SuperGlue
Learning Feature Matching with Graph Neural Networks

CVPR

ETH *zürich*

magic leap

psarlin.com/superglue

Paul-Edouard Sarlin[1]   Daniel DeTone[2]   Tomasz Malisiewicz[2]   Andrew Rabinovich[2]

[1]ETH Zurich   [2]Magic Leap

---

1. A Learnable Middle-end

Our problem:
• Inputs: two sets of **local features** - keypoints & local descriptors
• Output: strong & outlier-free **matches**
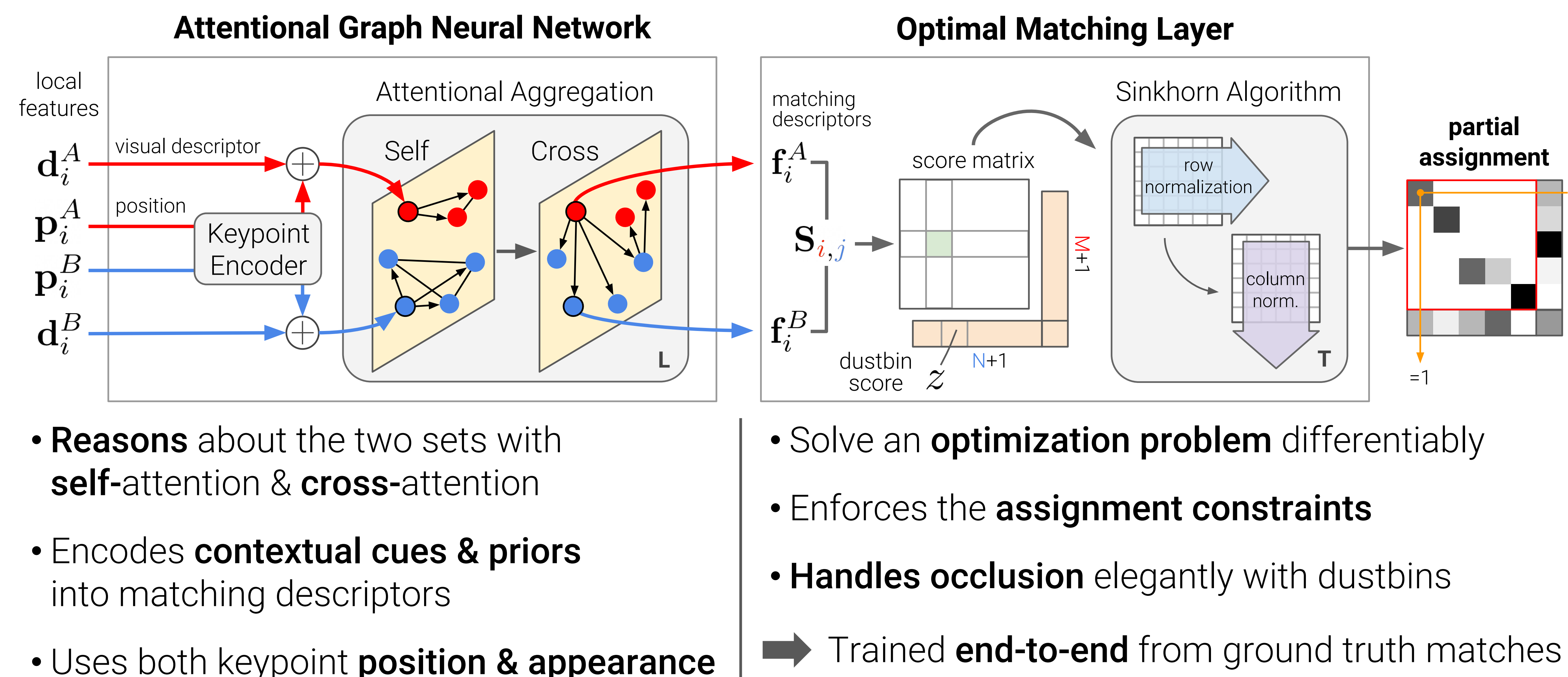• Applications: SLAM, SfM, visual localization

Super Glue

image pair

local features

CNN

Graph Neural Network

Back-End Optimizer

strong matches

Detector & Descriptor
*Deep **Front-End***

SuperGlue
*Deep **Middle-End** Matcher*
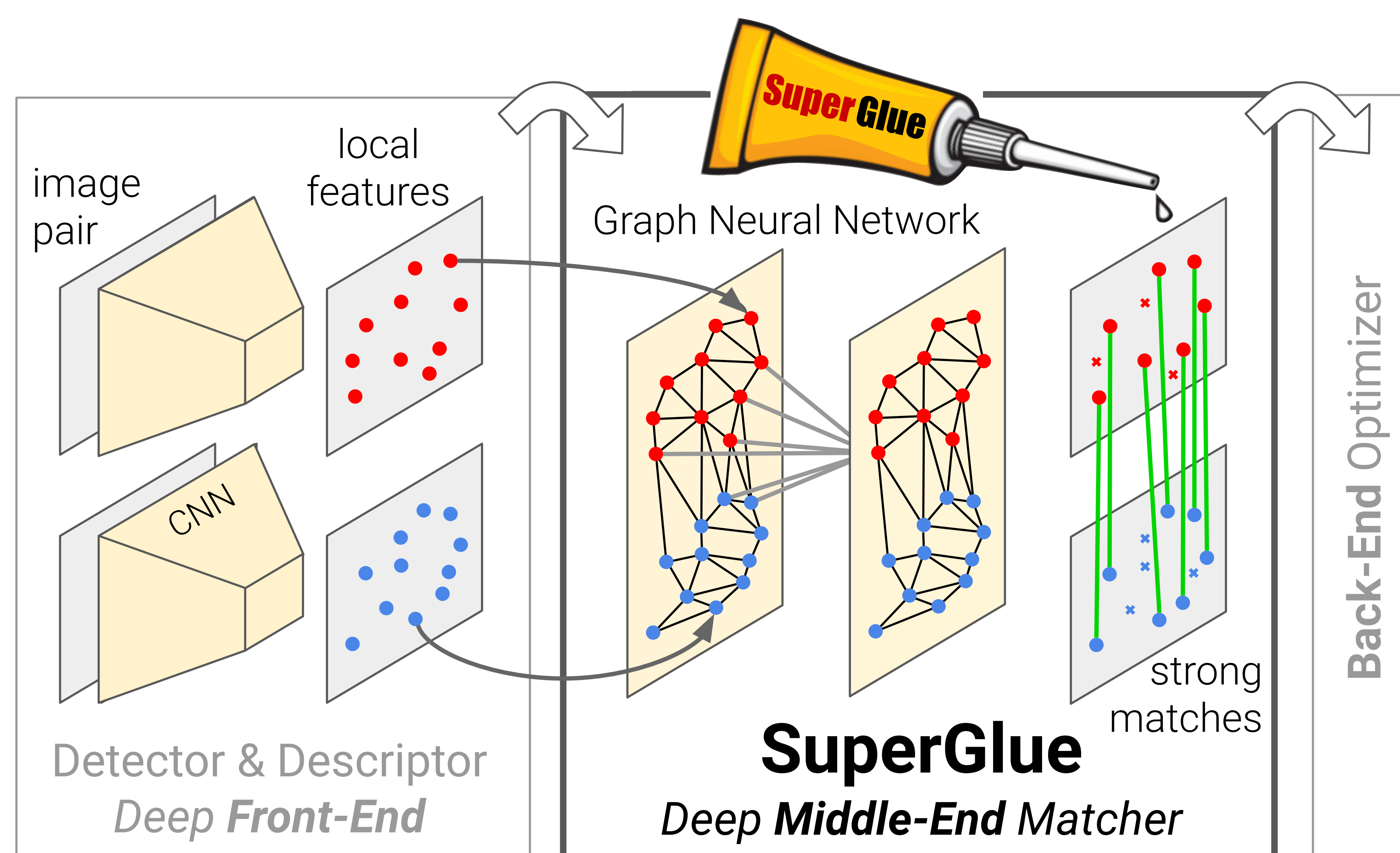
• Combine deep learning with classical optimization
  → Graph Neural Networks, Attention, Optimal Transport

• <u>Robust:</u> handles **extreme wide-baseline** image pairs

• <u>Real-time:</u> 15 FPS on GPU, scales to 1000s of keypoints

• <u>SOTA matching</u> for **indoor+outdoor** with **SIFT** & **SuperPoint**

• <u>Interpretable:</u> learn **complex reasoning & priors** about the world

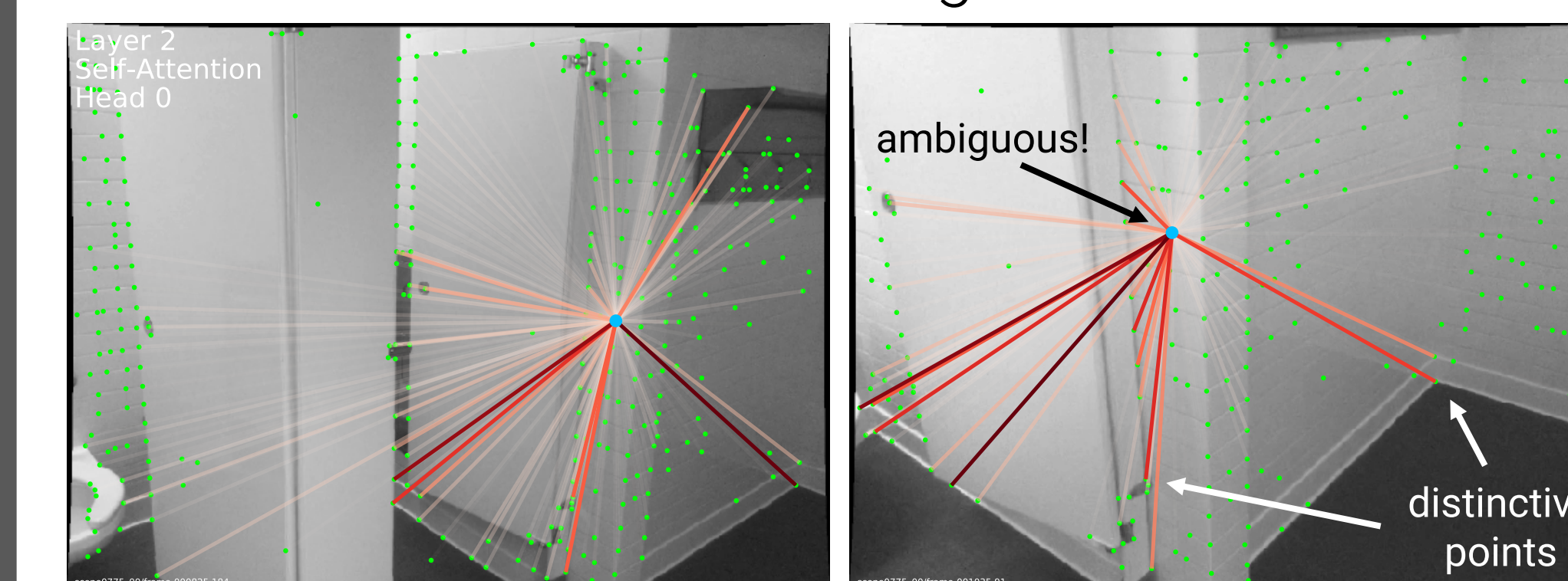SuperGlue: context aggregation + matching + filtering

detection → description → feature matching → outlier filtering → pose estimation

DoG+SIFT/HardNet, SuperPoint, D2-Net, R2D2 | Nearest Neighbor Matching | Ratio test, mutual check, GMS, OA-Net | RANSAC, Bundle Adjustment

---

2. Graph Neural Network + Optimal Transport

Attentional Graph Neural Network

Attentional Aggregation

local features

visual descriptor

$\mathbf{d}_i^A$

position $\mathbf{p}_i^A$

$\mathbf{p}_i^B$

$\mathbf{d}_i^B$

Keypoint Encoder

Self   Cross

L

Optimal Matching Layer

matching descriptors

$\mathbf{f}_i^A$

$\mathbf{S}_{i,j}$

$\mathbf{f}_i^B$

score matrix

M+1

dustbin score z N+1

Sinkhorn Algorithm

row normalization

column norm.

T

partial assignment

=1

• **Reasons** about the two sets with **self**-attention & **cross**-attention

• Encodes **contextual cues & priors** into matching descriptors

• Uses both keypoint **position & appearance**

• Solve an **optimization problem** differentiably

• Enforces the **assignment constraints**

• **Handles occlusion** elegantly with dustbins

➡ Trained **end-to-end** from ground truth matches

---

3. Interpretable

Self-attention = intra-image information flow

ambiguous!

distinctive points

Cross-attention = inter-image communication

potential matches

highest attention weight
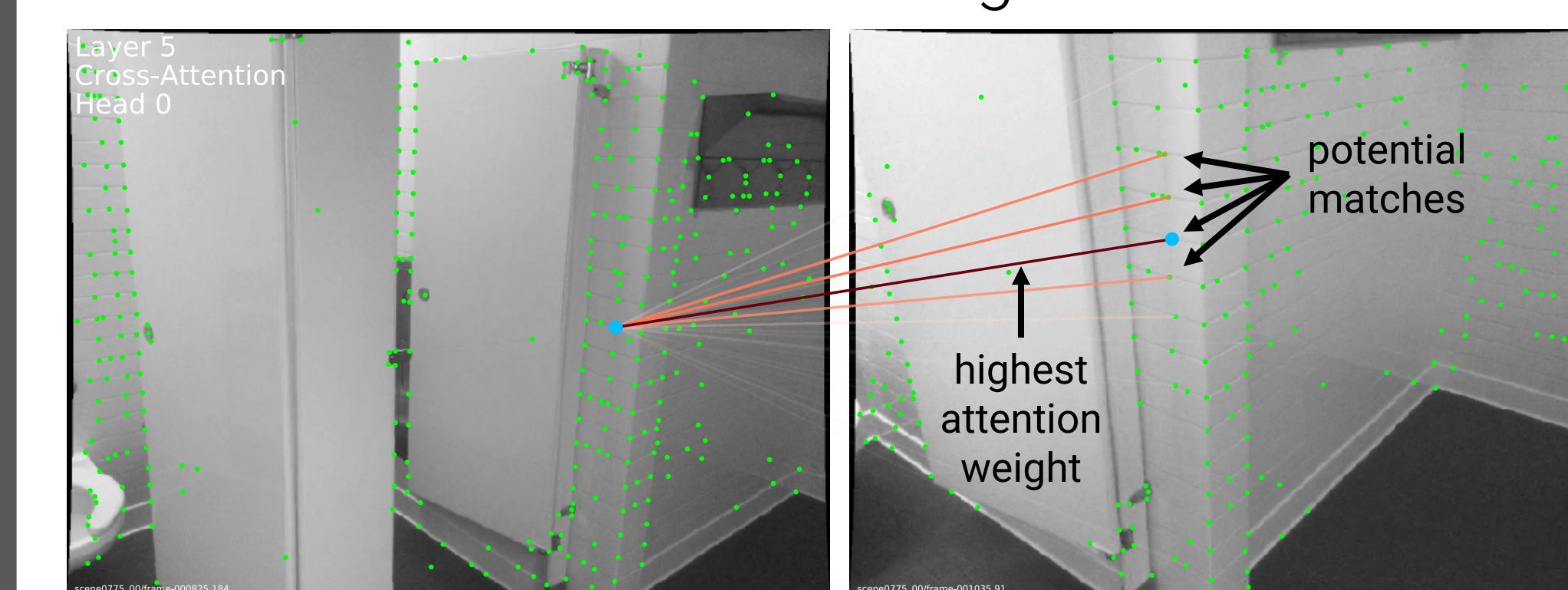
= a **soft**, **data-dependent**, **sparse graph**
to better understand feature matching

---

4. Results: State-of-the-art Relative Pose Estimation

indoor: ScanNet (RGBD)

outdoor: phototourism SfM

AUC@20° rotation+translation error

Indoor: 22.5, 25.9, 28.7, 36.4, 43.8, 51.8

Outdoor: 35.3, 40.3, 49.4, 30.9, 46.9, 64.2

SIFT + NN + ratio test
SIFT + NN + OANet
SIFT + **SuperGlue**
SuperPoint + NN + mutual
SuperPoint + NN + OANet
SuperPoint + **SuperGlue**
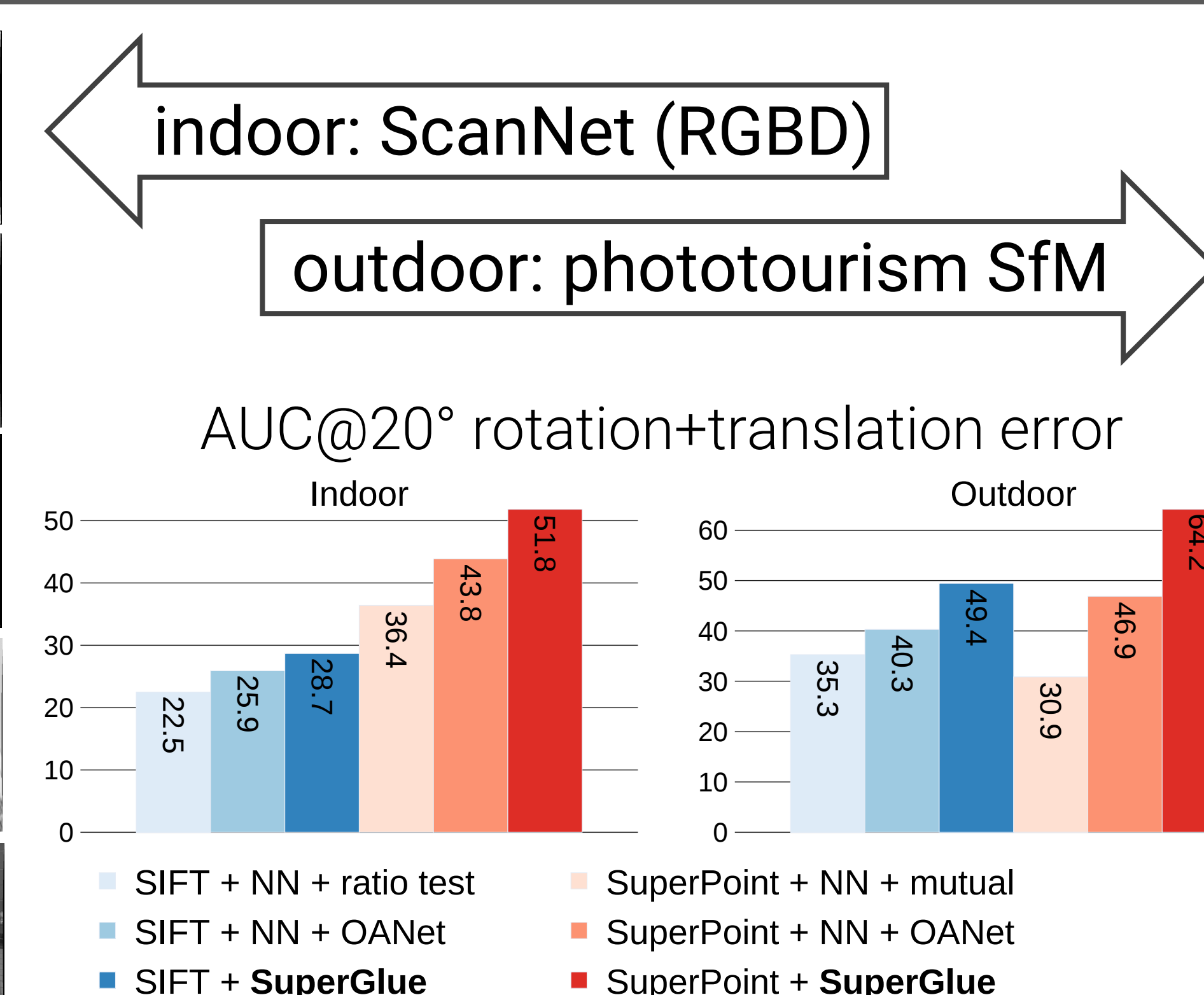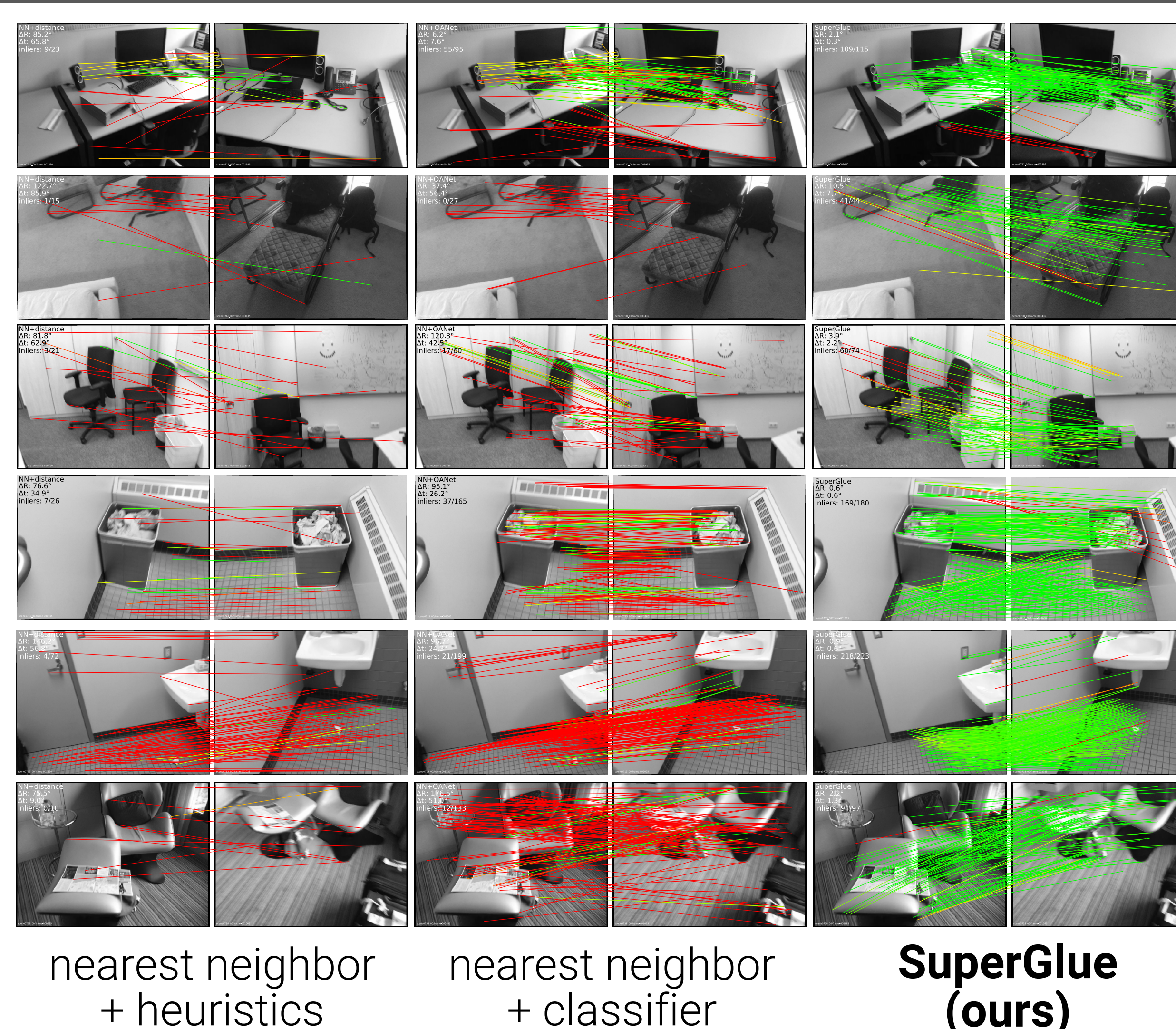
• SuperGlue > classifiers > heuristics
• SuperPoint + SuperGlue > all
• Generalizes well for SfM, localization

nearest neighbor + heuristics | nearest neighbor + classifier | **SuperGlue (ours)**

nearest neighbor + heuristics | nearest neighbor + classifier | **SuperGlue (ours)**