

HuPerFlow: A Comprehensive Benchmark for Human vs. Machine Motion Estimation Comparison

Supplementary Material

Appendix 1 Frame onset errors. In online psychophysical experiments on motion perception, presenting motion stimuli with precise onset timing is essential. Due to varying display configurations across devices, especially with refresh rates spanning 30 to 120 Hz and adaptive refresh rate (ARR) features on some modern devices, adjustments from typical offline experiments were required. To accommodate these differences without imposing strict device requirements, we presented visual stimuli at 30 Hz (33.33 ms per frame) and controlled frame onset timing using the `Clock.getTime` function in PsychoPy. To assess device stability among participants, frame drops and frame onset times were logged. Since each trial included at least three repetitions with varying lengths, we recorded every frame of image presentation only during the third repetition of each trial in the main experiment (see the Procedure section). Fig. S1 shows the distribution of frame onset errors, calculated as deviations from ideal frame onsets (e.g., 33.33 ms, 66.67 ms). The frame drop rate was 0.041%, and the average frame onset error was -0.055 ms (SD = 3.664 ms), indicating consistent and accurate visual stimulus presentation across devices.

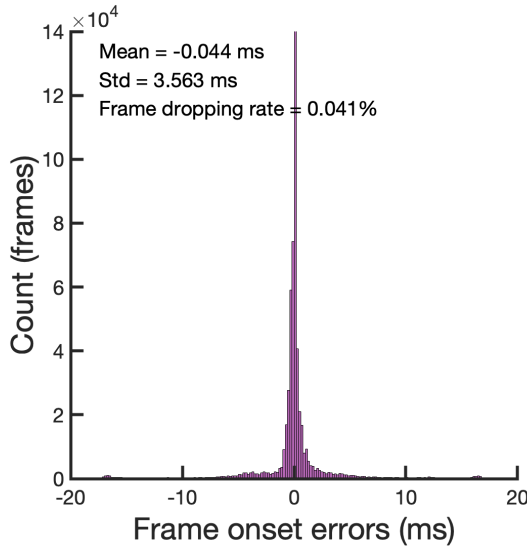


Figure S1. Distribution of frame onset errors

Appendix 2 Correlations between human responses and ground truth. Fig. S2 presents scatter plots comparing

human responses to ground truth data across each dataset. Each point represents the average response from the same locations across four repetitions within an observer.

Appendix 3 Response Consistency Index (RCI). Response Consistency Index (RCI) is an index from [47] to evaluate the similarity between model performance and human flow illusions at each probed location.

The RCI is defined as the product of $A \cdot B \cdot C$ in Equations S1, S2, and S3, measuring the relative alignment of ground truth (G), human response (R), model prediction (M), and the origin (O).

- A (Equation S1) quantifies the deviation of human responses from the ground truth.
- B (Equation S2) indicates the directional similarity between the response error vector \vec{GR} and the model error vector \vec{GM} relative to the ground truth.
- C (Equation S3) compares the distance between model prediction and ground truth $\|\vec{GM}\|$ with the distance between model prediction and response $\|\vec{RM}\|$.

The RCI approaches +1 when the model's prediction aligns closely with human flow illusions and approaches -1 when the prediction diverges in the opposite direction.

$$A = \frac{\|\vec{GR}\|}{\|\vec{OG}\| + \|\vec{OR}\|}. \quad (S1)$$

$$B = \frac{\vec{GR} \cdot \vec{GM}}{\|\vec{GR}\| \|\vec{GM}\|}. \quad (S2)$$

$$C = 0.5 \left(\frac{\|\vec{GM}\| - \|\vec{RM}\|}{\|\vec{GM}\| + \|\vec{RM}\|} + 1 \right) = \frac{\|\vec{GM}\|}{\|\vec{GM}\| + \|\vec{RM}\|}. \quad (S3)$$

Appendix 4 The predictions of optical flow algorithms in each dataset.

To assess which datasets align better between optical flow algorithms and human responses, and which show greater discrepancies, we compare algorithm predictions with human responses and ground truth across datasets. Following the approach in Table 2, we calculated indices for each dataset separately and presented them in Tables S1–S12. Overall, the findings align with the trends in Table 2: biologically inspired models (e.g., FFV1MT, V1Attention-StageI) show stronger partial correlations with human responses, while multi-frame models like VideoFlow align more closely with ground truth in dynamic datasets.

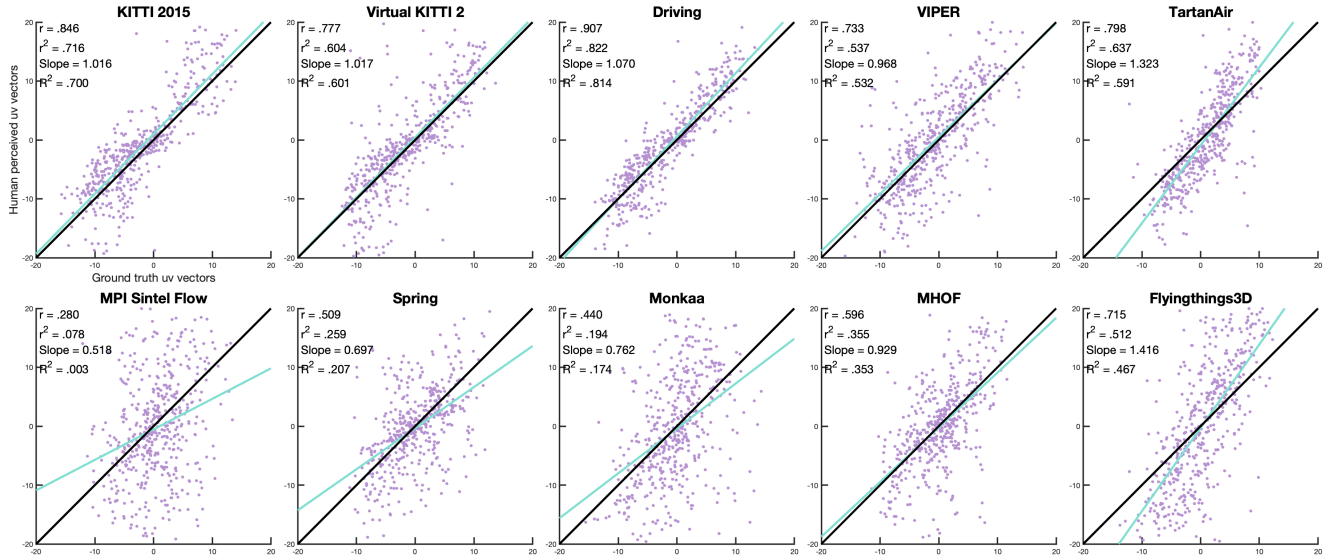


Figure S2. The scatter plots display the relationship between human-perceived UV vectors and ground truth UV vectors from each dataset. In the top-left corner, the Pearson correlation coefficient (r), the squared correlation (r^2), the slope of the linear regression line, and the coefficient of determination (R^2) relative to the ideal response (i.e., the ground truth) are shown.

Video S1 Demonstration of the procedure. This video shows how observers use a mouse to control the direction and speed of circular Brownian noise to match the target motion. After each response, vectors indicating ground truth and observers' responses are displayed during practice trials.

Video S2 Movies of human-perceived flows. These videos display human-perceived flows across the image sequences used in the ten datasets. The video speed (5 Hz) is slower than the actual perceived speed (30Hz) to aid visualization

Table S1. Partial Correlation of uv vectors across Models and Datasets.

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Farnebäck	0.0635	-0.0029	0.1029	0.1806	0.2684	0.1056	0.0440	0.0605	0.2838	0.1504	0.1670
FFV1MT	0.2068	0.2308	0.2246	0.3973	0.3350	0.1568	0.2839	0.2534	0.3670	0.3168	0.3063
RAFT	-0.0251	0.0957	0.1695	0.2817	0.1287	0.0815	0.1224	0.1632	0.2924	0.1480	0.1090
FlowNet2	0.0003	0.0454	0.1514	0.2712	0.2071	0.0401	0.0235	0.1050	0.3020	0.1538	0.1012
FlowFormer	-0.0229	0.0112	0.2137	0.2811	0.0657	0.0597	0.0698	0.1375	0.2771	0.1153	0.0748
VideoFlow	0.0079	0.0352	0.1798	0.2893	0.0582	0.1390	-0.0014	0.2205	0.2945	0.0847	0.0784
Perceiver IO	0.0402	0.0959	0.1373	0.3256	0.2104	0.0249	-0.0719	0.0121	0.2875	0.1858	0.0860
AGFlow	0.0581	0.1231	0.1954	0.2570	0.1030	0.0554	0.2545	0.1528	0.3603	0.1975	0.1589
V1Attention-Stage1	0.0471	-0.0588	0.1094	0.2835	0.3661	0.1646	0.1899	0.1802	0.3507	0.3052	0.2285
DualModel	0.2677	-0.0026	0.0578	0.2371	0.3457	0.1980	0.0016	0.1612	0.2495	0.2708	0.1973

Table S2. Partial Correlation of Direction across Models and Datasets.

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Farnebäck	-0.0867	0.0625	0.0529	0.0245	0.0206	0.1323	0.1681	0.1537	0.3345	0.1414	0.1424
FFV1MT	0.2775	0.4425	0.0464	0.4500	0.2075	0.1959	0.2863	0.1635	0.3175	0.2327	0.2756
RAFT	0.2411	0.4133	-0.0667	0.2428	-0.0364	-0.0532	0.1160	0.2310	0.3679	0.3528	0.2219
FlowNet2	0.3628	0.4406	0.0527	0.1431	0.0625	0.1289	0.2039	0.2164	0.4359	0.2249	0.2600
FlowFormer	0.0828	0.2904	0.1116	0.2819	0.0831	-0.1337	0.1933	0.1942	0.2785	0.1884	0.1878
VideoFlow	0.0644	0.2906	0.1597	0.2035	0.0403	0.0331	0.2285	0.2047	0.3675	0.2460	0.2188
Perceiver IO	-0.1114	0.2931	0.1857	0.0971	-0.0912	-0.0480	0.1029	0.1359	0.3593	0.2845	0.1698
AGFlow	0.2389	0.4662	0.2403	0.4238	0.0157	0.0935	0.0869	0.1815	0.2575	0.0952	0.2417
V1Attention-Stage1	0.1863	0.3324	0.0216	0.3750	0.2954	0.1674	0.2747	0.2562	0.3980	0.2972	0.2845
DualModel	0.3623	0.1425	0.0722	0.3110	0.2883	0.2721	0.1341	0.1630	0.3050	0.3391	0.2451

Table S3. Partial Correlation of Speed across Models and Datasets.

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Farnebäck	0.0768	-0.0505	0.1221	0.0382	0.2171	-0.0052	-0.0321	0.0777	0.0314	-0.0303	0.0410
FFV1MT	0.0679	-0.0050	0.0643	0.1513	0.2450	0.0194	0.1046	-0.1823	0.0908	0.1341	0.0604
RAFT	-0.0024	0.2023	0.2298	0.0132	0.1562	0.0879	0.0092	-0.0868	0.0503	-0.0086	0.0128
FlowNet2	0.0044	0.0446	0.2232	0.0373	0.3099	0.0367	0.0232	-0.0058	0.0581	-0.0007	0.0002
FlowFormer	-0.0322	0.0996	0.2542	0.0402	0.0029	0.0848	-0.1446	-0.0478	0.0382	-0.0445	-0.0232
VideoFlow	0.0155	0.0997	0.2192	-0.0130	0.0329	0.1594	-0.1301	-0.0887	0.0044	-0.0379	-0.0249
Perceiver IO	0.1009	0.2050	0.1456	0.0838	0.1999	0.1165	0.0281	-0.0225	0.0755	0.0381	0.0296
AGFlow	0.0026	0.1580	0.2288	0.0735	0.0872	0.0950	0.0750	-0.1143	0.1255	0.1451	0.0139
V1Attention-Stage1	-0.0401	-0.0801	-0.0868	0.1769	0.3956	-0.0274	-0.0580	-0.0360	0.1400	0.0870	0.0202
DualModel	0.1981	0.1170	0.0436	0.0697	0.3298	0.0240	-0.1591	-0.1164	0.0561	0.1613	0.0589

Table S4. Response Consistent Index (RCI) across Models and Datasets. The RCI is introduced from [47] to represent the model's similarity to human perception (the larger, the more human-aligned).

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Farnebäck	0.0388	0.0725	0.0493	0.0764	0.0722	0.2515	0.1297	0.2342	0.1569	0.1973	0.1279
FFV1MT	0.0706	0.0855	0.0553	0.1472	0.0697	0.2452	0.1749	0.2082	0.2274	0.2152	0.1499
RAFT	0.0121	0.0275	0.0134	0.0250	0.0238	0.0645	0.0731	0.0593	0.1290	0.0480	0.0476
FlowNet2	0.0250	0.0439	0.0190	0.0419	0.0322	0.1289	0.0793	0.0813	0.1544	0.0757	0.0682
FlowFormer	0.0106	0.0308	0.0131	0.0258	0.0189	0.0414	0.0702	0.0605	0.1137	0.0390	0.0424
VideoFlow	0.0128	0.0280	0.0118	0.0242	0.0151	0.0439	0.0553	0.0512	0.1108	0.0270	0.0380
Perceiver IO	0.0233	0.0370	0.0162	0.0293	0.0306	0.0995	0.0629	0.0579	0.1305	0.0477	0.0535
AGFlow	0.0498	0.0597	0.0326	0.0895	0.0439	0.1729	0.1359	0.1315	0.2226	0.1397	0.1078
V1Attention-Stage1	0.0857	0.0927	0.0747	0.1635	0.1530	0.3061	0.1870	0.2647	0.2379	0.2426	0.1808
DualModel	0.0546	0.0692	0.0441	0.1011	0.0635	0.2226	0.1247	0.2065	0.2265	0.1576	0.1270

Table S5. Correlation of uv vectors across Models and Datasets (v.s. Human).

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	0.8461	0.7769	0.9068	0.7330	0.7979	0.2802	0.5089	0.4403	0.5962	0.7154	0.6605
Farneback	0.7894	0.6119	0.7424	0.6813	0.7861	0.2267	0.3610	0.2925	0.5689	0.4227	0.5394
FFV1MT	0.7976	0.7518	0.8447	0.7343	0.8155	0.2620	0.5330	0.4472	0.5622	0.5838	0.6263
RAFT	0.8431	0.7614	0.9069	0.7576	0.8006	0.2902	0.5098	0.4641	0.6217	0.7086	0.6425
FlowNet2	0.8321	0.7470	0.8999	0.7526	0.8075	0.2083	0.4359	0.3674	0.6166	0.6621	0.5851
FlowFormer	0.8436	0.7276	0.9068	0.7576	0.7948	0.2857	0.4971	0.4564	0.6233	0.7046	0.6362
VideoFlow	0.8428	0.7466	0.9098	0.7587	0.7934	0.3030	0.4938	0.4786	0.6261	0.7158	0.6427
Perceiver IO	0.8393	0.7500	0.8975	0.7652	0.8002	0.2423	0.4361	0.4029	0.6110	0.7031	0.6188
AGFlow	0.7717	0.7237	0.9000	0.7227	0.7854	0.2033	0.5159	0.3310	0.5320	0.5178	0.5372
V1Attention-Stage1	0.7036	0.6271	0.6877	0.6462	0.7523	0.2434	0.4364	0.3676	0.5930	0.5809	0.5478
DualModel	0.8507	0.6638	0.6501	0.7127	0.8146	0.3046	0.3369	0.3660	0.5025	0.6204	0.5704

Table S6. Correlation of Direction across Models and Datasets (v.s. Human).

Models / Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	0.5358	0.4939	0.4850	0.4927	0.4856	0.1962	0.3040	0.1912	0.2899	0.6483	0.4286
Farneback	0.4220	0.3683	0.3635	0.4151	0.4386	0.2057	0.3112	0.2327	0.4141	0.3728	0.3762
FFV1MT	0.4746	0.6095	0.4577	0.5869	0.5179	0.2560	0.4017	0.2348	0.3542	0.4515	0.4534
RAFT	0.5693	0.6105	0.4429	0.5359	0.4675	0.1451	0.3079	0.2865	0.4508	0.7017	0.4694
FlowNet2	0.6136	0.6243	0.4744	0.4983	0.4794	0.2266	0.3522	0.2848	0.5079	0.6314	0.4841
FlowFormer	0.5223	0.5404	0.4943	0.5434	0.4796	0.1484	0.3552	0.2682	0.3847	0.6523	0.4560
VideoFlow	0.5127	0.5486	0.5038	0.5234	0.4750	0.1980	0.3567	0.2712	0.4488	0.6738	0.4711
Perceiver IO	0.4872	0.5476	0.4991	0.4947	0.4387	0.1059	0.2996	0.2329	0.4482	0.6628	0.4400
AGFlow	0.5188	0.6348	0.5286	0.5961	0.4577	0.1830	0.2410	0.2555	0.2924	0.5326	0.4468
V1Attention-Stage1	0.4620	0.5392	0.3376	0.5549	0.5334	0.2124	0.3851	0.3048	0.4459	0.5100	0.4473
DualModel	0.6169	0.4556	0.4106	0.5497	0.5466	0.3281	0.2775	0.2420	0.3833	0.5656	0.4510

Table S7. Correlation of Speed across Models and Datasets (v.s. Human).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	0.3432	0.3674	0.5136	0.2869	0.5685	0.2559	0.3954	0.0888	0.3622	0.2659	0.2876
Farneback	0.2860	0.1420	0.3264	0.1726	0.5013	0.0753	0.1426	0.0961	0.2261	0.0169	0.1476
FFV1MT	0.2687	0.2547	0.4008	0.2797	0.5750	0.1321	0.2992	-0.1305	0.2161	0.2205	0.2033
RAFT	0.3383	0.4113	0.5500	0.2627	0.5799	0.2627	0.3561	0.0263	0.3145	0.2274	0.2337
FlowNet2	0.3177	0.3348	0.5441	0.2677	0.6215	0.1811	0.3303	0.0203	0.2728	0.1495	0.1760
FlowFormer	0.3350	0.3562	0.5547	0.2709	0.5422	0.2673	0.2911	0.0439	0.2964	0.2041	0.2133
VideoFlow	0.3396	0.3732	0.5444	0.2599	0.5464	0.2949	0.3314	0.0385	0.2639	0.2434	0.2186
Perceiver IO	0.3552	0.4115	0.5254	0.2947	0.5796	0.2687	0.3581	0.0492	0.2961	0.2365	0.2327
AGFlow	0.2167	0.3680	0.5438	0.2360	0.5280	0.2119	0.2130	-0.0811	0.2048	0.2010	0.1452
V1Attention-Stage1	0.1407	0.1624	0.1529	0.2742	0.5649	0.0522	0.1148	-0.0085	0.2860	0.1539	0.1381
DualModel	0.3885	0.3271	0.3070	0.2232	0.6132	0.1480	0.0793	-0.0746	0.1697	0.2671	0.2000

Table S8. End-Point Error (v.s. Human).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	5.1482	5.5497	3.8906	6.4759	5.8027	10.3036	6.7171	9.7504	6.7639	9.2317	6.9634
Farneback	5.5556	6.3728	5.4047	6.3994	5.2625	8.2336	9.0176	7.5378	5.7985	9.5553	6.9138
FFV1MT	5.3739	5.5053	4.3577	6.1055	5.5255	9.5621	7.3843	8.3927	5.2278	8.3308	6.5766
RAFT	5.1886	5.6502	3.8650	6.2569	5.6337	10.1466	12.3883	9.4856	5.9235	9.1406	7.3679
FlowNet2	5.2660	5.6927	3.9814	6.1794	5.4872	11.0885	12.3818	10.2964	5.6786	9.4780	7.5530
FlowFormer	5.1739	5.8828	3.8653	6.2027	5.7910	10.3444	12.2990	9.4048	6.0709	9.2916	7.4326
VideoFlow	5.1594	5.8054	3.8246	6.3017	5.8205	10.2044	12.6882	9.4029	6.2865	9.2238	7.4717
Perceiver IO	5.2073	5.7400	4.2661	6.5664	5.7617	10.4737	13.1479	10.1592	6.1923	10.0008	7.7515
AGFlow	5.2642	6.0433	3.9644	6.8273	5.9579	13.3879	9.2022	13.1371	5.6402	9.9727	7.9397
V1Attention-Stage1	6.1725	6.1028	5.9365	6.4236	4.7982	8.0337	6.4705	7.4639	4.5903	7.6485	6.3640
DualModel	5.0696	6.1217	6.2542	6.5805	5.7466	9.9290	9.8038	8.6702	5.2384	9.6581	7.3072

Table S9. Correlation of uv vectors across Models and Datasets (v.s. GT).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground Truth	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Farneback	0.9170	0.7890	0.7893	0.8381	0.8947	0.4947	0.6531	0.5624	0.6709	0.4604	0.6769
FFV1MT	0.8811	0.8783	0.8824	0.7637	0.9273	0.4576	0.7085	0.6038	0.5213	0.5596	0.6994
RAFT	0.9975	0.9577	0.9878	0.9689	0.9890	0.9424	0.9214	0.9500	0.8141	0.9420	0.9259
FlowNet2	0.9834	0.9499	0.9775	0.9429	0.9852	0.6377	0.8346	0.6767	0.7793	0.8453	0.8201
FlowFormer	0.9980	0.9332	0.9804	0.9672	0.9887	0.9701	0.9350	0.9437	0.8473	0.9496	0.9325
VideoFlow	0.9956	0.9522	0.9949	0.9723	0.9874	0.9745	0.9710	0.9580	0.8274	0.9877	0.9437
Perceiver IO	0.9880	0.9385	0.9757	0.9671	0.9564	0.8155	0.9080	0.9045	0.7848	0.9060	0.8929
AGFlow	0.8958	0.8852	0.9707	0.8672	0.9634	0.5697	0.7112	0.4780	0.4621	0.5647	0.6811
V1Attention-Stage1	0.8143	0.8335	0.7233	0.6915	0.7646	0.3377	0.6009	0.5214	0.6265	0.5663	0.6270
DualModel	0.9560	0.8555	0.6976	0.8601	0.9166	0.4997	0.6599	0.5585	0.5659	0.6710	0.7045

Table S10. Correlation of direction across Models and Datasets (v.s. GT).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground Truth	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Farneback	0.8578	0.6635	0.6795	0.8176	0.8859	0.4613	0.6039	0.5672	0.4334	0.4246	0.6497
FFV1MT	0.5095	0.6277	0.9088	0.5053	0.9516	0.4131	0.6109	0.5022	0.1931	0.4528	0.5882
RAFT	0.8877	0.7907	0.9505	0.9090	0.9768	0.8702	0.7902	0.8207	0.5113	0.9218	0.8455
FlowNet2	0.8091	0.8201	0.9478	0.9024	0.9532	0.6869	0.7066	0.7286	0.5858	0.8244	0.8022
FlowFormer	0.9254	0.7648	0.9709	0.9547	0.9343	0.9547	0.8699	0.7850	0.5749	0.9190	0.8663
VideoFlow	0.9162	0.8127	0.9759	0.9317	0.9570	0.9674	0.9502	0.8264	0.4912	0.9745	0.8804
Perceiver IO	0.9590	0.7970	0.8558	0.9510	0.9531	0.7089	0.7865	0.7916	0.5788	0.8422	0.8237
AGFlow	0.6989	0.6953	0.9185	0.6266	0.9323	0.5386	0.5687	0.5719	0.1709	0.7473	0.6595
V1Attention-Stage1	0.6356	0.6443	0.6671	0.5927	0.7415	0.2787	0.5446	0.3750	0.2747	0.4807	0.5384
DualModel	0.8732	0.7592	0.7622	0.7564	0.8663	0.4785	0.5667	0.5900	0.3989	0.5364	0.6676

Table S11. Correlation of Speed across Models and Datasets (v.s. GT).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground Truth	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Farneback	0.6792	0.4975	0.4538	0.4901	0.6405	0.3128	0.4281	0.2353	0.5572	0.1720	0.3874
FFV1MT	0.6401	0.7025	0.7040	0.5543	0.7976	0.4508	0.5546	0.4023	0.3805	0.3798	0.5374
RAFT	0.9867	0.9251	0.9257	0.8959	0.9486	0.8532	0.8910	0.8337	0.7886	0.8707	0.7865
FlowNet2	0.9211	0.8525	0.8875	0.8719	0.9390	0.5962	0.8033	0.2909	0.6382	0.5644	0.6113
FlowFormer	0.9892	0.8283	0.9574	0.8804	0.9526	0.9162	0.8897	0.8095	0.7537	0.8523	0.7890
VideoFlow	0.9814	0.9133	0.9679	0.9224	0.9458	0.9277	0.9407	0.8887	0.7207	0.9560	0.8088
Perceiver IO	0.9432	0.8330	0.9403	0.9127	0.8847	0.7684	0.8742	0.7270	0.6739	0.8083	0.7435
AGFlow	0.6258	0.7275	0.8710	0.6326	0.8656	0.5217	0.3775	0.3071	0.2527	0.2461	0.4640
V1Attention-Stage1	0.5048	0.6036	0.4286	0.4192	0.4968	0.3028	0.4131	0.2899	0.4719	0.2754	0.4192
DualModel	0.8283	0.6707	0.5363	0.5897	0.7798	0.5001	0.5170	0.3714	0.3324	0.4971	0.5288

Table S12. End-Point Error (v.s. GT).

Models/Datasets	KITTI 2015	Virtual KITTI 2	Driving	VIPER	TartanAir	MPI Sintel	Spring	Monkaa	MHOF	Flythings3D	All Datasets
Ground truth	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Farneback	2.3448	4.3035	4.0169	3.3315	2.2899	8.0080	7.3517	7.7063	5.2443	10.5800	5.5177
FFV1MT	3.3863	3.5651	2.9159	5.6433	2.3130	9.3565	6.8494	7.7386	7.0360	9.6041	5.8408
RAFT	0.4900	1.1488	0.5472	0.8620	0.6431	2.1057	8.1461	2.0466	3.8601	2.0892	2.1939
FlowNet2	1.1011	1.8750	0.9257	1.4860	0.7700	5.8037	8.4985	3.7901	4.4807	3.7035	3.2434
FlowFormer	0.4159	1.6521	0.5475	0.8451	0.6086	1.4738	7.9602	1.9914	3.5032	1.7794	2.0777
VideoFlow	0.4811	1.3266	0.4025	0.7785	0.5140	1.4016	7.7573	1.7247	3.6526	1.0431	1.9082
Perceiver IO	0.9487	1.7568	1.0664	1.2783	1.0709	4.0563	8.5428	2.7422	4.1413	3.0831	2.8687
AGFlow	2.1481	2.9862	1.3214	4.0459	1.6407	10.2528	7.7348	9.6162	7.3885	7.2425	5.4377
V1Attention-Stage1	4.9537	4.9182	5.9380	6.8169	5.0388	10.1230	6.5684	8.8315	6.4919	10.3447	7.0025
DualModel	2.2578	3.7184	4.5595	4.3126	2.3973	8.8834	7.8088	7.6007	6.7577	8.2937	5.6590