UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss

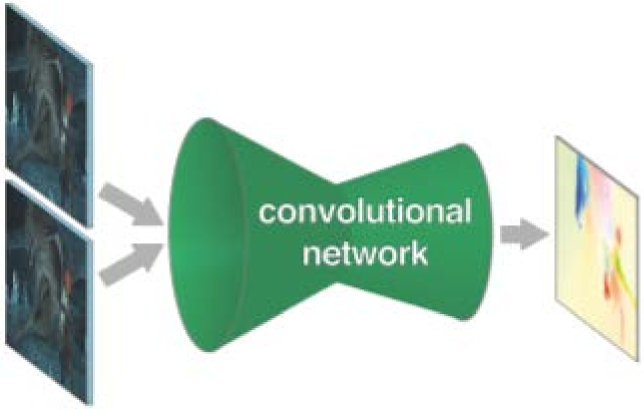
AAAI 2018, New Orleans, USA

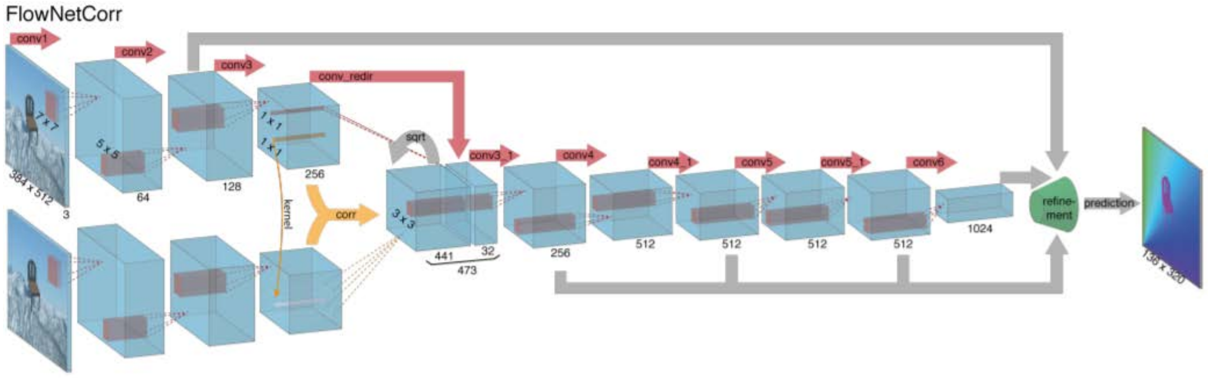
Simon Meister, Junhwa Hur, and Stefan Roth

Department of Computer Science, TU Darmstadt

Deep Networks for Optical Flow

- Deep CNNs for flow
  - e.g. FlowNet: Encoder-Decoder network
    - Given two images, outputs dense flow
  - Real-time inference with high accuracy
  - **Supervision from synthetic datasets**





Fischer et al. (2015), "FlowNet: Learning Optical Flow with Convolutional Networks"

Domain Mismatch

Training domains



Domains of interest



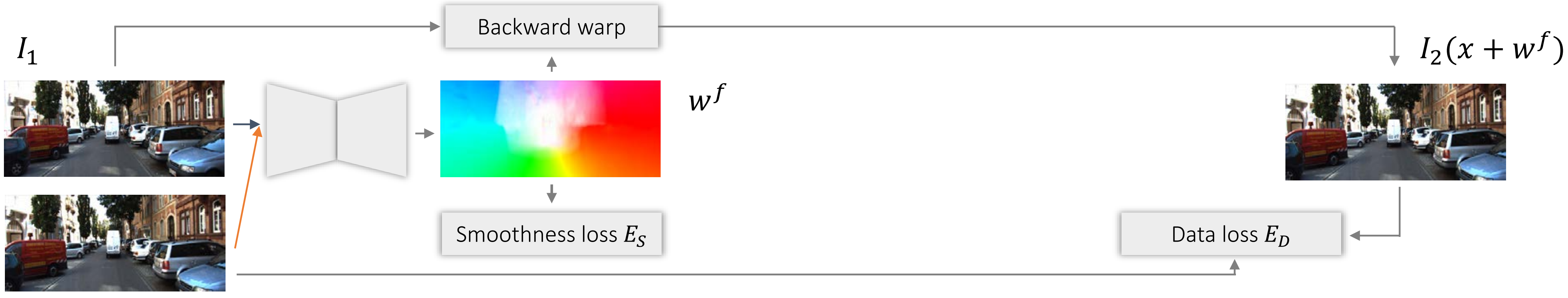
Training with Realistic Data

- **Unsupervised deep learning for optical flow**
  - Train on the target domain
  - No ground truth flow
  - Unlabeled frame pairs (e.g. from video)
  - Design **proxy loss**

...

Unsupervised Loss (Baseline)

- Use classical optical flow constraints [Yu et al.]
  - Backward-warp I_2 (using w^f) with **Bilinear sampling** [Jaderberg et al.] of I_2 at $w^f(x)$
  - Data loss E_D: brightness difference of I_1 and backward-warped I_2
  - First order smoothness loss E_S: difference of neighboring flows

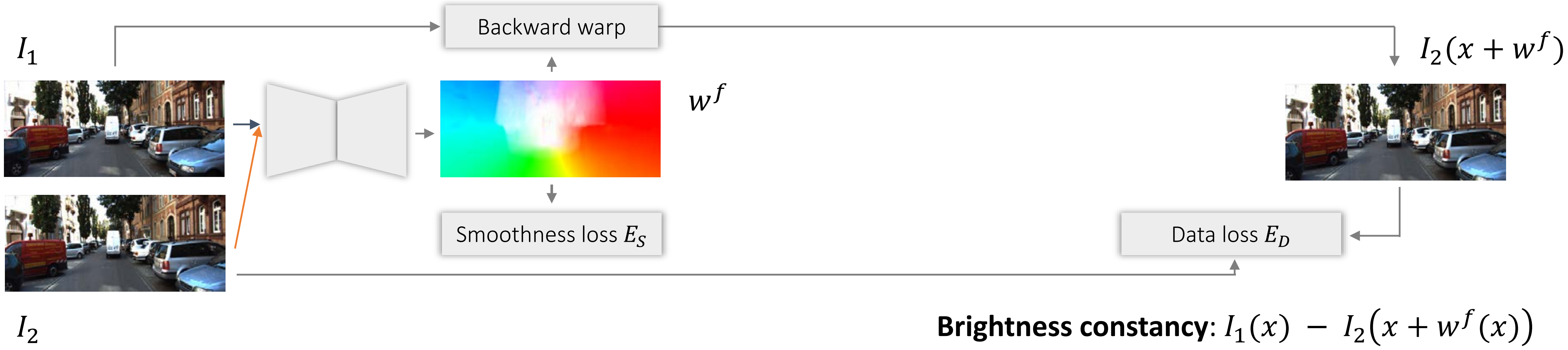


Brightness constancy: $I_1(x) - I_2\left(x + w^f(x)\right)$

Jaderberg et al. (2015), "Spatial Transformer Networks" Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

Unsupervised Loss (Baseline)

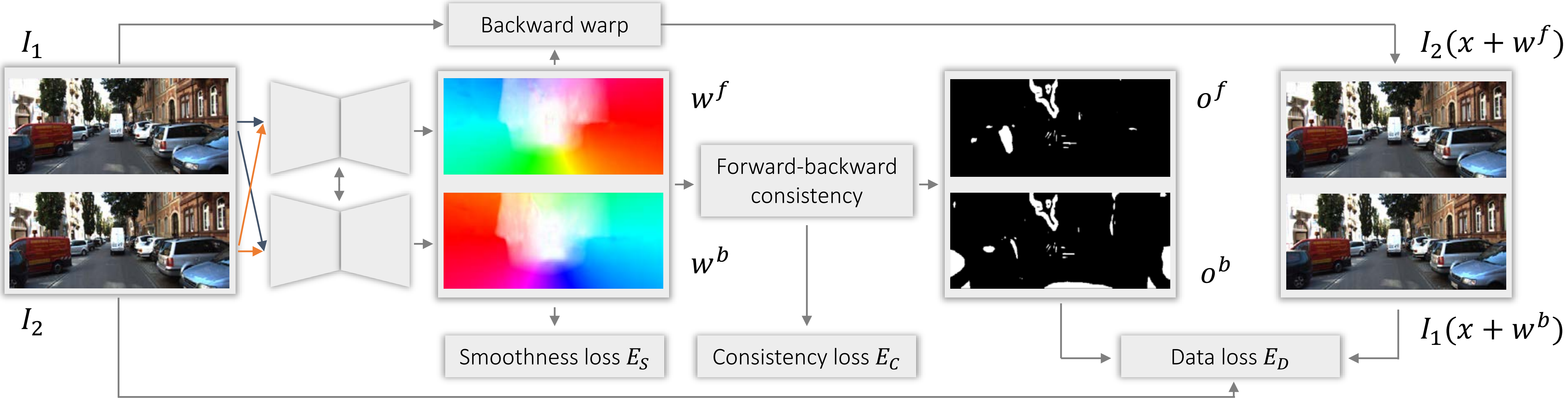
- Issues with this most basic loss
  - Lighting changes → brightness constancy violated
  - Occlusions → can't compare $I_1(x)$ and $I_2(x + w^f)$
  - First-order smoothness may be limiting



Brightness constancy: $I_1(x) - I_2(x + w^f(x))$

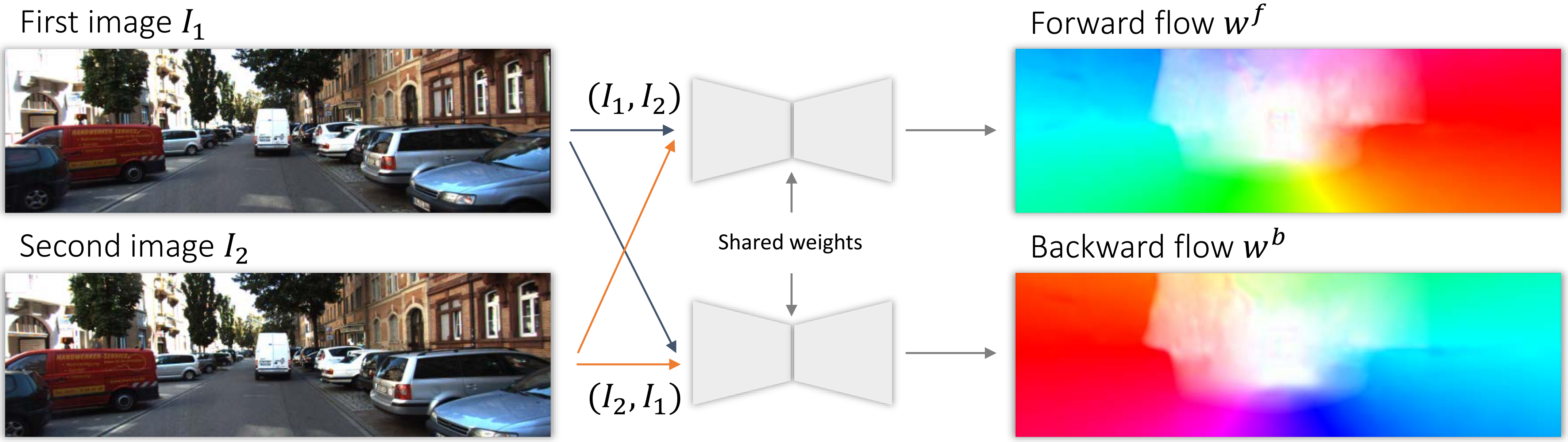
UnFlow

- Apply advanced ideas from classical optical flow to deep learning
  - Robustness to lighting changes (census transform)
  - Occlusion handling (bidirectional flow)
  - Second-order smoothness



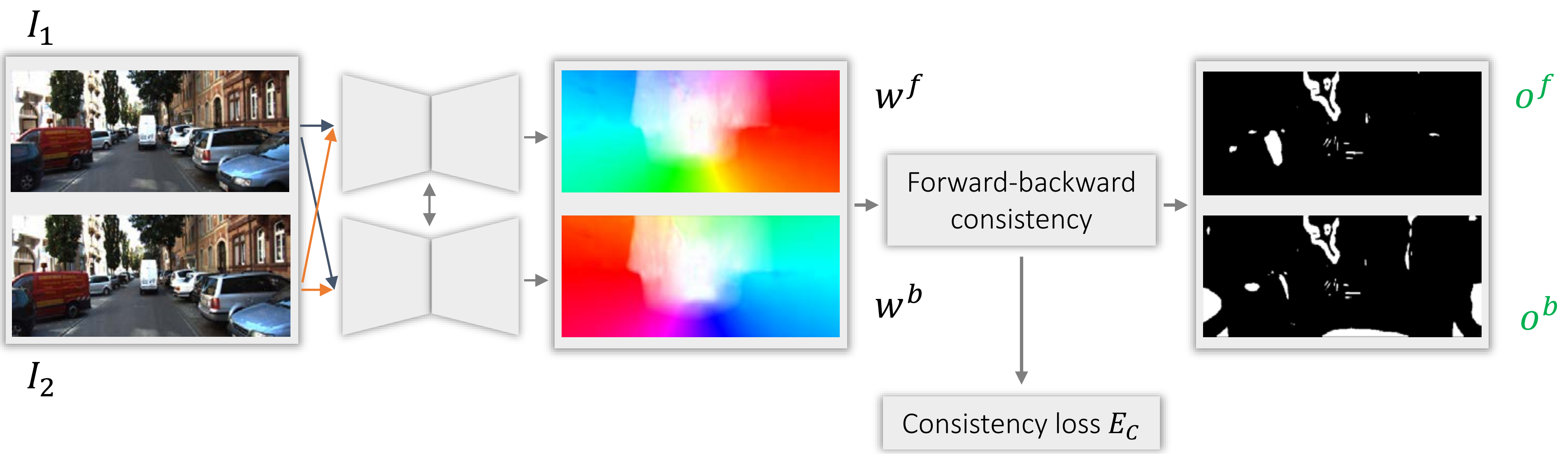
UnFlow

- Compute **bidirectional** flow (w^f, w^b) with CNN
  - FlowNetC (or any other optical flow network)



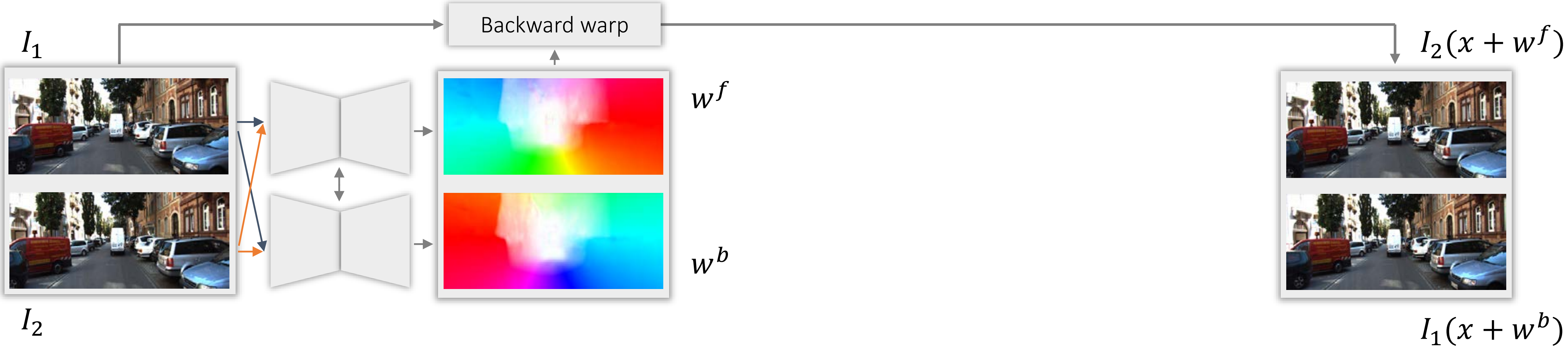
UnFlow

- Given bidirectional flow
  - Forward-backward check [Sundaram et al.]: compare $w^f(x)$ and $w^b\left(x + w^f(x)\right)$
    - Should be inverse to each other for non-occluded x
  - Threshold → Occlusion flag o^f (swap f/b for o^b)
  - Below? Consistency loss E_C for difference



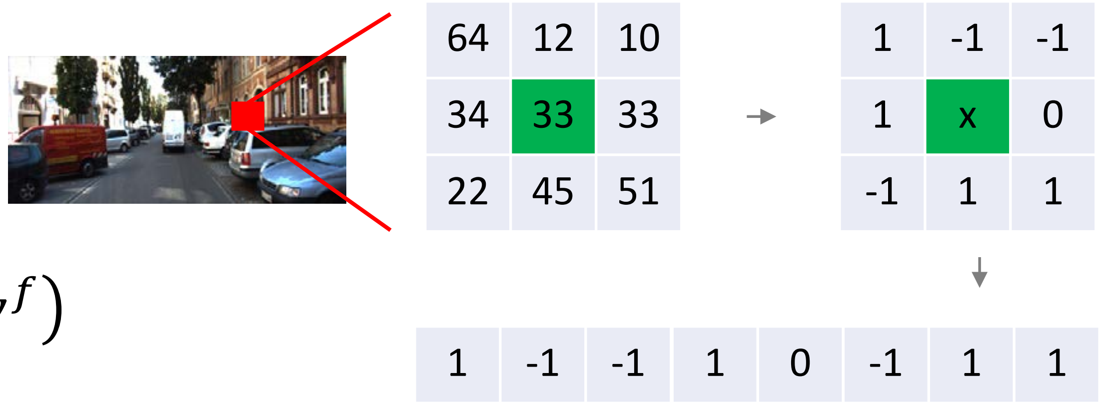
UnFlow

- **Bidirectional** image-based loss
  - Compare I_1 and backward-warped I_2 (using w^f)
    - Bilinear sampling [Jaderberg et al.] at $w^f(x)$
  - Compare I_2 and backward-warped I_1 (using w^b)

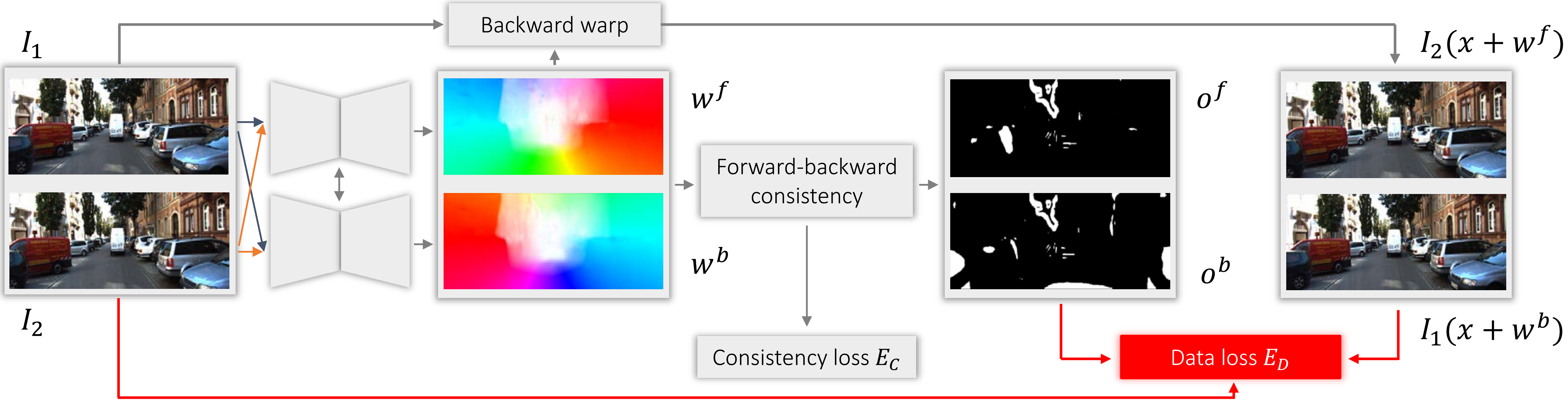


Jaderberg et al. (2015), "Spatial Transformer Networks"

UnFlow



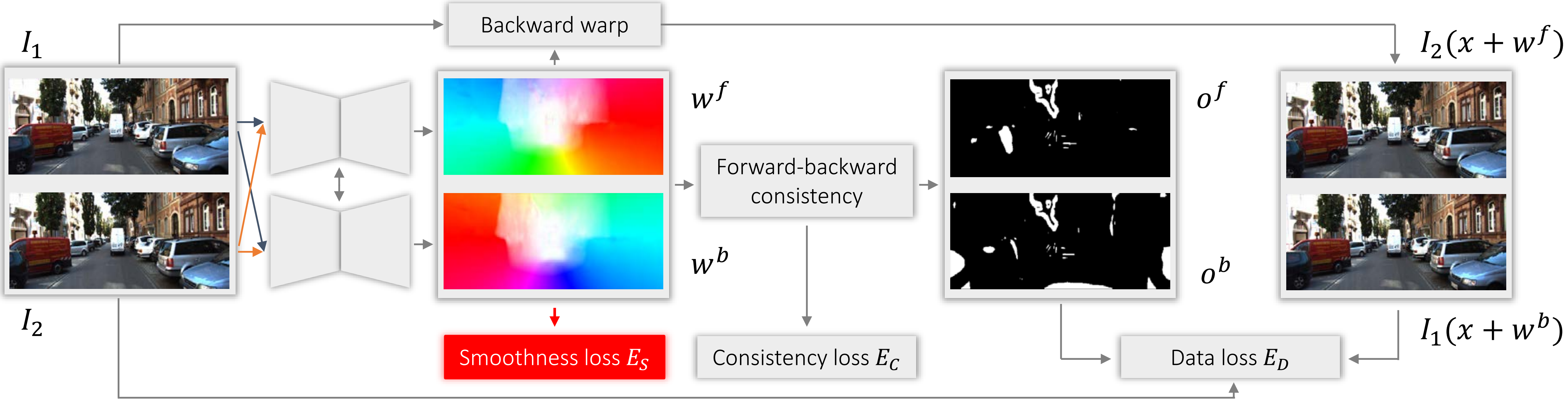
- Data loss E_D
  - Census transform [Stein] of I_1 and $I_2(x + w^f)$
    - Invariant to many changes due to lighting
  - Only at non-occluded pixels ($o^f = 0$)
  - Same for I_2 and $I_1(x + w^b)$



Stein (2004), "Efficient computation of optical flow using the census transform"

UnFlow

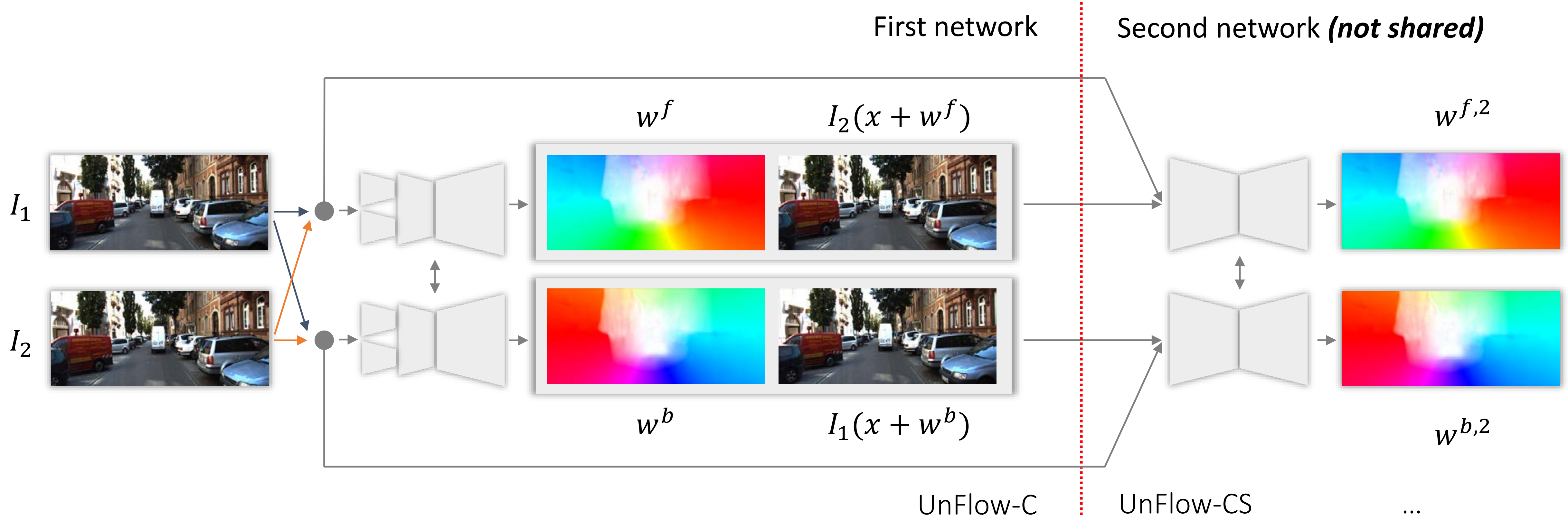
- Smoothness loss E_S
  - Second-order regularizer [Trobin et al.]
    - Penalizes large second derivatives of the flow w^f (or w^b)
  - Encourages collinear neighbors
  - The only loss at occluded pixels ($o^f = 1$)



Trobin et al. (2008), „An unbiased second-order prior for high-accuracy motion estimation"

Iterative refinement

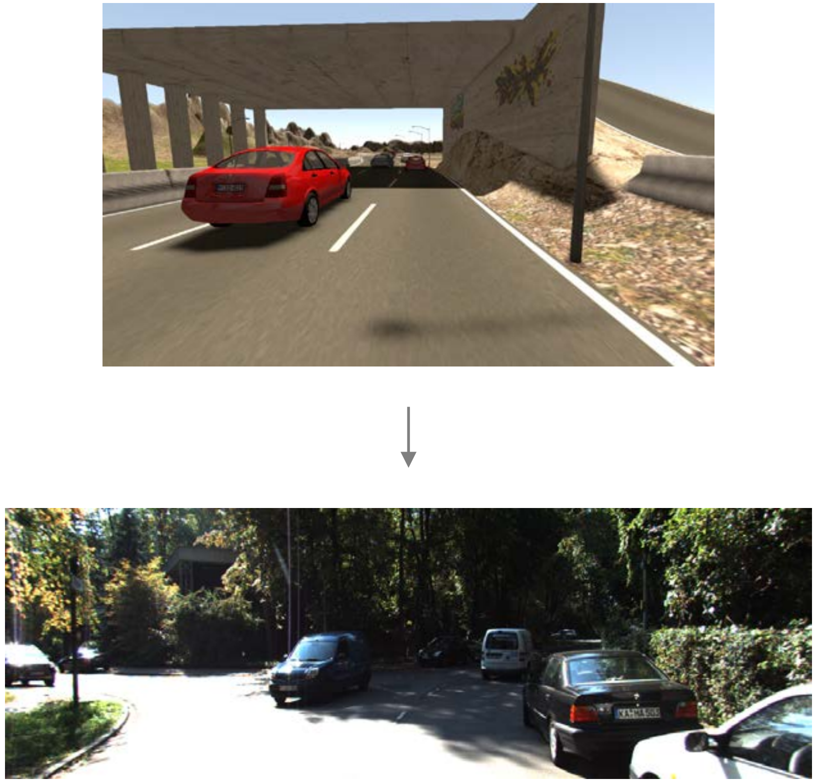
- Network stacking [Ilg et al.]
- FlowNet**C** → FlowNet**S** → …
- Input first flow estimates and warped images



Ilg et al. (2017), "FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks"

Training Schedule

- Curriculum (100% unsupervised)
  1. SYNTHIA *pre-training*
     - Large synthetic dataset with simple lighting
  2. **KITTI raw**
     - Large real-world driving dataset
     - Excluding small number of frames with ground truth flow



- No need for *specifically generated* synthetic optical flow datasets
  - FlyingChairs, FlyingThings3D, …

Results

Metrics

- Average Endpoint Error (AEE)
  - Average euclidean distance of prediction to ground truth flow vectors
- KITTI Outliers
  - Ratio of pixels where flow estimate is wrong by both 3 pixels and 5% (at least)

Loss Ablation – KITTI 2012

- Comparing Baseline [Yu et al.] vs. UnFlow-C
  - Brightness constancy → census loss
    - Reduces AEE by **35%**

Data loss	Smoothness	Occlusion	AEE (All)	Outliers (All)
Brightness	1st-order		7.20	31.93%
Census	1st-order		**4.66**	**20.85%**

Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

Loss Ablation – KITTI 2012

- Comparing Baseline [Yu et al.] vs. UnFlow-C
  - 1st → 2nd order smoothness
    - Reduces AEE by **5%** and outliers by **17%**

Data loss	**Smoothness**	Occlusion	**AEE (All)**	**Outliers (All)**
Brightness	1st-order		7.20	31.93%
Census	1st-order		4.66	20.85%
Census	**2nd-order**		**4.40**	**17.22%**

Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

Loss Ablation – KITTI 2012

- Comparing Baseline [Yu et al.] vs. UnFlow-C
  - Forward-backward mechanisms (occlusion masking & consistency)
    - Reduces AEE by **14%**

Data loss	Smoothness	Occlusion	AEE (All)	Outliers (All)
Brightness	1st-order		7.20	31.93%
Census	1st-order		4.66	20.85%
Census	2nd-order		4.40	17.22%
Census	2nd-order	**Forward-backward check**	**3.78**	**16.44%**

Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

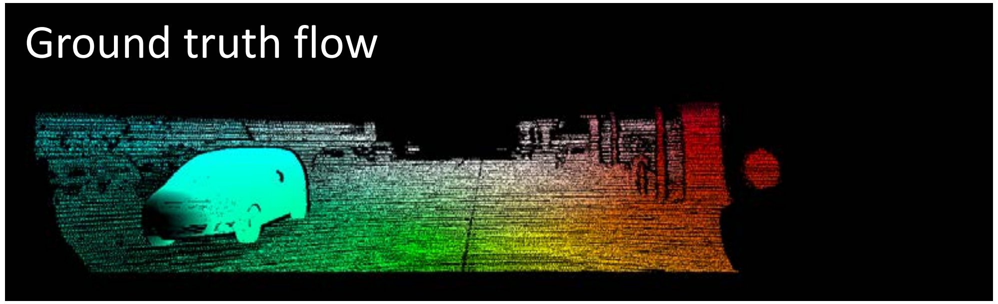
Loss Ablation – KITTI 2012

- Comparing Baseline [Yu et al.] vs. UnFlow-C
  - UnFlow reduces AEE and outliers by **48%**
  - Similar observations on KITTI 2015

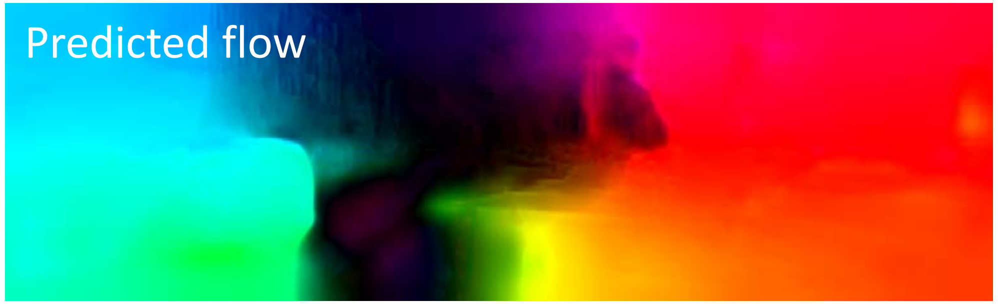
Data loss	Smoothness	Occlusion	AEE (All)	Outliers (All)
Brightness	1st-order		7.20	31.93%
Census	1st-order		4.66	20.85%
Census	2nd-order		4.40	17.22%
Census	**2nd-order**	**Forward-backward check**	**3.78**	**16.44%**

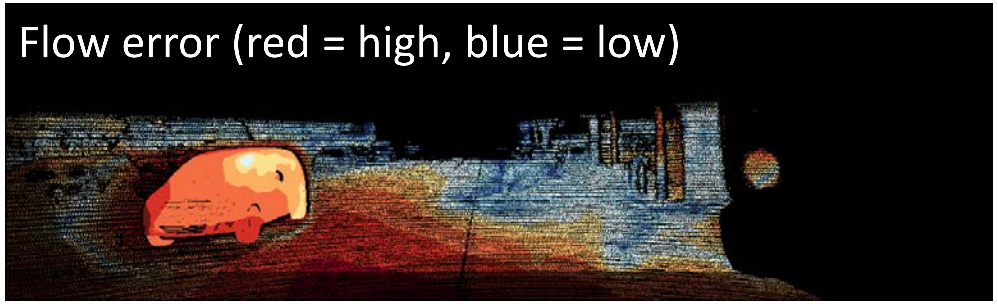
Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

Baseline vs. UnFlow (KITTI 2015)

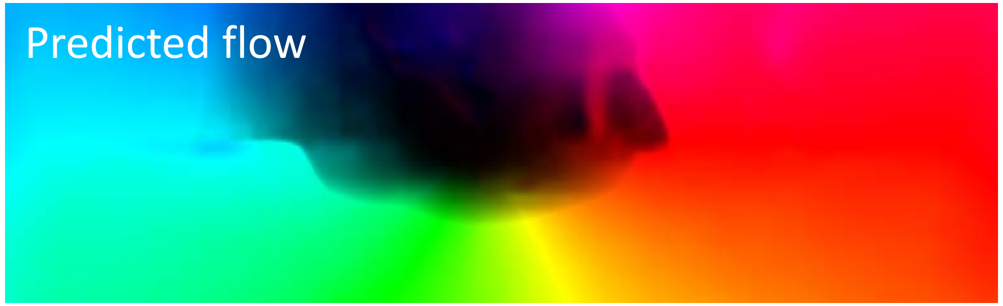


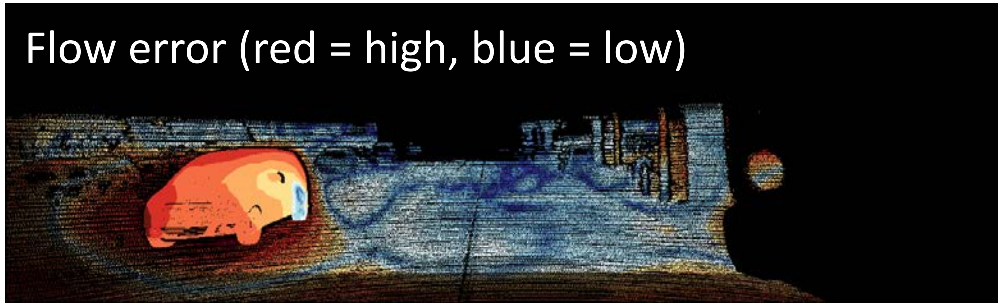
Baseline





UnFlow





Benchmarks (KITTI) – non-finetuned

- Comparing supervised networks vs. UnFlow
  - Similar networks trained on synthetic domains
  - UnFlow reduces AEE by up to **49%** (FlowNetS, 2012)

Method	**AEE (All) 2012 train**	**AEE (All) 2015 train**
FlowNetS+ft [Dosovitskiy et al.]	7.5	
FlowNet2-C [Ilg et al.]		11.36
UnFlow-C [ours]	**3.78**	**8.80**

Dosovitsky et al. (2015), "FlowNet: Learning optical flow with convolutional networks"
Ilg et al. (2017), "FlowNet 2.0: Evolution of optical flow estimation with deep networks"

Benchmarks (KITTI) – non-finetuned

- Comparing supervised networks vs. UnFlow
  - UnFlow even performs slightly better on off-domain data

Method	AEE (All) 2012 train	AEE (All) 2015 train	AEE (All) Middlebury
FlowNetS+ft [Dosovitskiy et al.]	7.5		0.98
FlowNet2-C [Ilg et al.]		11.36	
UnFlow-C [ours]	**3.78**	**8.80**	**0.88**

Dosovitsky et al. (2015), "FlowNet: Learning optical flow with convolutional networks"

Conclusion

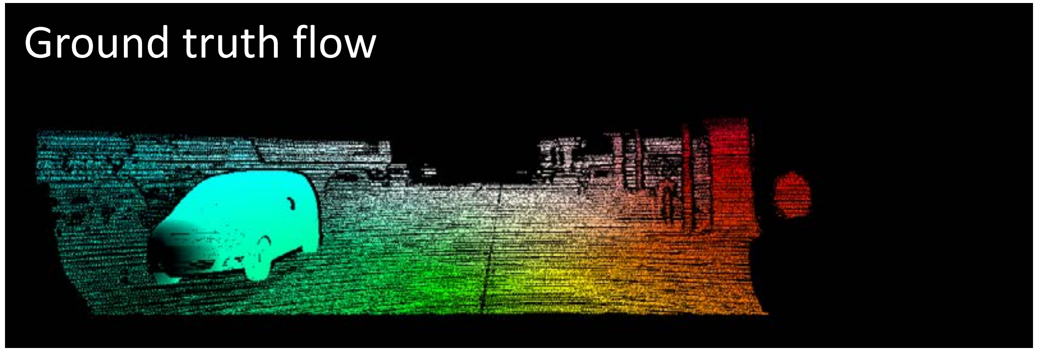
- UnFlow
  - Comprehensive unsupervised proxy loss
  - **48%** improvement over brightness constancy baseline
  - Outperforms synthetic off-domain supervision

Code open-sourced at
https://github.com/simonmeister/UnFlow

Supplementary slides

Supervised Fine-tuning

1. Unsupervised training
2. *(optional)* Supervised fine-tuning
   - KITTI 2012 & 2015 train



Method	**AEE (All) 2012 test**	**Outliers 2015 test**
FlowNet2-ft-kitti [Ilg et al.]	1.8	10.41%
UnFlow-CSS-ft	**1.7**	**11.11%**

- Competetive fine-tuning performance without pre-training with special synthetic datasets

Ilg et al. (2017), "FlowNet 2.0: Evolution of optical flow estimation with deep networks"

Benchmarks (KITTI) – non-finetuned

- Comparing previous unsupervised networks vs. UnFlow
  - Similar networks & training schedules
  - UnFlow reduces AEE by up to **66%**

Method	**AEE (All) 2012 train**
UnsupFlownet [Yu et al.]	11.3
DSTFlow [Ren et al.]	10.43
UnFlow-C [ours]	**3.78**

Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

Ren et al. (2017), "Unsupervised deep learning for optical flow estimation"

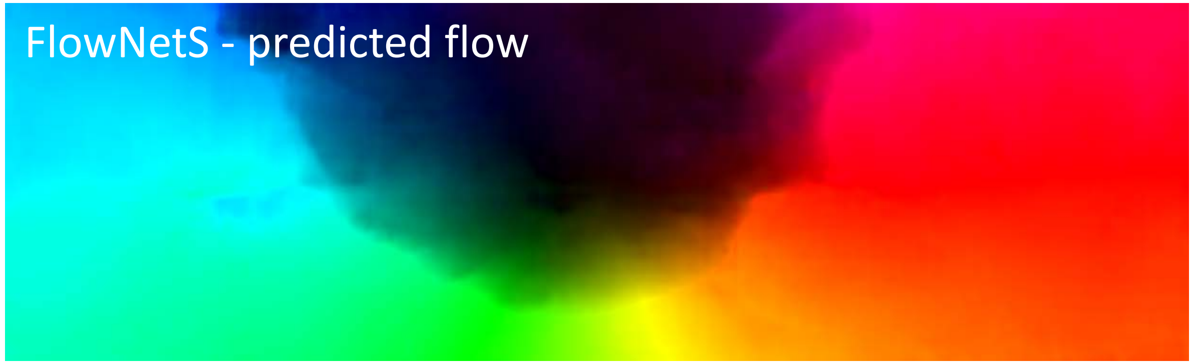
Loss Ablation – KITTI 2012

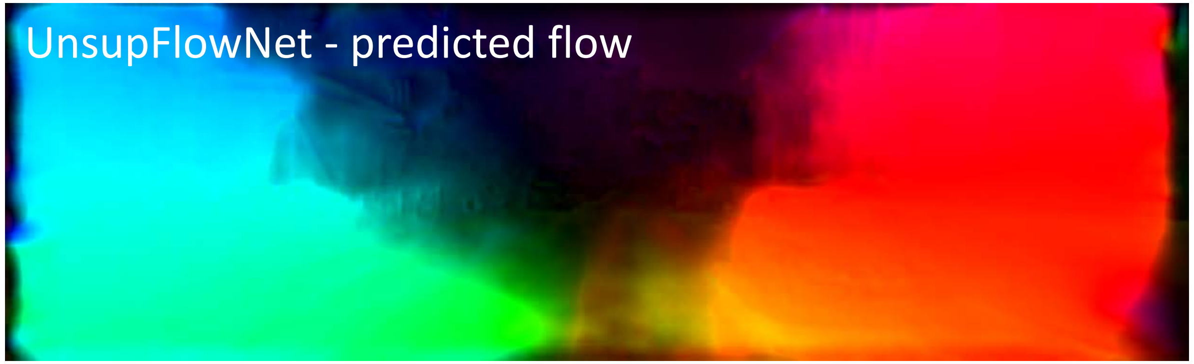
- Comparing Baseline [Yu et al.] vs. UnFlow
  - Training on SYNTHIA instead of FlyingChairs slightly improves AEE
  - Our baseline re-implementation is more accurate than the results reported by Yu et al. (AEE of **11.3** vs. our **8.26**)

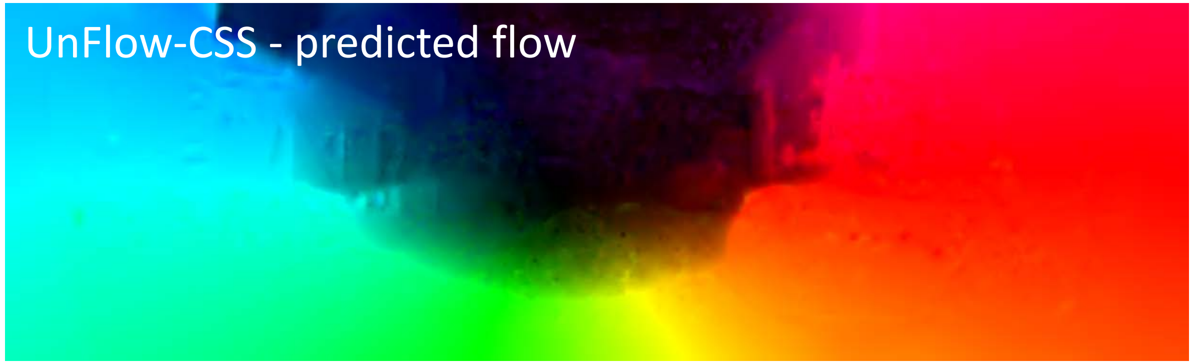
Data loss	Smoothness	Occlusion	AEE (All)	Outliers (All)
Brightness	*1st-order*		*8.26*	
Brightness	**1st-order**		**7.20**	31.93%
Census	1st-order		4.66	20.85%
Census	2nd-order		4.40	17.22%
Census	2nd-order	Forward-backward check	3.78	16.44%

Yu et al. (2016), "Unsupervised learning of optical flow via brightness constancy and motion smoothness"

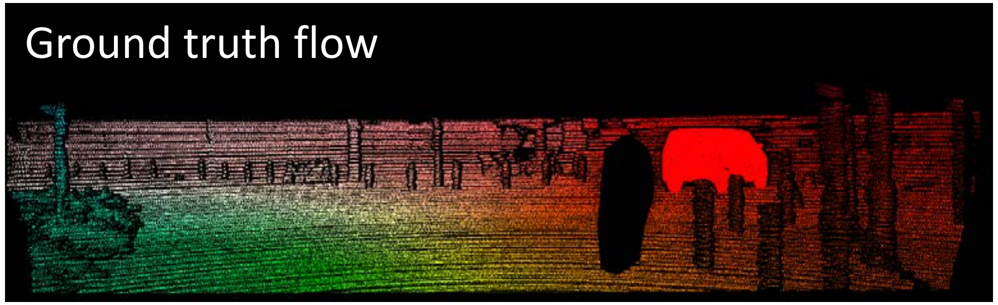
FlowNetS / UnsupFlownet / UnFlow-CSS



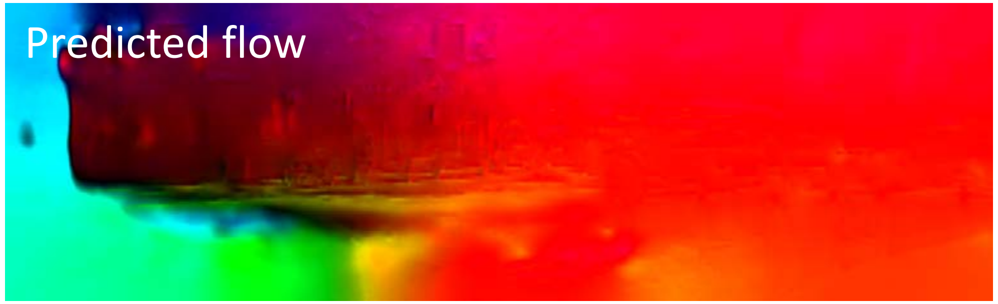


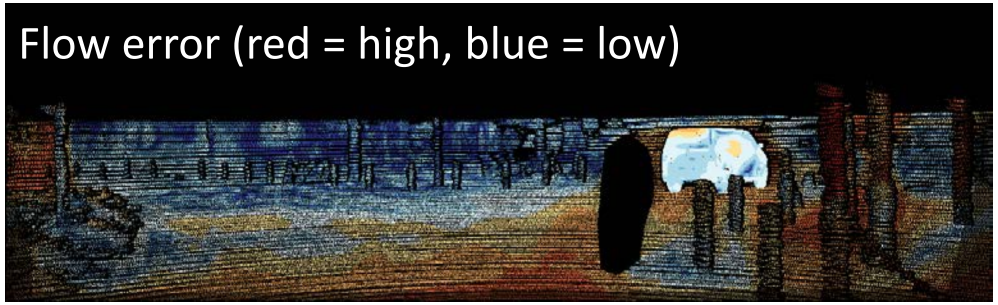


Baseline vs. UnFlow (KITTI 2015)

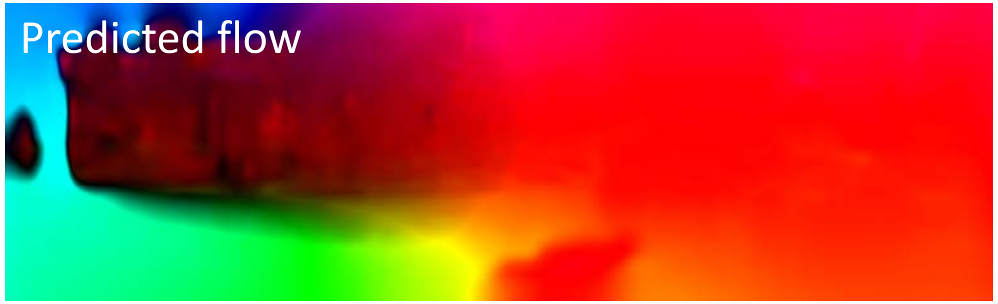


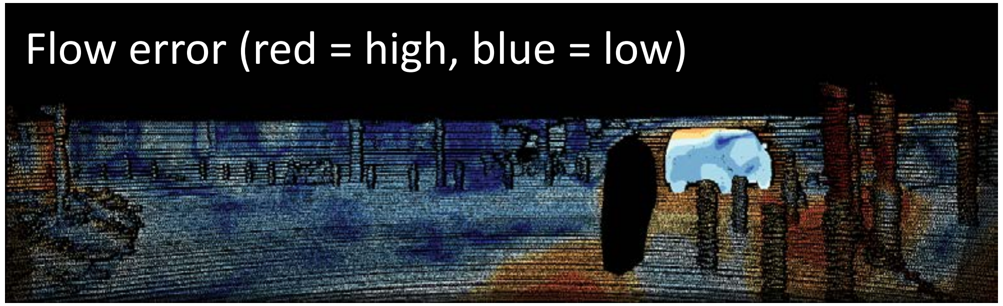
Baseline



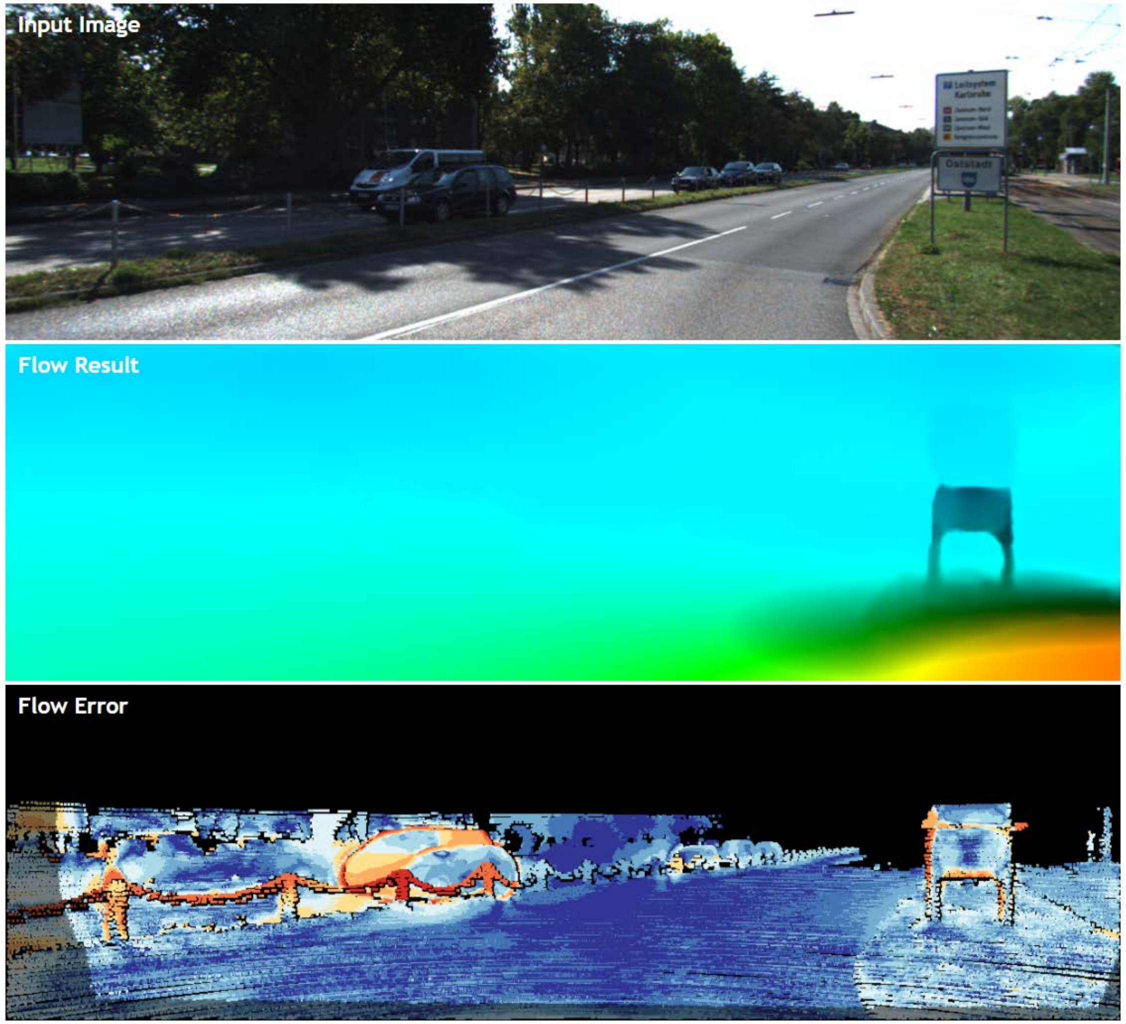


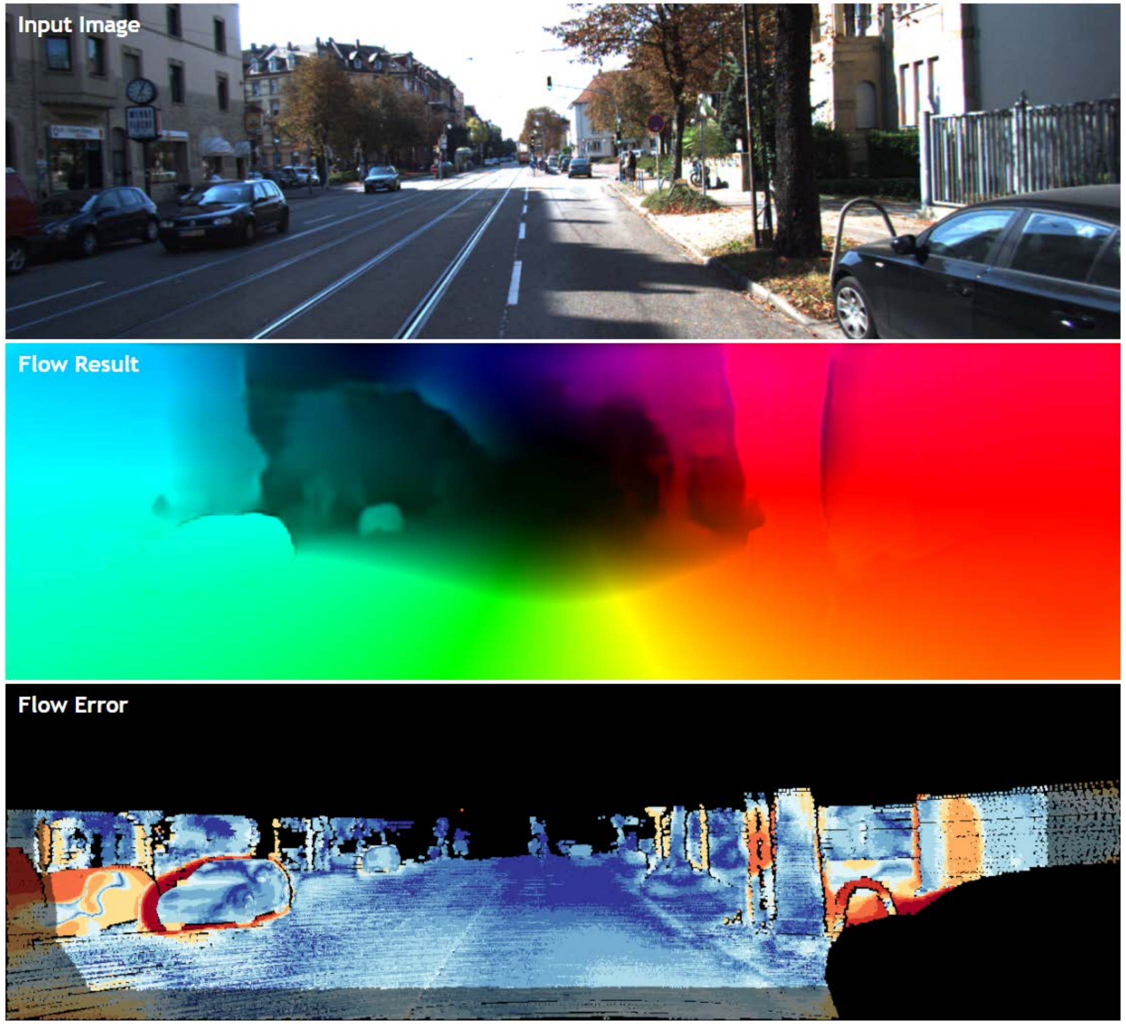
UnFlow



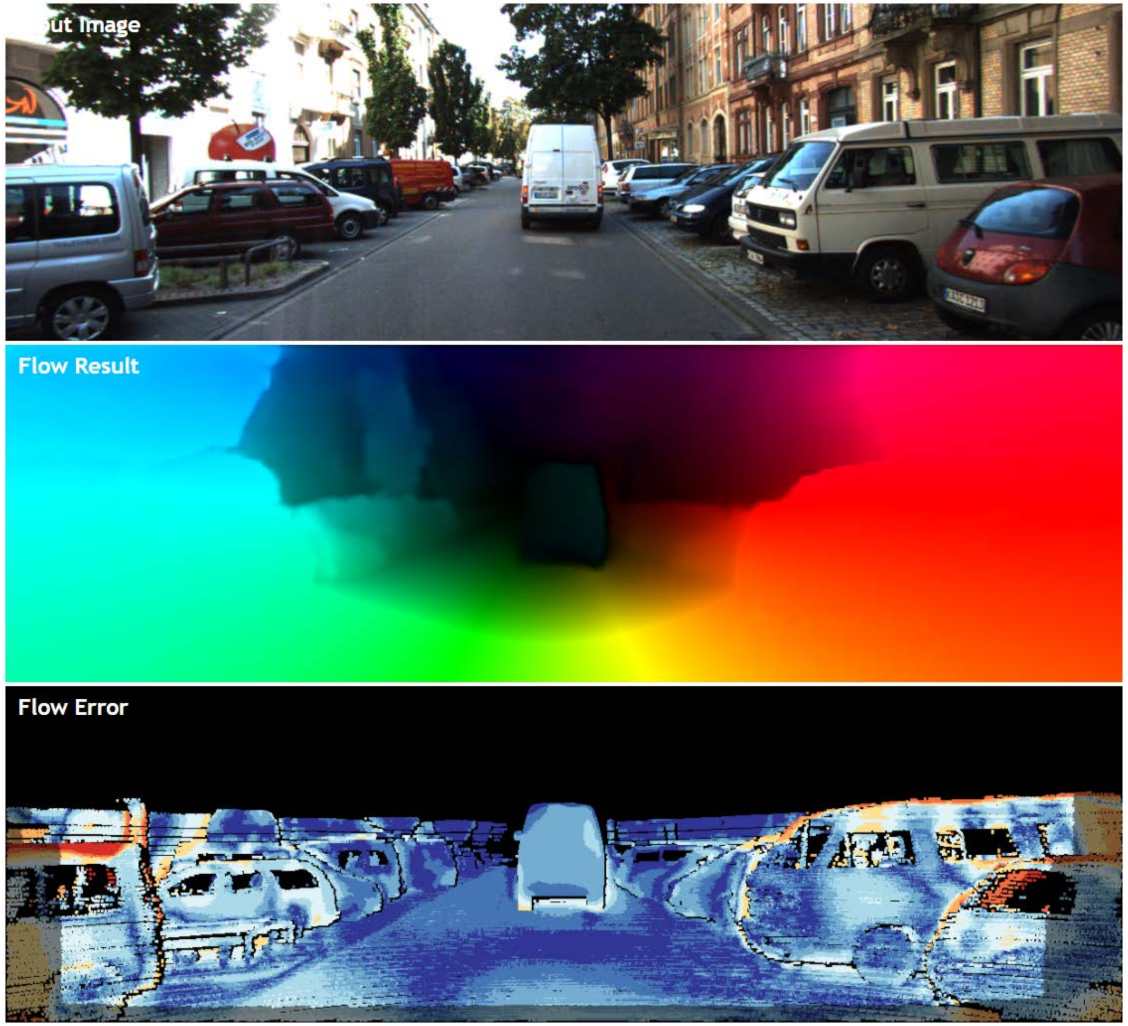


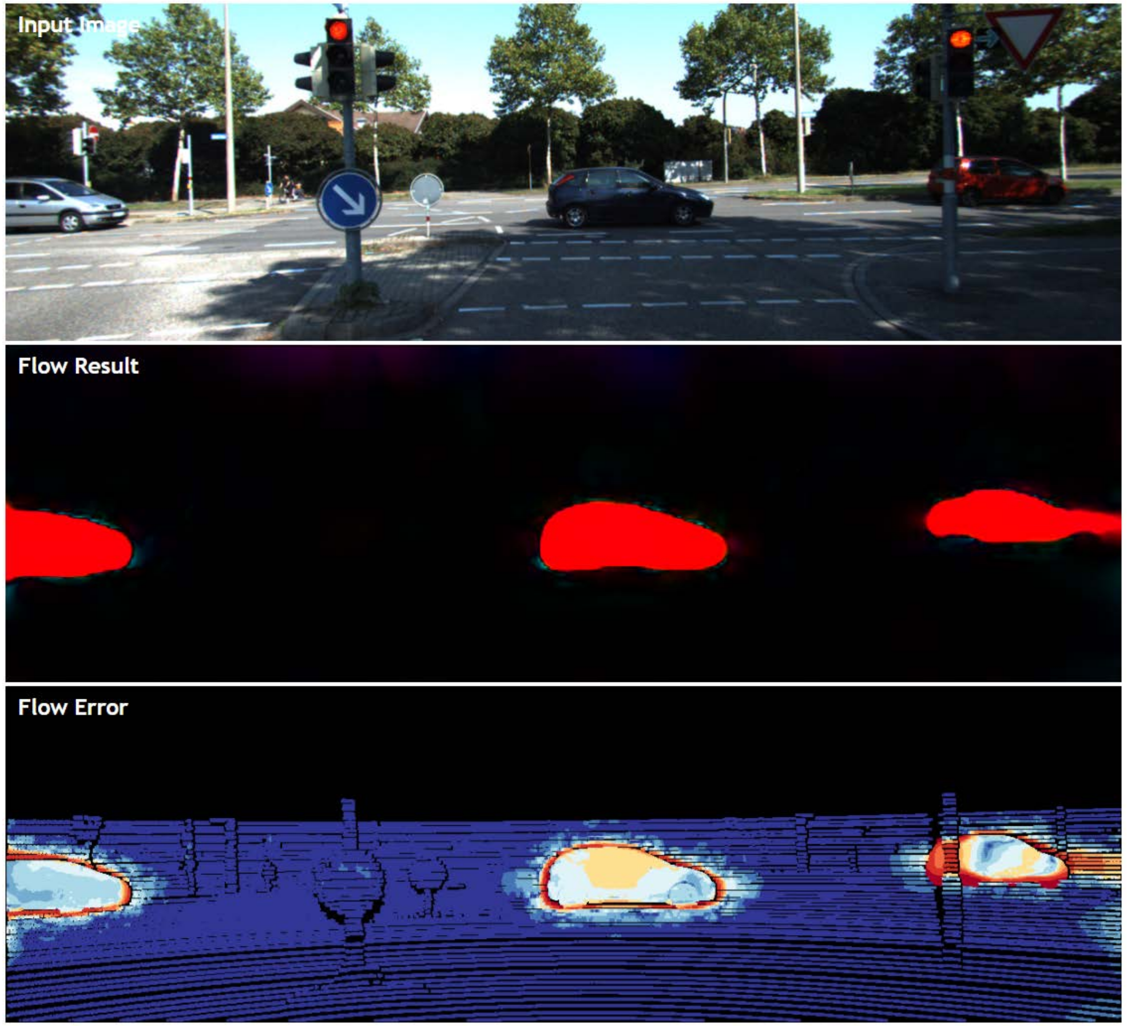
UnFlow-CSS-ft (KITTI 2015 *test*)





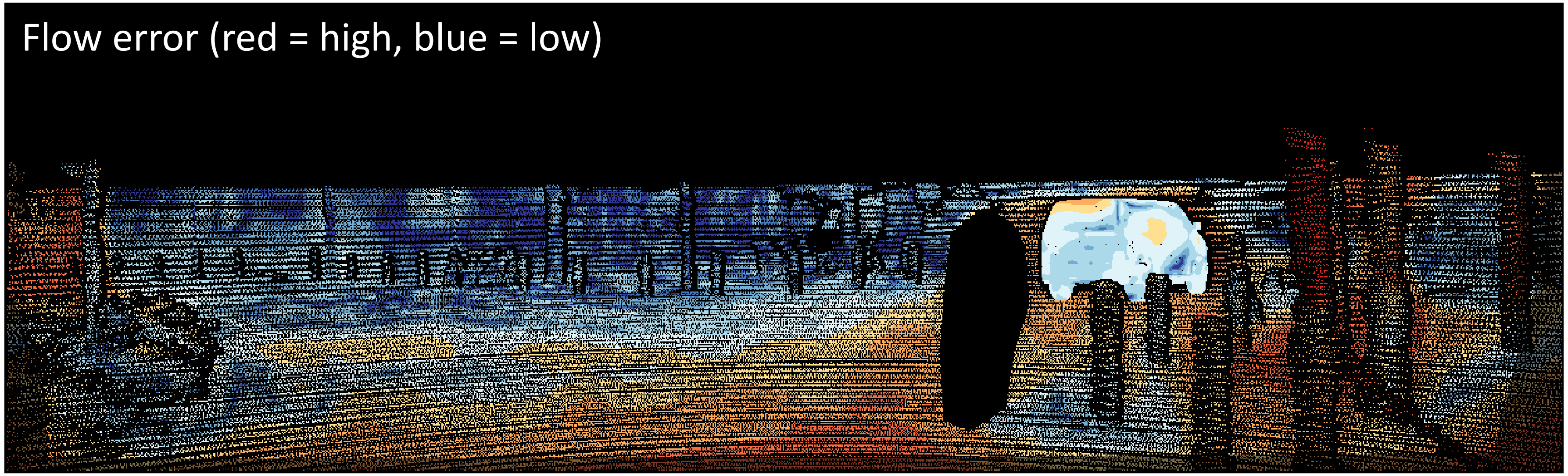
UnFlow-CSS-ft (KITTI 2015 *test*)



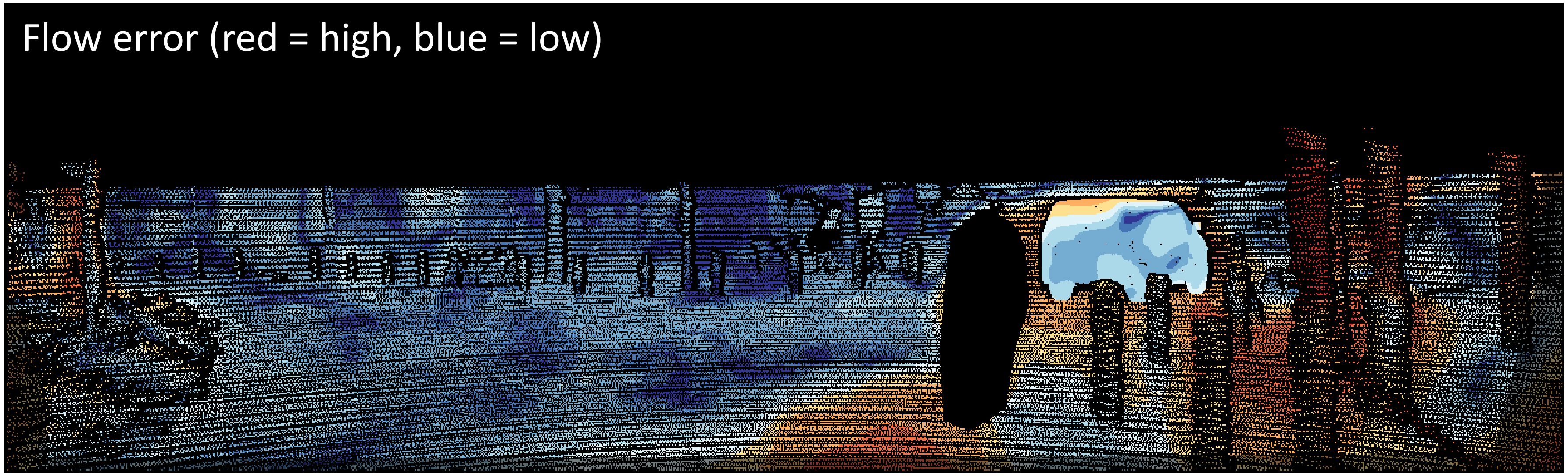


Baseline vs. UnFlow (KITTI 2015)

Baseline

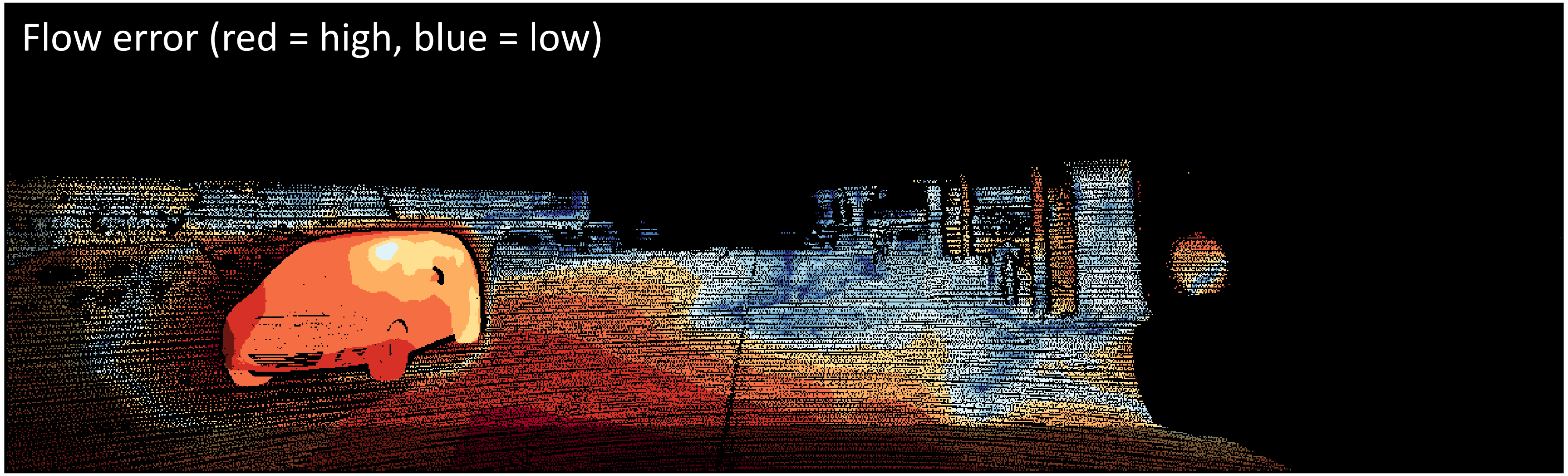


UnFlow



Baseline vs. UnFlow (KITTI 2015)

Baseline



UnFlow