On the Generalization of Optical Flow: Quantifying Robustness to Dataset Shifts

Supplementary Material

A. More Implementation Details

A.1. Dataset Splits

Tab. A1 lists the datasets and which of their splits we used for evaluation. As we need ground truth for evaluation, we used the validation split where available and the training split on the other datasets. Note that no C+T checkpoint was trained on any OOD data which we used for evaluation. Moreover, regarding our ID dataset Things, we only evaluate C+T checkpoints on the validation split of Things which was also not used for training.

The datasets Things, Sintel and Driving offer several rendering passes. We only use the final pass which is the most realistic and difficult as it includes various atmospheric effects, realistic illumination and motion blur.

	Available Splits			Rendering		
Dataset	Train	Validation	Test	Clean	Final	**Motion**
Chairs [4]	✓	✓				objects
Things [14]	✓	$\boxed{✓}$		✓	$\boxed{✓}$	objects
Sintel [1]	$\boxed{✓}$		✓	✓	$\boxed{✓}$	movie
KITTI [16]	$\boxed{✓}$		✓			automotive
HD1K [13]	$\boxed{✓}$		✓			automotive
Driving [14]	$\boxed{✓}$			✓	$\boxed{✓}$	automotive
VIPER [18]	✓	$\boxed{✓}$	✓			automotive
Spring [15]	$\boxed{✓}$		✓			movie

Table A1. Overview of our evaluation datasets and available splits. Train and validation splits both offer ground truth while only the former is intended for training. The test split refers to data without public ground truth which can only be evaluated by submitting results to a public leaderboard. ✓ indicates the availability of a certain split. $\boxed{✓}$ indicates that we used the respective split for evaluation. For the datasets that offer a clean and a final rendering pass we only use the final pass for evaluation.

A.2. Model Testbed

Tab. A2 lists the models and their variants evaluated for this study as well as links to their source code, available checkpoints, number of weights and publication year. For model-specific settings, *e.g.* the padding of frames and the number of iterations, we apply the Sintel default settings to all datasets.

In the following, we comment on some architectural details of some model variants. For more architecture and training details, we refer to the respective publications.

GMFlow+. GMFlow+ [26] provides three variants which we refer to GMFlow+ (s1), GMFlow+ (s2) and the complete GMFlow+ model. By GmFlow+ (s1), we refer to the transformer backbone. GmFlow+ (s2) adds a post-processing step which the authors call hierarchical matching refinement. The complete GMFlow+ model uses the output of GmFlow+ (s2) as initialization to the RAFT model and performs six refinement steps. While both post-processing extensions improve the average effective robustness, the transformer-based initialization leads to an overall low effective robustness for all variants.

Architecture	**GitHub Repository**	**C**	**C+T**	**S**	**K**	**Weights**	**Year**
FlowNet2 [9]	NVIDIA/flownet2-pytorch						2017
FlowNet2C		✗	✓	✗	✗	39 175 298	
FlowNet2S		✗	✓	✗	✗	38 676 506	
FlowNet2		✗	✓	✗	✗	162 518 834	
IRR-PWC [8]	visinf/irr	✓	✓	✓	✓	6 362 092	2019
RAFT [23]	princeton-vl/RAFT	✓	✓	✓	✓	5 257 536	2020
GMA [12]	zacjiang/GMA	✓	✓	✓	✓	5 879 873	2021
SKFlow [22]	littlespray/SKFlow	✗	✓	✓	✓	6 273 149	2022
MemFlow [2]	DQiaole/MemFlow	✗	✓	✓	✓	6 273 149	2024
RPKNet [17]	hmorimitsu/ptlflow	✓	✓	✓	✓	2 846 624	2024
FlowFormer [7]	drinkingcoder/FlowFormer-Official	✓	✓	✓	✓	16 168 113	2022
MatchFlow [3]	DQiaole/MatchFlow						2023
MatchFlow (R)		✗	✓	✓	✗	15 446 225	
MatchFlow (G)		✓	✓	✓	✓	15 446 225	
GMFlow+ [26]	LiheYoung/UniMatch						2023
GMFlow-scale1 (s1)		✗	✓	✗	✗	4 680 288	
GMFlow-scale2 (s2)		✗	✓	✓	✗	4 716 720	
GMFlow-scale2-regrefine6 (GMFlow+)		✗	✓	✓	✓	7 360 688	
SEA-RAFT [25]	princeton-vl/SEA-RAFT						2024
SEA-RAFT (S)		✓	✓	✓	✓	8 883 460	
SEA-RAFT (M)		✓	✓	✓	✓	19 663 876	
SEA-RAFT (L)		✓	✓	✓	✓	19 663 876	
MS-RAFT+ [11]	cv-stuttgart/MS_RAFT_plus	✗	(✓)	✗	✗	16 177 316	2023
CCMR+ [10]	cv-stuttgart/CCMR	✗	✓	✓	✓	11 523 620	2024

Table A2. Model architectures and their variants in our model testbed. All model implementations were taken from the respective GitHub repositories. The columns C, C+T, S and K indicate whether checkpoints from the respective training stage are available. For MS-RAFT+, we used an author-provided C+T checkpoint.

SEA-RAFT. SEA-RAFT [25] provides three different variants. The only difference between SEA-RAFT(S) and SEA-RAFT(M) is the feature encoder. While SEA-RAFT(S) uses the first six layers of ResNet-18, SEA-RAFT(M) uses 13 layers of ResNet-34 as feature and context encoders. SEA-RAFT(L) uses the same weights as SEA-RAFT(M) but increases the number of refinement iterations. While SEA-RAFT(M) applies four iterations, SEA-RAFT(L) applies twelve. Both modifications to the architecture improve the general accuracy of the model but decrease its effective robustness.

MemFlow. MemFlow [2] extends on SKFlow [22] adding the ability to perform predictions in a multi-frame setting. For our evaluations, we only apply MemFlow on pairs of two frames to ensure comparability with all other models.

Training Stage	OOD Dataset	Linear Fit in Logit Space $\beta(\mathrm{WAUC_{ID}}, a, b) = \mathrm{expit}(a \cdot \mathrm{logit}(\mathrm{WAUC_{ID}}) + b)$		R^2
		a	b	
C+T	Sintel	0.670	0.627	0.980
	KITTI	0.741	−0.507	0.964
	HD1K	0.807	0.701	0.873
	Driving	0.749	−0.670	0.983
	VIPER	0.521	−0.271	0.686
	Spring	0.707	1.518	0.533
S	Sintel	1.084	0.534	0.953
	KITTI	0.925	−0.078	0.840
	HD1K	1.212	0.550	0.810
	Driving	0.671	−0.235	0.967
	VIPER	0.517	−0.049	0.600
	Spring	0.671	1.760	0.357
K	Sintel	0.823	0.696	0.944
	KITTI	−0.208	1.791	0.116
	HD1K	0.466	1.700	0.706
	Driving	0.336	0.302	0.715
	VIPER	0.405	0.223	0.529
	Spring	0.653	1.569	0.579

Table A3. Parameters for the linear fits to the distribution shift from Things to each OOD dataset. Fits to the C+T checkpoints are visualized in Main Fig. 4 and used as baselines for the effective robustness evaluation. For completeness, we also provide the linear fits for checkpoints fine-tuned to Sintel (S) and KITTI (K) visualized in Main Fig. 5 for Driving and in Figs. A2 and A3 for the other datasets. A gray background indicates that the listed dataset was used for training and can therefore not be used for OOD robustness evaluation. The R^2 score [5] in the range $[0, 1]$ indicates the quality of the fit in logit space.

Training Details. For all models except MS-RAFT+, we use publicly available checkpoints. For MS-RAFT+ [11], we use an author-provided C+T checkpoint, which was not part of the original publication. All models were trained following the curriculum learning schedule outlined in Appendix 4.1 with two exceptions: Firstly, SEA-RAFT [25] was additionally pre-trained on TartanAir [24] before applying the curriculum learning. We still include SEA-RAFT in our analysis, as it does not violate the requirement that it has not seen the OOD datasets and also follows the linear trend. Secondly, CCMR+ [10] includes the VIPER dataset in the KITTI fine-tuning. This, however, does not affect our analysis for the C+T checkpoints.

A.3. Baseline Parameters

Tab. A3 lists the parameters for the logit-linear fits. The fit to the C+T checkpoints was used to compute the effective robustness in Main Tab. 3 and can also be used to evaluate the robustness of future models.

B. Other Robustness Concepts

In this section, we compare the effective robustness to dataset shifts to corruption and adversarial robustness.

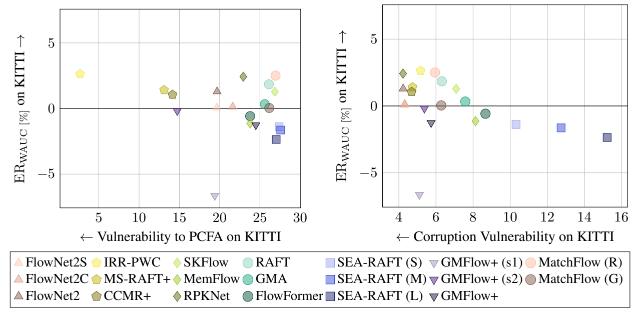

Figure A1. Effective robustness of each model over vulnerability to the perturbation-constrained flow attack (PCFA) [19] and common corruptions [6, 21]. Note that for adversarial and corruption robustness, smaller values (←) indicate a more robust model. The effective robustness weakly correlates with corruption robustness but not with adversarial robustness.

B.1. Setup for Corruption Robustness

For evaluating corruption robustness, we selected 16 corruptions from RobustSpring [21] (namely Gaussian noise, shot noise, impulse noise, speckle noise, Gaussian blur, defocus blur, glass blur, zoom blur, brightness, contrast, saturate, elastic transform, pixelate, JPEG compression, spatter, frost) as well as the three weather corruptions snow, rain and fog from Schmalfuss et al. [20]. For each corruption, we calculate the corruption robustness proposed in [21] as end point error between the prediction on the perturbed and the unperturbed frames:

$$\mathrm{EPE}(f_{\mathrm{clean}}, f_{\mathrm{corrupted}}).$$

In Fig. A1, we visualize the median corruption robustness over all 19 corruptions.

B.2. Setup for Adversarial Robustness

We evaluate the adversarial robustness using the implementation at https://github.com/cv-stuttgart/FlowUnderAttack for the perturbation-constrained optical flow attack [19] using the default parameters with zero target on KITTI (train). As proposed by Schmalfuss et al. [19], we measure adversarial robustness as the end-point error between the prediction on the attacked and the unattacked frames:

$$\mathrm{EPE}(f_{\mathrm{clean}}, f_{\mathrm{attacked}}).$$

B.3. Comparison of Robustness Concepts

Fig. A1 compares effective robustness to adversarial and corruption robustness. Contrary to the clear accuracy-robustness tradeoff observed in [19], there is no notable correlation between effective and adversarial robustness. While Fig. A1 *[right]* suggests a weak correlation between robustness to common corruptions and the effective robustness to dataset shifts, not every model follows this trend necessitating the separate evaluation of all robustness types.

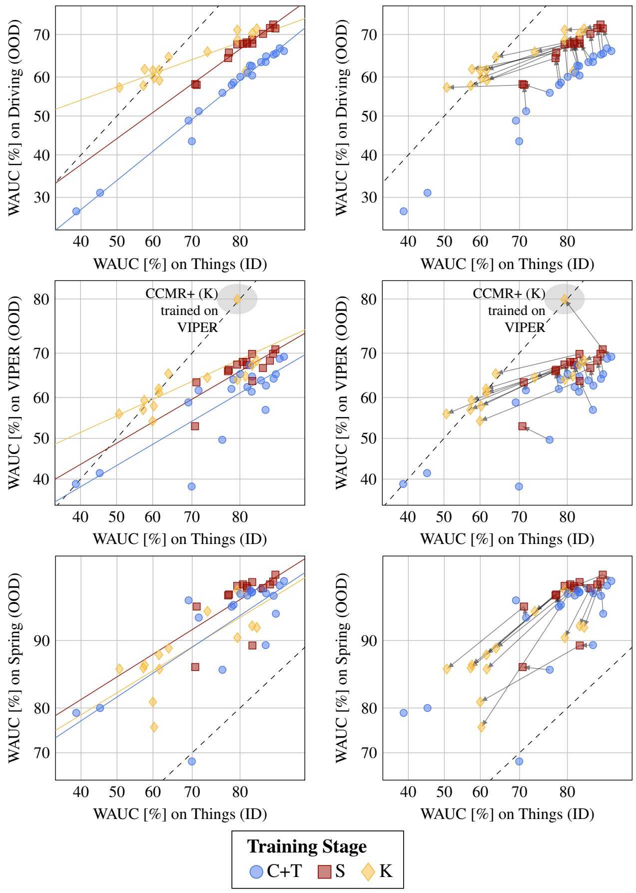

Figure A2. Distribution shift from Things to the datasets Driving *[top]*, VIPER *[middle]* and Spring *[bottom]* for checkpoints at the training stages C+T, S and K. *[Left]* Lines represent the logit-linear fit to the WAUC at a fixed training stage. *[Right]* Arrows connect checkpoints of the same architecture at different training stages.

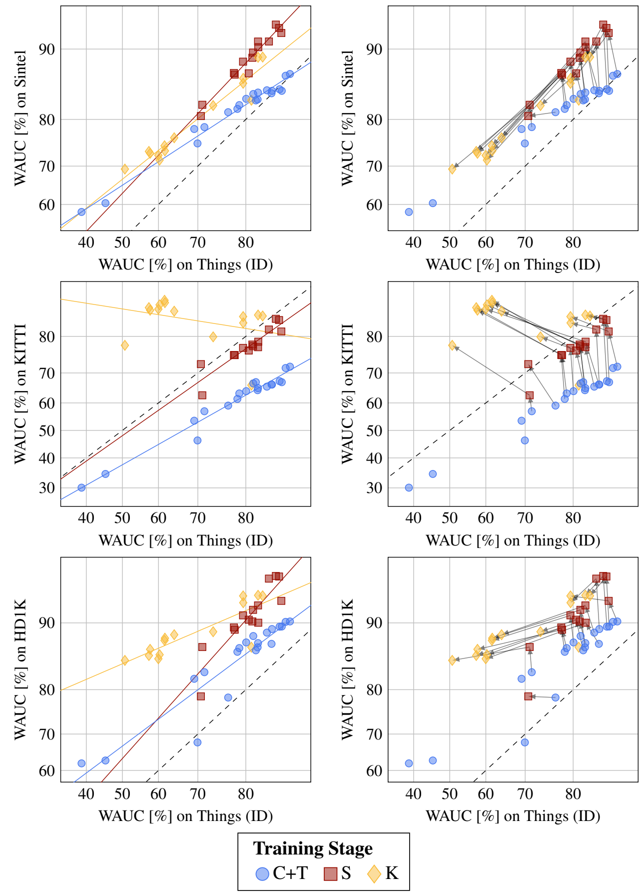

Figure A3. Distribution shift from Things to the datasets Sintel *[top]*, KITTI *[middle]*, HD1K *[bottom]* for checkpoints at the training stages C+T, S and K. *[Left]* Lines represent the logit-linear fit to the WAUC at a fixed training stage. *[Right]* Arrows connect checkpoints of the same architecture at different training stages. Checkpoints S and K were trained on Sintel, KITTI and HD1K which means these datasets are not OOD for these checkpoints and only given here for the sake of completeness.

C. More Visualizations

C.1. Distribution Shift Plots For All Datasets

Analogously to Main Fig. 5, Fig. A2 *[left]* visualizes the WAUC on the distribution shift from Things to each of the other OOD datasets. Moreover, Fig. A2 *[right]* visualizes the training trajectories. The training trajectories show that, for most models, fine-tuning to Sintel improves the OOD accuracy while fine-tuning to KITTI degrades overall model accuracy. Fig. A3 provides analogous visualizations for Sintel, KITTI and HD1K, though one should keep in mind that these datasets were used for training the visualized model checkpoints.

C.2. Distribution Shift Plots of RAFT Checkpoints

Analogously to Main Fig. 6, Fig. A4 visualizes the WAUC of our RAFT checkpoints on the distribution shift from Things to each of the OOD datasets. While the trend of the RAFT checkpoints does not always parallel the robustness baseline, the variance of the OOD WAUC for checkpoints with similar accuracy is smaller than the effective robustness of the official RAFT model. This shows that the deviation from the robustness baseline, *i.e*. our effective robustness metric, is not due to randomness in training.

C.3. Distribution Shift Plots For Architectures

Fig. A5 visualizes the WAUC of the C+T checkpoints, similarly to Figs. A2 and A3 but with different markers for each architecture. On VIPER, the ID-OOD correlation is slightly weaker, which indicates that model architecture has a larger influence on the performance on this dataset than on the other ones.

Fig. A6 visualizes the effective robustness of the C+T checkpoints over their ID accuracy. By definition, the ef-

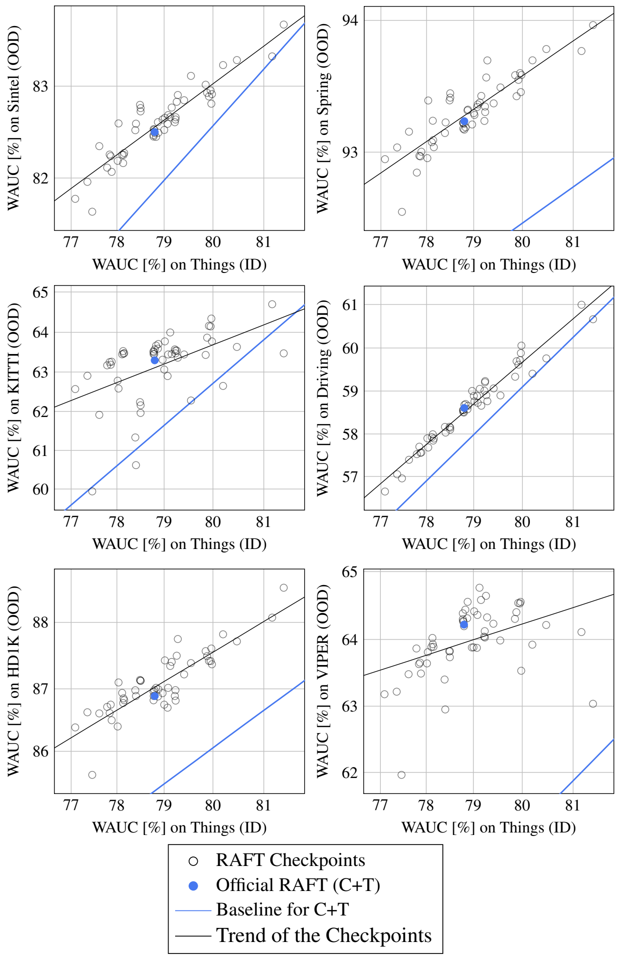

Figure A4. Distribution shift plot of our trained RAFT checkpoints for the dataset shift from Things to each of the OOD datasets. The shift to Driving is the same as Main Fig. 6 and included here for sake of completeness.

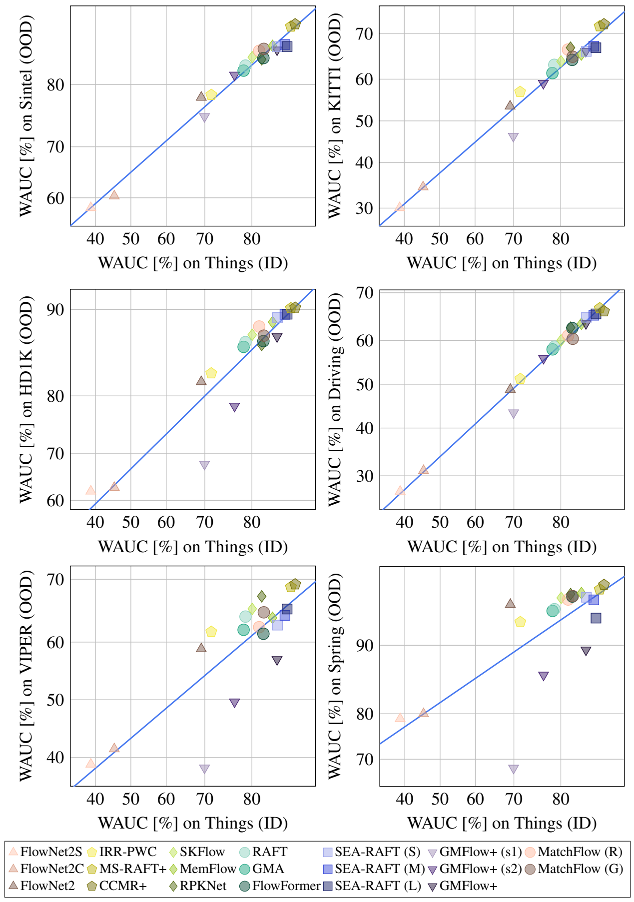

Figure A5. Distribution shift plot from Things to each of the OOD datasets.

fective robustness metric decouples robustness and model accuracy. Therefore, the high robustness of the most accurate models MS-RAFT+ and CCMR+ is a stronger finding than just their high OOD accuracy.

C.4. Publication Year and Number of Weights

Fig. A7 shows the effective robustness of each model for each dataset over its publication year. Fig. A8 shows the effective robustness of each model over its number of weights. The corresponding Main Fig. 7 only showed the effective robustness averaged over all OOD datasets.

D. Raw Results

D.1. Model Accuracies

Tabs. A4 to A8 list the WAUC, inlier rates with thresholds 1, 3, and 5 pixels as well as the EPE of all evaluated model checkpoints on all datasets. For our robustness analysis we used the WAUC, as it exhibits the largest correlation, *cf*. Main Tab. 2.

D.2. RAFT Checkpoints

Tab. A9 lists the training parameters of all RAFT checkpoints which we trained as well as the WAUC on each evaluated dataset.

References

[1] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In *Proc. European Conference on Computer Vision (ECCV)*, pages 611–625, 2012. 1

[2] Qiaole Dong and Yanwei Fu. MemFlow: optical flow estimation and prediction with memory. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19068–19078, 2024. 1

[3] Qiaole Dong, Chenjie Cao, and Yanwei Fu. Rethinking optical flow from geometric matching consistent perspective. In

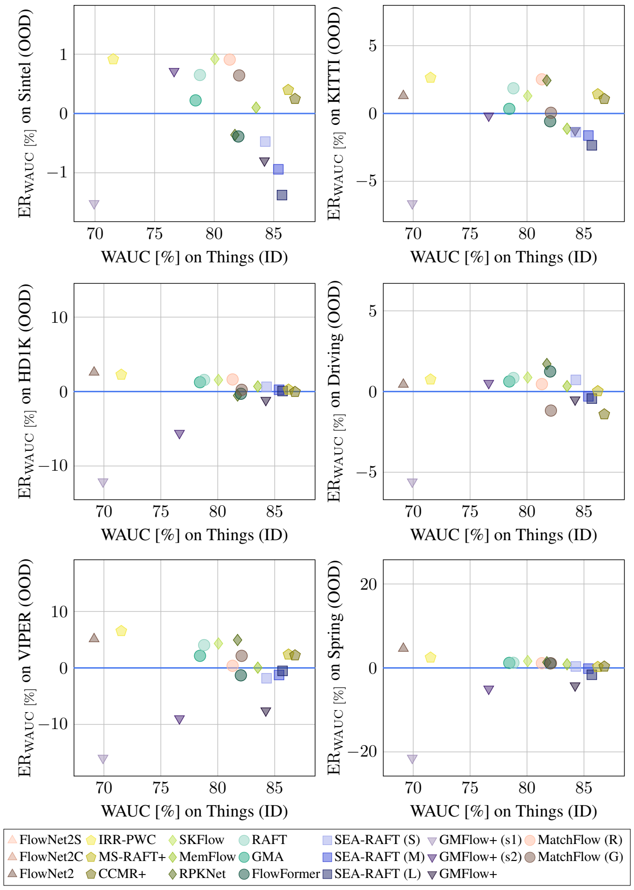

Figure A6. Effective robustness over the WAUC on the ID dataset Things.

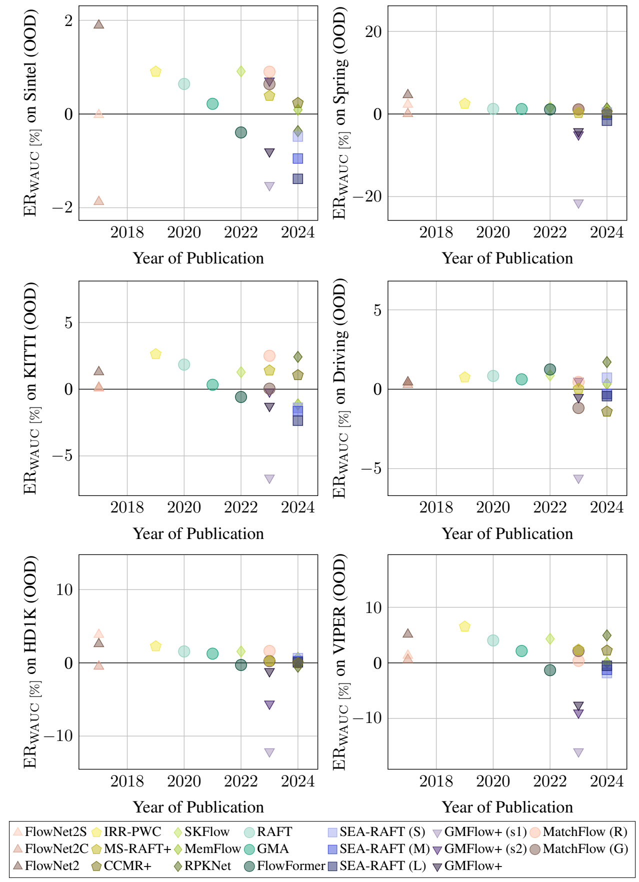

Figure A7. Effective robustness over the year of publication for each architecture.

Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1337–1347, 2023. 1

[4] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: learning optical flow with convolutional networks. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2758–2766, 2015. 1

[5] Norman R Draper and Harry Smith. *Applied regression analysis*. John Wiley & Sons, 1998. 2

[6] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 2

[7] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. FlowFormer: a transformer architecture for optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 668–685, 2022. 1

[8] Junhwa Hur and Stefan Roth. Iterative residual refinement for joint optical flow and occlusion estimation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5754–5763, 2019. 1

[9] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: evolution of optical flow estimation with deep networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2462–2470, 2017. 1

[10] Azin Jahedi, Maximilian Luz, Marc Rivinius, and Andrés Bruhn. CCMR: high resolution optical flow estimation via coarse-to-fine context-guided motion reasoning. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6899–6908, 2024. 1, 2

[11] Azin Jahedi, Maximilian Luz, Marc Rivinius, Lukas Mehl, and Andrés Bruhn. MS-RAFT+: high resolution multi-scale RAFT. *International Journal of Computer Vision (IJCV)*, 132(5):1835–1856, 2024. 1, 2

[12] Shihao Jiang, Dylan Campbell, Yao Lu, Hongdong Li, and Richard Hartley. Learning to estimate hidden motions with global motion aggregation. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9772–9781, 2021. 1

[13] Daniel Kondermann, Rahul Nair, Katrin Honauer, Karsten Krispin, Jonas Andrulis, Alexander Brock, Burkhard Gussefeld, Mohsen Rahimimoghaddam, Sabine Hofmann, Claus Brenner, et al. The HCI benchmark suite: stereo and flow ground truth with uncertainties for urban autonomous driv-

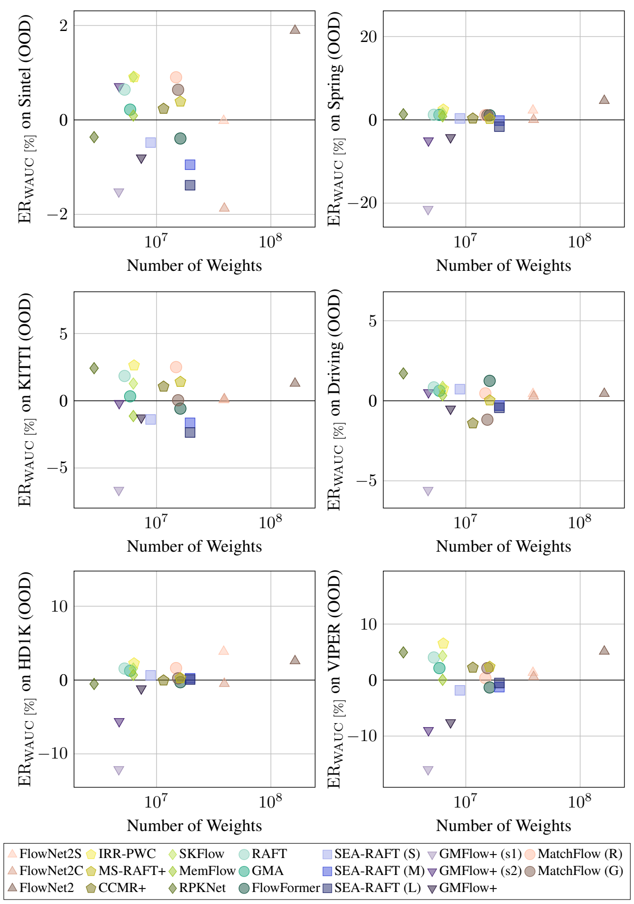

Figure A8. Effective robustness over the number of weights for each architecture.

ing. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 19–28, 2016. 1

[14] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4040–4048, 2016. 1

[15] Lukas Mehl, Jenny Schmalfuss, Azin Jahedi, Yaroslava Nalivayko, and Andrés Bruhn. Spring: a high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4981–4991, 2023. 1

[16] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3070, 2015. 1

[17] Henrique Morimitsu, Xiaobin Zhu, Xiangyang Ji, and Xu-Cheng Yin. Recurrent partial kernel network for efficient optical flow estimation. In *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pages 4278–4286, 2024. 1

[18] Stephan R. Richter, Zeeshan Hayder, and Vladlen Koltun. Playing for benchmarks. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2213–2222, 2017. 1

[19] Jenny Schmalfuss, Philipp Scholze, and Andrés Bruhn. A perturbation-constrained adversarial attack for evaluating the robustness of optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 183–200, 2022. 2

[20] Jenny Schmalfuss, Lukas Mehl, and Andrés Bruhn. Distracting downpour: adversarial weather attacks for motion estimation. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10106–10116, 2023. 2

[21] Jenny Schmalfuss, Victor Oei, Lukas Mehl, Madlen Bartsch, Shashank Agnihotri, Margret Keuper, and Andrés Bruhn. RobustSpring: Benchmarking robustness to image corruptions for optical flow, scene flow and stereo. *arXiv preprint arXiv:2505.09368*, 2025. 2

[22] Shangkun Sun, Yuanqi Chen, Yu Zhu, Guodong Guo, and Ge Li. SKFlow: learning optical flow with super kernels. *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, 35:11313–11326, 2022. 1

[23] Zachary Teed and Jia Deng. RAFT: recurrent all-pairs field transforms for optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 402–419, 2020. 1

[24] Wenshan Wang, Delong Zhu, Xiangwei Wang, Yaoyu Hu, Yuheng Qiu, Chen Wang, Yafei Hu, Ashish Kapoor, and Sebastian Scherer. TartanAir: a dataset to push the limits of visual slam. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4909–4916, 2020. 2

[25] Yihan Wang, Lahav Lipson, and Jia Deng. SEA-RAFT: simple, efficient, accurate RAFT for optical flow. In *Proc. European Conference on Computer Vision (ECCV)*, pages 36–54, 2024. 1, 2

[26] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, Fisher Yu, Dacheng Tao, and Andreas Geiger. Unifying flow, stereo and depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023. 1

Training Stage **C**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	WAUC [%]						
IRR-PWC	**68.45**	**77.61**	**51.93**	**80.26**	**45.79**	**58.05**	**92.09**
RAFT	64.05	71.48	42.27	72.06	36.72	48.61	88.20
GMA	64.20	72.45	44.14	75.00	41.41	49.72	89.61
RPKNet	64.36	72.97	45.45	74.23	40.45	54.43	90.83
FlowFormer	67.02	74.46	45.33	74.96	42.66	50.42	89.44
MatchFlow (G)	66.68	73.98	41.09	73.89	36.90	43.68	87.16
SEA-RAFT (S)	65.53	71.04	42.89	73.52	37.35	50.98	89.59
SEA-RAFT (M)	65.21	71.51	44.89	76.45	40.69	52.11	90.13
SEA-RAFT (L)	64.84	69.79	41.65	74.02	38.48	51.22	89.22

Training Stage **C+T**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	WAUC [%]						
FlowNet2S	38.73	57.90	30.06	62.04	27.06	38.84	79.03
FlowNet2C	45.26	60.36	34.48	62.86	31.01	41.43	79.99
FlowNet2	69.14	78.16	53.56	82.03	48.79	58.84	93.59
IRR-PWC	71.53	78.54	57.01	83.18	51.24	61.73	92.22
RAFT	78.80	82.51	63.30	86.89	58.61	64.22	93.25
GMA	78.44	81.87	61.40	86.36	58.00	62.05	93.11
SKFlow	80.06	83.53	64.07	87.65	60.04	65.46	94.04
MemFlow	83.52	84.84	65.59	88.88	63.64	64.02	94.36
RPKNet	81.74	83.28	67.08	86.58	62.83	67.42	94.29
FlowFormer	82.03	83.42	64.40	86.99	62.71	61.40	94.13
MatchFlow (R)	81.30	84.27	66.67	88.47	61.06	62.49	93.93
MatchFlow (G)	82.10	84.50	65.10	87.56	60.37	64.91	94.16
GMFlow+ (s1)	69.92	75.20	46.31	67.82	43.45	38.24	67.76
GMFlow+ (s2)	76.63	81.29	59.04	78.43	55.98	49.63	86.36
GMFlow+	84.22	84.39	66.31	87.46	63.70	57.03	89.49
SEA-RAFT (S)	84.29	84.75	66.27	89.30	65.01	62.83	94.08
SEA-RAFT (M)	85.39	84.99	67.39	89.58	65.44	64.41	93.91
SEA-RAFT (L)	85.69	84.75	67.06	89.59	65.72	65.43	92.55
MS-RAFT+	86.21	86.87	71.50	90.11	**66.90**	68.83	94.58
CCMR+	**86.78**	**87.09**	**71.90**	**90.15**	66.27	**69.25**	**94.85**

Training Stage **S**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	WAUC [%]						
IRR-PWC	71.05	82.56	62.66	87.03	58.00	63.59	93.13
RAFT	77.93	87.02	75.21	89.23	65.93	66.44	94.00
GMA	77.77	87.23	75.24	89.48	64.62	66.11	93.94
SKFlow	79.50	88.62	77.14	90.79	67.79	67.57	94.56
MemFlow	82.10	90.75	77.33	91.75	68.87	69.84	94.78
RPKNet	80.52	87.18	76.45	90.34	67.93	68.15	94.65
FlowFormer	81.26	89.63	77.58	91.34	67.95	67.48	94.44
MatchFlow (G)	81.14	89.06	78.00	90.16	68.26	68.14	94.55
GMFlow+ (s2)	70.72	80.61	72.65	78.66	58.21	53.01	86.74
GMFlow+	82.14	90.19	78.71	90.02	67.97	63.88	89.45
SEA-RAFT (S)	83.83	90.76	81.57	93.90	70.14	66.82	94.42
SEA-RAFT (M)	84.86	**92.27**	**83.82**	**94.08**	71.40	68.42	94.65
SEA-RAFT (L)	85.31	91.99	83.57	94.05	**72.14**	69.87	94.82
CCMR+	**85.63**	91.55	81.13	92.16	71.31	**70.80**	**95.21**

Training Stage **K**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	WAUC [%]						
IRR-PWC	50.70	69.28	77.81	85.09	57.35	56.05	86.48
RAFT	57.72	73.01	85.47	86.71	61.93	59.33	87.02
GMA	57.39	73.44	85.94	85.82	57.83	57.07	86.61
SKFlow	61.62	74.69	**87.20**	87.92	61.71	62.07	88.35
MemFlow	64.15	76.39	85.29	88.63	64.86	65.55	89.14
RPKNet	73.49	82.43	79.89	88.99	66.11	64.66	92.75
FlowFormer	61.71	73.56	86.80	88.20	59.14	61.15	86.46
MatchFlow (G)	60.30	71.44	86.43	85.95	59.75	57.97	76.04
GMFlow+	59.97	72.54	85.67	85.34	61.61	54.27	81.12
SEA-RAFT (S)	79.53	86.53	82.99	92.00	68.79	64.21	90.30
SEA-RAFT (M)	82.11	89.11	84.57	**92.64**	70.25	67.43	91.48
SEA-RAFT (L)	**82.85**	**89.14**	84.38	92.59	**71.23**	68.46	91.32
CCMR+	79.51	85.84	84.27	92.59	70.98	**79.95**	**94.28**

Table A4. WAUC of all model checkpoints on all datasets. Best and second-best per column are **bold** and underlined respectively. Our analysis focused on C+T checkpoints.

Training Stage **C**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	IR_{1px} **[%]**						
IRR-PWC	**70.64**	**80.47**	**49.18**	**82.89**	**42.69**	**57.65**	**95.09**
RAFT	65.19	73.16	34.34	71.98	29.09	44.34	90.81
GMA	66.46	75.62	37.88	75.05	36.95	46.13	92.33
RPKNet	65.07	74.31	39.02	74.25	34.19	51.32	92.69
FlowFormer	69.27	77.73	39.73	75.93	38.76	48.07	92.36
MatchFlow (G)	68.34	76.67	36.57	74.31	33.19	41.28	90.33
SEA-RAFT (S)	66.84	71.22	35.99	72.81	32.40	48.04	91.79
SEA-RAFT (M)	66.98	72.35	38.31	78.52	35.54	49.61	92.07
SEA-RAFT (L)	64.98	68.64	34.02	74.02	33.08	48.99	91.25

Training Stage **C+T**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	IR_{1px} **[%]**						
FlowNet2S	31.99	57.35	23.55	61.42	18.84	34.74	86.57
FlowNet2C	39.83	60.26	27.38	62.86	23.01	37.20	86.79
FlowNet2	70.62	80.43	52.43	85.31	47.20	59.24	95.48
IRR-PWC	73.67	81.24	55.84	86.58	49.49	62.11	95.11
RAFT	80.67	85.03	63.14	91.20	59.09	65.06	95.47
GMA	80.30	84.52	60.92	90.47	58.27	62.60	95.38
SKFlow	81.76	85.85	63.11	92.00	60.10	66.01	95.67
MemFlow	85.09	86.99	65.18	93.49	63.92	64.18	95.73
RPKNet	83.77	85.66	67.27	90.79	63.26	68.29	**96.38**
FlowFormer	83.75	86.03	63.98	91.59	63.18	61.84	96.01
MatchFlow (R)	82.97	86.37	66.63	93.00	60.97	63.31	95.52
MatchFlow (G)	83.59	86.54	65.15	91.37	60.96	65.50	95.99
GMFlow+ (s1)	73.78	79.05	41.61	71.38	38.71	28.92	59.61
GMFlow+ (s2)	81.90	85.33	58.19	83.08	56.03	46.82	93.39
GMFlow+	86.54	86.96	67.34	92.36	64.48	56.57	94.41
SEA-RAFT (S)	85.57	86.80	66.19	93.68	65.27	62.97	95.64
SEA-RAFT (M)	86.59	86.91	67.83	93.92	65.40	64.57	95.36
SEA-RAFT (L)	86.93	86.65	67.07	93.92	65.38	65.63	93.88
MS-RAFT+	87.33	88.32	73.06	93.96	**67.12**	69.39	96.25
CCMR+	**87.71**	**88.40**	**73.28**	**93.98**	66.50	**69.57**	96.33

Training Stage **S**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	IR_{1px} **[%]**						
IRR-PWC	73.42	85.10	62.83	91.96	57.95	64.47	95.74
RAFT	79.88	88.77	76.28	93.79	66.04	67.11	95.60
GMA	79.74	89.01	76.56	94.01	64.92	66.74	95.81
SKFlow	81.33	90.16	78.69	95.45	68.26	68.15	95.97
MemFlow	83.83	92.31	78.91	96.47	69.32	70.21	96.26
RPKNet	82.70	89.03	77.94	94.87	68.06	69.15	96.35
FlowFormer	83.10	91.25	79.37	96.14	68.39	67.83	96.14
MatchFlow (G)	82.77	90.53	79.75	94.44	68.62	68.43	95.87
GMFlow+ (s2)	78.08	88.57	74.30	88.87	61.21	52.16	92.99
GMFlow+	84.49	91.87	80.78	94.98	68.17	64.43	93.90
SEA-RAFT (S)	85.20	92.37	83.82	98.04	70.88	67.38	95.84
SEA-RAFT (M)	86.13	**93.92**	**86.19**	**98.12**	72.11	68.86	95.99
SEA-RAFT (L)	86.62	93.63	85.98	98.10	**73.22**	70.28	96.21
CCMR+	**86.65**	92.54	83.31	96.17	71.28	**70.98**	**96.46**

Training Stage **K**

Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
	IR_{1px} **[%]**						
IRR-PWC	51.18	71.53	80.00	89.67	56.87	56.69	90.76
RAFT	58.88	76.15	87.85	91.33	61.71	59.97	91.32
GMA	58.77	76.59	88.24	90.13	57.99	57.74	90.78
SKFlow	62.63	77.21	**89.48**	92.92	61.81	62.50	91.97
MemFlow	65.87	78.92	87.72	93.46	65.45	66.19	92.75
RPKNet	76.13	85.00	82.12	93.89	66.23	65.49	95.36
FlowFormer	63.48	76.78	89.07	93.37	58.56	61.62	91.06
MatchFlow (G)	61.22	74.07	88.68	90.57	59.52	58.11	77.38
GMFlow+	61.59	75.83	88.48	90.63	62.42	53.91	85.83
SEA-RAFT (S)	81.58	88.64	85.33	96.38	69.57	65.12	93.17
SEA-RAFT (M)	83.89	91.32	86.86	**97.18**	71.13	68.27	94.62
SEA-RAFT (L)	**84.68**	**91.35**	86.73	97.15	**72.64**	69.29	94.44
CCMR+	81.01	87.45	86.89	96.58	71.08	**80.87**	**95.48**

Table A5. IR_{1px} of all model checkpoints on all datasets.

Training Stage **C**

	IR$_{3px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	80.60	88.45	**68.89**	95.09	**62.01**	**70.86**	**98.21**
RAFT	82.12	88.08	62.79	93.54	56.48	64.92	96.77
GMA	81.46	88.52	63.53	94.32	59.99	65.22	97.24
RPKNet	81.56	88.67	65.24	93.36	60.80	70.56	98.05
FlowFormer	81.63	88.86	65.17	94.06	60.08	65.31	97.11
MatchFlow (G)	82.29	87.85	57.09	92.49	51.87	56.27	95.73
SEA-RAFT (S)	**83.83**	**89.07**	62.67	94.14	54.14	65.95	96.91
SEA-RAFT (M)	82.99	88.49	65.02	**96.01**	58.93	66.93	96.87
SEA-RAFT (L)	83.25	87.83	61.36	94.92	56.25	65.81	96.55

Training Stage **C+T**

	IR$_{3px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
FlowNet2S	60.11	79.69	47.80	85.72	45.69	57.03	96.23
FlowNet2C	67.51	81.66	54.17	85.53	50.84	60.43	96.40
FlowNet2	81.72	88.64	68.81	95.11	63.86	70.80	98.13
IRR-PWC	83.24	89.49	73.15	96.34	67.33	74.72	98.35
RAFT	89.09	91.72	80.60	97.80	72.64	76.22	98.35
GMA	88.59	91.30	78.76	97.56	72.13	74.72	98.45
SKFlow	90.24	92.43	82.56	98.12	75.03	77.80	98.65
MemFlow	92.03	92.96	83.53	98.33	77.78	77.08	98.69
RPKNet	91.22	92.37	84.45	97.33	77.45	**80.24**	**98.81**
FlowFormer	91.13	92.55	82.38	97.95	76.61	74.64	98.65
MatchFlow (R)	90.35	92.48	83.70	98.32	75.19	72.84	98.35
MatchFlow (G)	90.88	92.61	81.65	97.97	73.77	76.70	98.60
GMFlow+ (s1)	87.64	90.21	65.27	89.53	63.44	59.02	96.46
GMFlow+ (s2)	90.50	92.27	76.77	95.26	72.30	67.79	98.34
GMFlow+	**92.47**	92.85	82.04	97.94	75.80	71.35	98.33
SEA-RAFT (S)	91.38	92.36	83.30	98.49	77.39	76.25	98.08
SEA-RAFT (M)	92.00	92.34	84.06	98.41	77.77	77.38	97.98
SEA-RAFT (L)	92.40	92.58	85.22	**98.50**	**78.54**	78.83	97.37
MS-RAFT+	91.93	92.86	84.79	98.00	76.83	78.33	98.60
CCMR+	92.39	**93.18**	**85.52**	97.95	76.60	79.34	98.68

Training Stage **S**

	IR$_{3px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	81.61	91.75	78.78	97.77	73.49	76.15	98.49
RAFT	88.39	94.70	91.73	98.57	80.43	77.80	98.37
GMA	88.00	94.80	91.50	98.63	78.90	77.80	98.53
SKFlow	89.73	95.84	93.14	99.28	82.43	78.59	98.58
MemFlow	91.11	97.22	93.39	99.51	82.95	81.47	98.65
RPKNet	90.45	94.70	92.79	98.88	82.62	80.03	**98.83**
FlowFormer	90.51	96.35	92.82	99.38	81.87	79.68	98.50
MatchFlow (G)	90.24	95.83	93.01	98.73	82.03	79.54	98.61
GMFlow+ (s2)	89.43	95.25	89.59	98.41	75.72	70.03	98.08
GMFlow+	91.04	96.65	93.54	98.84	81.12	77.36	98.24
SEA-RAFT (S)	91.00	96.60	94.77	**99.80**	82.56	78.75	98.36
SEA-RAFT (M)	91.52	**97.46**	**95.70**	99.80	83.48	80.54	98.57
SEA-RAFT (L)	**92.06**	97.36	95.65	99.79	**84.08**	**82.42**	98.74
CCMR+	91.58	96.40	94.43	99.24	82.81	81.17	98.79

Training Stage **K**

	IR$_{3px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	64.38	81.96	92.42	97.00	73.37	68.06	95.88
RAFT	72.42	85.75	97.02	97.74	78.48	71.70	96.09
GMA	71.17	86.02	97.04	97.40	73.16	68.76	95.78
SKFlow	76.78	87.44	**97.69**	98.11	78.21	74.66	97.25
MemFlow	78.01	88.19	97.00	98.16	80.23	78.02	97.30
RPKNet	86.33	92.38	94.68	98.34	81.23	76.96	**98.37**
FlowFormer	76.21	87.11	97.18	98.26	76.05	73.94	95.78
MatchFlow (G)	75.30	86.11	97.00	96.34	75.22	70.41	91.25
GMFlow+	75.81	85.10	97.22	97.34	78.02	69.44	95.55
SEA-RAFT (S)	88.44	94.09	95.35	99.18	81.94	75.28	97.49
SEA-RAFT (M)	89.93	96.12	96.05	**99.60**	82.84	79.20	98.04
SEA-RAFT (L)	**90.80**	**96.27**	95.98	99.54	83.38	80.65	97.86
CCMR+	87.30	92.63	95.87	99.41	**84.24**	**90.00**	98.35

Table A6. IR$_{3px}$ of all model checkpoints on all datasets.

Training Stage **C**

	IR$_{5px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	83.73	90.96	**75.21**	97.50	**68.73**	75.27	**98.89**
RAFT	85.77	90.97	72.15	97.62	65.09	70.72	97.82
GMA	84.91	91.14	71.81	97.90	67.07	71.14	98.29
RPKNet	85.17	**91.63**	73.42	97.13	68.67	**75.68**	98.84
FlowFormer	84.56	91.38	73.63	97.06	66.65	69.86	98.21
MatchFlow (G)	85.56	90.26	64.12	96.66	58.64	61.03	96.93
SEA-RAFT (S)	**86.57**	91.38	70.89	97.70	60.89	71.30	97.82
SEA-RAFT (M)	85.51	90.64	72.86	**98.05**	65.40	71.52	97.75
SEA-RAFT (L)	85.79	90.04	69.92	97.63	63.06	70.34	97.46

Training Stage **C+T**

	IR$_{5px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
FlowNet2S	69.28	85.36	57.85	91.80	56.53	64.51	97.73
FlowNet2C	76.00	86.91	64.22	91.30	61.40	67.95	97.81
FlowNet2	85.27	91.27	73.72	97.43	69.32	74.62	98.72
IRR-PWC	86.33	91.94	78.33	98.16	72.80	78.68	99.01
RAFT	91.50	93.66	85.05	98.99	77.05	79.42	98.92
GMA	91.03	93.26	83.44	98.91	76.64	78.29	99.04
SKFlow	92.61	94.28	87.57	99.21	80.18	81.27	99.18
MemFlow	93.98	**94.70**	88.05	99.26	82.52	81.12	99.22
RPKNet	93.30	94.21	88.89	98.56	82.18	**83.64**	**99.26**
FlowFormer	93.21	94.46	86.95	99.08	81.30	78.23	99.17
MatchFlow (R)	92.44	94.26	88.08	99.27	80.38	75.42	98.91
MatchFlow (G)	92.99	94.38	86.05	99.10	78.33	79.81	99.12
GMFlow+ (s1)	91.14	92.90	72.66	94.11	69.90	67.33	98.48
GMFlow+ (s2)	92.91	94.22	81.83	97.84	77.18	73.91	98.99
GMFlow+	**94.23**	94.56	85.99	99.01	79.97	76.46	98.96
SEA-RAFT (S)	93.08	93.93	87.72	**99.38**	82.02	79.90	98.62
SEA-RAFT (M)	93.59	93.89	88.23	99.29	82.51	80.93	98.53
SEA-RAFT (L)	93.98	94.20	**89.67**	99.37	**83.54**	82.39	97.99
MS-RAFT+	93.39	94.29	88.11	98.93	80.70	81.06	99.17
CCMR+	93.86	94.63	88.85	98.89	80.48	82.37	99.18

Training Stage **S**

	IR$_{5px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	84.22	93.76	82.68	98.81	78.68	79.50	99.09
RAFT	90.80	96.34	95.05	99.44	84.67	80.90	99.05
GMA	90.41	96.47	94.71	99.46	83.34	81.09	99.12
SKFlow	92.09	97.33	96.12	99.71	86.17	82.00	99.20
MemFlow	93.14	**98.39**	96.18	99.83	86.58	84.68	99.27
RPKNet	92.56	96.22	95.79	99.62	86.79	82.90	**99.32**
FlowFormer	92.58	97.72	95.57	99.82	85.87	82.95	99.09
MatchFlow (G)	92.39	97.24	95.86	99.49	85.72	82.55	99.18
GMFlow+ (s2)	92.00	96.85	93.04	99.45	80.66	74.47	98.96
GMFlow+	92.99	97.86	96.12	99.59	85.33	80.91	99.07
SEA-RAFT (S)	92.69	97.63	96.84	99.96	86.15	81.94	98.95
SEA-RAFT (M)	93.09	98.28	**97.48**	**99.96**	86.86	83.67	99.15
SEA-RAFT (L)	**93.61**	98.21	97.44	99.96	**87.28**	**85.65**	99.23
CCMR+	93.13	97.54	96.70	99.70	86.29	84.14	99.25

Training Stage **K**

	IR$_{5px}$ [%]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	67.70	84.92	95.07	98.37	79.08	71.81	96.89
RAFT	76.08	88.55	98.57	98.91	83.51	75.53	97.09
GMA	74.68	88.69	98.59	98.83	78.53	72.46	96.80
SKFlow	80.59	90.41	**98.97**	99.13	82.89	78.90	98.35
MemFlow	81.28	90.83	98.60	99.11	84.38	81.81	98.32
RPKNet	89.06	94.43	97.07	99.25	85.58	80.33	**99.00**
FlowFormer	79.54	89.96	98.57	99.19	81.29	77.75	96.84
MatchFlow (G)	79.08	89.38	98.50	97.84	80.84	74.57	94.61
GMFlow+	79.74	87.76	98.72	98.63	82.75	73.93	97.43
SEA-RAFT (S)	90.40	95.59	97.26	99.74	85.86	78.46	98.41
SEA-RAFT (M)	91.63	97.33	97.76	**99.93**	86.23	82.35	98.78
SEA-RAFT (L)	**92.47**	**97.45**	97.72	99.91	86.63	83.91	98.59
CCMR+	89.21	94.14	97.78	99.81	**88.32**	**92.38**	98.97

Table A7. IR$_{5px}$ of all model checkpoints on all datasets.

Training Stage **C**							
	EPE [px]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	10.15	3.91	10.42	**0.86**	29.53	**55.48**	0.68
RAFT	9.94	4.28	9.93	1.04	29.06	58.08	1.31
GMA	10.02	4.03	11.16	0.94	28.98	61.38	0.95
RPKNet	9.87	**3.77**	10.84	1.06	**25.48**	56.72	**0.65**
FlowFormer	13.48	4.50	10.22	1.04	33.10	65.44	1.05
MatchFlow (G)	**9.72**	4.90	16.14	1.05	34.70	72.10	1.63
SEA-RAFT (S)	11.59	4.72	9.61	0.99	28.94	56.95	1.08
SEA-RAFT (M)	13.44	5.75	**9.16**	0.89	27.98	56.21	1.27
SEA-RAFT (L)	15.79	9.26	10.45	1.03	32.25	55.52	2.81

Training Stage **C+T**							
	EPE [px]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
FlowNet2S	12.45	5.05	14.84	1.79	31.87	59.11	1.29
FlowNet2C	9.35	4.60	12.09	1.81	29.00	56.08	1.16
FlowNet2	7.66	3.55	10.64	0.78	27.20	54.01	0.70
IRR-PWC	7.24	3.36	8.92	0.70	26.38	51.63	0.61
RAFT	4.64	2.82	5.45	0.49	21.46	50.03	0.66
GMA	4.69	2.86	6.07	0.51	23.21	54.28	0.54
SKFlow	3.66	2.44	4.30	0.45	15.95	44.36	0.47
MemFlow	2.90	**2.33**	4.17	0.43	13.41	43.54	0.43
RPKNet	3.55	2.55	**3.73**	0.56	13.08	**40.10**	**0.42**
FlowFormer	3.60	2.36	6.19	0.48	14.64	51.44	0.53
MatchFlow (R)	4.12	2.61	4.71	0.43	18.53	63.88	1.17
MatchFlow (G)	3.42	2.45	4.72	0.46	18.11	53.58	0.51
GMFlow+ (s1)	3.83	2.95	10.49	1.38	25.94	58.41	1.32
GMFlow+ (s2)	3.29	2.47	6.93	0.82	20.68	53.67	0.68
GMFlow+	**2.69**	2.74	5.04	0.46	15.43	50.24	0.71
SEA-RAFT (S)	3.73	3.71	4.48	0.38	15.26	51.91	1.30
SEA-RAFT (M)	3.48	3.87	4.32	0.38	14.84	49.27	1.73
SEA-RAFT (L)	3.22	3.46	3.76	**0.37**	**12.69**	45.60	3.45
MS-RAFT+	3.95	2.63	4.53	0.41	17.01	47.64	0.45
CCMR+	3.66	2.49	4.21	0.41	15.71	42.98	0.43

Training Stage **S**							
	EPE [px]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	10.32	2.41	7.84	0.51	22.37	53.08	0.67
RAFT	5.87	1.19	1.53	0.37	12.69	45.52	0.61
GMA	5.78	1.13	1.58	0.36	16.15	48.49	0.54
SKFlow	4.50	0.77	1.23	0.31	9.74	41.22	0.46
MemFlow	3.89	**0.51**	1.22	0.27	10.69	38.11	**0.39**
RPKNet	4.39	1.46	1.32	0.32	8.97	41.80	0.41
FlowFormer	4.55	0.65	3.04	0.27	11.40	46.30	0.61
MatchFlow (G)	4.06	0.82	1.27	0.34	10.50	50.92	0.48
GMFlow+ (s2)	4.04	1.11	2.00	0.65	16.70	51.33	0.62
GMFlow+	**3.56**	0.63	1.18	0.33	11.51	43.66	0.53
SEA-RAFT (S)	4.71	0.78	1.06	0.19	13.87	46.33	0.51
SEA-RAFT (M)	4.71	0.58	**0.88**	**0.19**	11.59	43.22	0.52
SEA-RAFT (L)	4.28	0.61	0.90	0.19	**8.66**	**35.92**	0.44
CCMR+	4.35	0.74	1.03	0.27	11.36	39.57	0.43

Training Stage **K**							
	EPE [px]						
Method	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
IRR-PWC	57.63	8.23	1.56	0.59	17.00	106.74	24.50
RAFT	18.46	5.97	0.63	0.50	12.17	48.68	3.36
GMA	20.39	5.71	0.60	0.52	18.66	59.11	3.19
SKFlow	13.00	3.94	**0.53**	0.44	11.42	42.60	0.91
MemFlow	14.56	4.06	0.62	0.42	10.45	39.09	1.37
RPKNet	7.16	1.93	1.00	0.39	9.53	46.73	**0.58**
FlowFormer	14.20	4.22	0.61	0.42	13.92	51.19	2.17
MatchFlow (G)	14.38	4.92	0.60	0.60	13.62	55.93	3.14
GMFlow+	11.96	5.05	0.58	0.54	12.61	49.58	1.09
SEA-RAFT (S)	7.16	1.76	0.94	0.27	10.30	50.98	1.08
SEA-RAFT (M)	6.05	0.96	0.81	**0.23**	10.80	43.67	0.81
SEA-RAFT (L)	**5.38**	**0.93**	0.82	0.23	8.25	38.13	0.85
CCMR+	8.45	2.72	0.74	0.24	**4.50**	**28.51**	0.60

Table A8. EPE of all model checkpoints on all datasets.

Training Parameters				**WAUC [%]**						
Steps $[10^3]$	Batch-size	LR $[10^{-3}]$	Weight Decay $[10^{-3}]$	Things	Sintel	KITTI	HD1K	Driving	VIPER	Spring
25	12	0.125	0.100	77.09	81.77	62.57	86.39	56.64	63.18	92.94
50	1	0.125	0.100	77.62	82.35	61.91	86.61	57.39	63.48	93.16
50	6	0.062	0.100	77.36	81.96	62.91	86.63	57.05	63.22	93.04
50	6	0.125	0.010	77.84	82.26	63.25	86.62	57.70	63.63	92.97
50	6	0.125	0.100	77.79	82.12	63.17	86.72	57.52	63.86	92.83
50	6	0.125	1.000	77.87	82.24	63.18	86.75	57.56	63.65	92.97
50	6	0.250	0.100	78.02	82.19	62.78	86.40	57.90	63.64	92.95
50	12	0.125	0.100	77.89	82.07	63.26	86.51	57.56	63.49	93.00
100	1	0.062	0.100	78.04	82.60	62.58	87.10	57.68	63.81	93.41
100	1	0.125	0.010	78.51	82.73	61.96	87.12	58.10	63.73	93.46
100	1	0.125	0.100	78.49	82.80	62.24	87.14	58.17	63.82	93.41
100	1	0.125	1.000	78.51	82.76	62.16	87.13	58.15	63.82	93.43
100	1	0.250	0.100	78.41	82.60	60.62	86.89	58.03	62.95	93.24
100	1	0.250	1.000	78.39	82.53	61.33	86.98	58.16	63.40	93.15
100	6	0.062	0.010	78.14	82.26	63.46	86.85	57.84	63.93	93.10
100	6	0.062	0.100	78.11	82.26	63.45	86.94	57.92	63.89	93.08
100	6	0.062	1.000	78.16	82.28	63.48	86.82	57.89	63.89	93.04
100	6	0.125	0.010	78.80	82.53	63.45	86.98	58.50	64.26	93.23
100	6	0.125	0.050	78.78	82.54	63.50	86.89	58.51	64.27	93.19
100	6	0.125	0.100	78.78	82.47	63.52	86.96	58.55	64.39	93.23
100	6	0.125	0.200	78.78	82.46	63.45	86.82	58.56	64.30	93.23
100	6	0.125	1.000	78.81	82.50	63.57	86.89	58.68	64.20	93.18
100	6	0.250	0.010	79.00	82.53	63.06	86.75	58.89	63.88	93.31
100	6	0.250	0.100	78.97	82.64	63.31	86.93	59.00	63.88	93.22
100	6	0.250	1.000	79.07	82.60	62.90	86.70	58.88	63.87	93.36
100	6	0.500	0.100	77.46	81.62	59.94	85.60	56.95	61.96	92.50
100	6	1.000	0.100	65.87	74.71	47.73	79.36	45.12	53.91	87.66
100	12	0.062	0.100	78.13	82.17	63.53	86.78	57.99	64.02	93.25
100	12	0.125	0.100	78.83	82.45	63.66	87.02	58.70	64.44	93.20
100	12	0.125	1.000	78.88	82.50	63.71	86.99	58.66	64.56	93.34
100	12	0.250	0.010	79.24	82.64	63.54	86.98	59.21	64.12	93.36
100	12	0.250	0.100	79.22	82.61	63.57	86.89	59.01	64.04	93.25
100	12	0.250	1.000	79.24	82.67	63.47	86.82	59.23	64.03	93.32
120	5	0.100	0.100	78.87	82.61	63.59	86.88	58.58	64.31	93.33
120	5	0.125	0.100	79.08	82.69	63.44	87.02	58.73	64.36	93.38
150	4	0.083	0.100	79.02	82.66	63.77	87.44	58.76	64.41	93.30
150	4	0.125	0.100	79.41	82.85	63.44	87.22	59.06	64.23	93.39
200	1	0.125	0.100	79.55	83.11	62.28	87.39	58.90	63.98	93.51
200	3	0.062	0.100	79.16	82.77	63.45	87.41	58.66	64.58	93.44
200	3	0.125	0.100	79.85	83.01	63.43	87.57	59.33	64.30	93.60
200	6	0.062	0.100	79.13	82.67	64.00	87.35	59.05	**64.76**	93.39
200	6	0.125	0.010	79.92	82.89	64.16	87.43	59.61	64.54	93.57
200	6	0.125	0.100	79.88	82.93	63.86	87.50	59.69	64.40	93.44
200	6	0.125	1.000	79.95	82.95	64.15	87.43	59.72	64.53	93.62
200	6	0.250	0.100	79.98	82.91	63.77	87.38	60.05	63.53	93.60
200	12	0.125	0.100	79.97	82.81	64.34	87.61	59.88	64.56	93.47
300	1	0.125	0.100	80.20	83.22	62.65	87.82	59.40	63.92	93.71
300	2	0.042	0.100	79.26	82.83	63.54	87.51	58.90	64.65	93.58
300	2	0.125	0.100	80.47	83.28	63.63	87.72	59.76	64.21	93.79
400	6	0.125	0.100	81.16	83.31	**64.70**	88.07	**60.99**	64.11	93.78
600	1	0.021	0.100	79.28	82.91	63.36	87.76	58.76	64.33	93.71
600	1	0.125	0.100	**81.38**	**83.64**	63.47	**88.49**	60.66	63.04	**93.97**

Table A9. Training parameters and WAUC of our RAFT checkpoints. Accuracy on the dataset shift to Driving is visualized in Fig. 6 in the main paper and to the other OOD datasets in Fig. A4. As our goal was to generate many RAFT checkpoints of varying accuracy, each combination of training parameters was only evaluated once and not all combinations were tested. Rows are sorted by training parameters.