DSAC – Differentiable RANSAC for Camera Localization

Eric Brachmann[1], Alexander Krull[1], Sebastian Nowozin[2], Jamie Shotton[2], Frank Michel[1], Stefan Gumhold[1], Carsten Rother[1]
[1]TU Dresden, [2]Microsoft

Code and trained models:

Motivation

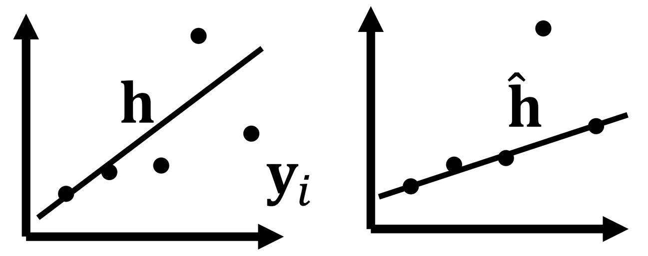
RANSAC (Random Sample Consensus) is an important algorithm in robust optimization. It cannot be used in deep learning pipelines because it is non-differentiable.

Our Goal: Use RANSAC in end-to-end learning.

RANSAC in a Nutshell:

1) Sample multiple hypotheses $\mathbf{h}_j$
2) Score each hypothesis: $s(\mathbf{h}_j)$
3) Take best one (and refine): $\hat{\mathbf{h}} = \boxed{\operatorname{argmax}_{\mathbf{h}_j} s(\mathbf{h}_j)}$

Not differentiable



Naïve Fit   RANSAC Fit

Contributions

- We explore **two ways** of making RANSAC differentiable: **soft** argmax **selection** and **probabilistic selection**.
- We put both options in a **new** end-to-end trainable **camera localization pipeline**.
- We show experimentally that **probabilistic selection is superior.** We call RANSAC with this option **DSAC**.
- We **exceed state-of-the-art** on camera localization **by 7.2%** (new results: by 12.8%).

Differentiable Alternatives to argmax

Our hypotheses depend on learnable parameters $\mathbf{w}$: $\mathbf{h}_j^{\mathbf{w}} \coloneqq \mathbf{h}_j(\mathbf{w})$

We wish to minimize the loss ℓ of selected hypotheses over training images I:

$$\widehat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}} \sum_I \ell(\hat{\mathbf{h}}^{\mathbf{w}})$$

$$\frac{\partial}{\partial \mathbf{w}} \ell(\hat{\mathbf{h}}^{\mathbf{w}}) = \frac{\partial \ell}{\partial \hat{\mathbf{h}}^{\mathbf{w}}} \frac{\partial \hat{\mathbf{h}}^{\mathbf{w}}}{\partial \mathbf{w}} \quad \text{but} \quad \frac{\partial}{\partial \mathbf{w}} \operatorname*{argmax}_{\mathbf{h}_j^{\mathbf{w}}} s(\mathbf{h}_j^{\mathbf{w}})$$

Selection depends on all hypotheses and all scores!

We define a softmax score distribution: $\boxed{P(j|\mathbf{w}) \coloneqq \exp\left(s(\mathbf{h}_j^{\mathbf{w}})\right) / \sum_k \exp(s(\mathbf{h}_k^{\mathbf{w}}))}$

Soft argmax Selection

$$\mathbf{h}_{\mathrm{SoftAM}}^{\mathbf{w}} = \sum_j P(j|\mathbf{w}) \mathbf{h}_j^{\mathbf{w}}$$

Soft, deterministic decision

Probabilistic Selection

$$\mathbf{h}_{\mathrm{DSAC}}^{\mathbf{w}} = \mathbf{h}_j^{\mathbf{w}}, \text{where } j \sim P(j|\mathbf{w})$$

Hard, probabilistic decision

Derivatives

Soft argmax **Selection:**

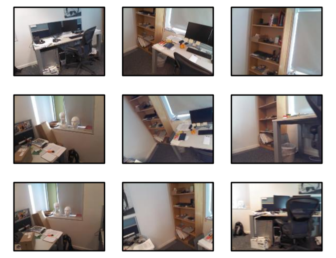
$$\frac{\partial}{\partial \mathbf{w}} \mathbf{h}_{\mathrm{SoftAM}}^{\mathbf{w}} = \sum_j \left(\mathbf{h}_j^{\mathbf{w}} \frac{\partial}{\partial \mathbf{w}} P(j|\mathbf{w}) + P(j|\mathbf{w}) \frac{\partial}{\partial \mathbf{w}} \mathbf{h}_j^{\mathbf{w}} \right)$$

Probabilistic Selection: We derive the expectation of the task loss.

$$\frac{\partial}{\partial \mathbf{w}} \mathbb{E}_{j \sim P(j|\mathbf{w})}\left[\ell(\mathbf{h}_j^{\mathbf{w}})\right] = \mathbb{E}_{j \sim P(j|\mathbf{w})}\left[\ell(\mathbf{h}_j^{\mathbf{w}}) \frac{\partial}{\partial \mathbf{w}} \log P(j|\mathbf{w}) + \frac{\partial}{\partial \mathbf{w}} \ell(\mathbf{h}_j^{\mathbf{w}})\right]$$

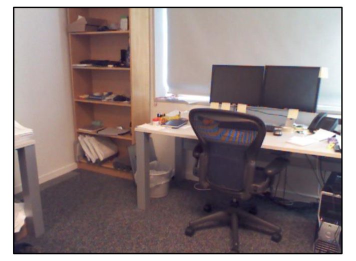
Application: Camera Localization

Training Data:



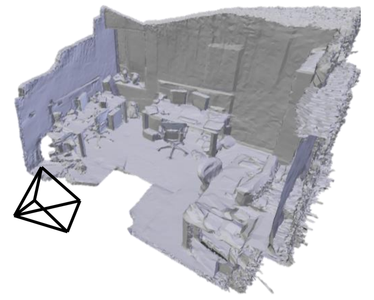
Annotated Images

Test Input:



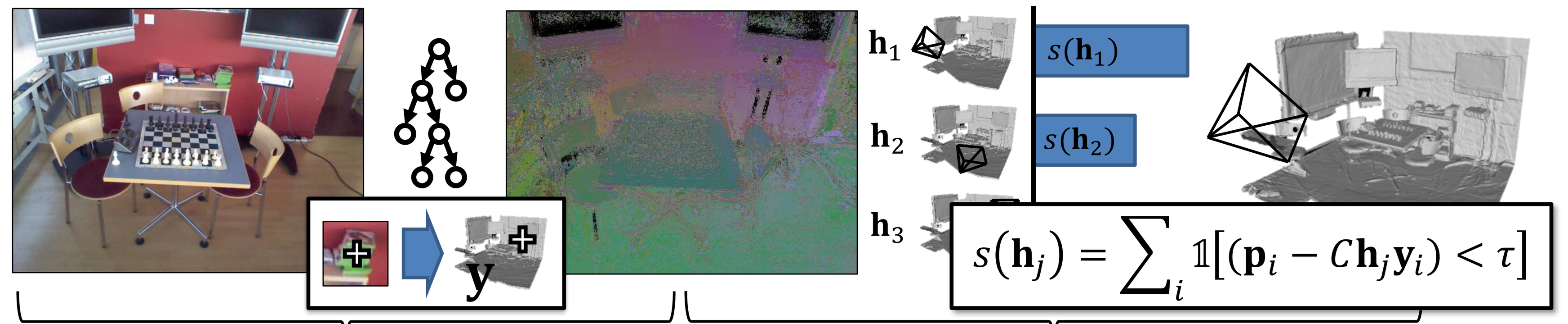
RGB Image

Test Output:



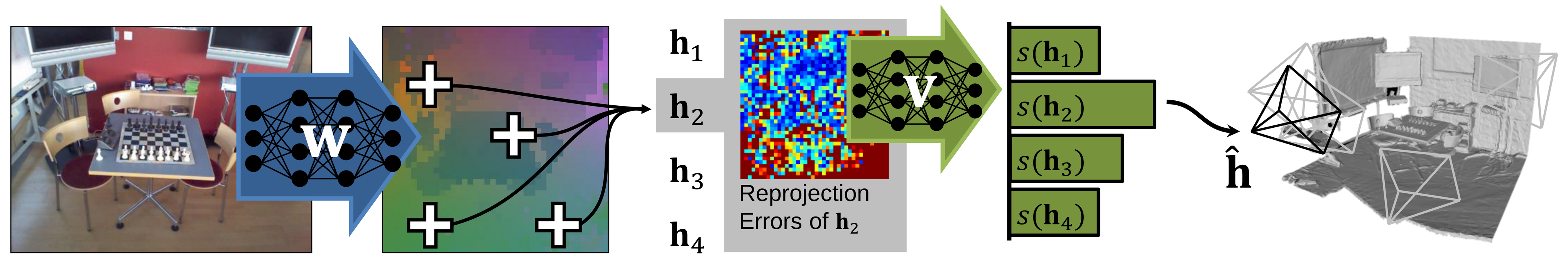
$\hat{\mathbf{h}}$: 6D Pose $\begin{cases} \text{3D Translation } \mathbf{t} \\ \text{3D Rotation } \boldsymbol{\theta} \end{cases}$

Previous Work: Scene Coordinate Regression [Sho13, Bra16]



Scene Coordinate Regression   RANSAC + Refinement

Our Pipeline



Input RGB | Scene Coordinate (**y**) Regression | Hypothesis Sampling | Scoring (s) | Hypothesis Selection | Refinement (**R**)

We learn two CNNs with parameters **w** and **v** jointly by minimizing:

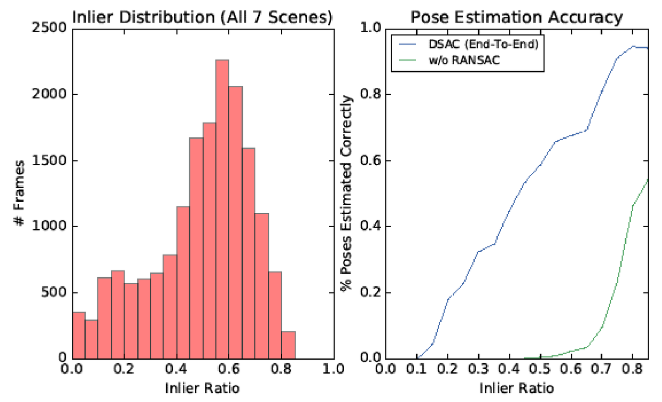
$$\mathbb{E}_{j \sim P(j|\mathbf{w},\mathbf{v})}\left[\ell(\mathbf{R}(\mathbf{h}_j^{\mathbf{w}}, \mathbf{w}), \mathbf{h}^*)\right], \text{with}$$
$$\ell(\mathbf{h}, \mathbf{h}^*) = \max(\measuredangle(\boldsymbol{\theta}, \boldsymbol{\theta}^*), \|\mathbf{t} - \mathbf{t}^*\|)$$

Experiments

Dataset: 7-Scenes [Sho13]

RGB-D images of 7 indoor scenes with pose annotations, 1k-5k training resp. test images. We use **only RGB** at test time.

Difficult frames of 7-Scenes.



Initialization: Componentwise Training

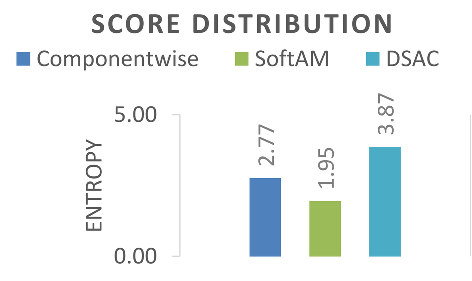
Scene Coordinate Regression: Generate ground truth scene coordinates $\mathbf{y}^*$ and minimize:

$$\widehat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}} \sum_i \|\mathbf{y}_i - \mathbf{y}_i^*\|.$$

Score Regression: Optimize a dummy score of hypotheses sampled around ground truth poses.

Results:

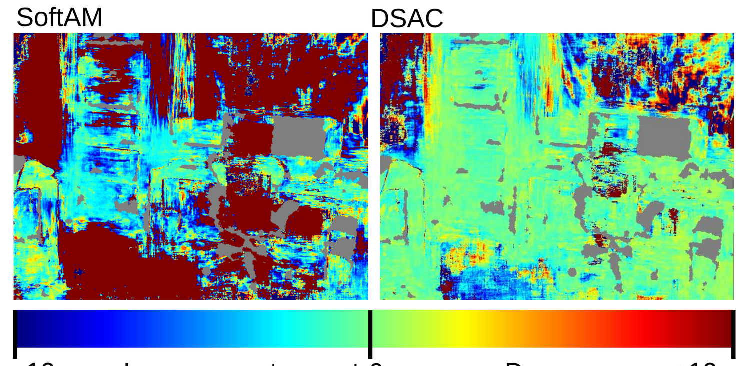
% Frames with Error <5°,5cm	Trained Componentwise	Trained End-to-End
Shotton et al. [Sho13]	38.6%	-
Brachmann et al. [Bra16]	55.2%	-
Ours, RANSAC	61.0%	-
Ours, SoftAM	61.6%	59.6% (-2.0%)
Ours, DSAC	60.3%	**62.4%** (+2.1%)



Input RGB   Scene Coordinate Ground Truth   Componentwise Training, Prediction

Change in Prediction Error after End-to-End Training w.r.t. Componentwise Training:

SoftAM   DSAC

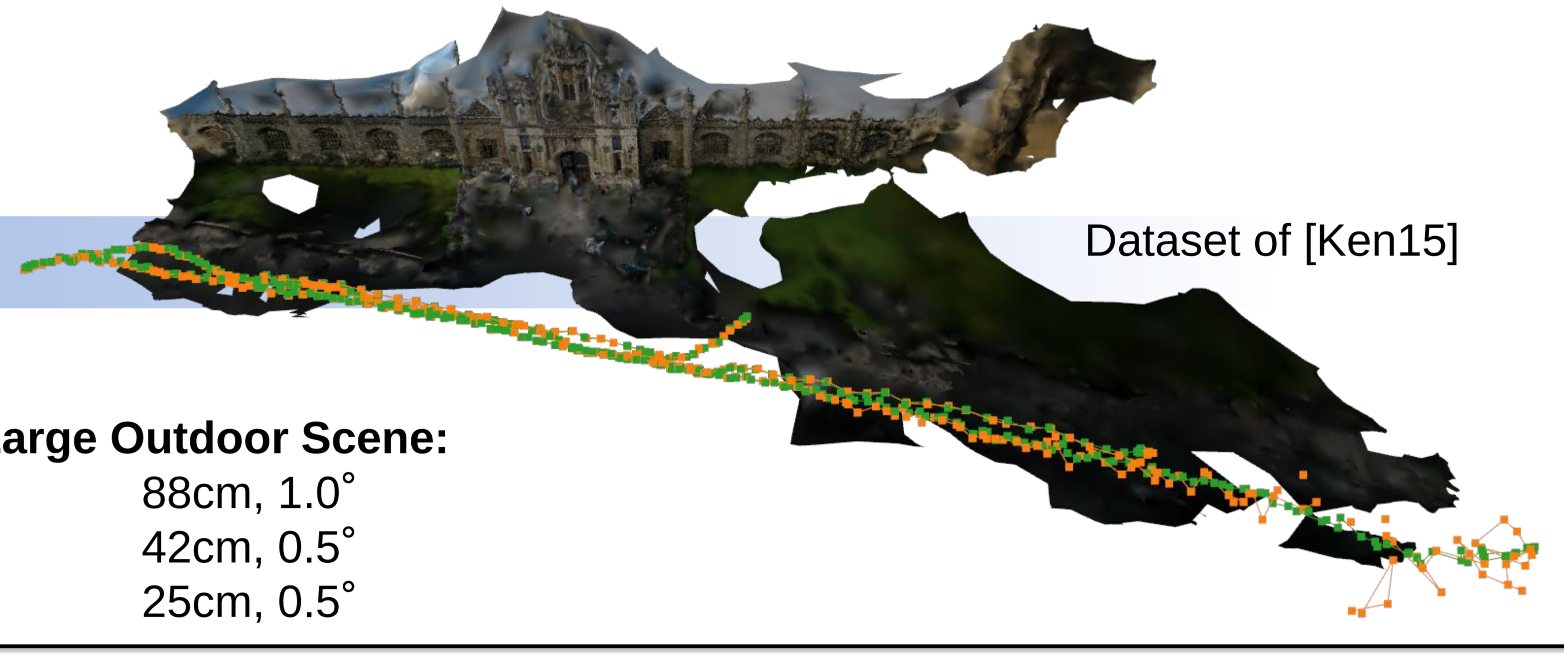


New Results

7-Scenes: 68.0%

Median Accuracy on a Large Outdoor Scene:

PoseNet [Ken17]:	88cm, 1.0°
Sparse Features [Sat16]:	42cm, 0.5°
Our Results:	25cm, 0.5°



Dataset of [Ken15]

[Sho13] "Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images", Shotton et al., CVPR'13
[Bra16] "Uncertainty-Driven 6D Pose Estimation of Objects and Scenes from a Single RGB Image", Brachmann et al., CVPR'16
[Ken15] "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization" Kendall et al., ICCV'15
[Ken17] "Geometric Loss Functions for Camera Pose Regression with Deep Learning" Kendall and Cipolla, CVPR 2017
[Sat16] "Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization" Sattler et al., PAMI 2016

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 647769)

European Research Council
Established by the European Commission
European Commission
Horizon 2020 European Union funding for Research & Innovation