JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Self-supervised AutoFlow

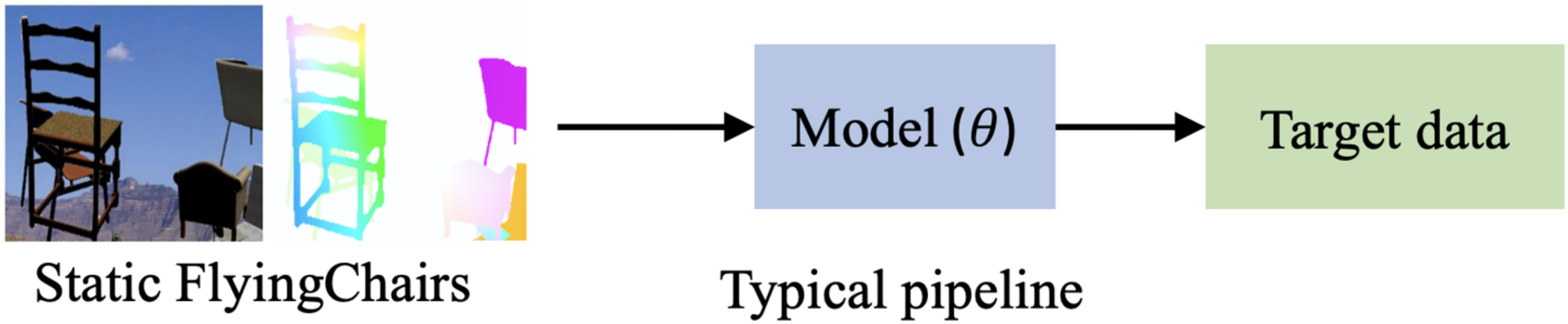
Hsin-Ping Huang[1,2] Charles Herrmann[1] Junhwa Hur[1] Erika Lu[1]
Kyle Sargent[1] Austin Stone[1] Ming-Hsuan Yang[1,2] Deqing Sun[1]

[1]Google Research [2]University of California, Merced

Poster: WED-AM-303

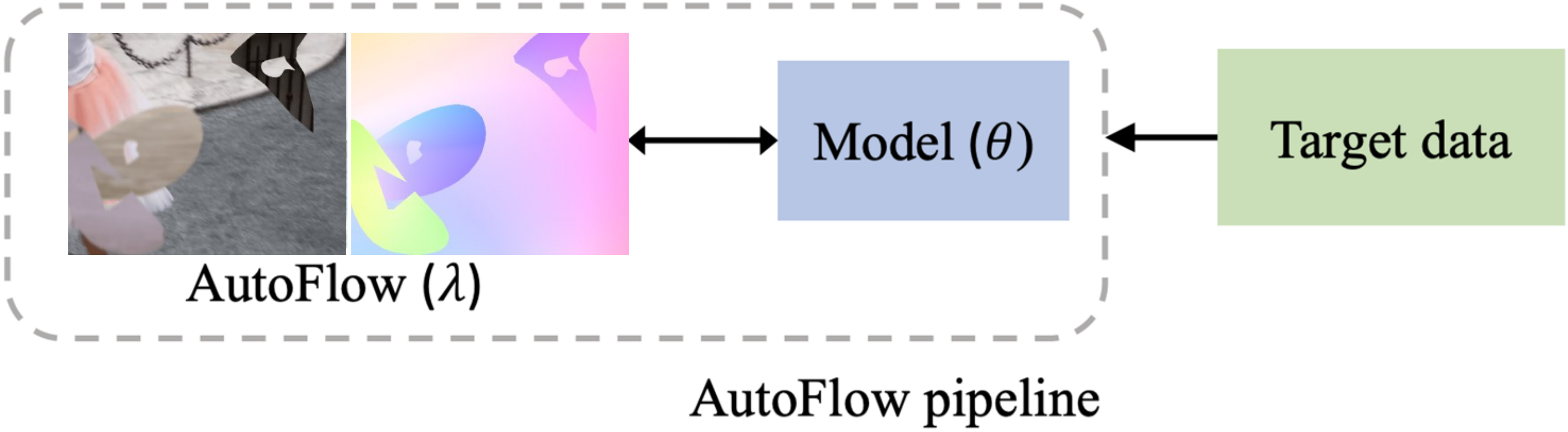
Webpage: autoflow-google.github.io

Typical pipeline for learning optical flow



- Pretrain on large-scale synthetic datasets
- Issue: exist a domain gap between the synthetic and target data

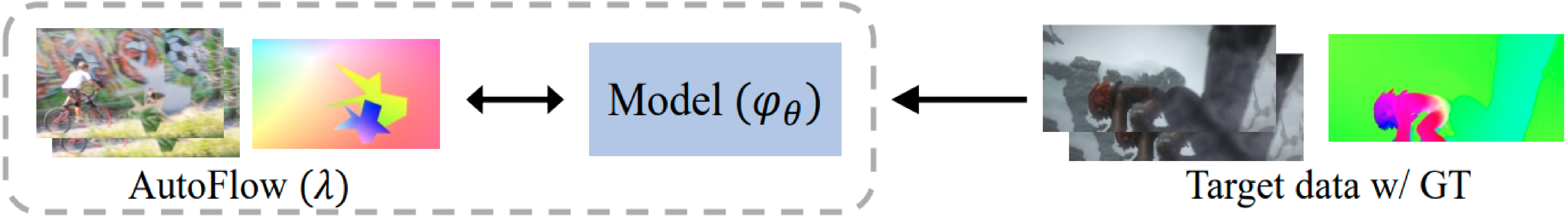
AutoFlow: learning a training set for optical flow



- Learn a training set to optimize performance on a target dataset

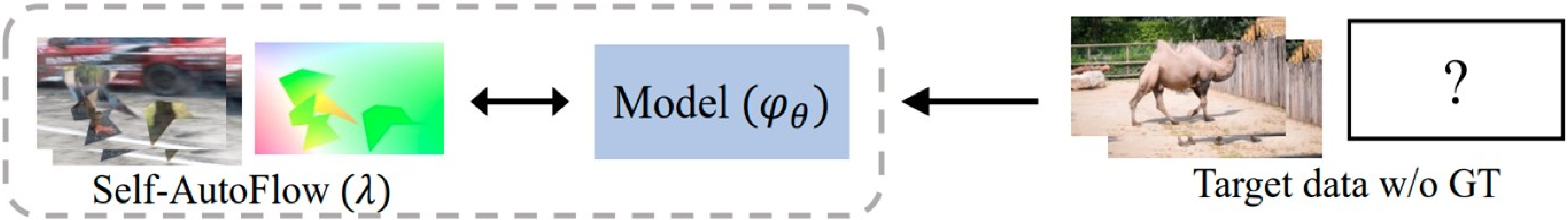
Sun et al. AutoFlow: Learning a Better Training Set for Optical Flow. CVPR 2021

Issues: rely on ground truth from target domain



(Supervised) AutoFlow

Can we remove the reliance on ground truth?

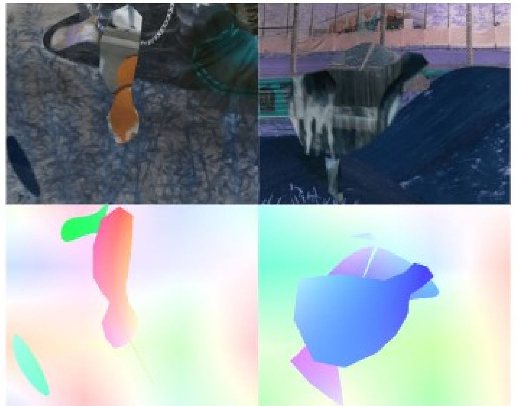


Self-supervised AutoFlow
Self-supervised learning + learning to render

Self-supervised learning for optical flow

		Sintel Clean [3]		Sintel Final [3]		KITTI 2015 [22]			
		EPE		EPE		EPE	EPE (noc)	ER in %	
	Method	*train*	*test*	*train*	*test*	*train*	*train*	*train*	*test*
Supervised in domain	FlowNet2-ft [9]	(1.45)	4.16	(2.01)	5.74	(2.30)	–	(8.61)	11.48
	PWC-Net-ft [28]	(1.70)	3.86	(2.21)	5.13	(2.16)	–	(9.80)	9.60
	SelFlow-ft [17] (MF)	(1.68)	[3.74]	(1.77)	{4.26}	(1.18)	–	–	8.42
	VCN-ft [37]	(1.66)	2.81	(2.24)	4.40	(1.16)	–	(4.10)	6.30
	RAFT-ft [29]	(0.76)	**1.94**	(1.22)	**3.18**	(0.63)	–	(1.5)	**5.10**
Supervised out of domain	FlowNet2 [9]	2.02	**3.96**	3.14	**6.02**	9.84	–	28.20	–
	PWC-Net [28]	2.55	–	3.93	–	10.35	–	33.67	–
	VCN [37]	2.21	–	3.62	–	8.36	–	25.10	–
	RAFT [29]	**1.43**	–	**2.71**	–	**5.04**	–	**17.4**	–
Unsupervised	EPIFlow [42]	3.94	7.00	5.08	8.51	5.56	2.56	–	16.95
	DDFlow [16]	{2.92}	6.18	{3.98}	7.40	[5.72]	[2.73]	–	14.29
	SelFlow [17] (MF)	[2.88]	[6.56]	{3.87}	{6.57}	[4.84]	[2.40]	–	14.19
	UnsupSimFlow [10]	{2.86}	5.92	{3.57}	6.92	[5.19]	–	–	[13.38]
	ARFlow [14] (MF)	{2.73}	{4.49}	{3.69}	{5.67}	[2.85]	–	–	[11.79]
	UFlow [12]	3.01	5.21	4.09	6.50	2.84	1.96	9.39	11.13
	SMURF-test (ours)	**1.99**	–	**2.80**	–	**2.01**	**1.42**	**6.72**	–
	SMURF-train (ours)	{1.71}	**3.15**	{2.58}	**4.18**	{2.00}	{1.41}	{6.42}	**6.83**

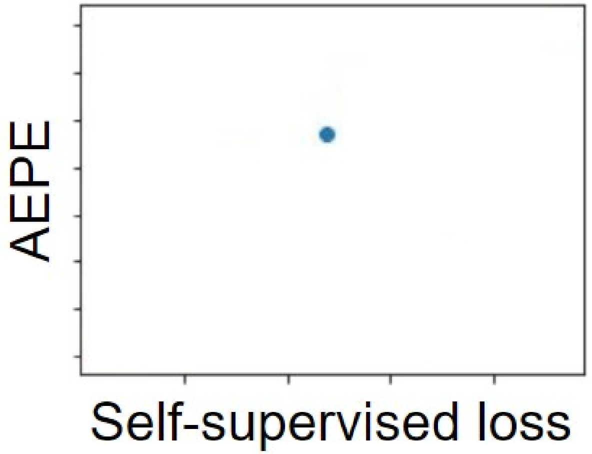
Are self-supervised losses related to ground truth errors?



A set of hyperparameters of AutoFlow

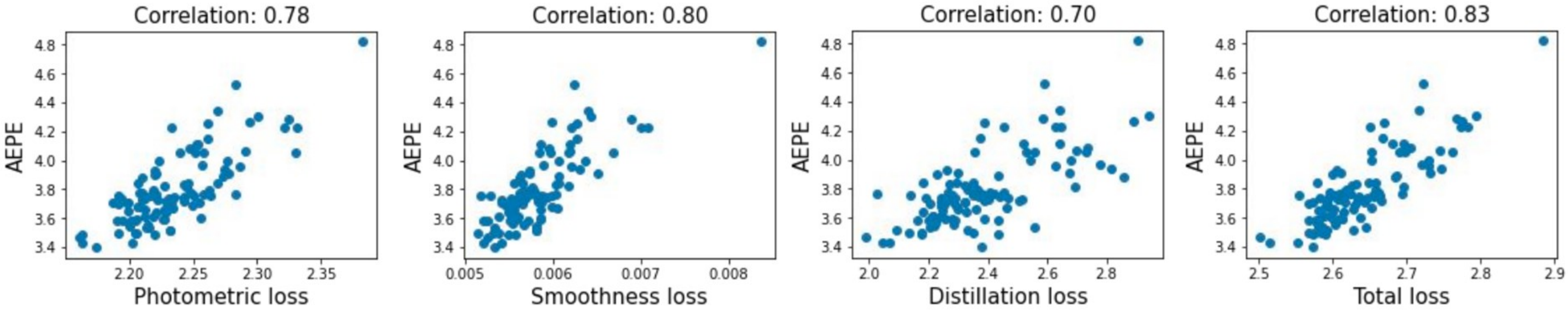
→ Model (φ_θ) →

Train a model



Compute self-supervised loss and AEPE on target dataset

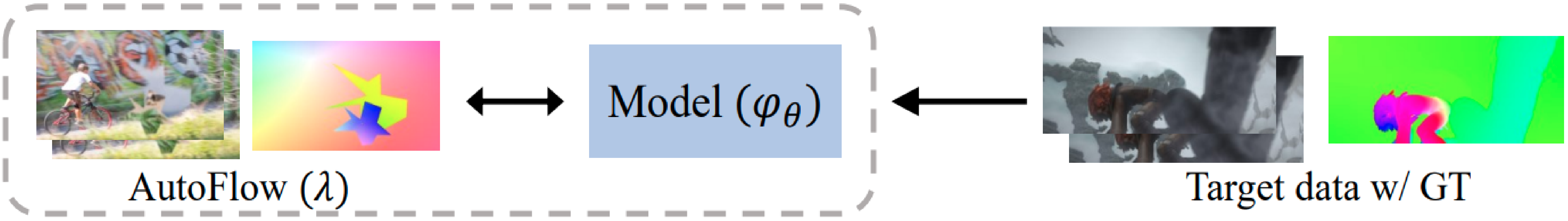
Are self-supervised losses related to ground truth errors?



$$\mathcal{L} = \mathcal{L}_{\text{photo}} + \omega_{\text{smooth}}\mathcal{L}_{\text{smooth}} + \omega_{\text{distill}}\mathcal{L}_{\text{distill}}$$

- Strong correlation: self-supervised losses can be search metric for AutoFlow

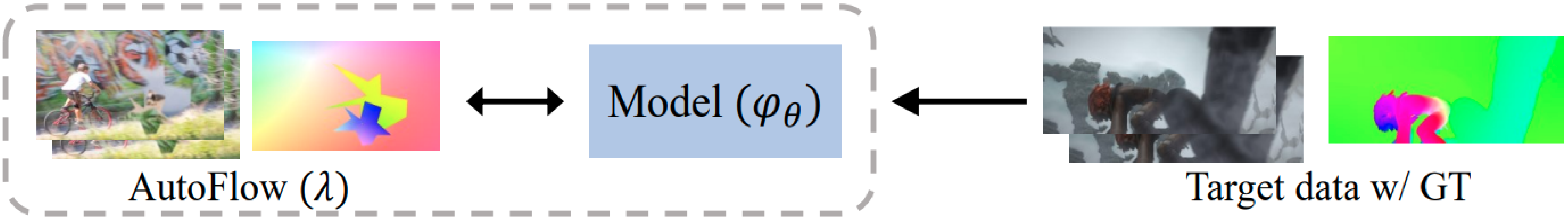
AutoFlow



$$\lambda^* = \underset{\lambda \in \Lambda}{\operatorname{argmin}}\, \Omega\left(\phi_\theta(\lambda)\right)$$

- Search for an optimal set of hyperparameters λ so that the optical flow network $\phi_\theta(\lambda)$ trained on the dataset rendered with λ minimize a search metric Ω on the target dataset

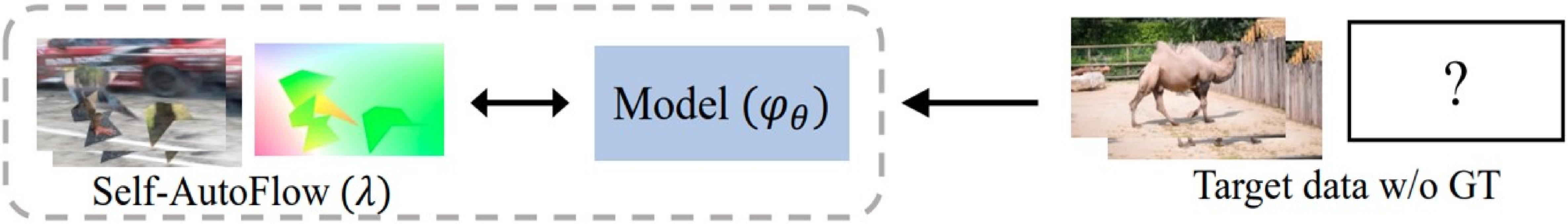
(Supervised) AutoFlow



$$\Omega_{\mathrm{AF}}(\phi_\theta(\lambda)) = \mathrm{AEPE}$$

- Learn a training dataset to optimize the performance in the target domain with labels by minimizing the ground truth error

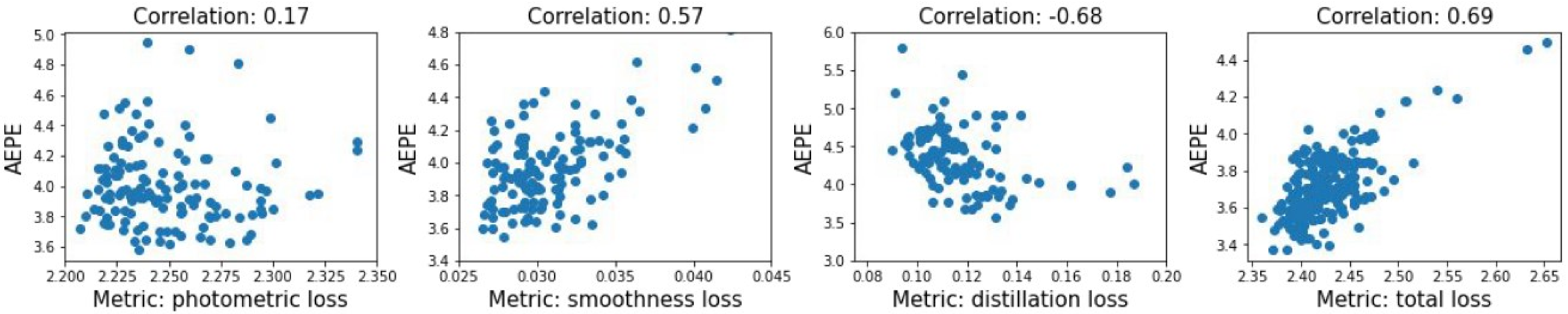
Self-supervised AutoFlow



$$\Omega_{\text{S-AF}}(\phi_\theta(\lambda)) = \mathcal{L}_{\text{photo}} + \omega_{\text{smooth}}\mathcal{L}_{\text{smooth}} + \omega_{\text{distill}}\mathcal{L}_{\text{distill}}$$

- Learn a training dataset to approximately optimize the performance in the unlabeled target domain by minimizing the self-supervision metric

Self-supervised AutoFlow

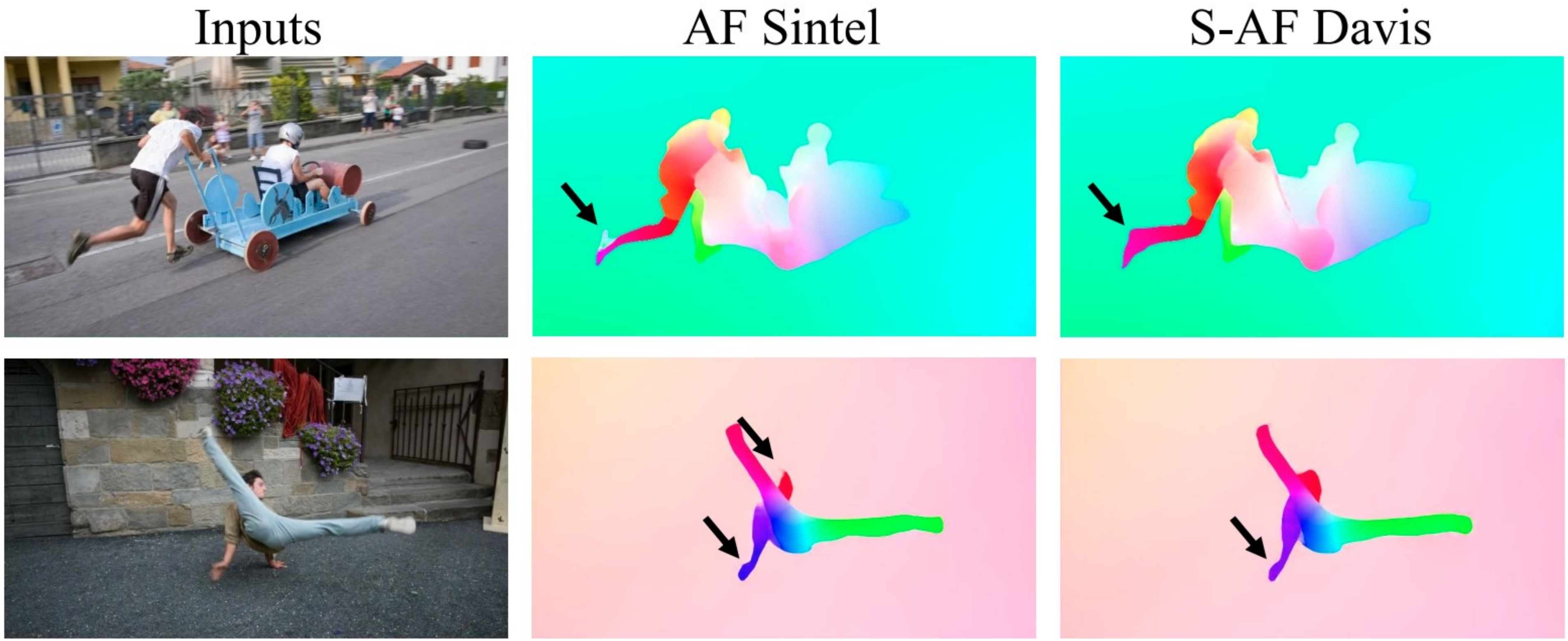


- Combination of three self-supervised signals acts as effective search metric
- Mixing data generated by top-3 hyperparameter sets increases robustness

Comparison of (self-)supervised pre-training approaches

Dataset and Method	Sintel Clean (AEPE ↓)	Sintel Final (AEPE ↓)	KITTI (AEPE ↓)
Supervised			
RAFT Chairs [43]	2.27	3.76	7.63
AF Sintel (3.2M) [40]	**1.74**	**2.41**	4.18
AF-mix Sintel (3.2M)	1.85	2.53	3.92
AF KITTI (0.8M) [41]	2.09	2.82	4.33
AF-mix KITTI (0.8M)	1.87	2.77	**3.86**
Self-supervised			
SMURF Chairs [38]	2.19	3.35	7.94
S-AF Sintel (3.2M)	**1.83**	**2.59**	5.22
S-AF KITTI (0.2M)	2.20	3.01	4.58
S-AF KITTI (0.8M)	1.99	3.00	4.29
S-AF KITTI (3.2M)	1.88	2.85	**4.22**

Results on Davis data w/o ground truth



Combining S-AF with Self-supervised Optical Flow

Pre-training on Self-Autoflow

Self-supervised fine-tuning

$$\mathcal{L} = \mathcal{L}_{\text{photo}} + \omega_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \omega_{\text{distill}} \mathcal{L}_{\text{distill}}$$

Multiframe fine-tuning

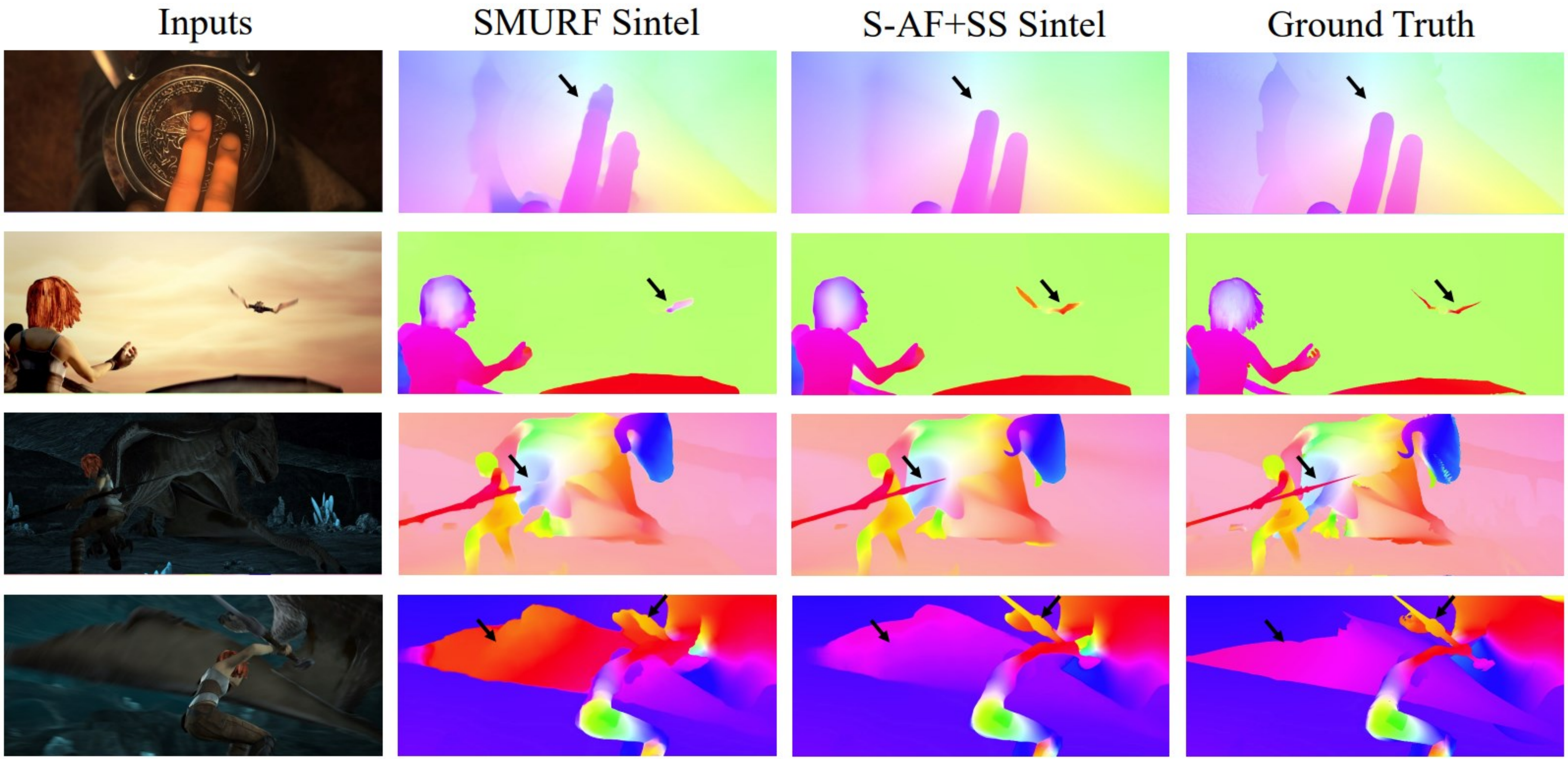
$$\mathcal{L} = \sum_{n} \gamma^{N-n} \rho_F(\mathbf{W}_{\text{pseudo}} - \mathbf{W}^n)$$

Stone et al. SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping. CVPR 2021

Comparison of self-supervised learning approaches

Method	Sintel Clean [4] AEPE ↓		Sintel Final [4] AEPE ↓		KITTI 2015 [28] AEPE ↓	AEPE (noc) ↓	Fl-all (%) ↓	
	train	*test*	*train*	*test*	*train*	*train*	*train*	*test*
EPIFlow [54]	3.94	7.00	5.08	8.51	5.56	2.56	–	16.95
UFlow [20]	3.01	5.21	4.09	6.50	2.84	1.96	9.39	11.13
SemiFlow [16]	**1.30**	–	2.46	–	3.35	–	11.12	–
SMURF test [38]	1.99	–	2.80	–	2.01	1.42	6.72	–
S-AF+SS test	1.65	–	**2.40**	–	**1.94**	**1.37**	**6.56**	–
DDFlow [24]	{2.92}	6.18	{3.98}	7.40	[5.72]	[2.73]	–	14.29
SelFlow [25] (MF)	[2.88]	[6.56]	{3.87}	{6.57}	[4.84]	[2.40]	–	14.19
UnsupSimFlow [15]	{2.86}	5.92	{3.57}	6.92	[5.19]	–	–	[13.38]
ARFlow [23] (MF)	{2.73}	{4.49}	{3.69}	{5.67}	[2.85]	–	–	[11.79]
RealFlow [9]	{1.34}	–	{2.38}	–	{2.16}	–	–	–
SMURF train [38]	{1.71}	3.15	{2.58}	4.18	{2.00}	{1.41}	{6.42}	6.83
S-AF+SS train	{1.51}	**3.02**	{2.30}	**3.97**	{1.96}	{1.38}	{6.26}	**6.76**

Visual results



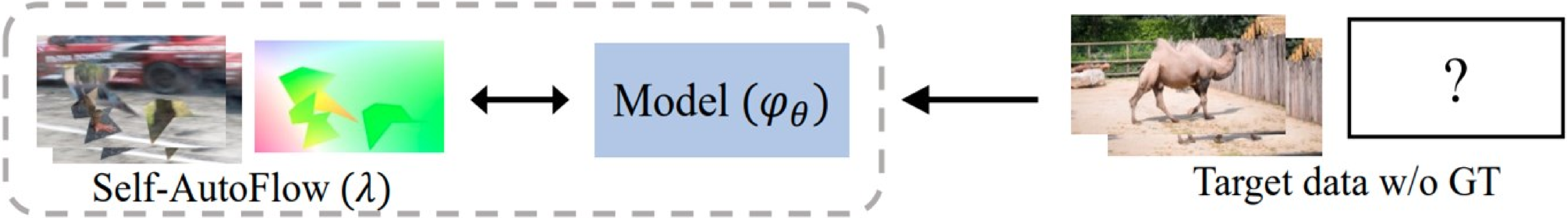
Visual results

Supervised fine-tuning on public benchmarks

Method	Sintel Clean	Sintel Final	KITTI
RealFlow [9]	-	-	4.63 %
SemiFlow (RAFT)* [16]	1.65	2.79	4.85 %
RAFT-it [40]	1.55	2.90	4.31 %
RAFT-S-AF	**1.42**	**2.75**	**4.12 %**

Sun et al. Disentangling Architecture and Training for Optical Flow. ECCV 2022

Self-supervised AutoFlow



Self-supervised AutoFlow
Self-supervised learning + learning to render