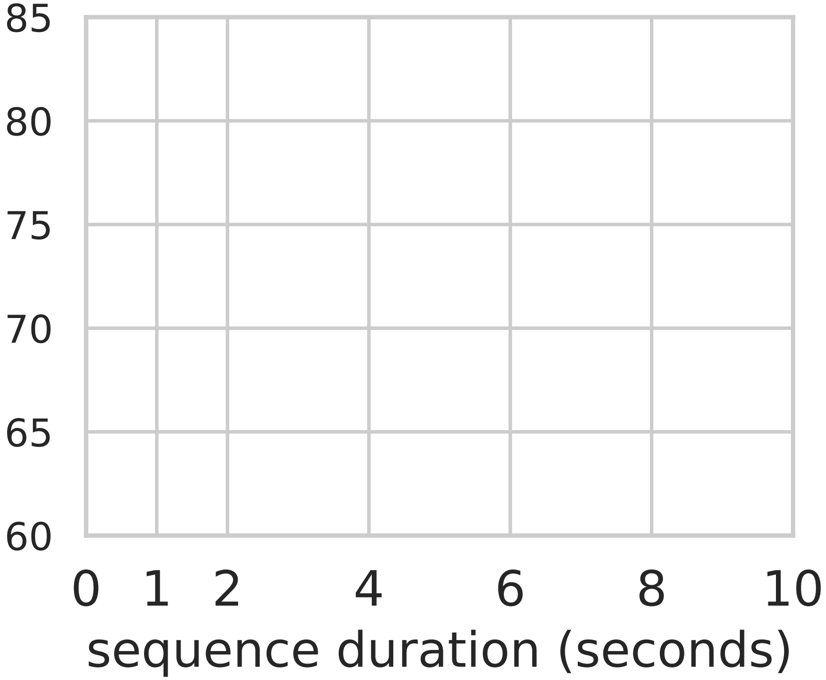
Sequence – Tracks



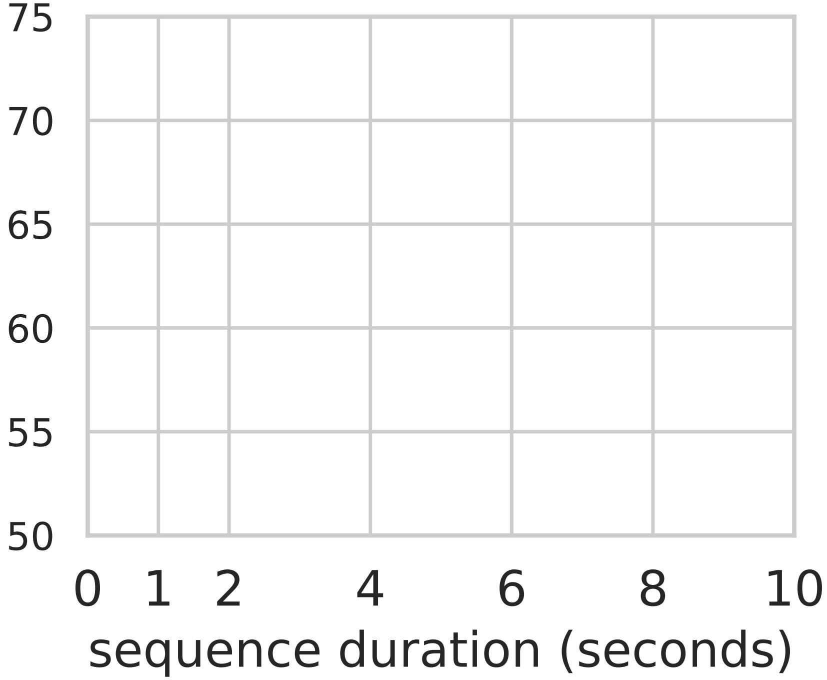
- **sequence duration**
- **+ radios**
- **localization device** — HL, phone
- **per scene** — HGE, LIN, CAB
- **per type of scene** — indoor vs outdoor
- **per query condition** — day vs night

Sequence Localization Results

AR localization recall (%) at 10cm, 1deg



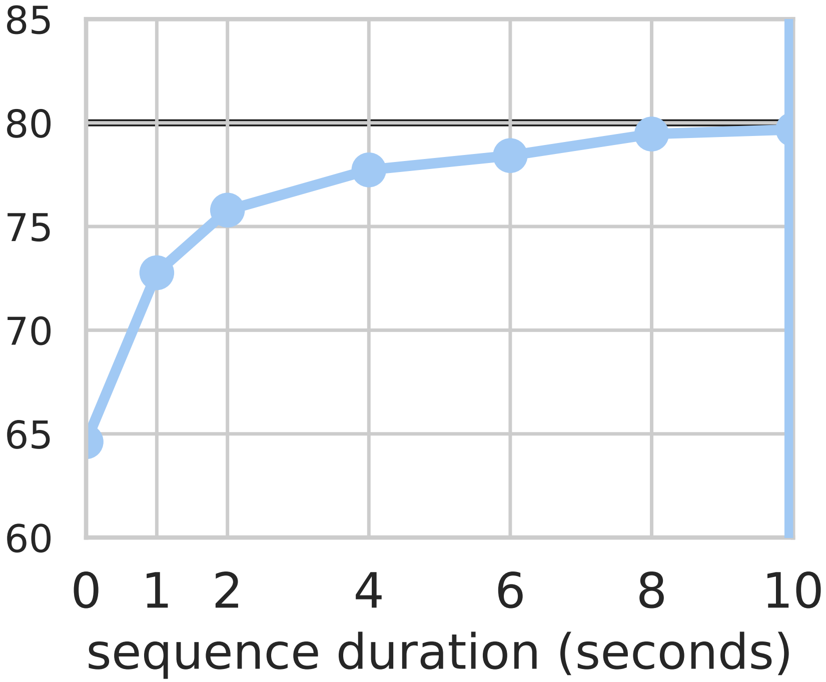
HoloLens 2



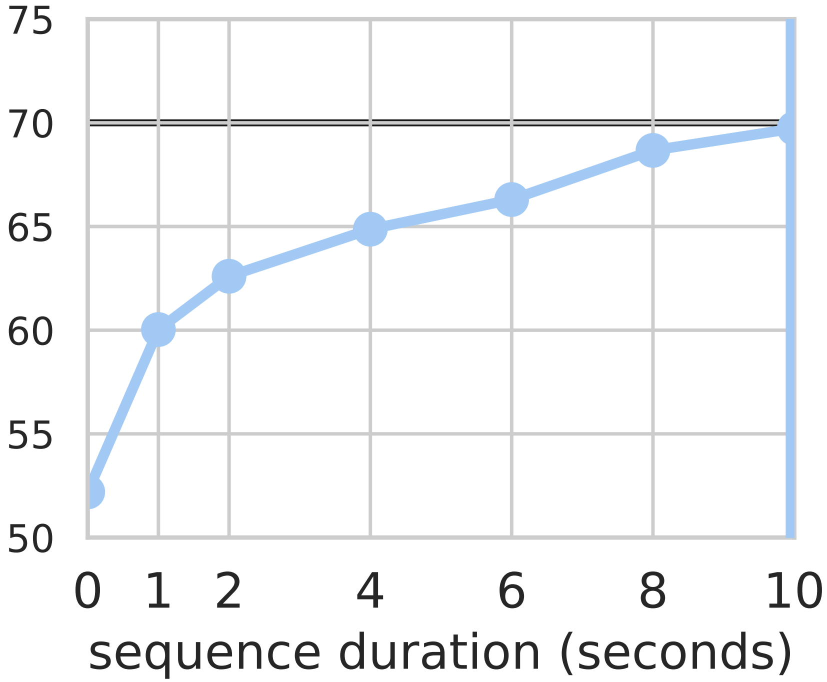
Phone

Sequence Localization Results

AR localization recall (%) at 10cm, 1deg



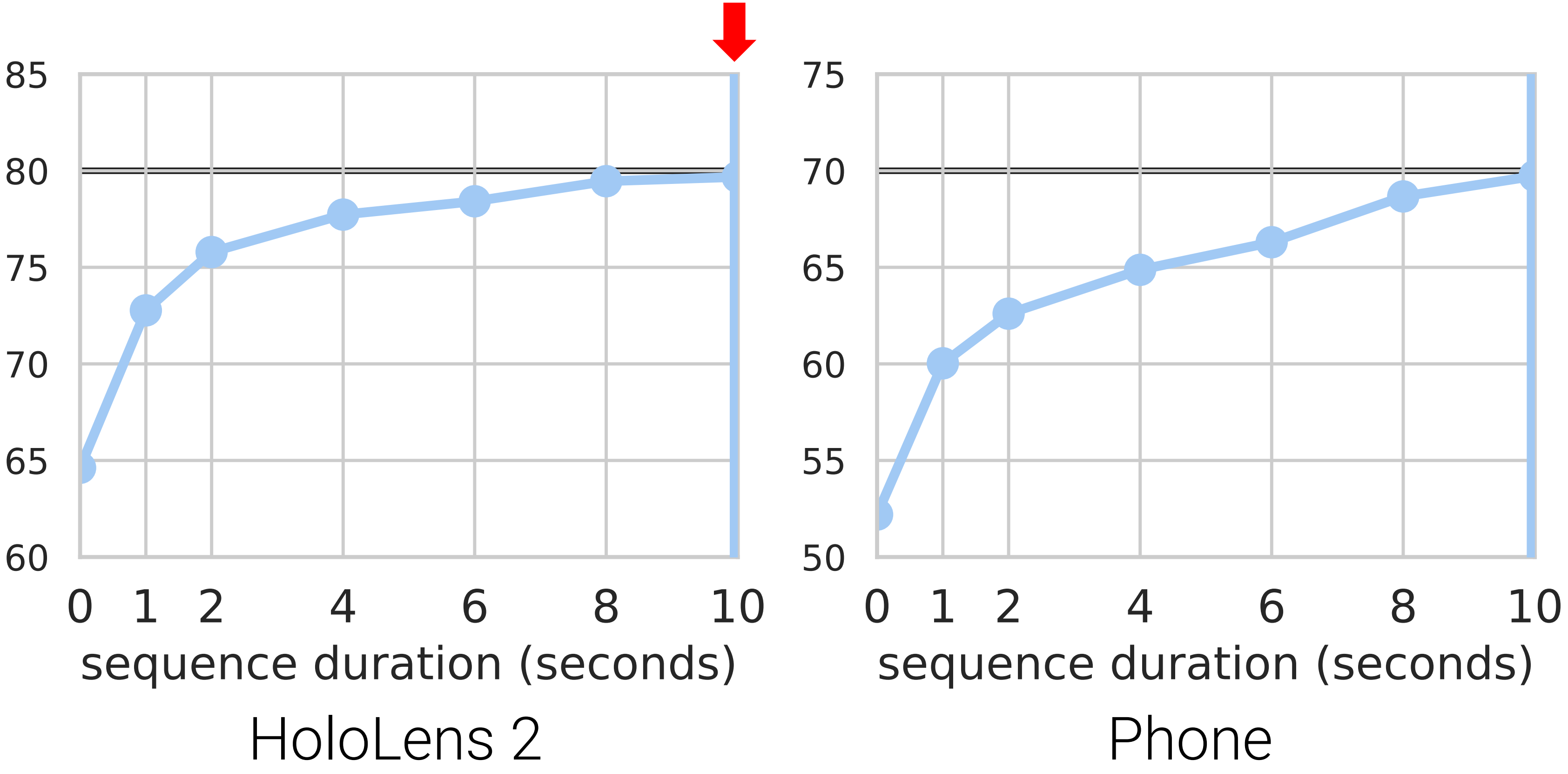
HoloLens 2



Phone

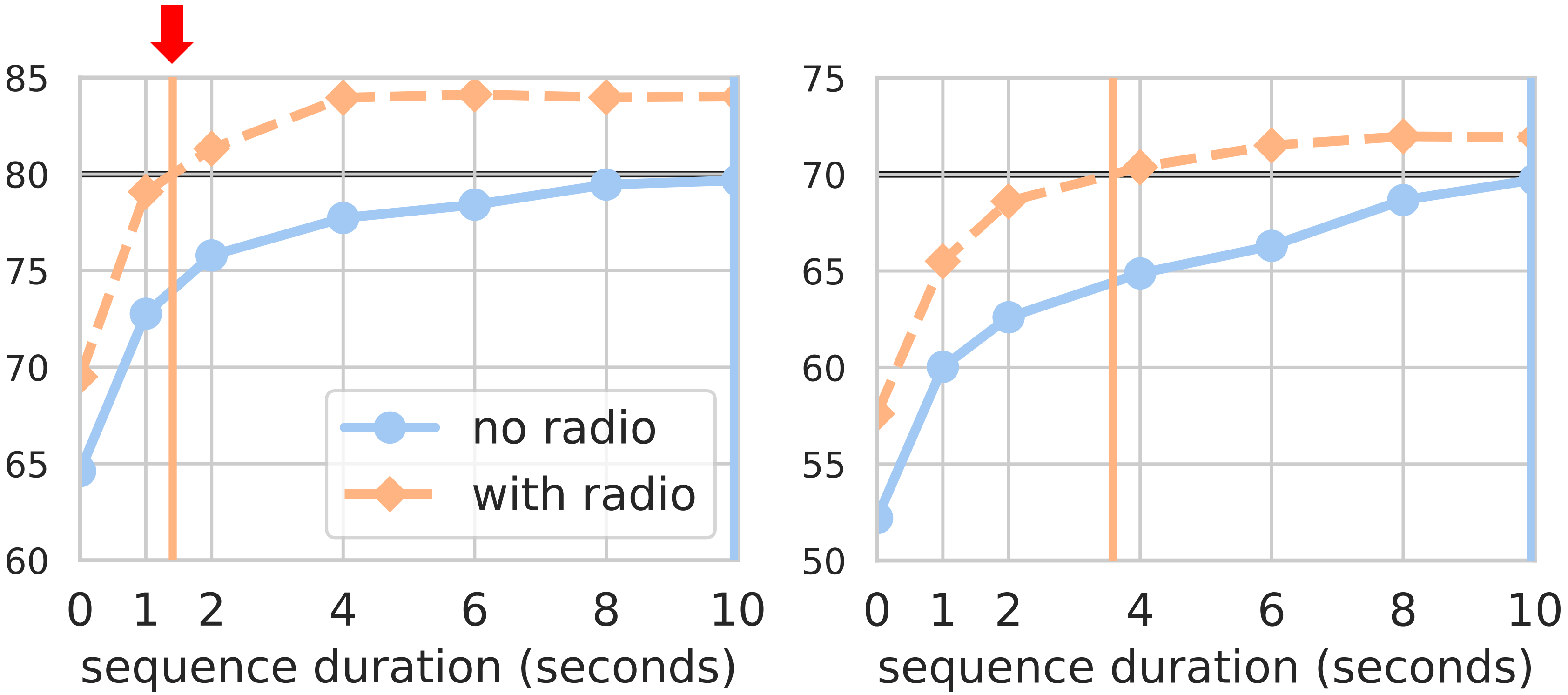
Sequence Localization Results

AR localization recall (%) at 10cm, 1deg



Sequence Localization Results

AR localization recall (%) at 10cm, 1deg

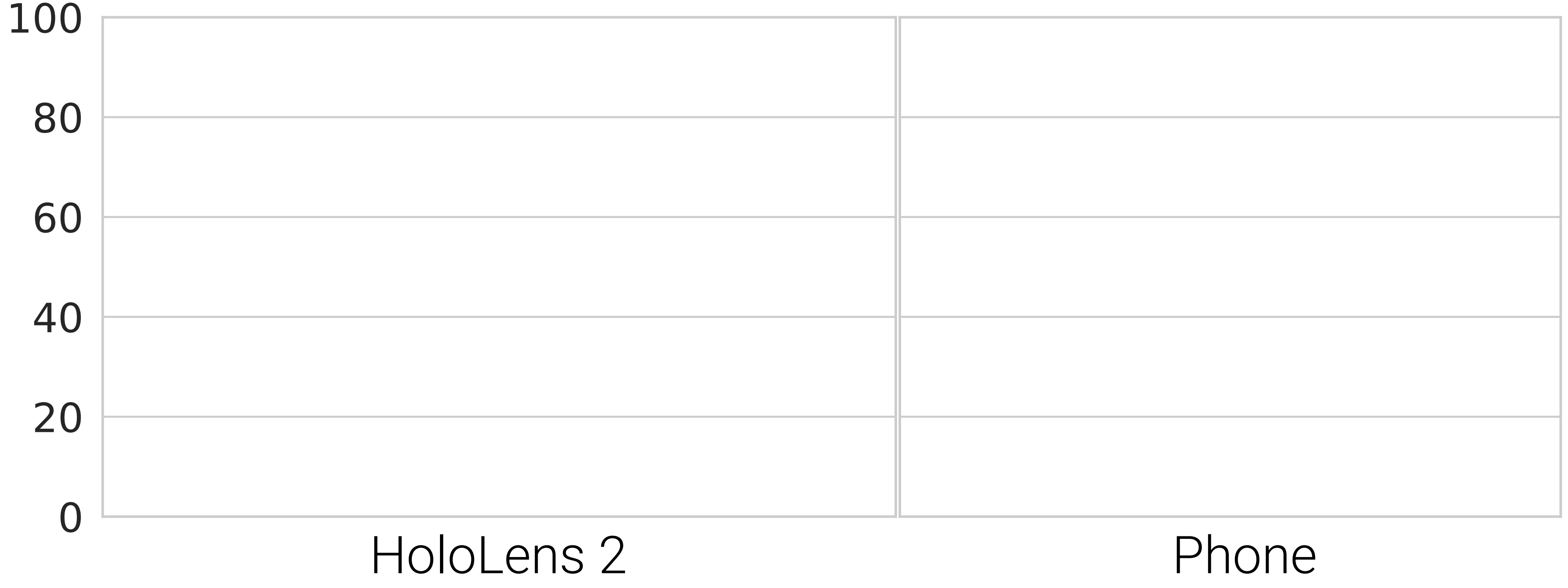


HoloLens 2

Phone

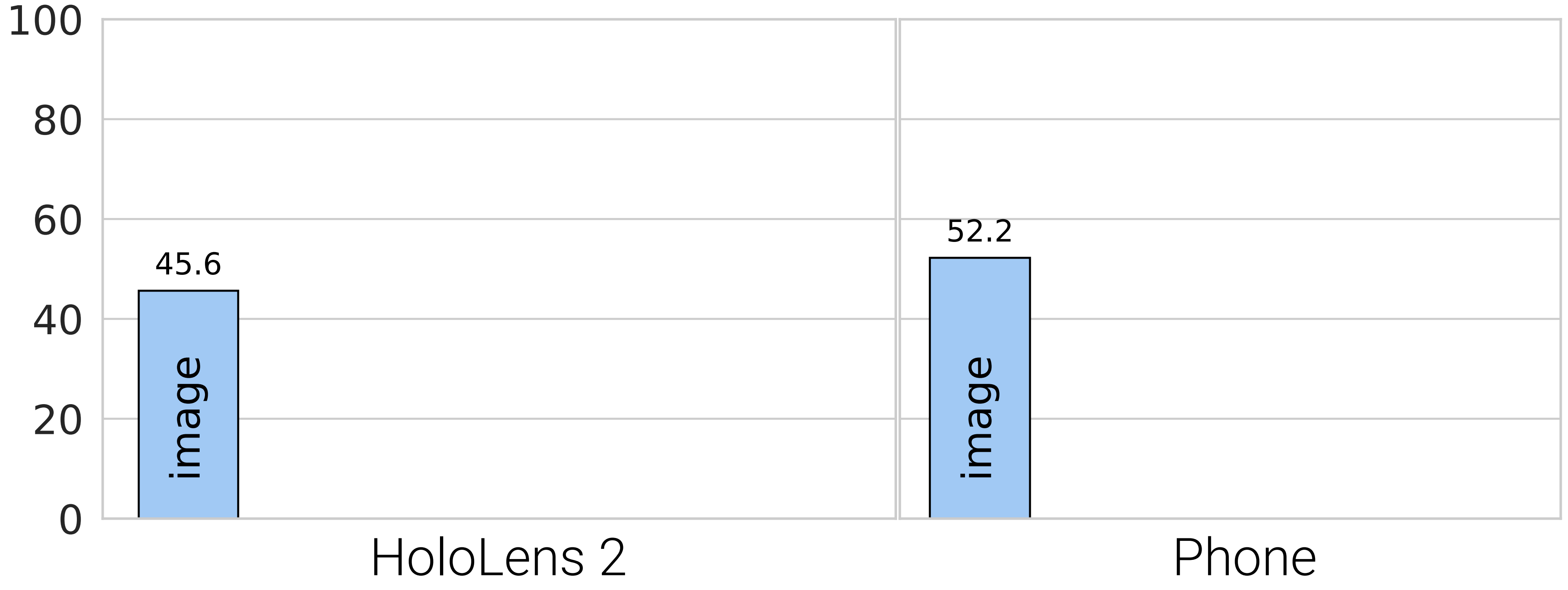
Overview of Results

AR localization recall (%) at 10cm, 1deg



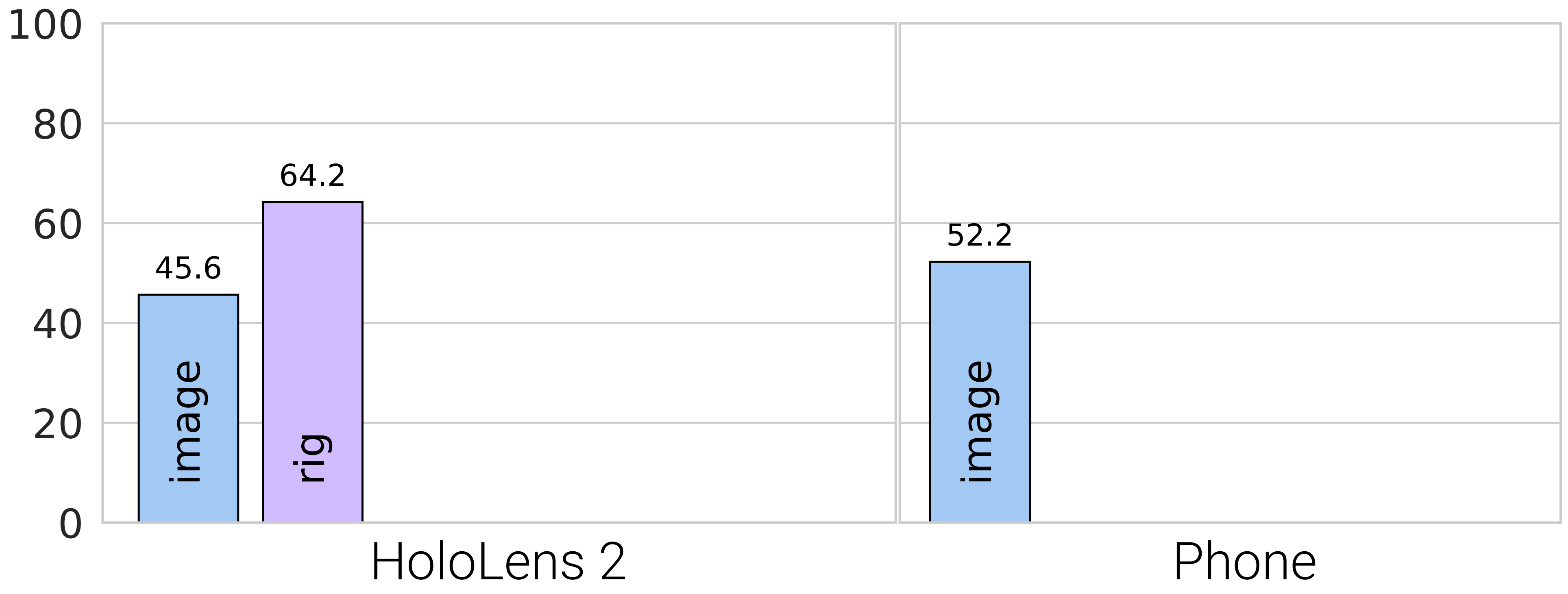
Overview of Results

AR localization recall (%) at 10cm, 1deg



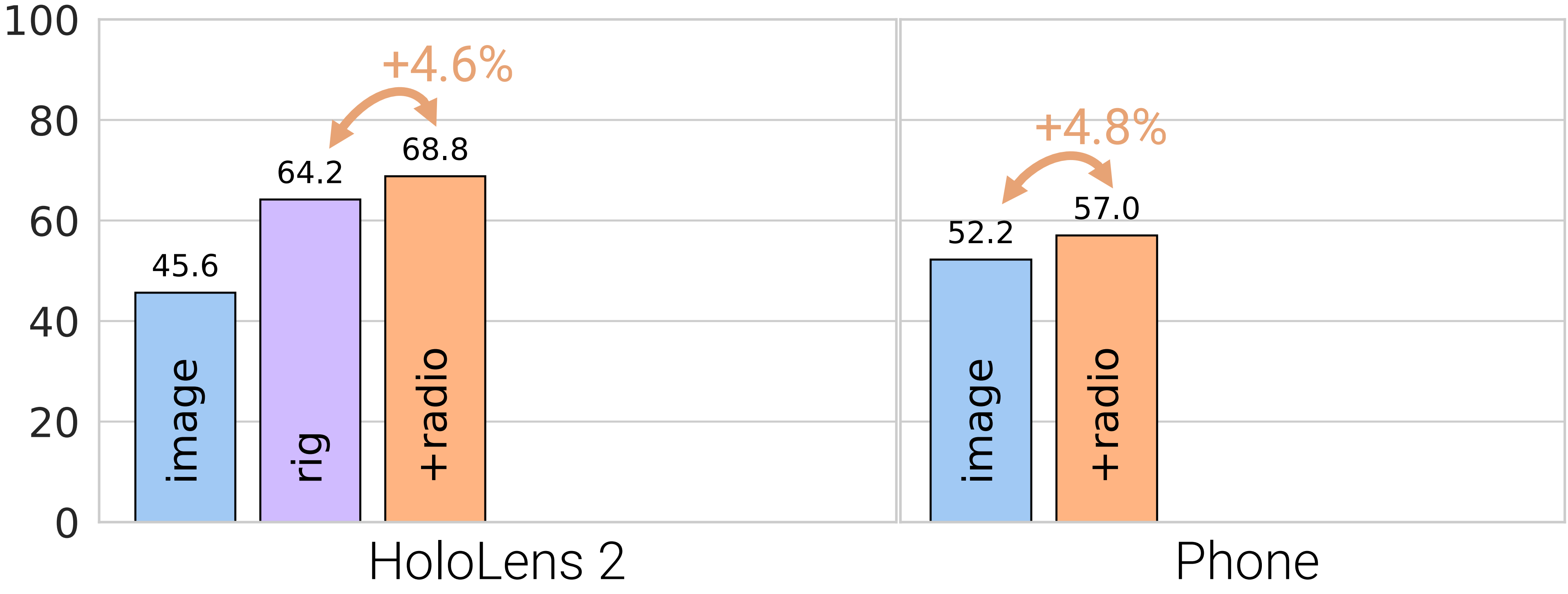
Overview of Results

AR localization recall (%) at 10cm, 1deg



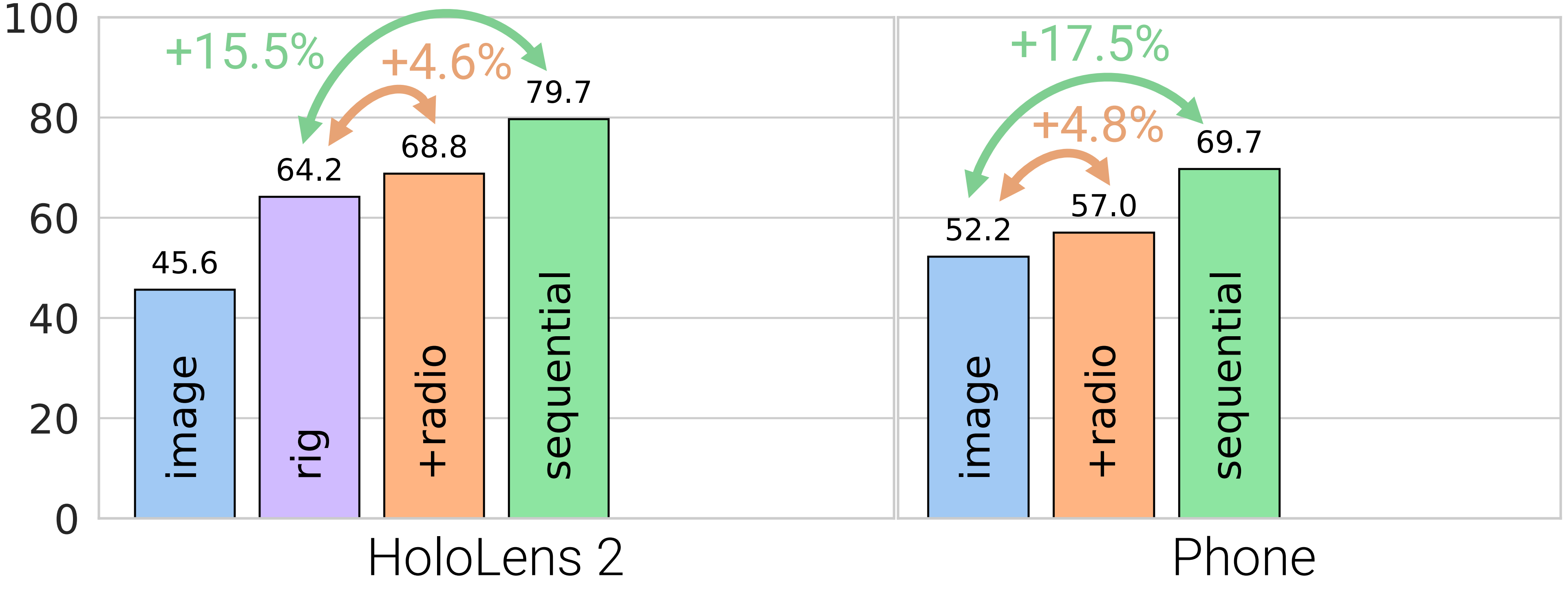
Overview of Results

AR localization recall (%) at 10cm, 1deg



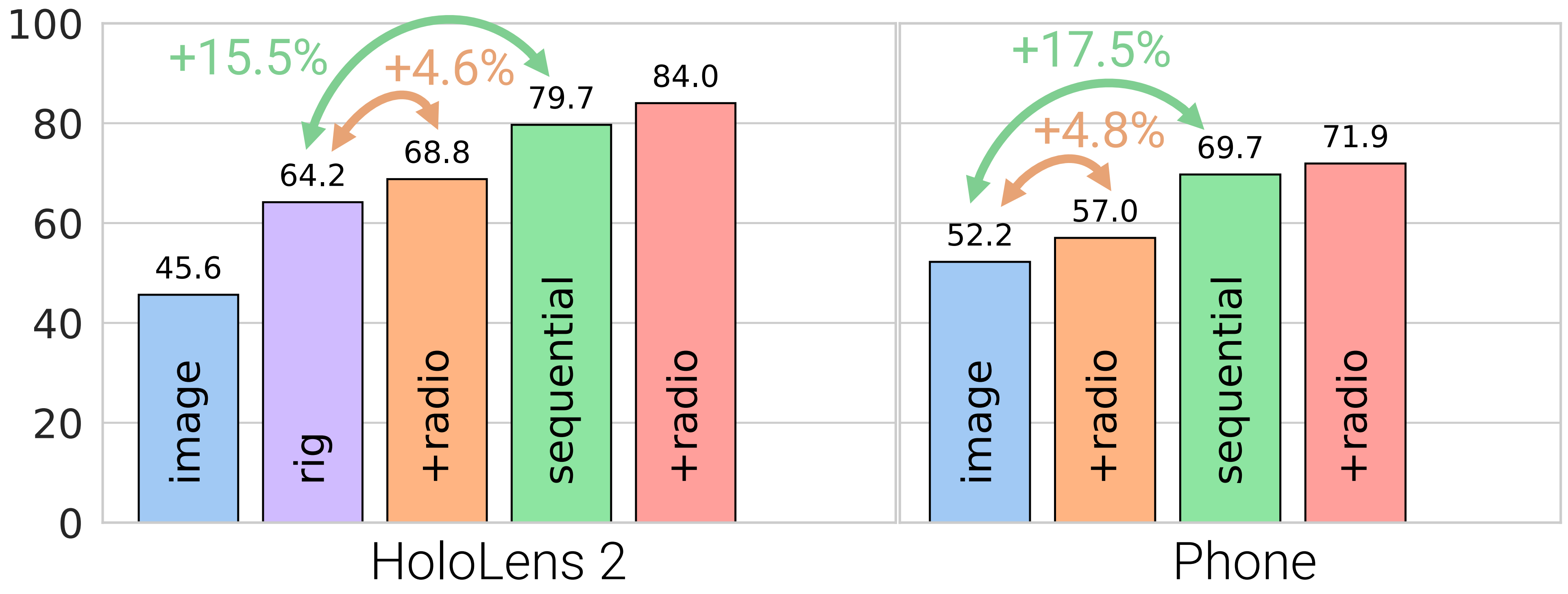
Overview of Results

AR localization recall (%) at 10cm, 1deg



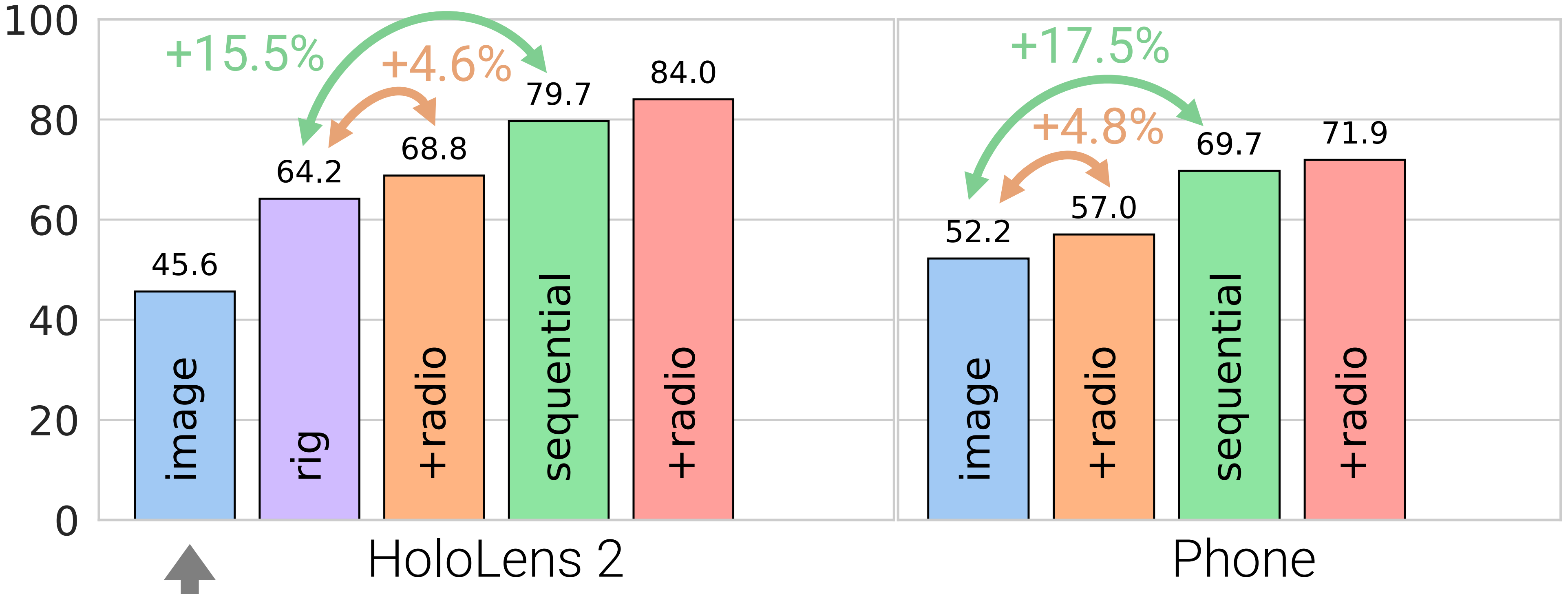
Overview of Results

AR localization recall (%) at 10cm, 1deg



Overview of Results

AR localization recall (%) at 10cm, 1deg

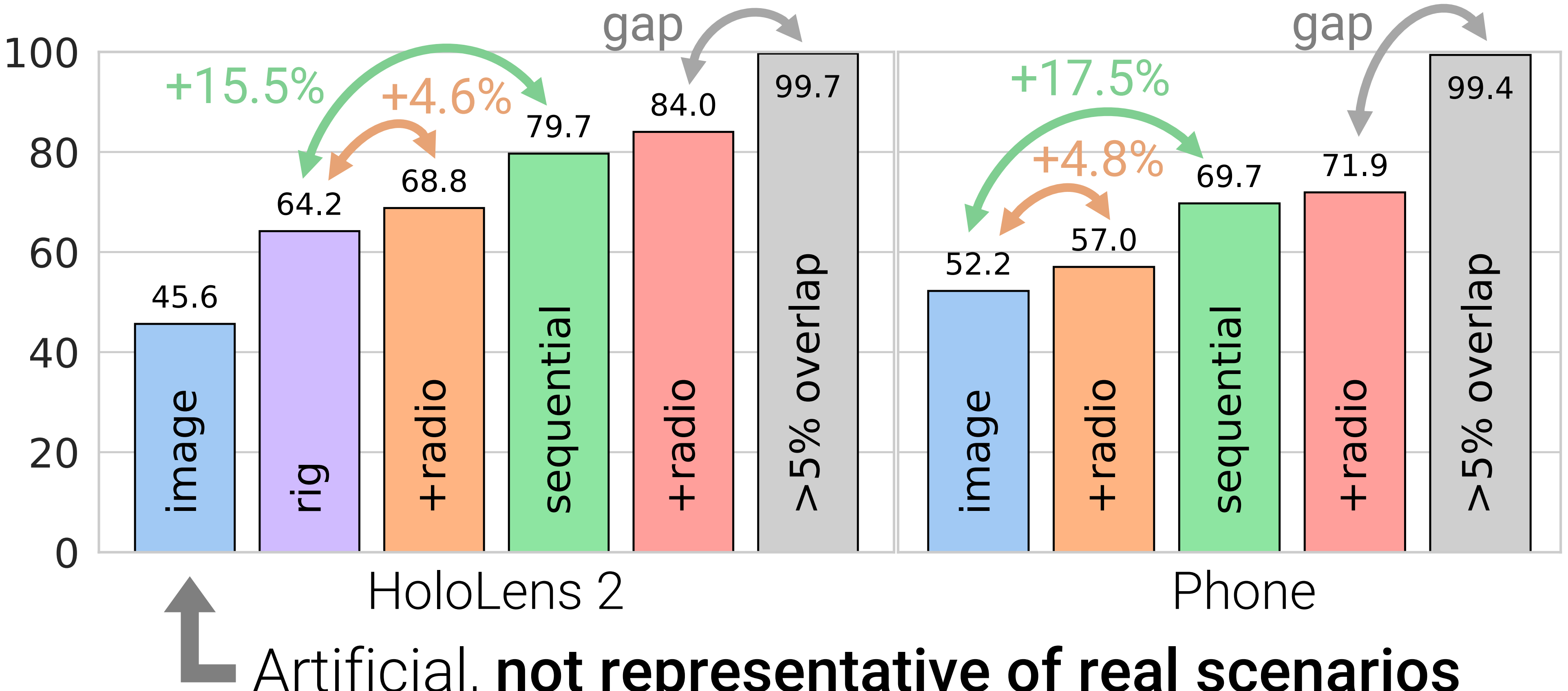


Artificial, **not representative of real scenarios**

ECCV TEL AVIV 2022

Overview of Results

AR localization recall (%) at 10cm, 1deg



Artificial, **not representative of real scenarios**

Further research needed for optimal multi-sensor localization!

c) Limitations and Open Problems

Limitations

- Radios:
  - Our goal: show the potential of exposing radios
  - Due to iOS limitations, we transfer radios from nearby HL trajectories
  - Crowd-sourced: different sensors / detection patterns / attenuation
- We only considered a subset of all possible baselines
  - Contributions are welcome!

Open Problems

- Global features
  - Learned works reasonably good
  - Bias towards imaging sensor – prefers same-device to same-scene
  - For real-world scenarios top 5 should be enough

Open Problems

- Global features
  - Learned works reasonably good
  - Bias towards imaging sensor – prefers same-device to same-scene
  - For real-world scenarios top 5 should be enough
- Local features
  - Need for a learned matcher (SG for other features)
  - Feature detection accuracy matters!

Open Problems

- Global features
  - Learned works reasonably good
  - Bias towards imaging sensor – prefers same-device to same-scene
  - For real-world scenarios top 5 should be enough
- Local features
  - Need for a learned matcher (SG for other features)
  - Feature detection accuracy matters!
- Mapping: more accurate, robust to multi-device

Open Problems

- Global features
  - Learned works reasonably good
  - Bias towards imaging sensor – prefers same-device to same-scene
  - For real-world scenarios top 5 should be enough
- Local features
  - Need for a learned matcher (SG for other features)
  - Feature detection accuracy matters!
- Mapping: more accurate, robust to multi-device
- MVS: compatible with multi-device data

Open Problems

- Global features
  - Learned works reasonably good
  - Bias towards imaging sensor – prefers same-device to same-scene
  - For real-world scenarios top 5 should be enough
- Local features
  - Need for a learned matcher (SG for other features)
  - Feature detection accuracy matters!
- Mapping: more accurate, robust to multi-device
- MVS: compatible with multi-device data
- The goal is <u>TTR@99.9</u>9% as low as possible (range of seconds)
  - Long way to go…

Q&A + Coffee Break

15 minutes