Supplementary Material: Learning to Find Good Models in RANSAC

Daniel Barath, Luca Cavalli, Marc Pollefeys
Computer Vision and Geometry Group, Department of Computer Science, ETH Zürich
dbarath@ethz.ch

1. MF-Net Architecture

We show the detailed architecture of MF-Net in Figure 1. We design MF-Net as a lightweight network architecture for early detection of degenerate minimal samples. The main design principle is the incorporation of many invariances in the task: we normalize the points in each image, then we process the correspondences from image I_1 to I_2 and the opposite from I_2 to I_1 in two parallel shared branches of the network. Each branch follows the PointNet [4] design to achieve correspondence ordering invariance. Finally, features from the two branches are merged in a global feature, and classification is performed by a final MLP. All MLPs in the figure have a single hidden layer and use leaky relu activation function with negative slope 0.01, except for the final activation function using the sigmoid function. The dimensionality of their representations, visible on top of each MLP in the figure, do not exceed 64 to keep the computation lightweight and, thus, fast.

We implement the normalization by subtracting the mean per-channel for essential matrix estimation, and additionally we divide by the per-channel standard deviation for fundamental matrix estimation since in this case the input points are expressed in pixel coordinates instead of standardized image coordinates. Note that, while the architecture supports general n-samples of correspondences, we train it consistently with minimal samples only (*i.e.*, $n = 5$ for essential matrix estimation and $n = 7$ for fundamental matrix estimation).

2. Generalization Experiment

In order to test how well the proposed methods generalizes to other datasets, we ran the tested methods (trained on the CVPR IMW 2020 PhotoTourism dataset) on the KITTI dataset. The results are reported in Table 1. The proposed MQ-Net + MF-Net achieves the best results.

In Table 2, the average rotation and translation errors and run-times of MF-Net combined with different scoring techniques are reported on the KITTI dataset. Independently of the used scoring, MF-Net always improves both the accuracy and processing times.

	RANSAC	MSAC	MAGSAC++	**MQ-Net**	**MQ-Net + MF-Net**
$\epsilon_{\mathbf{R}}$ (°)	1.37	1.39	**1.35**	**1.18**	**1.18**
$\epsilon_{\mathbf{t}}$ (°)	2.99	3.04	2.93	**2.41**	**2.26**
t (secs)	**0.088**	**0.089**	0.120	0.381	0.345

Table 1. Average rotation ($\epsilon_{\mathbf{R}}$) and translation ($\epsilon_{\mathbf{t}}$) errors and run-time (t) on 11 sequences (22090 pairs) from the KITTI dataset. The proposed MQ-Net + MF-Net achieves the best results.

	RANSAC		MSAC		MAGSAC++	
MF-Net	w/o	w/	w/o	w/	w/o	w/
$\epsilon_{\mathbf{R}}$ (°)	1.37	**1.30**	1.39	**1.31**	1.35	**1.29**
$\epsilon_{\mathbf{t}}$ (°)	2.99	**2.50**	3.04	**2.51**	2.93	**2.48**
t (secs)	0.088	**0.051**	0.089	**0.051**	0.120	**0.072**

Table 2. Average rotation ($\epsilon_{\mathbf{R}}$) and translation ($\epsilon_{\mathbf{t}}$) errors and run-time (t) on 11 sequences (22090 pairs) from the KITTI dataset. MF-Net always improves the results and run-times.

3. Additional Experiments

In this section, we show the fundamental and essential matrix estimation results without pre-filtering the correspondences by their SNN ratio [3].

Fundamental matrix estimation. Table 3 reports the rotation and translation mean Average Accuracy (mAA) at 10°; the median errors ($\epsilon_{\mathbf{R}}$ and $\epsilon_{\mathbf{t}}$) in degrees, and the run-times (t) in milliseconds of the entire robust fundamental matrix estimation. The mAA score is calculated as the area under the recall curve cropped at 10°. All three variants of MQ-Net lead to a significantly improved accuracy compared to the traditional techniques. The median rotation error is the 32% of the MAGSAC++ error. The median translation error is decreased to its half. MF-Net accelerates the method by almost an order-of-magnitude while improving the accuracy as well. MQ-Net combined with MF-Net is both *faster* and *more accurate* than the traditional methods.

Compared to Table 1 in the main paper, the difference between the proposed and traditional approaches is signifi-

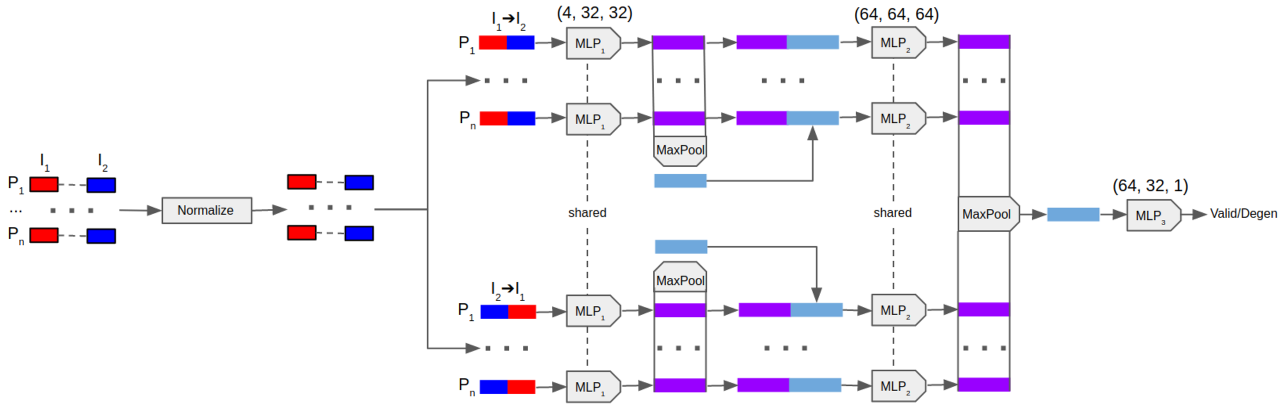

Figure 1. **MF-Net architecture**: we design a lightweight network architecture for early detection of degenerate minimal samples. The main design principle is the incorporation of many invariances in the task: we normalize the points in each image, then we process the correspondences from image I_1 to I_2 and the opposite from I_2 to I_1 in two parallel shared branches of the network. Each branch follows the PointNet [4] design to achieve correspondence ordering invariance. Finally, features from the two branches are merged in a global feature, and the classification is performed by a final MLP. All MLPs use leaky relu activation, except for the output layer of MLP_3 that uses sigmoid. The dimensionality of the input, hidden, and output layer are displayed on top of each MLP.

cantly larger. Consequently, the proposed MQ-Net and MF-Net are less affected by the reduced inlier ratio (due to not filtering by SNN) than the traditional algorithms.

Essential matrix estimation. Table 4 reports the rotation and translation mAA@10° scores, median errors ($\epsilon_\mathbf{R}$ and $\epsilon_\mathbf{t}$) in degrees, and the run-times (t) in milliseconds of the entire robust estimation procedure. All three variants of the proposed algorithm lead to better accuracy than the traditional techniques. Again, the best results are achieved by the network trained on both problems. Both median errors and mAA@10° are improved by a *large margin*. The median error of the proposed algorithm is lower than the half of the error of the traditional methods. MF-Net more than halves the run-time while also improving accuracy.

Compared to Table 2 in the main paper, the difference between the proposed and traditional approaches is significantly larger. Consequently, the proposed MQ-Net and MF-Net are less affected by the reduced inlier ratio (due to not filtering by SNN) than the traditional algorithms.

References

[1] D. Barath, J. Noskova, and J. Matas. Marginalizing sample consensus. *IEEE TPAMI*, 2021. 2, 3

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 2, 3

[3] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*. IEEE, 1999. 1

	mAA@10° ↑		Median (px) ↓		Time (ms) ↓	
Model scoring	**R**	**t**	$\epsilon_\mathbf{R}$	$\epsilon_\mathbf{t}$	AVG	MED
RANSAC [2]	0.43	0.14	6.52	29.78	**3.06**	**3.46**
MSAC [5]	0.43	0.13	6.46	29.57	**3.06**	3.58
MAGSAC++ [1]	0.43	0.14	6.58	28.11	3.39	3.80
MQ-Net (E)	0.47	0.20	5.09	24.10	28.31	31.66
MQ-Net (F)	**0.59**	0.28	2.45	15.22	26.57	29.62
MQ-Net (E & F)	**0.61**	**0.29**	**2.15**	**14.35**	26.33	30.59
MQ-Net + MF-Net	**0.61**	**0.30**	**2.10**	**14.01**	**3.13**	**2.53**

Table 3. **Fundamental matrix estimation**. The reported values are the rotation and translation mAA@10° scores; median errors ($\epsilon_\mathbf{R}$ and $\epsilon_\mathbf{t}$) in degrees; and the run-times in milliseconds. MQ-Net (**E**) and (**F**) are trained, respectively, on essential and fundamental matrix estimation. MQ-Net (**E & F**) is trained on both problems. The last row shows the results with MF-Net. No SNN pre-filtering.

[4] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, pages 652–660, 2017. 1, 2

[5] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *CVIU*, 2000. 2, 3

	mAA@10° ↑		Median (°) ↓		Time (ms) ↓	
Model scoring	**R**	**t**	$\epsilon_{\mathbf{R}}$	$\epsilon_{\mathbf{t}}$	AVG	MED
RANSAC [2]	0.61	0.41	2.35	7.12	**8.89**	**11.21**
MSAC [5]	0.62	0.42	2.33	7.08	**9.08**	11.40
MAGSAC++ [1]	0.62	0.43	2.23	6.87	9.61	12.31
MQ-Net (E)	**0.72**	**0.58**	**0.97**	**2.74**	19.47	13.96
MQ-Net (F)	**0.72**	0.56	1.11	3.18	19.37	13.85
MQ-Net (E & F)	**0.72**	**0.57**	1.04	2.94	19.86	14.66
MQ-Net + MF-Net	**0.73**	**0.57**	**1.01**	**2.87**	9.22	**5.73**

Table 4. **Essential matrix estimation**. The reported values are the rotation and translation mAA@10° scores; median errors ($\epsilon_{\mathbf{R}}$ and $\epsilon_{\mathbf{t}}$) in degrees; and the run-times in milliseconds. MQ-Net (**E**) and (**F**) are trained, respectively, on essential and fundamental matrix estimation. MQ-Net (**E & F**) is trained on both problems. The last row shows the results with MF-Net. No SNN pre-filtering.