Robust Pose Optimization Made Differentiable

Eric Brachmann

5th International Workshop on Recovering 6D Object Pose @ICCV19

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Visual Learning Lab Heidelberg

European Research Council
Established by the European Commission

Background

Dr.

Eric Brachmann

@eric_brachmann

2012-2017

PhD at TECHNISCHE UNIVERSITÄT DRESDEN

since 2018

Post-Doc at Visual Learning Lab Heidelberg

since 2019

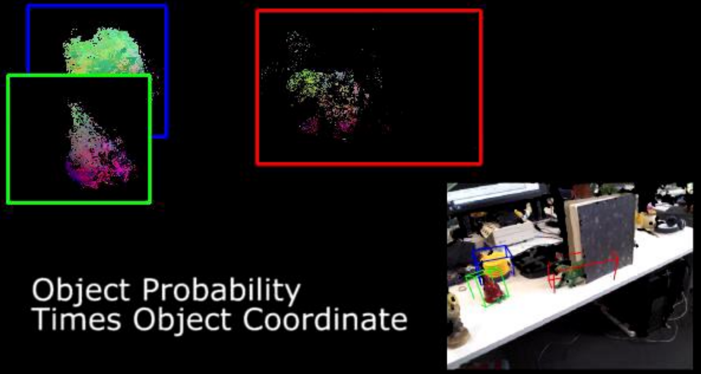
Guest at Leibniz Universität Hannover

Prof.

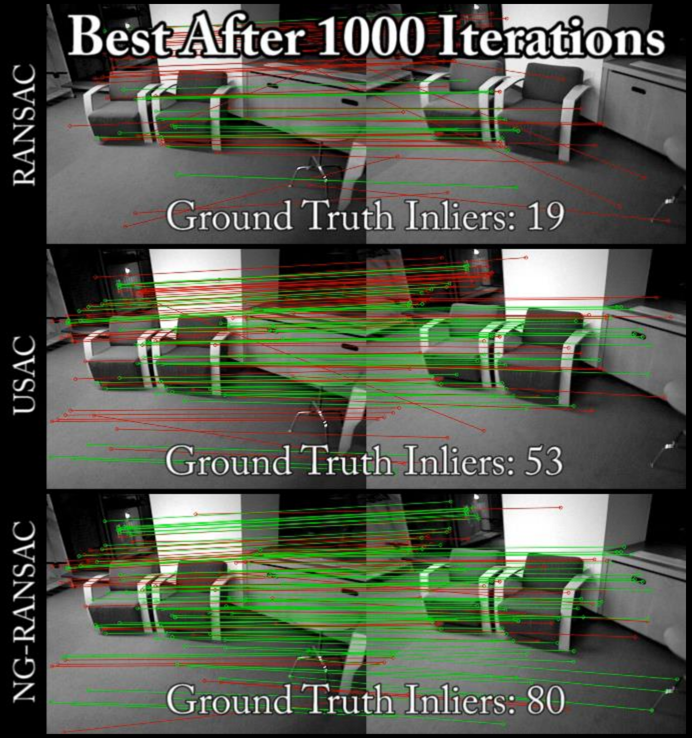
Carsten Rother

Main Research Interests

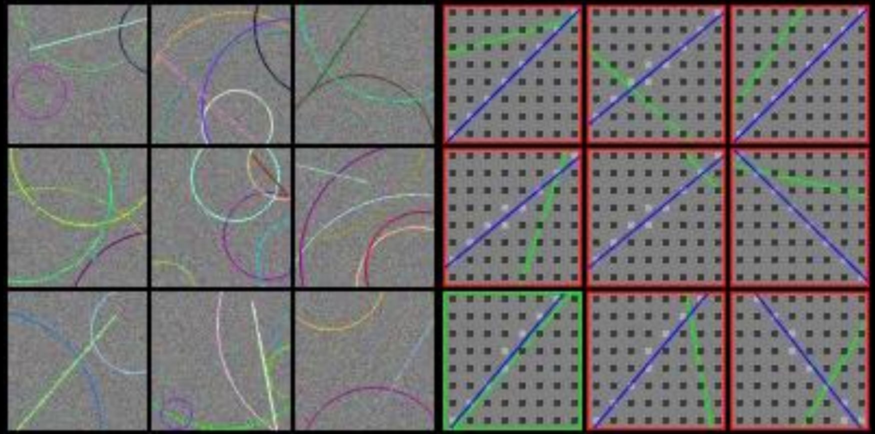
- Machine learning and projective geometry
- Robust fitting with (differentiable) RANSAC
    - Object poses
    - Camera poses
    - Lines
    - Epipolar Geometry



Object Coordinates – ECCV‘14



NG-RANSAC – ICCV‘19



DSAC – CVPR‘17

DSAC++ – CVPR‘18

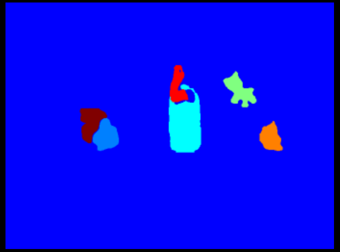
Goal

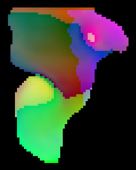
RGB(-D) Image I

Pose Estimation Pipeline

6D Poses $\hat{\mathbf{h}}_o$

Object Detection	Object Classification	Correspondence Prediction	RANSAC: Pose Solver, Pose Scoring	Pose Loss



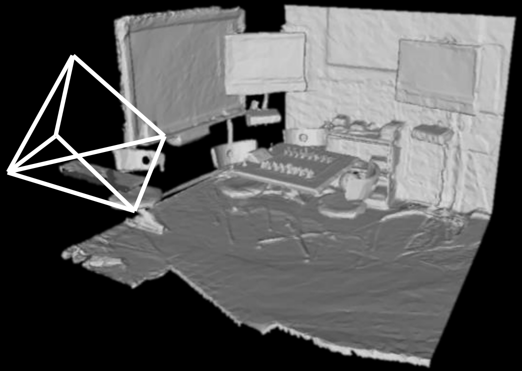

“Learning 6D object pose estimation using 3D object coordinates”, Brachmann et al., ECCV’14
“iPose: instance-aware 6D pose estimation of partly occluded objects”, Jafari et al., ACCV’18
“Segmentation-driven 6D Object Pose Estimation”, Hu et al., CVPR’19
“Pix2Pose: Pixel-Wise Coordinate Regression of Objects for 6D Pose Estimation”, Park et al., ICCV’19
“DPOD: 6D Pose Object Detector and Refiner”, Zakharov et al., ICCV’19
…

Why End-to-End?

RGB(-D) Image I

Pose Estimation Pipeline

6D Camera Pose $\hat{\mathbf{h}}$



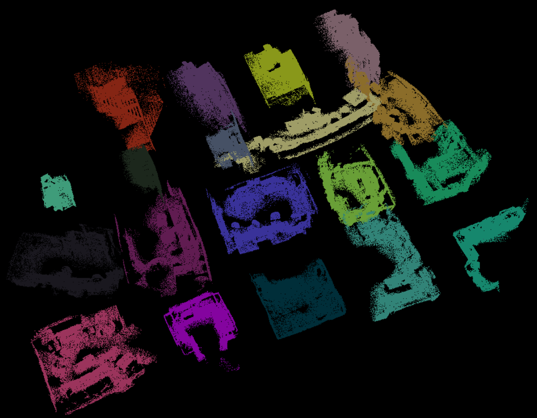
~~Object Detection~~

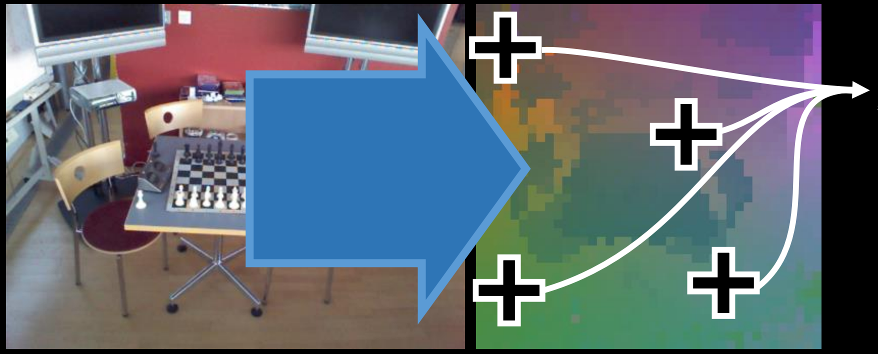
Object Classification

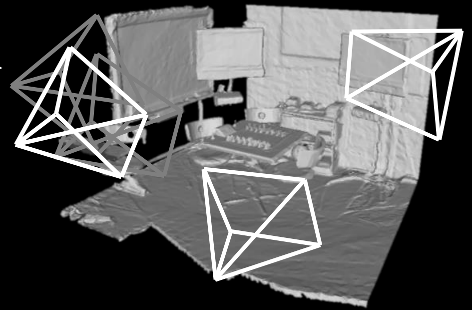
Correspondence Prediction

RANSAC
- Pose Solver
- Pose Scoring

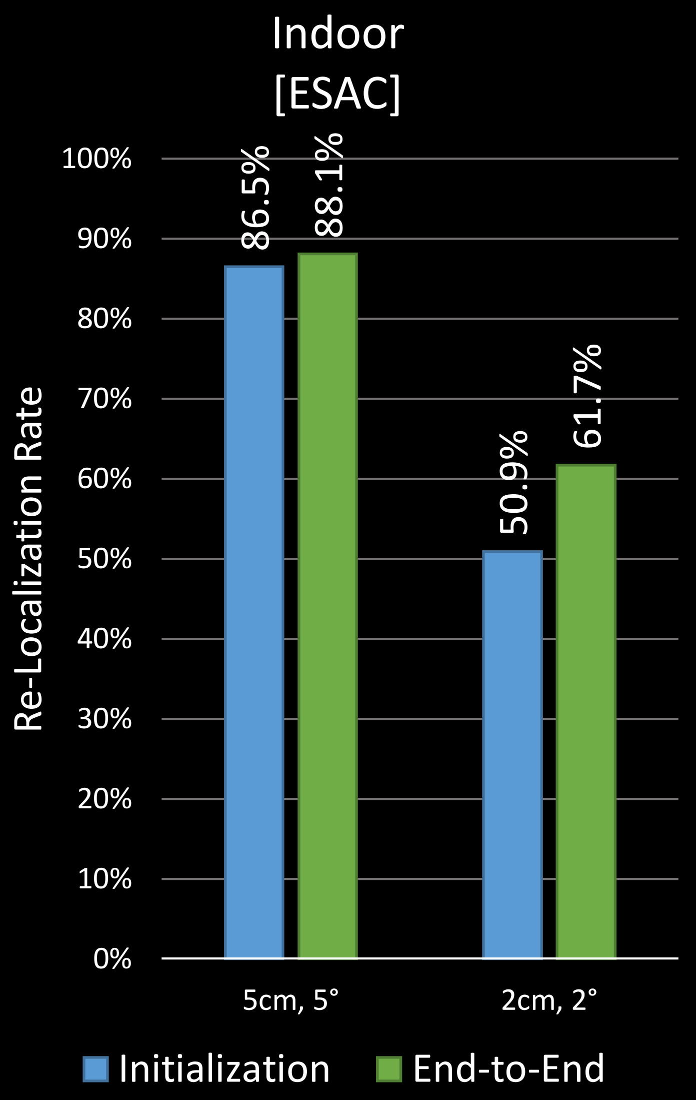
Pose Loss

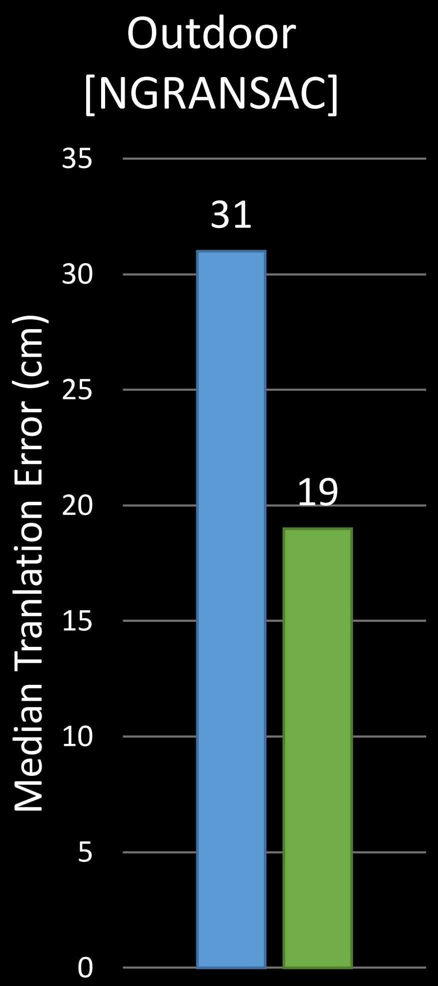


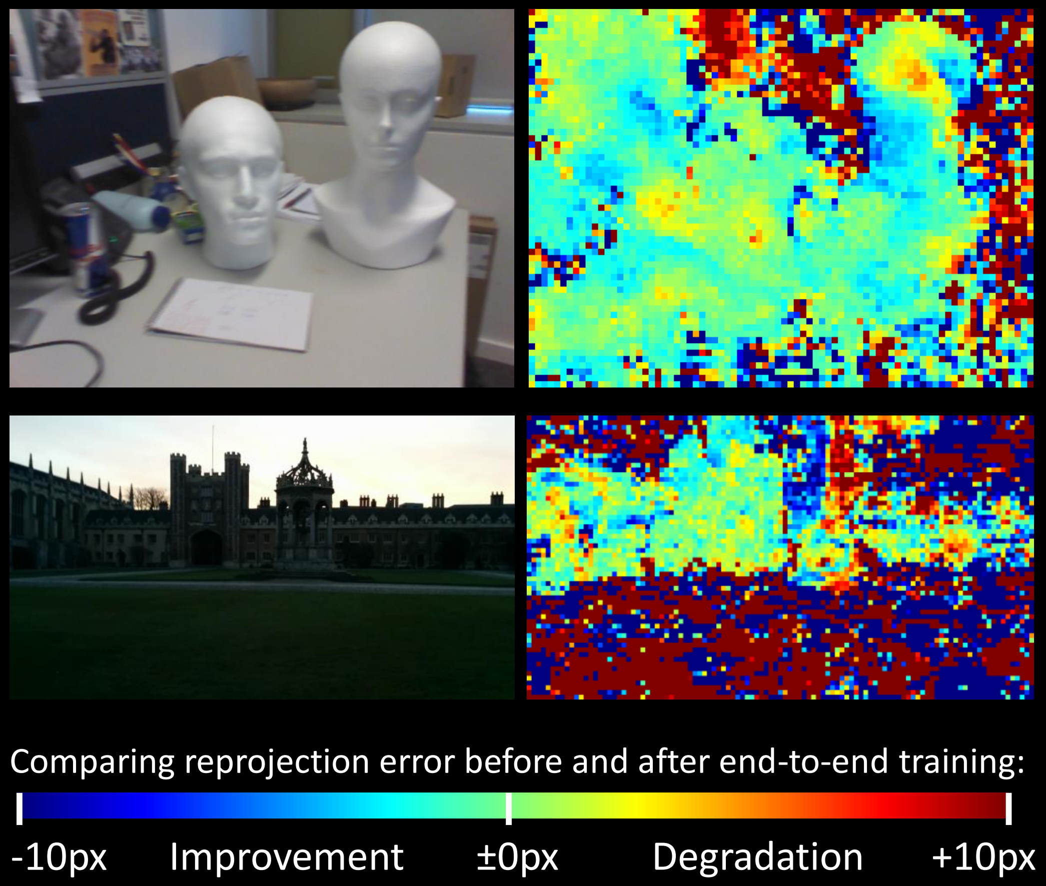




Why End-to-End?







[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19
[NGRANSAC] "Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses", Brachmann and Rother, ICCV19

Roadmap

Object Detection	Object Classification	Correspondence Prediction	RANSAC: Pose Solver / Pose Scoring	Pose Loss

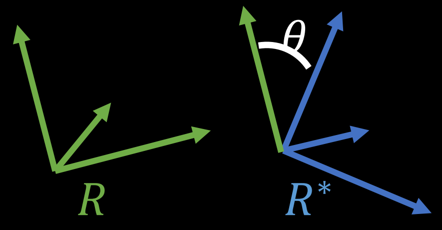
Pose Loss (RGB-D)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | **Pose Loss**

Input: RGB-D


$$\ell(\mathbf{t}, \mathbf{t}^*) + \alpha \ell(R, R^*) \text{ with } \mathbf{h} = (\mathbf{t}, R)$$

$$\|\mathbf{t} - \mathbf{t}^*\| \qquad \left\|\log(R^* R^{\mathrm{T}})\right\| \text{ with } \log(R): \mathbb{R}^{3\times 3} \rightarrow \mathbb{R}^3$$



in OpenCV:
cv2.Rodrigues()
incl. gradients

Pose Loss (RGB)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | **Pose Loss**

Input: RGB

$$\ell_\pi(\mathbf{h},\mathbf{h}^*) = \frac{1}{|\mathcal{V}|}\sum_{\mathbf{v}\in\mathcal{V}} ||C\mathbf{h}^*\mathbf{v} - C\mathbf{h}\mathbf{v}|| \text{ [Bra16]}$$

$\mathcal{V}$... Model vertices

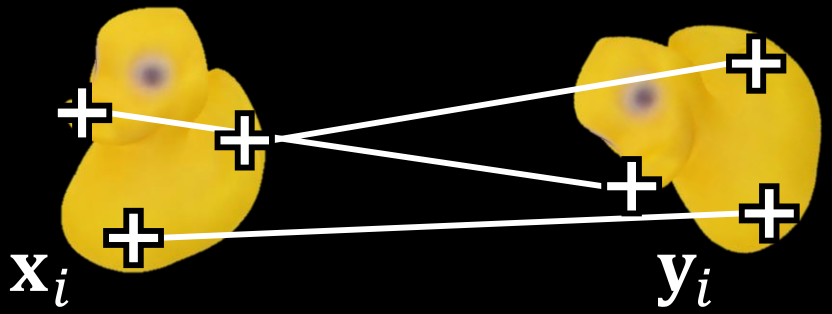
C... Camera calibration matrix

[Bra16] Brachmann et al., “Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image”, CVPR 2016

Pose Solver (RGB-D)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Input: RGB-D



$$(\hat{R}, \hat{\mathbf{t}}) = \underset{R,\mathbf{t}|RR^T=1}{\operatorname{argmin}} \sum_i \|\mathbf{x}_i - (R\mathbf{y}_i - \boldsymbol{t})\|^2$$

Kabsch Algorithm:

$$\operatorname{cov}[\mathbf{x}_\mathrm{i}, \mathbf{y}_\mathrm{i}] = \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^T$$

$$\operatorname{cov}[\mathbf{x}_\mathrm{i}, \mathbf{y}_\mathrm{i}] = U\Sigma V^T$$

$$\hat{R} = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(VU^T) \end{pmatrix} U^T$$

$$\hat{\mathbf{t}} = \hat{R}\bar{\mathbf{y}}\text{-}\bar{\mathbf{x}}$$

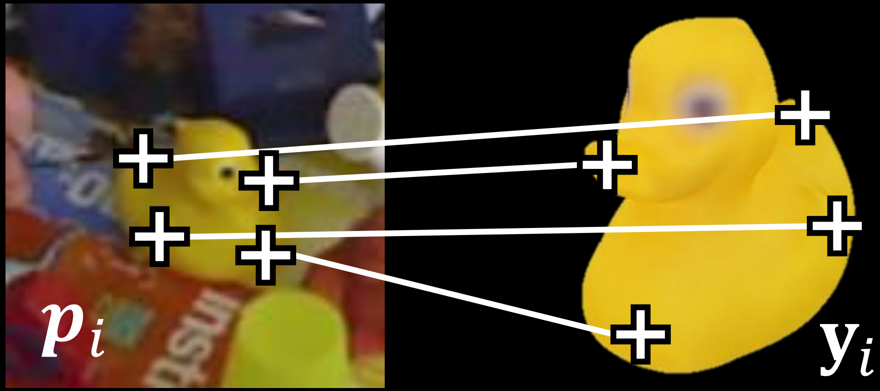
C++ code with PyTorch integration coming soon.

[Kab76] Kabsch, “A solution for the best rotation to relate two sets of vectors”, Acta Crystallographica, 1976

Pose Solver (RGB)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Input: RGB



$$(\hat{R}, \hat{\mathbf{t}}) = \underset{R,\mathbf{t}}{\operatorname{argmin}} \sum_i \|\mathbf{p}_i - C(R\mathbf{y}_i - \mathbf{t})\|^2$$

Solving Perspective-n-Point:

Initialization → Gauss-Newton

[Lep09] Lepetit et al., “EPnP: An Accurate O(n) Solution to the PnP Problem”, IJCV’09

[Gao03] Gao et al., “Complete Solution Classification for the Perspective-Three-Point Problem”, TPAMI’03

Pose Solver (RGB)

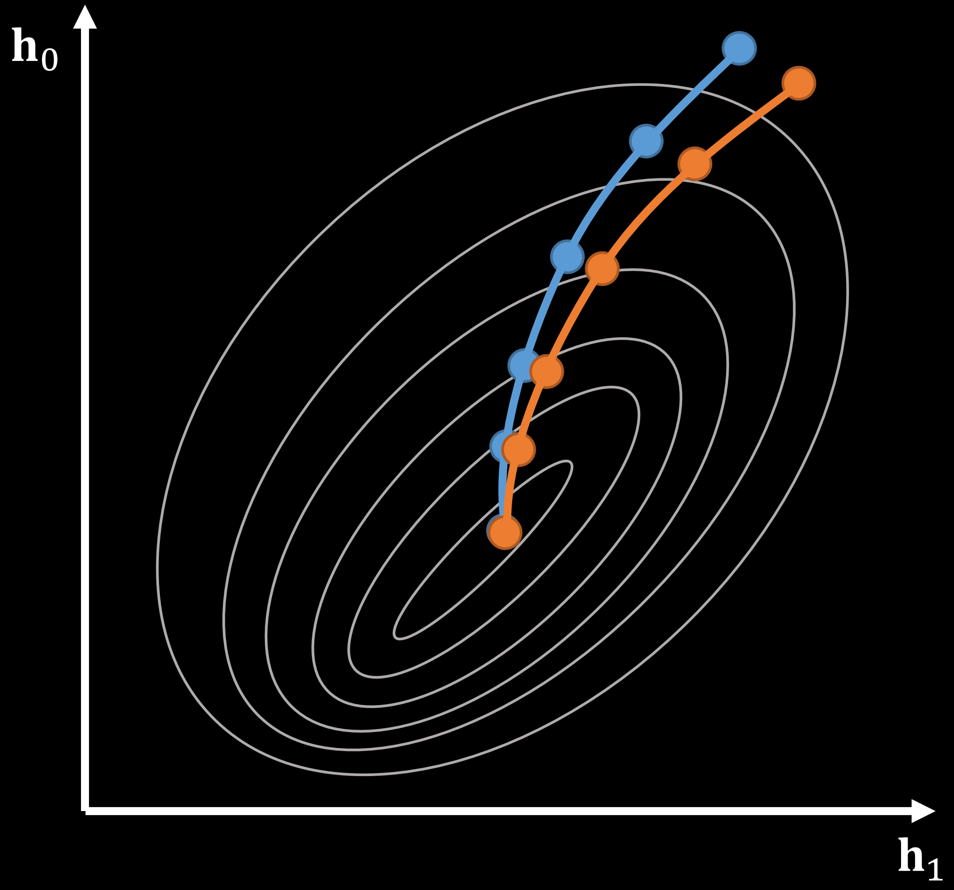
Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Initialization → Gauss-Newton

Residual vector: $[\mathbf{r}(\mathbf{h})]_i = \|\mathbf{p}_i - C\mathbf{h}\mathbf{y}_i\|^2$

Update Rule: $\mathbf{h}^{t+1} = \mathbf{h}^t - (J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \mathbf{r}(\mathbf{h}^t)$

Jacobean: $[J_\mathbf{r}]_{ij} = \frac{\partial[\mathbf{r}(\mathbf{h}^t)]_i}{\partial[\mathbf{h}^t]_j}$



Pose Solver (RGB)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Initialization → Gauss-Newton

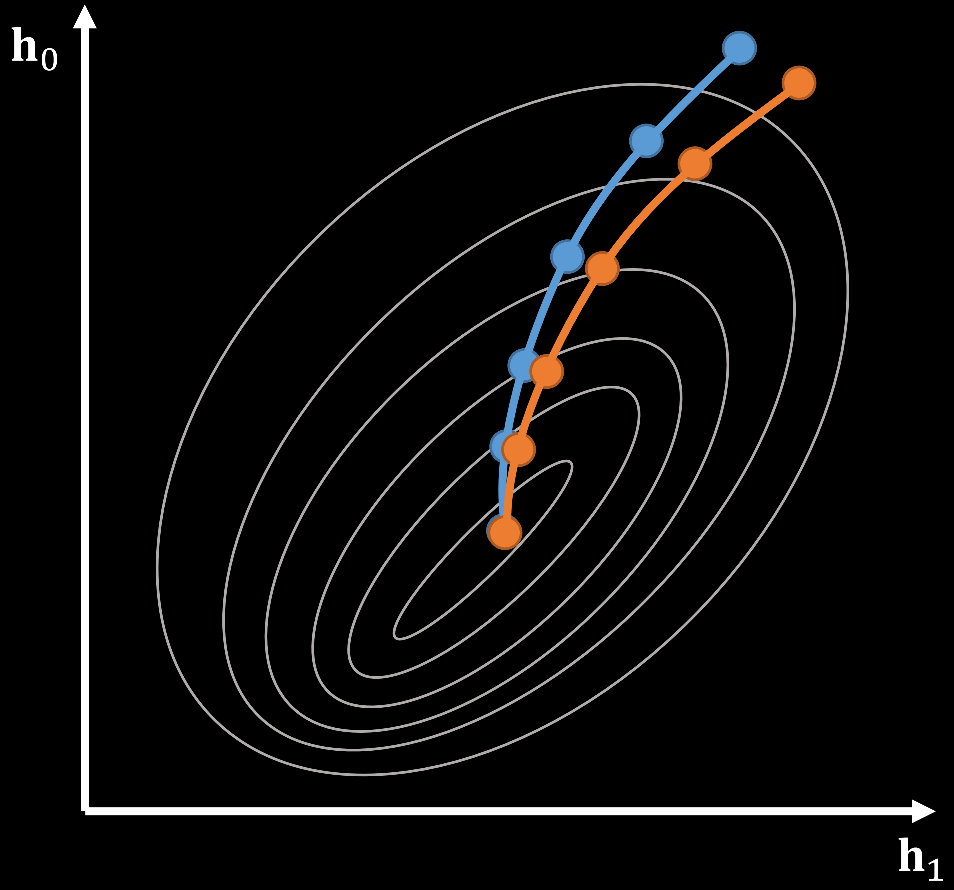
Residual vector: $[\mathbf{r}(\mathbf{h})]_i = \|\mathbf{p}_i - C\mathbf{h}\mathbf{y}_i\|^2$

Update Rule: $\mathbf{h}^{t+1} = \mathbf{h}^t - (J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \mathbf{r}(\mathbf{h}^t)$

Jacobean: $[J_\mathbf{r}]_{ij} = \frac{\partial [\mathbf{r}(\mathbf{h}^t)]_i}{\partial [\mathbf{h}^t]_j}$

Last update: $\hat{\mathbf{h}} = \mathbf{h}^\infty - (J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \mathbf{r}(\mathbf{h}^\infty)$

$\boxed{\text{Gradients: } \frac{\partial}{\partial \mathbf{y}_i} \hat{\mathbf{h}} \approx -(J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \frac{\partial}{\partial \mathbf{y}_i} \mathbf{r}(\mathbf{h}^\infty)}$



Pose Solver (RGB)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Initialization → Gauss-Newton

Residual vector: $[\mathbf{r}(\mathbf{h})]_i = \|\mathbf{p}_i - C\mathbf{h}\mathbf{y}_i\|^2$

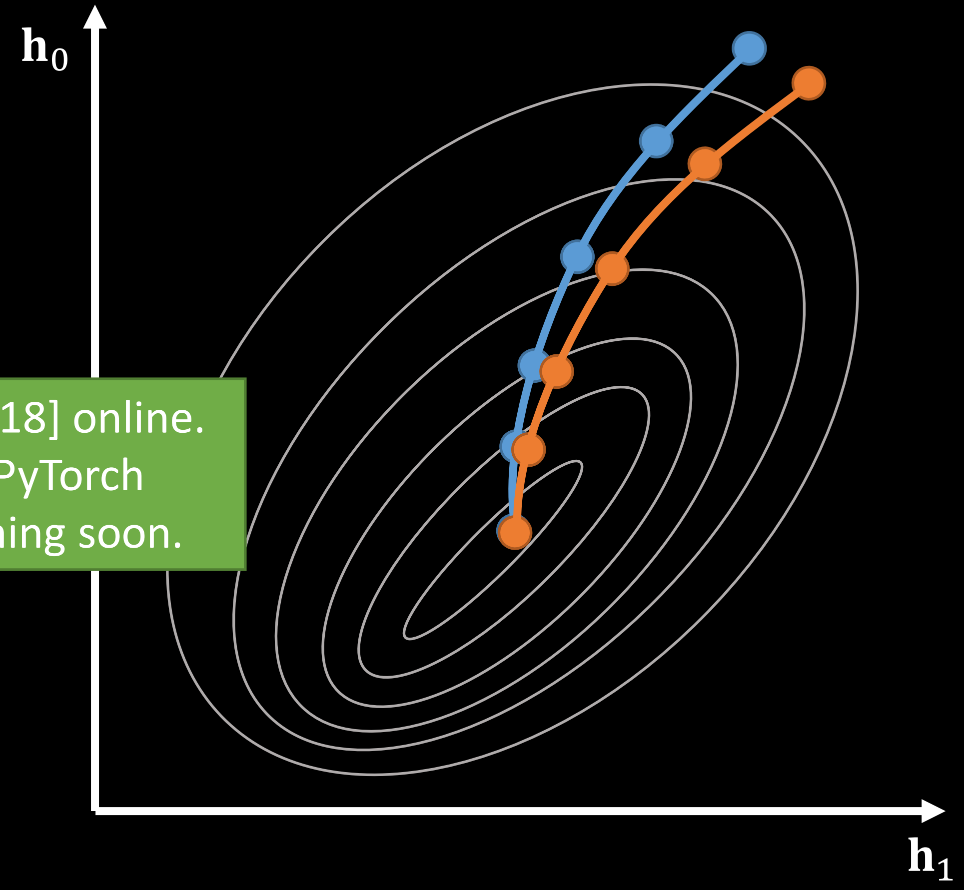
Update Rule: $\mathbf{h}^{t+1} = \mathbf{h}^t - (J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \mathbf{r}(\mathbf{h}^t)$

Jacobean: $[J_\mathbf{r}]_{ij} = \frac{\partial[\mathbf{r}(\mathbf{h}^t)]_i}{\partial[\mathbf{h}^t]_j}$

Last update: $\hat{\mathbf{h}} = \mathbf{h}^\infty - (J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \mathbf{r}(\mathbf{h}^\infty)$

Gradients: $\frac{\partial}{\partial \mathbf{y}_i}\hat{\mathbf{h}} \approx -(J_\mathbf{r}^T J_\mathbf{r})^{-1} J_\mathbf{r}^T \frac{\partial}{\partial \mathbf{y}_i}\mathbf{r}(\mathbf{h}^\infty)$

C++ code of [Bra18] online. Version with PyTorch integration coming soon.



[För16] Förstner and Wrobel, “Photogrammetric Computer Vision – Statistics, Geometry, Orientation and Reconstruction”, Springer’16
[Bra18] Brachmann and Rother, “Learning less is more - 6D camera localization via 3D surface regression”, CVPR’18

RANSAC

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

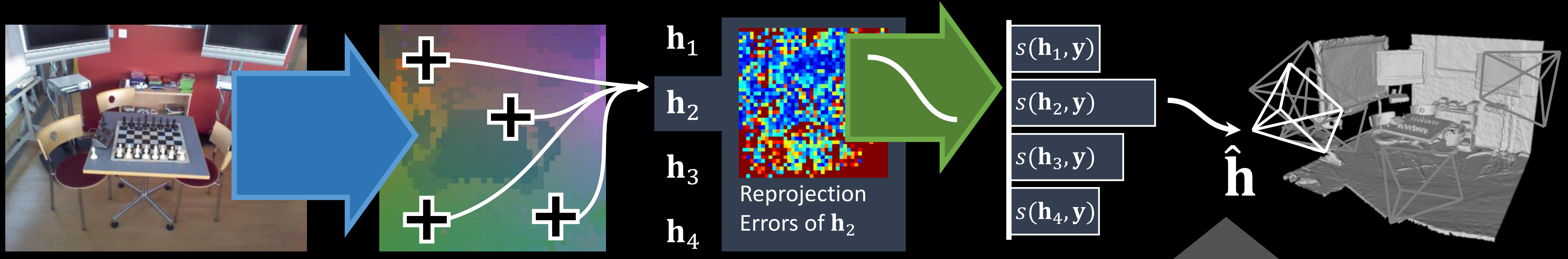


Image | Correspondence Prediction | Hypothesis Sampling | Scoring | Hypoth… …election | Result

Soft Inlier Counting [Bra18]:

$$s(\mathbf{h}, \mathbf{y}) = \sum_i \mathrm{sig}(\tau - \beta \|\mathbf{p}_i - C\mathbf{h}\mathbf{y}_i\|)$$

argmax Selection

$$\hat{\mathbf{h}} = \operatorname*{argmax}_{\mathbf{h}_j} s(\mathbf{h}_j, \mathbf{y})$$

hard decision

non-differentiable

Probabilistic Selection [Bra17]

$$\hat{\mathbf{h}} = \mathbf{h}_j, \text{where } j \sim \frac{\exp(s(\mathbf{h}_j \mathbf{y}))}{\sum_k \exp(s(\mathbf{h}_k \mathbf{y}))}$$

hard decision

differentiable

[Bra17] Brachmann et al., “DSAC - Differentiable RANSAC for camera localization”, CVPR’17

[Bra18] Brachmann and Rother, “Learning less is more - 6D camera localization via 3D surface regression”, CVPR’18

Differentiable RANSAC (DSAC)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Hypothesis selection: $\hat{\mathbf{h}} = \mathbf{h}_j, \text{where } j \sim \frac{\exp(s(\mathbf{h}_j, \mathbf{y}))}{\sum_k \exp(s(\mathbf{h}_k, \mathbf{y}))} = P(j; \mathbf{y})$

Learning objective: $\mathcal{L}(\mathbf{y}) = \mathbb{E}_{j \sim P(j;\mathbf{y})}\left[\ell(\mathbf{h}_j, \mathbf{h}^*)\right]$

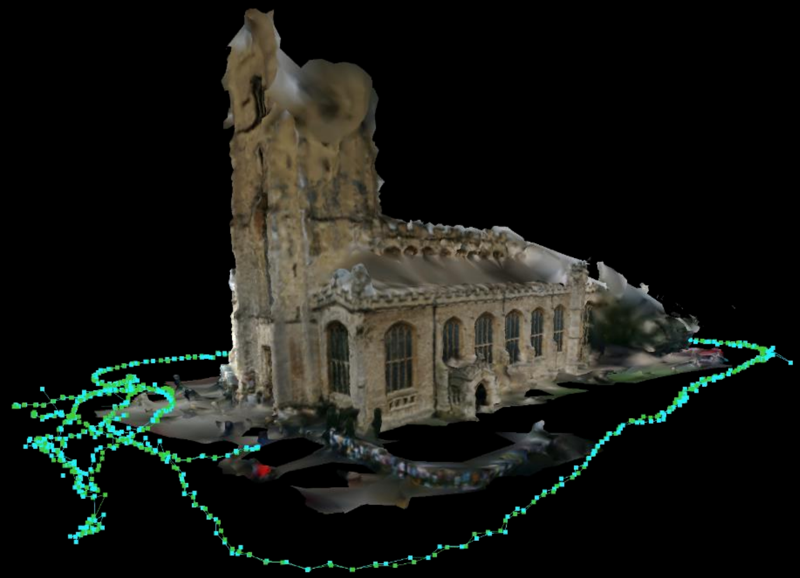
Gradients:

$$\frac{\partial}{\partial \mathbf{y}} \mathcal{L}(\mathbf{y}) = \mathbb{E}_{j \sim P(j;\mathbf{y})}\left[\underbrace{\ell(\mathbf{h}_j, \mathbf{h}^*) \frac{\partial}{\partial \mathbf{y}} \log P(j; \mathbf{y})}_{\text{derivative of selection probability}} + \underbrace{\frac{\partial}{\partial \mathbf{y}} \ell(\mathbf{h}_j, \mathbf{h}^*)}_{\text{derivative of task loss}}\right]$$

C++ code for camera re-localization online.
PyTorch code for DSAC line fitting also online.

[Bra17] Brachmann et al., “DSAC - Differentiable RANSAC for camera localization”, CVPR’17

Differentiable RANSAC (DSAC)



PoseNet	149cm, 3.4°
Active Search	19cm, 0.5°
DSAC++	13cm, 0.4°

Test Video
Cambridge St Mary's Church

Our Estimates
Trained with the 3D Model

[Posenet] "Geometric Loss Functions for Camera Pose Regression with Deep Learning" Kendall and Cipolla, CVPR '17
[Active Search] "Efficient & effective prioritized matching for large-scale image-based localization", Sattler et al., TPAMI'17
[DSAC] "DSAC - Differentiable RANSAC for Camera Localization", Brachmann et al., CVPR'17
[DSAC++] "Learning Less is More – 6D Camera Localization via 3D Surface Regression", Brachmann and Rother, CVPR'18

Correspondence Prediction

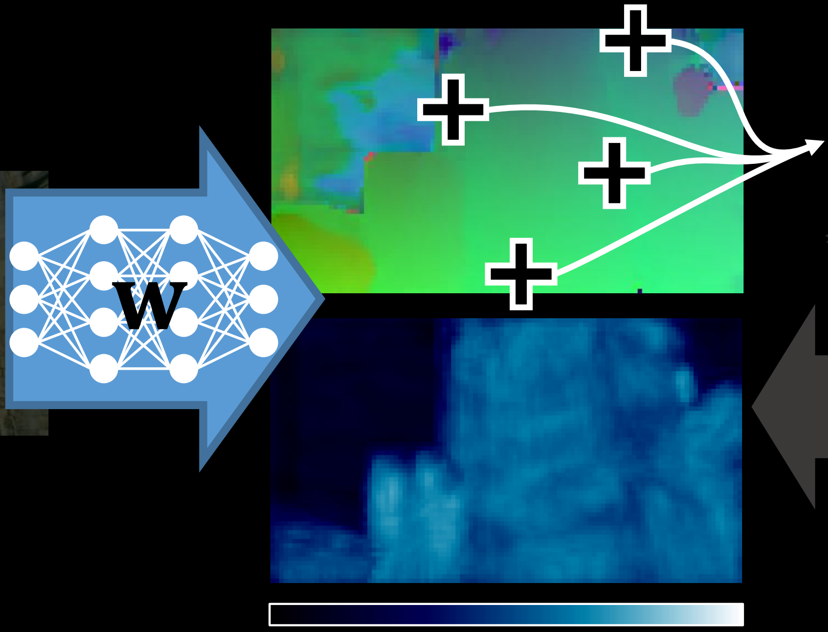
Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Input Image　　Dense Correspondences　　RANSAC / DSAC

Neural Guided RANSAC (NG-RANSAC)

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Input Image



0 Sampling Weight 1

RANSAC / DSAC

Selecting a scene coordinate: $p(\mathbf{y}) = g(I; \mathbf{w})$

Selecting a hypothesis: $p(\mathbf{h}) = \prod_{i=0}^{4} p(\mathbf{y}_i)$

Selecting a hypotheses pool: $p(\mathcal{H}) = \prod_j p(\mathbf{h}_j)$

Learning objective: $\mathbb{E}_{\mathcal{H}\sim p(\mathcal{H})}[\mathcal{L}(\mathbf{w})]$

$$= \underbrace{\mathbb{E}_{\mathcal{H}\sim p(\mathcal{H})}}_{\text{Neural Guidance}} \underbrace{\mathbb{E}_{j\sim P(j|\mathcal{H};\mathbf{w})}\left[\ell(\mathbf{h}_j, \mathbf{h}^*)\right]}_{\text{DSAC}}$$

Neural Guided RANSAC (NG-RANSAC)

Object Detection | Object Classification | **Correspondence Prediction** | **RANSAC** (Pose Solver, Pose Scoring) | **Pose Loss**

	PoseNet	ActiveSearch	DSAC++	NG-DSAC++
Great Court	700cm	-	40.3cm	**35.0cm**
Kings College	99cm	42cm	13.0cm	**12.6cm**
Old Hospital	217cm	44cm	22.4cm	**21.9cm**
Shop Facade	107cm	12cm	5.7cm	**5.6cm**
St M. Church	149cm	19cm	9.9cm	**9.8cm**

[PoseNet] "Geometric Loss Functions for Camera Pose Regression with Deep Learning"
Kendall and Cipolla, CVPR '17
[ActiveSearch] "Efficient & effective prioritized matching for large-scale image-based localization",
Sattler et al., TPAMI'17
[DSAC++] "Learning Less is More – 6D Camera Localization via 3D Surface Regression",
Brachmann and Rother, CVPR'18

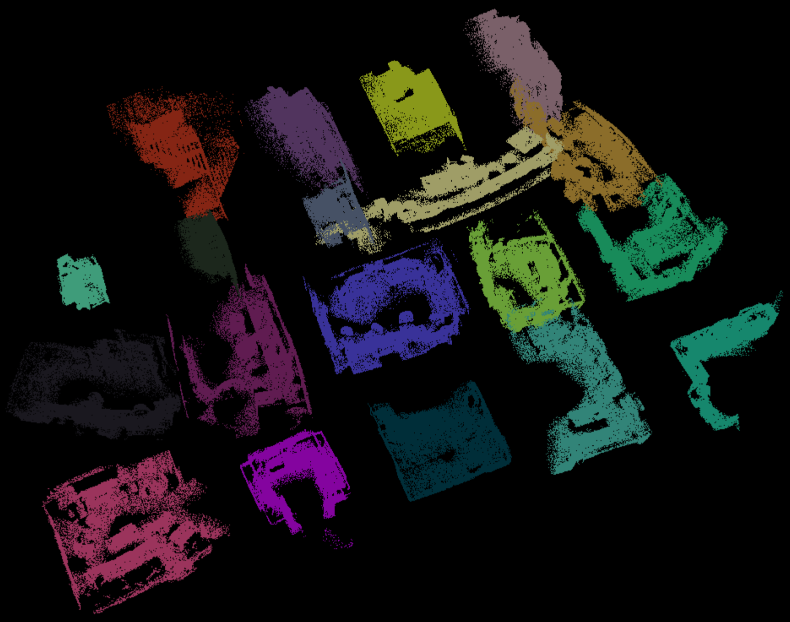
[NG-DSAC++] "Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses", Brachmann and Rother, ICCV19

Object Classification

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

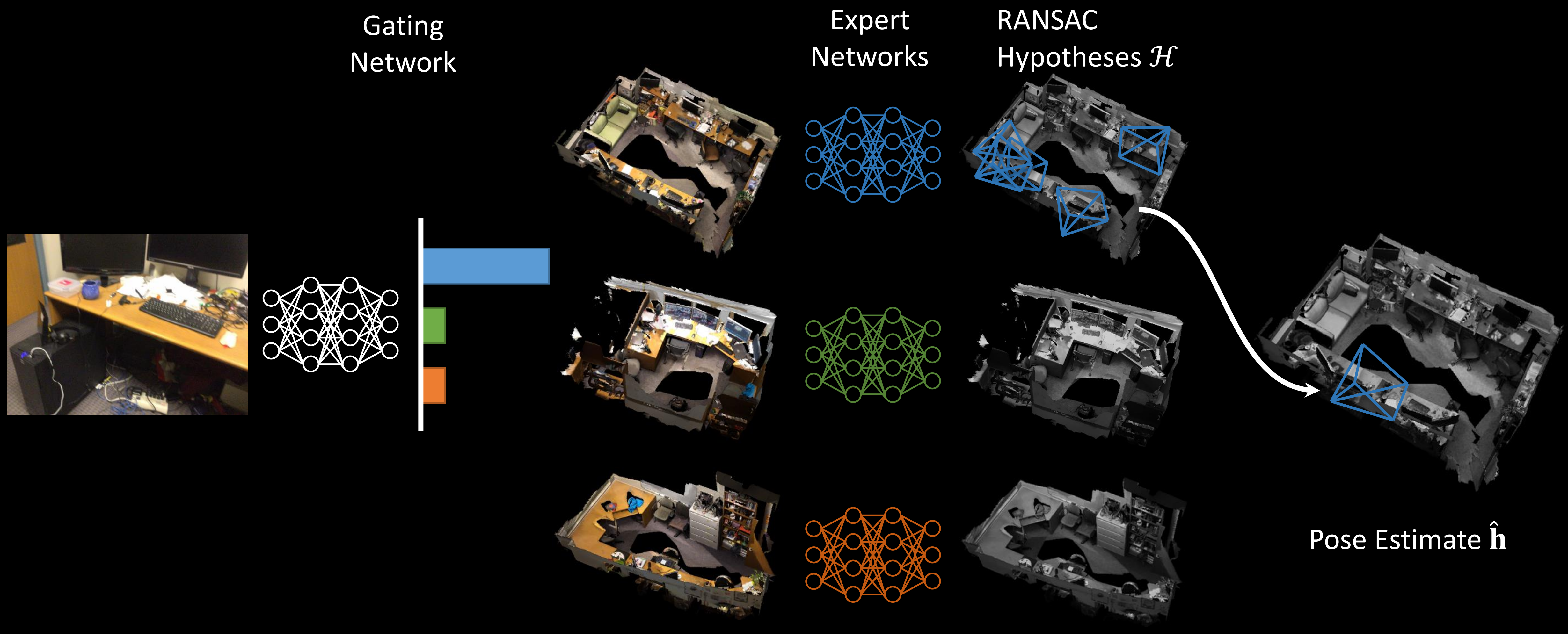
Environment

Classes



Query Image

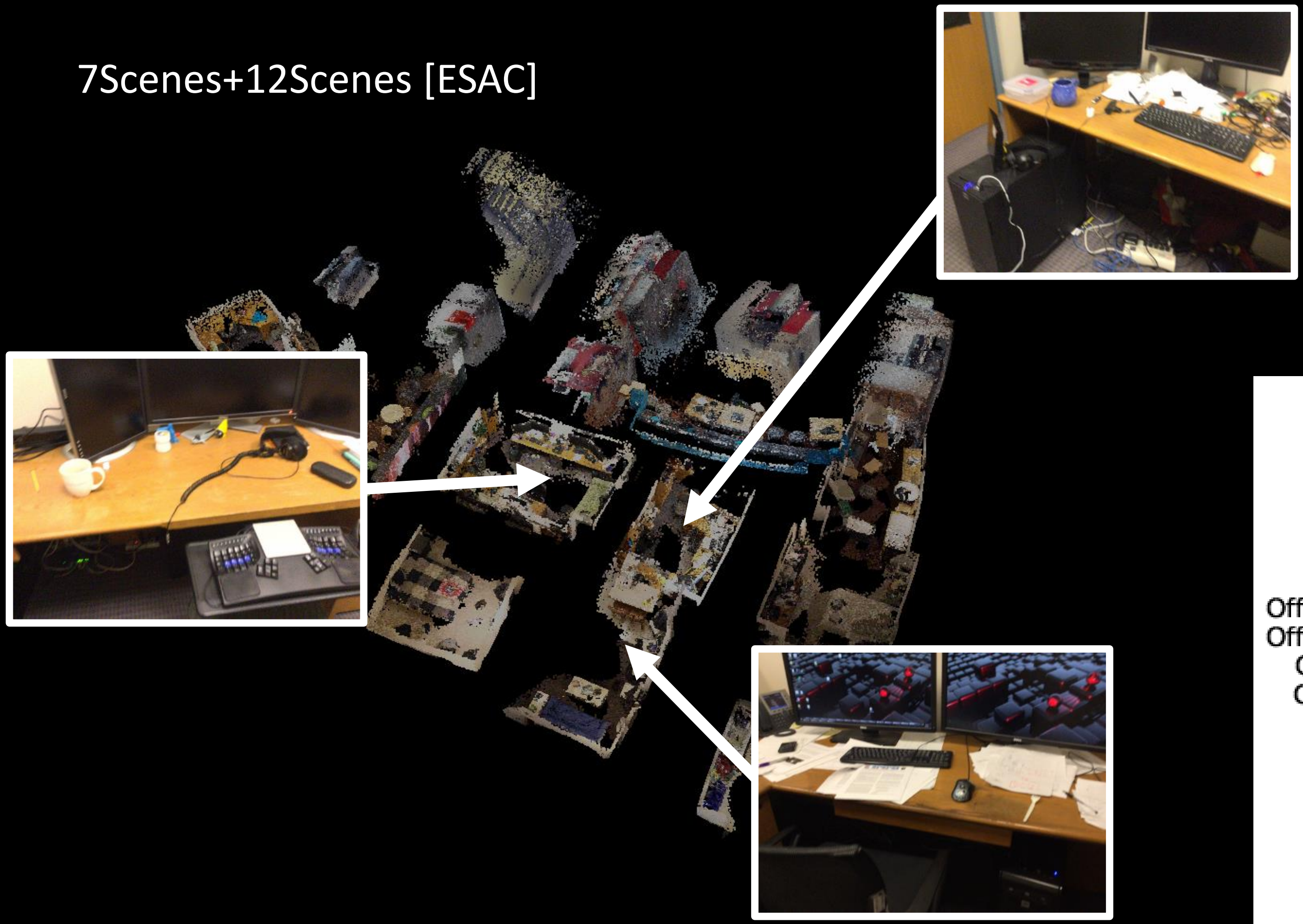
Object Classification



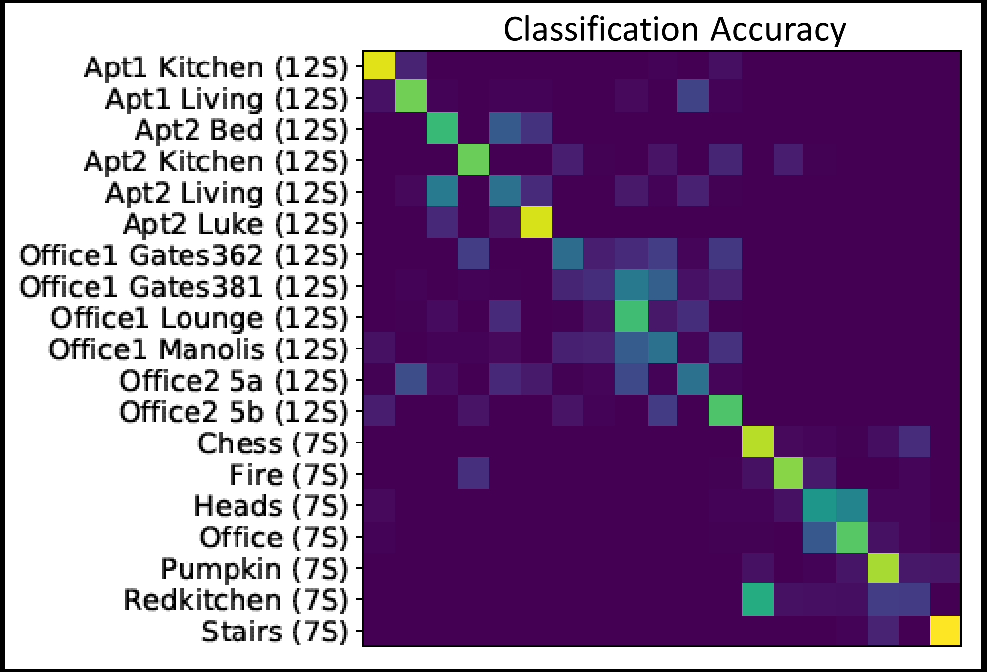
[Jacobs‘91] „Adaptive Mixtures of Local Experts“, Jacobs et al., Neural Computation, 1991
[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization”, Brachmann and Rother, ICCV’19

Object Classification

7Scenes+12Scenes [ESAC]



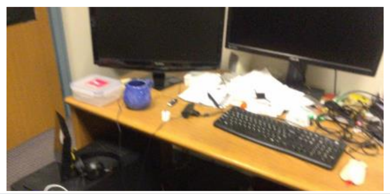
Average Accuracy (5cm,5°):
Classification + DSAC++: 47.5%
Oracle + DSAC++: 89.0%



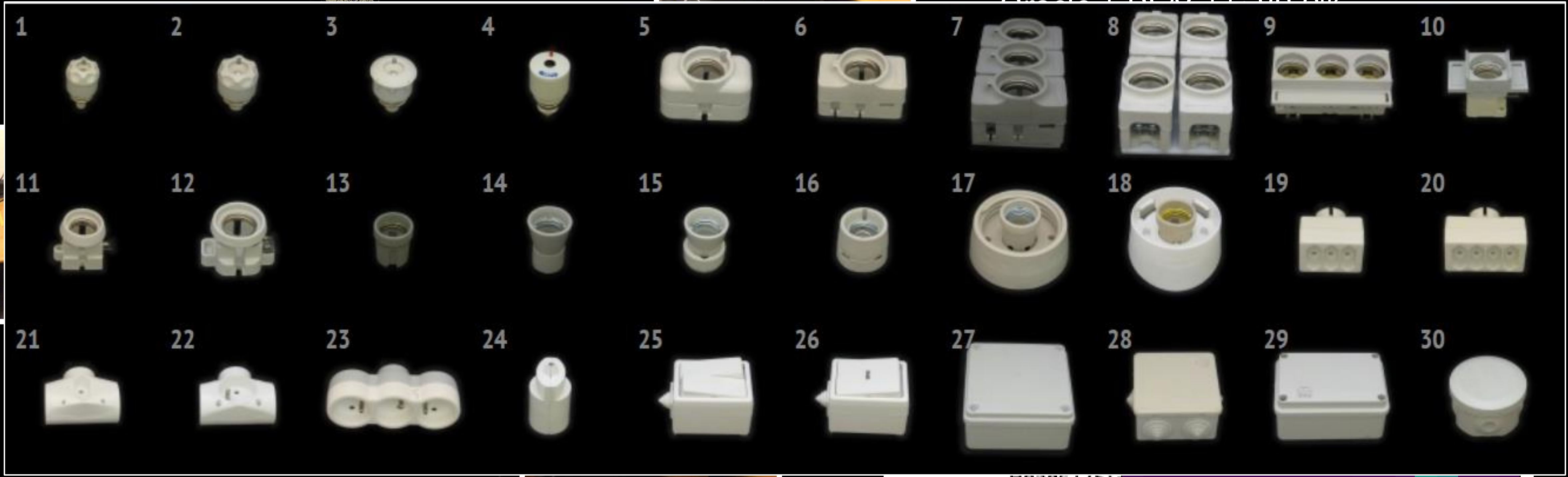
[DSAC++] Brachmann and Rother, "Learning less is more - 6D camera localization via 3D surface regression", CVPR'18
[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19

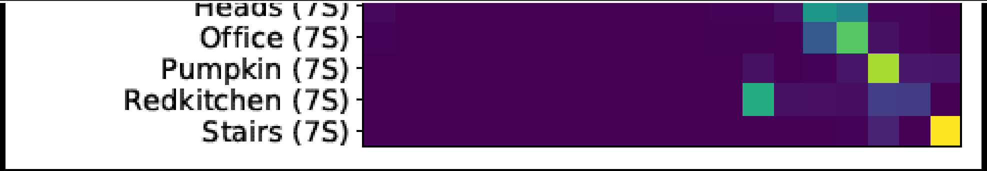
Object Classification

7Scenes+12Scenes [ESAC]



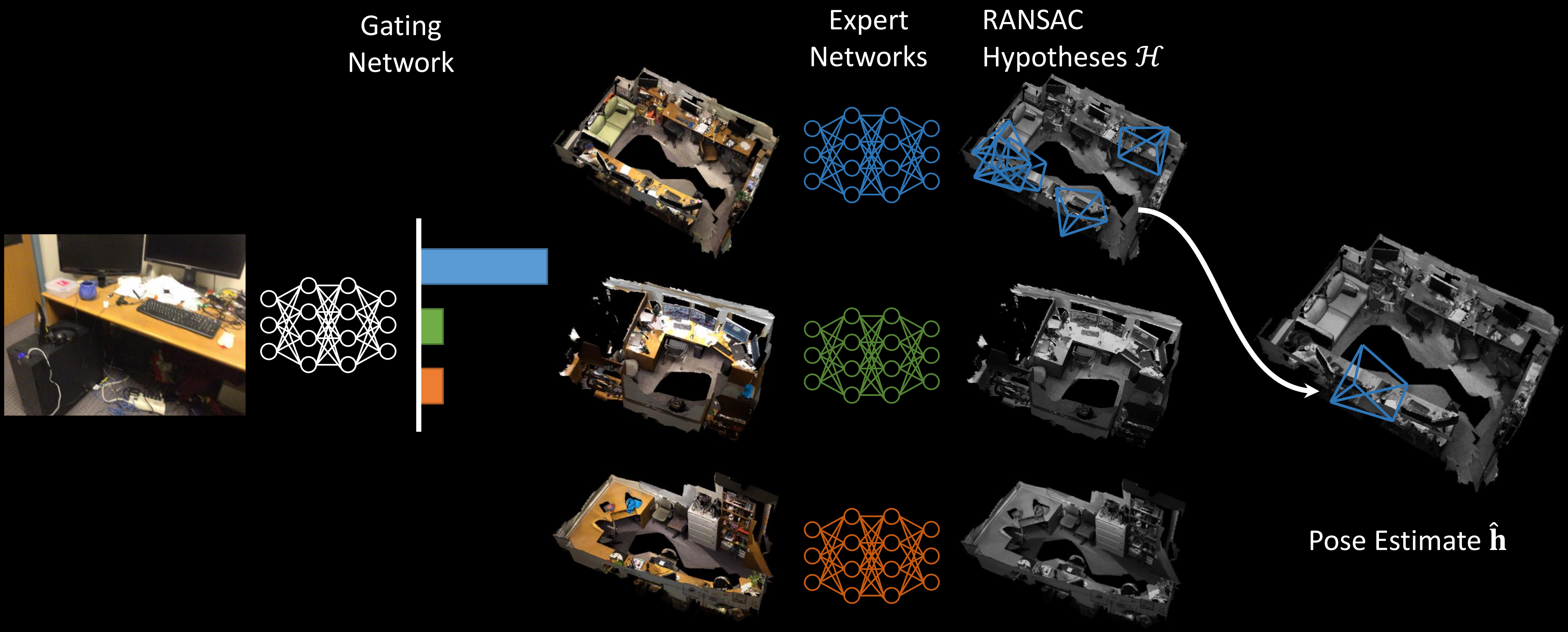
Average Accuracy (5cm,5°):
Classification + DSAC++: 47.5%





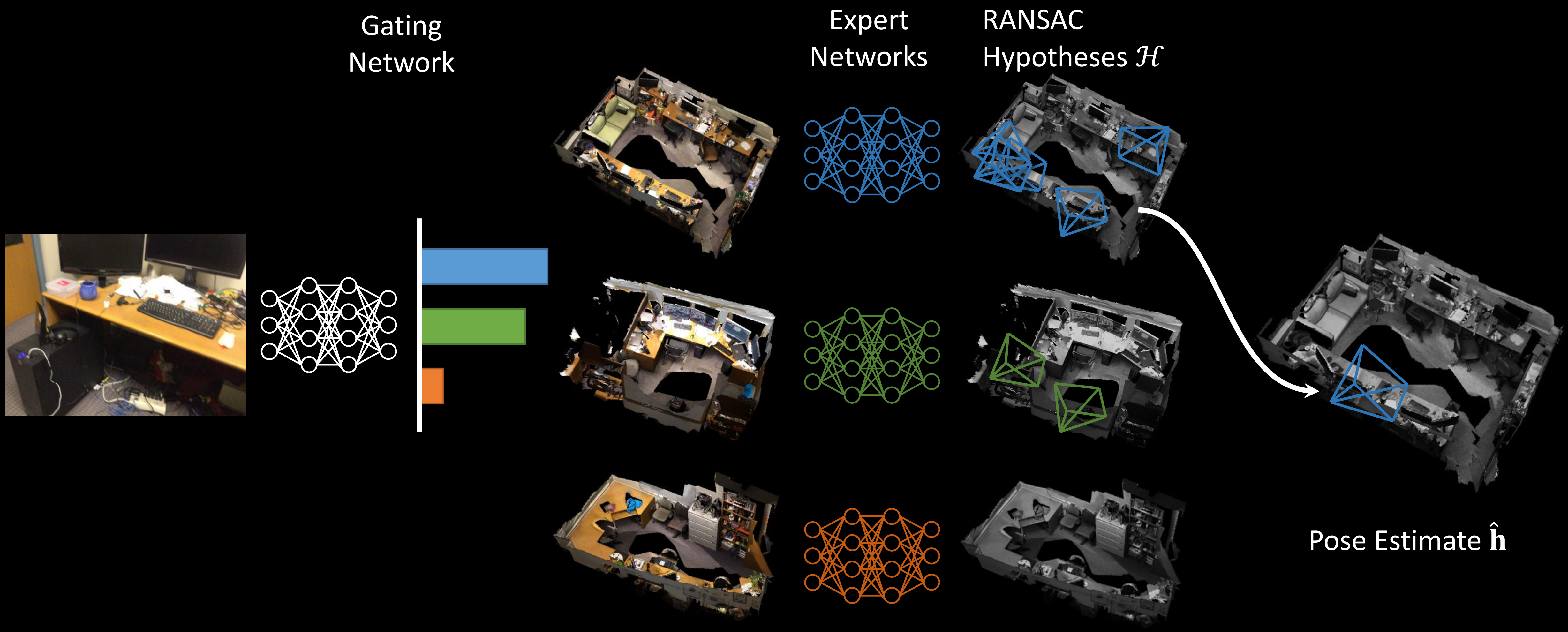
[DSAC++] Brachmann and Rother, “Learning less is more - 6D camera localization via 3D surface regression”, CVPR’18
[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization”, Brachmann and Rother, ICCV’19

Object Classification



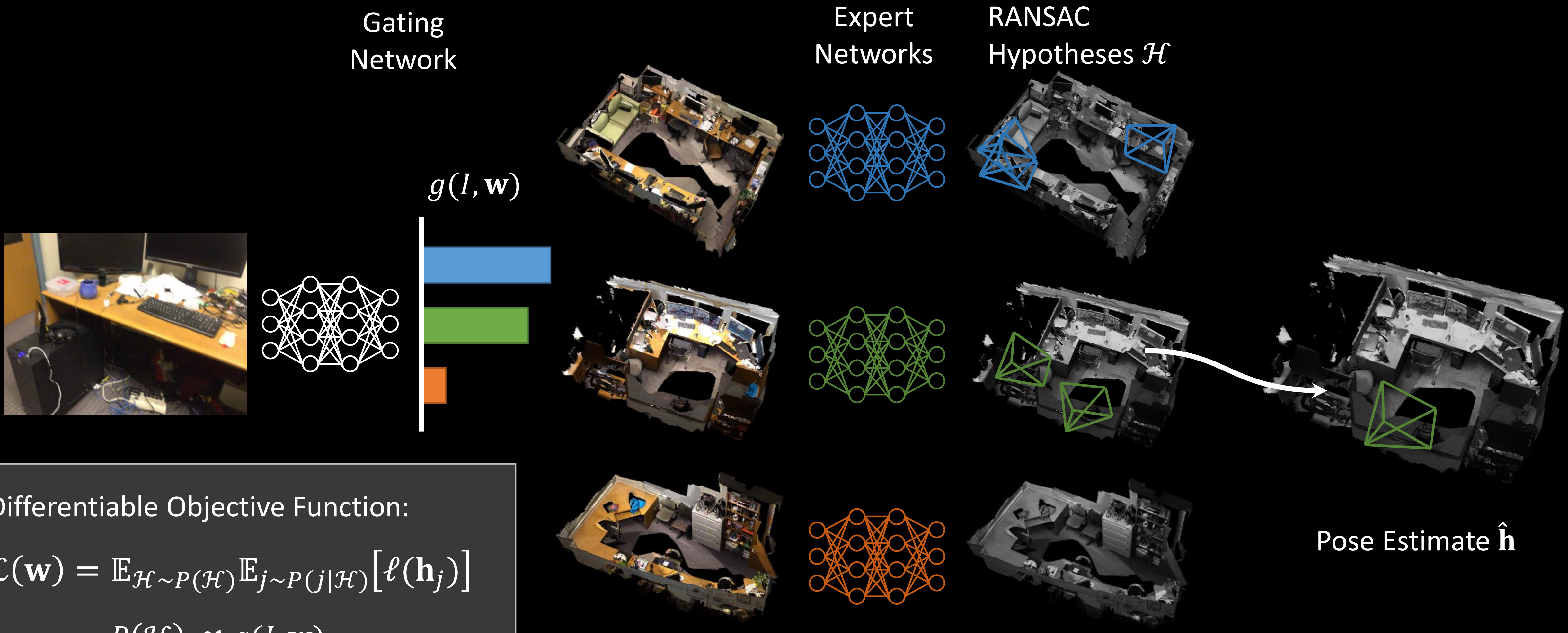
[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19

Expert Sample Consensus



[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19

Expert Sample Consensus



Differentiable Objective Function:

$$\mathcal{L}(\mathbf{w}) = \mathbb{E}_{\mathcal{H}\sim P(\mathcal{H})}\mathbb{E}_{j\sim P(j|\mathcal{H})}\left[\ell(\mathbf{h}_j)\right]$$

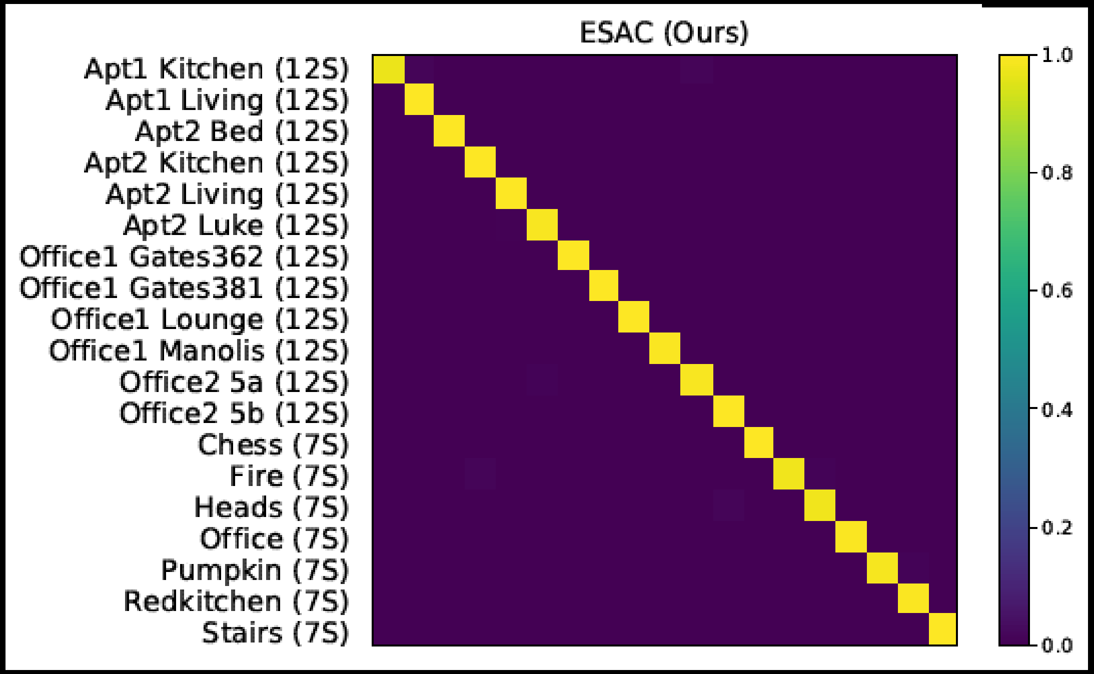
$$P(\mathcal{H}) \propto g(I, \mathbf{w})$$

[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19

Expert Sample Consensus

7Scenes+12Scenes [ESAC]

Average Accuracy (5cm,5°):
Classification + DSAC++: 47.5%
Oracle + DSAC++: 89.0%
ESAC: 88.1%



[DSAC++] Brachmann and Rother, "Learning less is more - 6D camera localization via 3D surface regression", CVPR'18
[ESAC] „Expert Sample Consensus Applied to Camera Re-Localization", Brachmann and Rother, ICCV'19

Object Detection

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Conclusion

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Conclusion:

- Differentiable PnP [Bra18]
- Differentiable RANSAC → [DSAC]
- Differentiable Correspondence Selection → [NG-RANSAC]
- Differentiable Expert Selection → [ESAC]

[Bra18] Brachmann and Rother, “Learning less is more - 6D camera localization via 3D surface regression”, CVPR’18
[DSAC] Brachmann et al., “DSAC - Differentiable RANSAC for camera localization”, CVPR’17
[NG-RANSAC] Brachmann and Rother, “Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses”, ICCV19
[ESAC] Brachmann and Rother, “Expert Sample Consensus Applied to Camera Re-Localization”, ICCV’19

Conclusion

Object Detection | Object Classification | Correspondence Prediction | RANSAC (Pose Solver, Pose Scoring) | Pose Loss

Conclusion:

- Differentiable PnP [Bra18]
- Differentiable RANSAC → [DSAC]
- Differentiable Correspondence Selection → [NG-RANSAC]
- Differentiable Expert Selection → [ESAC]

Code of many methods online:

DSAC for camera re-localization [Lua/Torch]: https://github.com/cvlab-dresden/DSAC
DSAC for Line Fitting [PyTorch]: https://github.com/vislearn/DSACLine
DSAC++ for Camera Re-Localization, incl. differentiable PnP [Lua/Torch]: https://github.com/vislearn/LessMore
DSAC*, improved DSAC++ incl. differentiable PnP and differentiable Kabsch [PyTorch]: Coming soon
ESAC, differentiable expert selection [PyTorch]: Coming soon (https://hci.iwr.uni-heidelberg.de/vislearn/research/scene-understanding/pose-estimation/#ICCV19)
NG-DSAC, differentiable correspondence selection [PyTorch]: Coming soon (https://hci.iwr.uni-heidelberg.de/vislearn/research/neural-guided-ransac/)

Thank You!