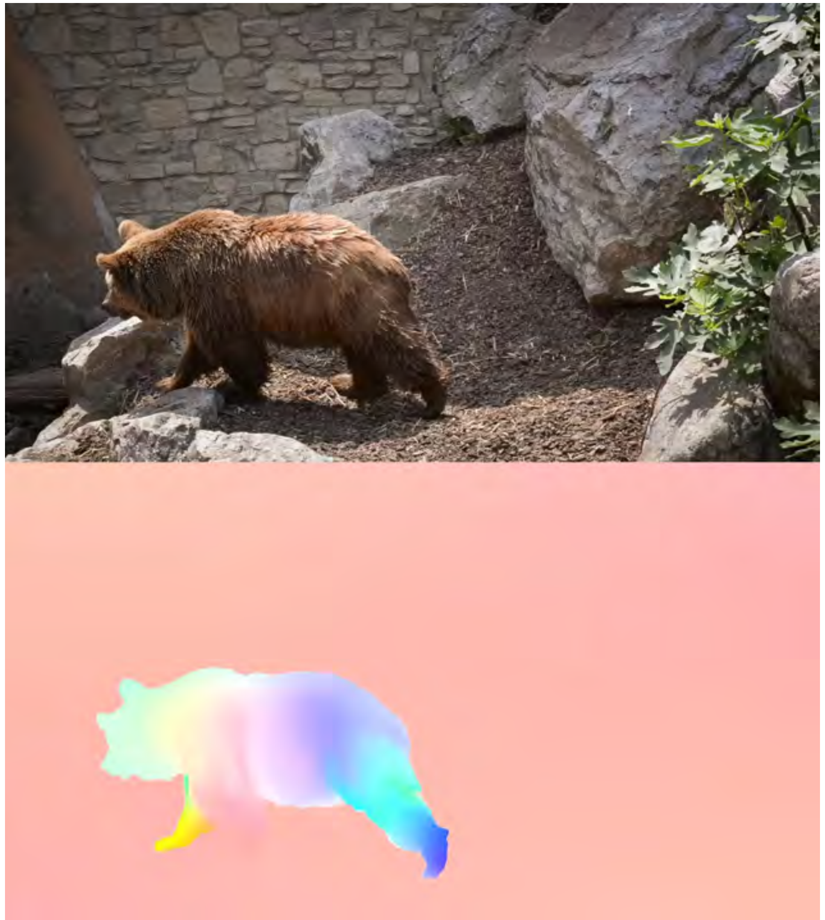
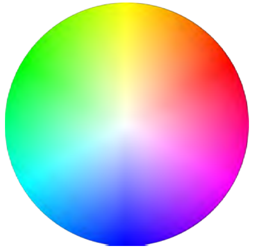
Optical Flow

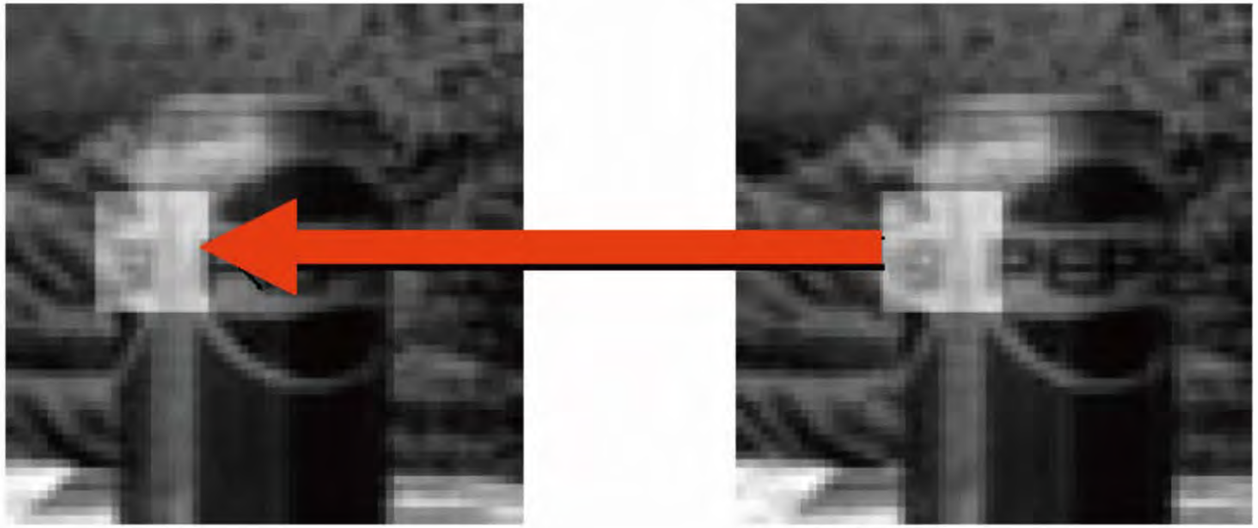
- Definition: optical flow is the *apparent* motion of *brightness patterns* in the image





Color wheel

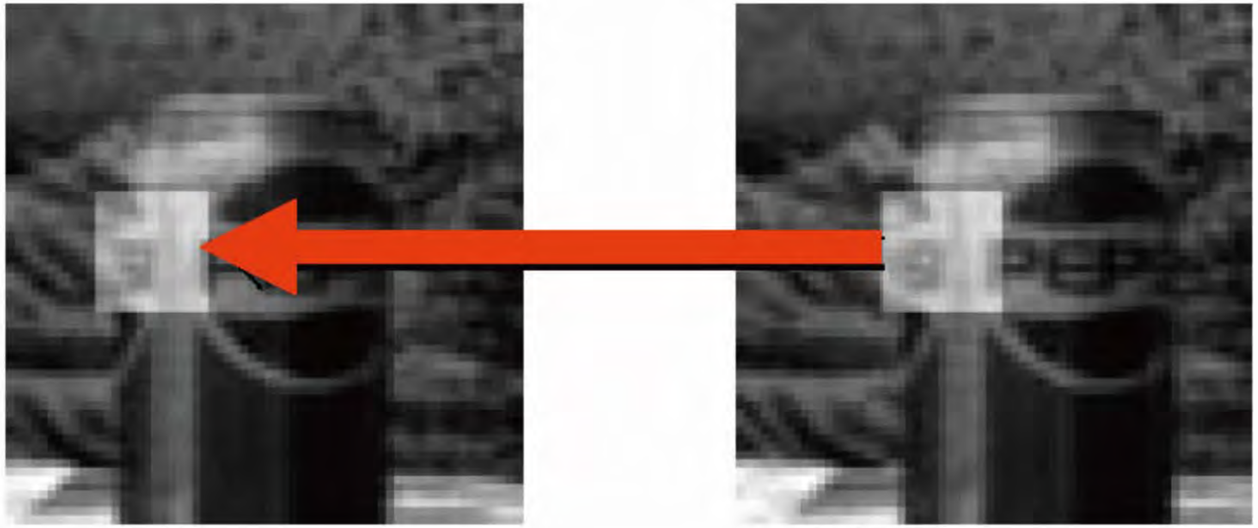
Key Assumptions: brightness Constancy



* Slide from Michael Black, CS143 2003

Credit: Juan Carlos Niebles and Ranjay Krishna @ Stanford Vision and Learning Lab

Key Assumptions: brightness Constancy



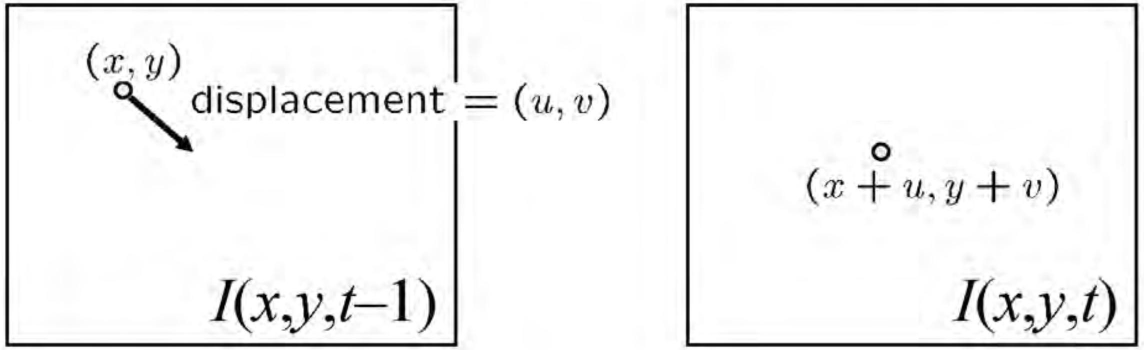
Assumption

Image measurements (e.g. brightness) in a small region remain the same although their location may change.

$$I(x+u, y+v, t+1) = I(x, y, t)$$

(assumption)

The brightness constancy constraint



- Brightness Constancy Equation:

$$I(x,y,t-1) = I(x+u(x,y), y+v(x,y), t)$$

Linearizing the right side using Taylor expansion:

Image derivative along x

$$I(x+u, y+v, t) \approx I(x,y,t-1) + \boxed{I_x} \cdot u(x,y) + I_y \cdot v(x,y) + \boxed{I_t}$$

$$I(x+u, y+v, t) - I(x,y,t-1) = I_x \cdot u(x,y) + I_y \cdot v(x,y) + I_t$$

Hence, $I_x \cdot u + I_y \cdot v + I_t \approx 0 \quad \rightarrow \nabla I \cdot [u \;\; v]^T + I_t = 0$

Source: Silvio Savarese

Filters used to find the derivatives

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \text{first image} \qquad \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \text{first image} \qquad \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \text{first image}$$

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \text{second image} \qquad \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \text{second image} \qquad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{second image}$$

$$I_x \qquad I_y \qquad I_t$$

The brightness constancy constraint

Can we use this equation to recover image motion (u,v) at each pixel?

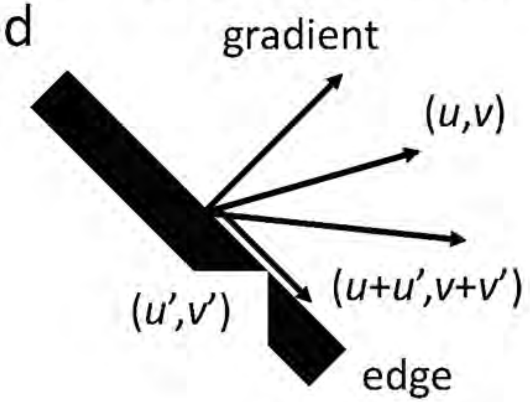
$$\nabla I \cdot [u \ v]^T + I_t = 0$$

- How many equations and unknowns per pixel?
  - One equation (this is a scalar equation!), two unknowns (u,v)

The component of the flow perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

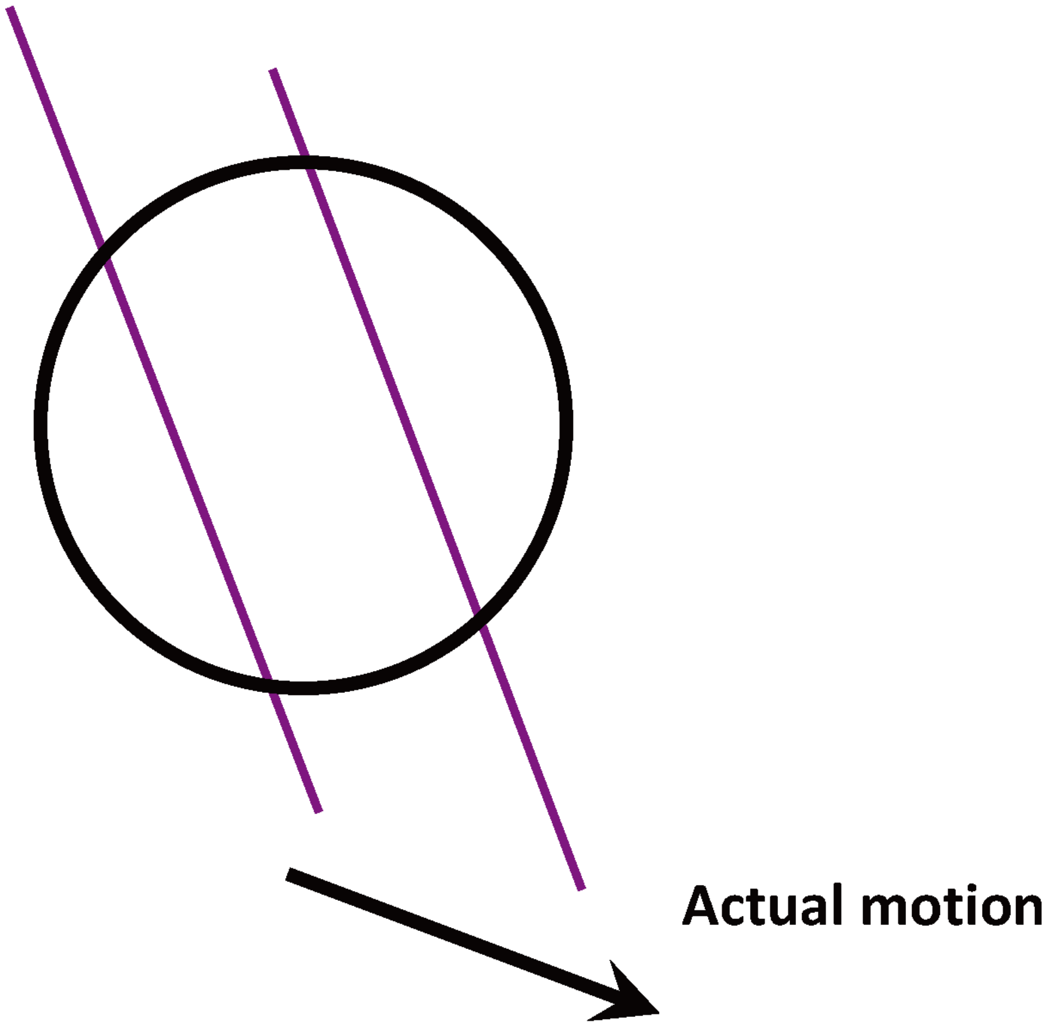
If (u, v) satisfies the equation, so does $(u+u', v+v')$ if

$$\nabla I \cdot [u' \ v']^T = 0$$



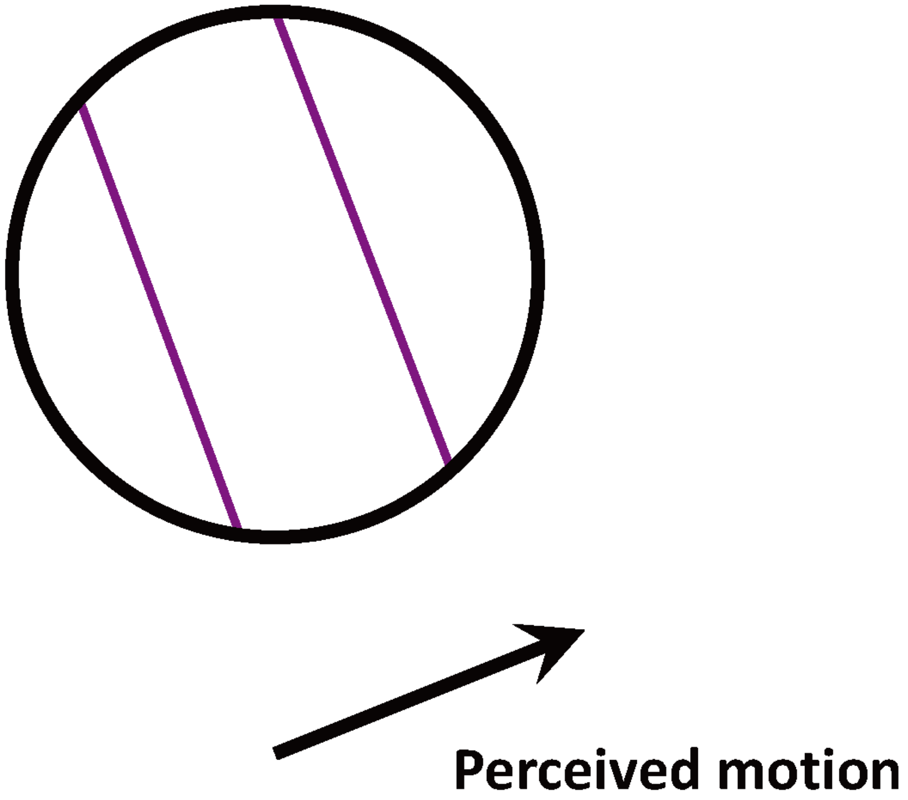
Source: Silvio Savarese

The aperture problem



Source: Silvio Savarese

The aperture problem



Source: Silvio Savarese

Credit: Juan Carlos Niebles and Ranjay Krishna @ Stanford Vision and Learning Lab

Solving the ambiguity…

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

- How to get more equations for a pixel?
- **Spatial coherence constraint:**
- Assume the pixel's neighbors have the same (u,v)
  - If we use a 5x5 window, that gives us 25 equations per pixel

$$0 = I_t(\mathbf{p_i}) + \nabla I(\mathbf{p_i}) \cdot [u \;\; v]$$

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix}$$

Source: Silvio Savarese

Horn-Schunk method for optical flow

- The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[\boxed{(I_x u + I_y v + I_t)}^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] \mathrm{d}x\mathrm{d}y$$

- The first part of the function is the brightness consistency.

Horn-Schunk method for optical flow

- The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\boxed{\|\nabla u\|^2 + \|\nabla v\|^2} \right) \right] \mathrm{d}x\mathrm{d}y$$

- The second part is the smoothness constraint. It's trying to make sure that the changes between frames are small.

Why do we need Optical Flow?

Without Optical Flow

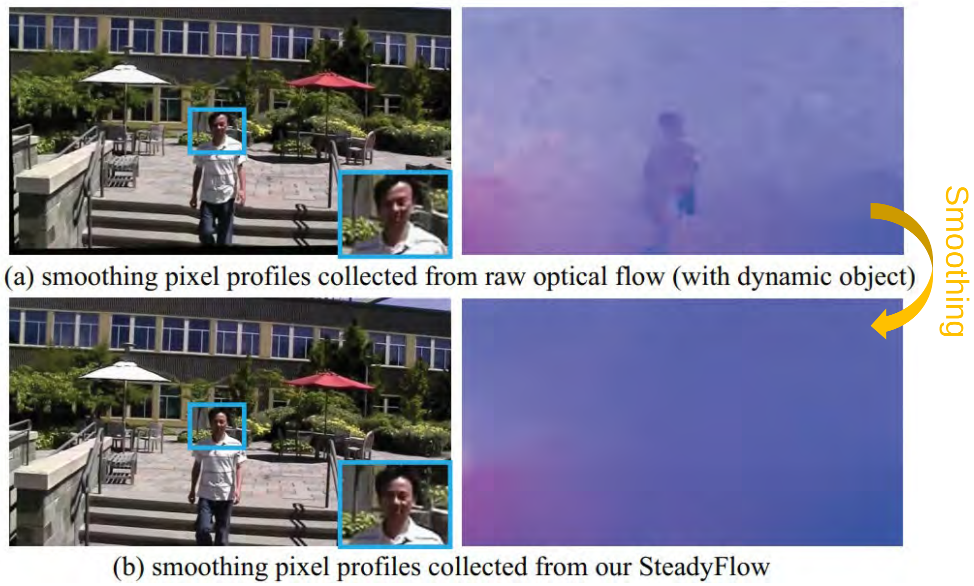
https://www.pinterest.com/pin/527836018822238481/

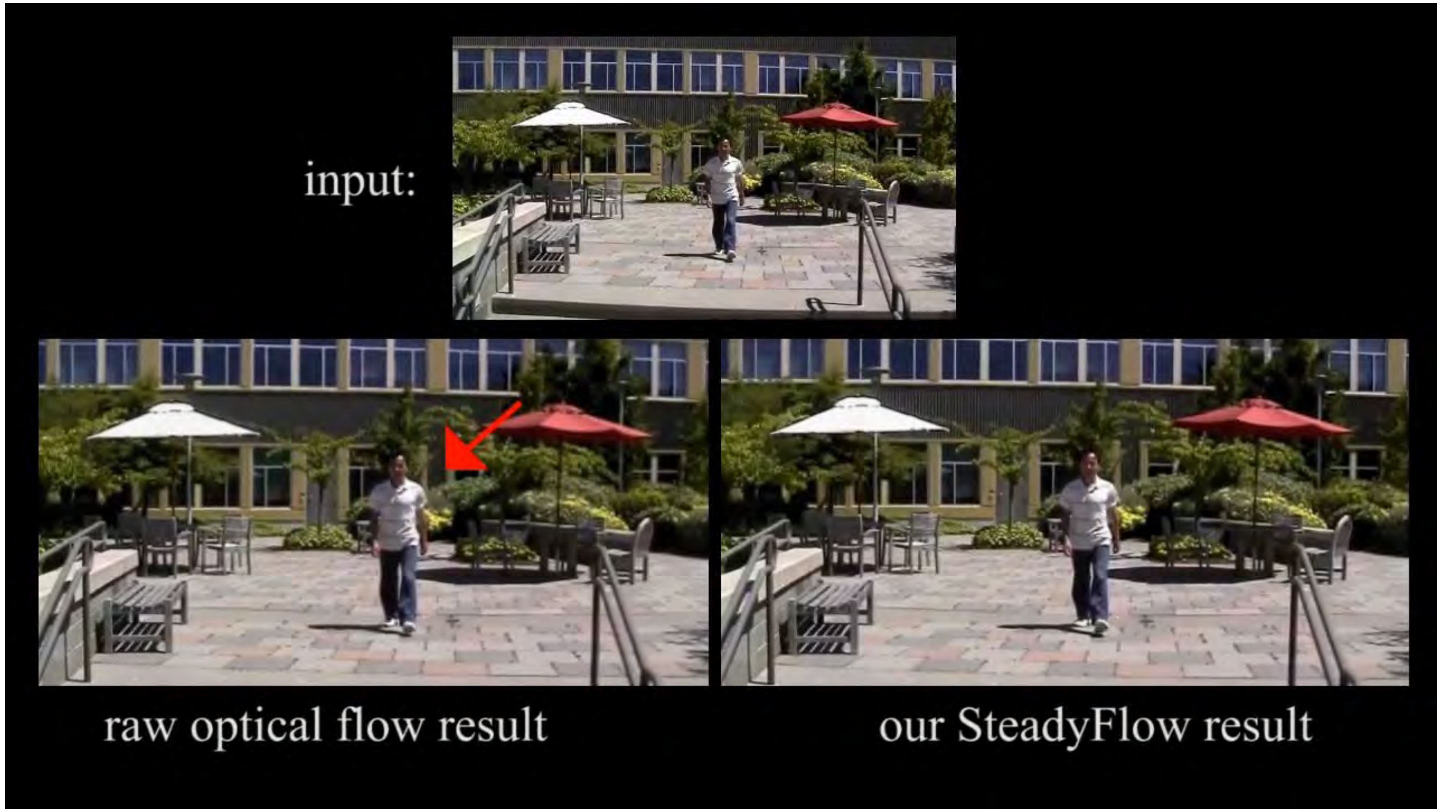
With Optical Flow

https://harrisburg.craigslist.org/sys/d/camp-hill-dell-n231-black-usb-optical/7720318568.html

Optical Flow in Computer Vision

- Video stabilization by **Spatially Smooth Optical Flow** (SteadyFlow; CVPR 2014)

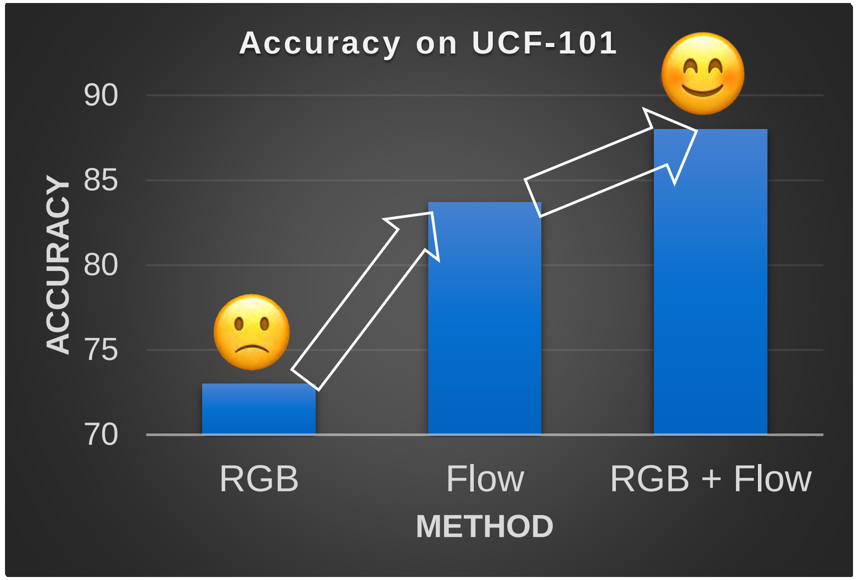


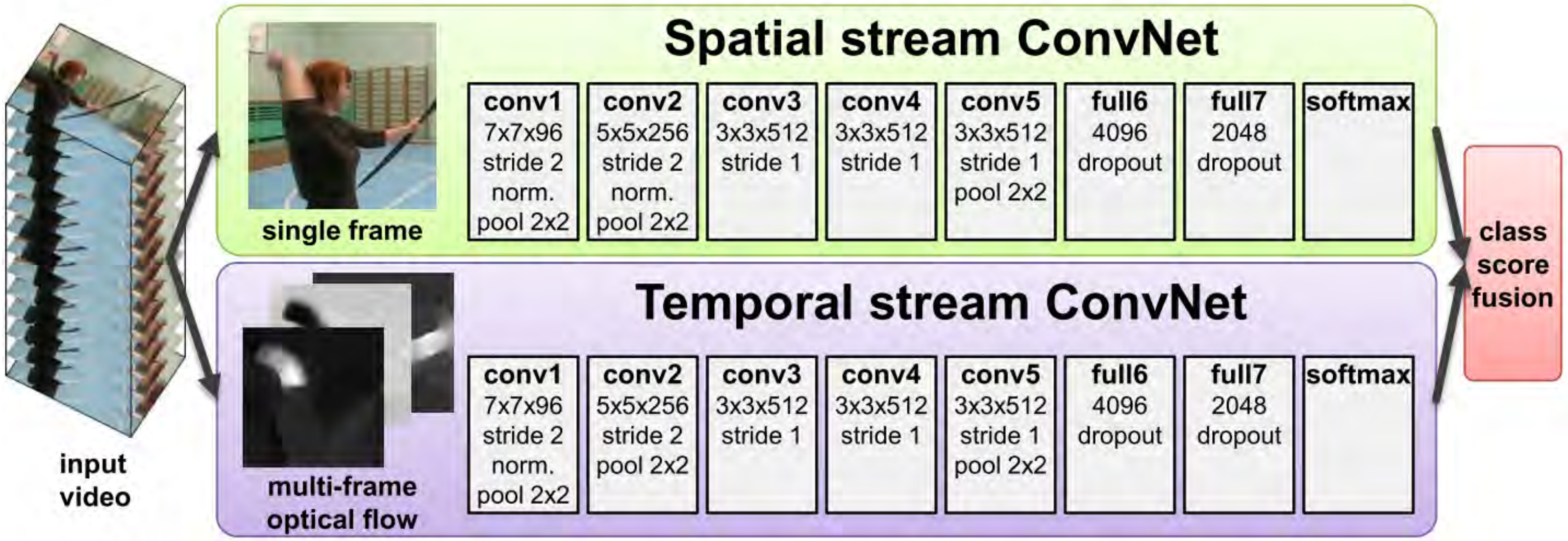


S. Liu et al., SteadyFlow, CVPR 2014

Optical Flow in Computer Vision

- Action recognition by **two-stream networks** (NIPS 2014)

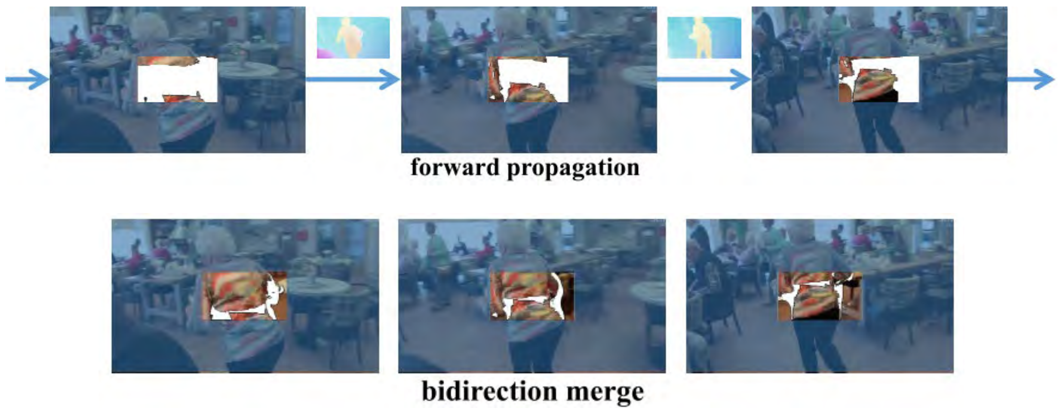




K. Symonyan et al., Two stream networks, NIPS 2014

Optical Flow in Computer Vision

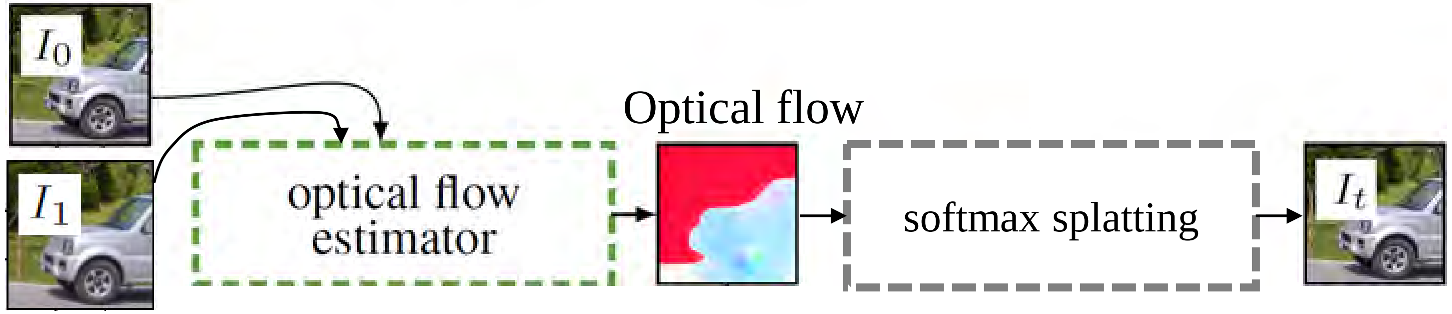
- Video inpainting by **optical flow-guided algorithm** (CVPR 2019)

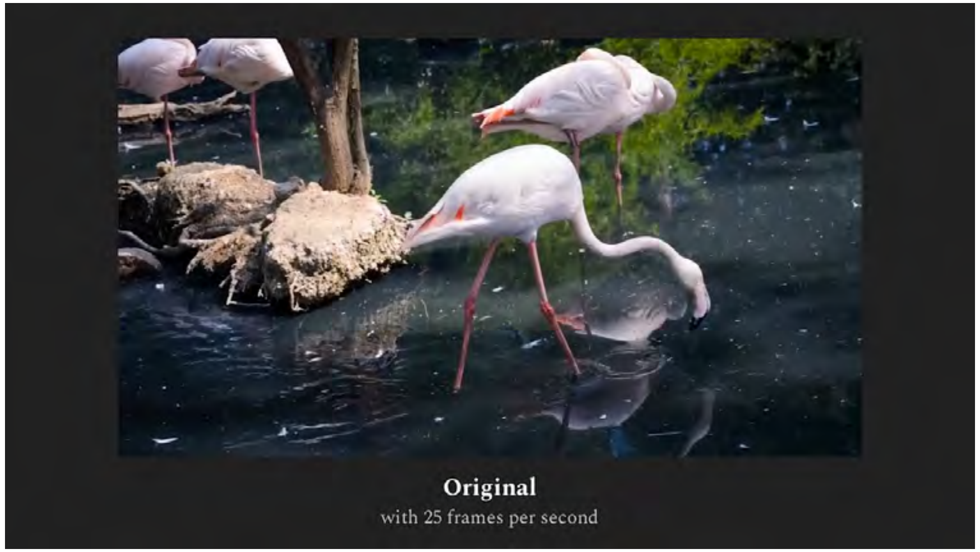


Deep Flow-Guided Video Inpainting (CVPR 2019)

Optical Flow in Computer Vision

- Video frame interpolation with **optical flow + splatting** (CVPR 2020)

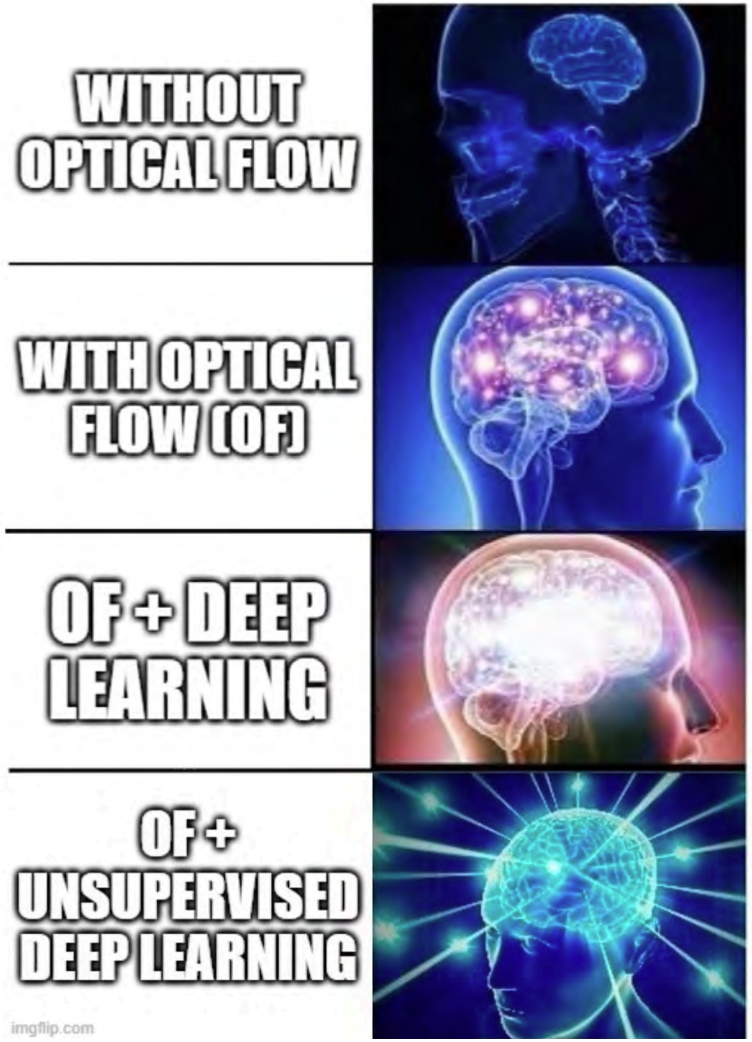




Softmax Splatting for Video Frame Interpolation (CVPR 2020)

In this talk...

AI Vision System



Deep Optical Flow Estimation

Overview

Limitation of Classical Methods

- **Classical Optical Flow**
  - **Optical** flow is the **apparent** motion of brightness patterns in the image
    - Motion can be caused by lighting changes **without any actual motion**
- **Deep Optical Flow**
  - **Optical** flow is not very **optical**
  - We understand optical flow as **actual motion made in a scene**
  - **Purely optical (classical) → Semantical inference (current)**

Frame 1

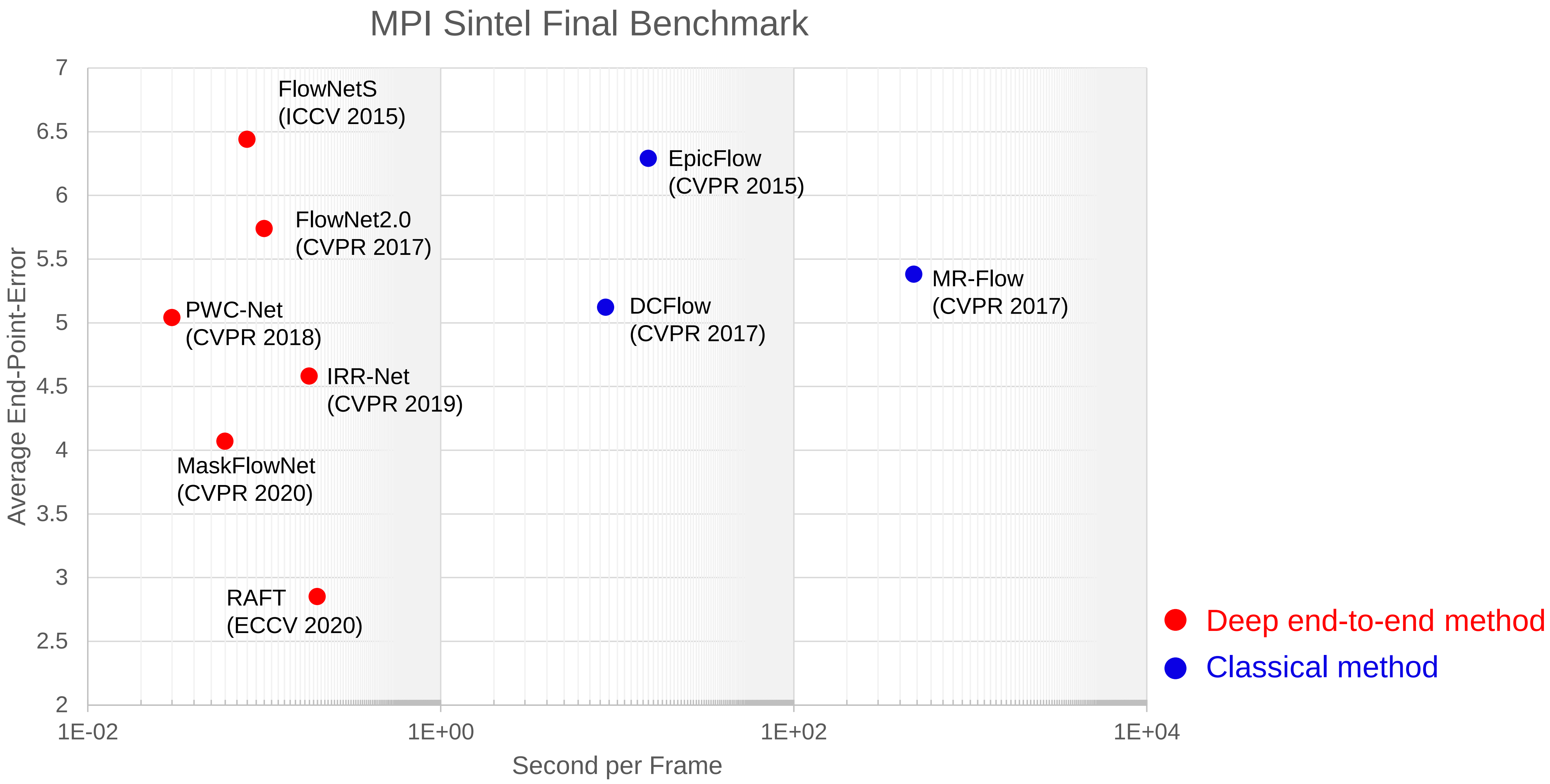
Frame 2

GT Optical Flow

SGVR Lab KAIST

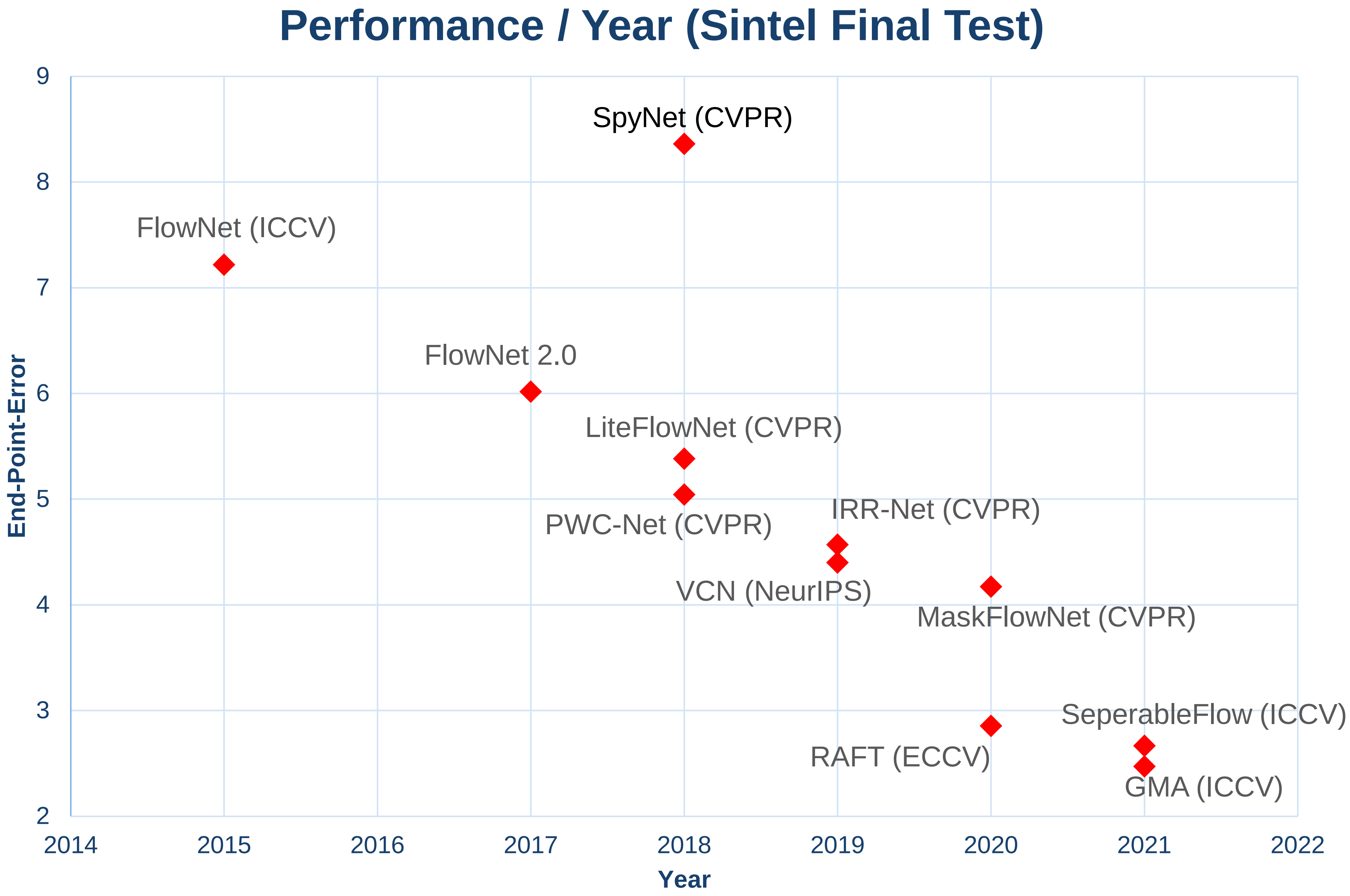
Image from: MPI Sintel dataset

Performance Difference



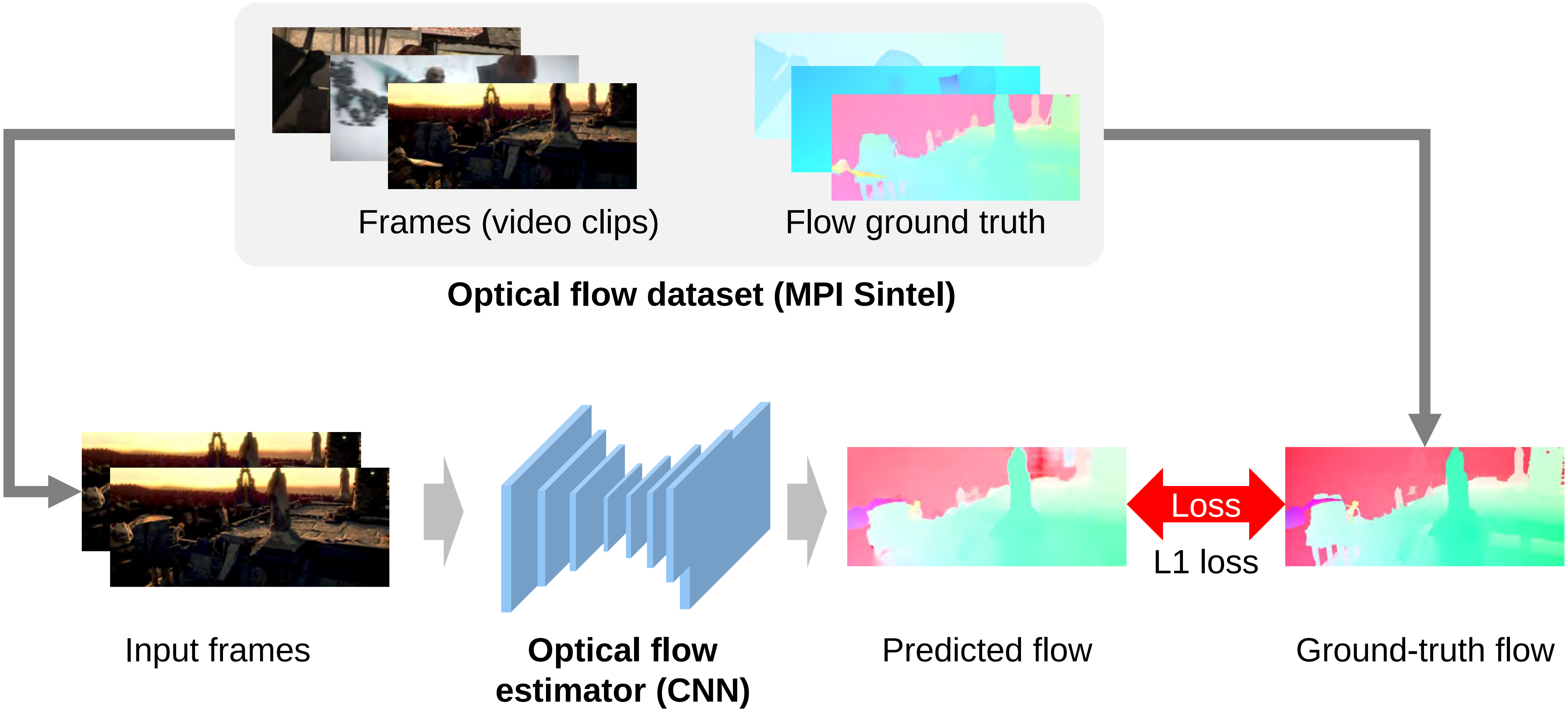
SGVR Lab
KAIST

Deep Architectures for Optical Flow



SGVR Lab
KAIST

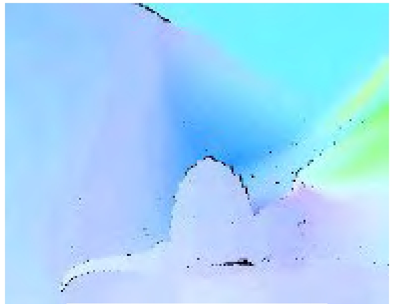
How to Learn Optical Flow? (end-to-end deep learning)



How to Make Optical Flow Datasets?

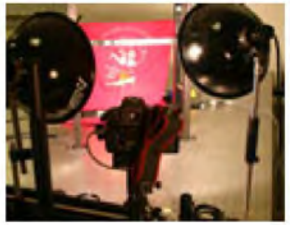
- Middlebury
  ① Spray some fluorescent paint to surfaces
  ② Take two pictures in different light types (visible / UV)
  ③ Move objects and repeat ①-②
- Fluorescent pattern in UV light gives optical flow (correspondence) ground truth!

Image



Flow

Setup

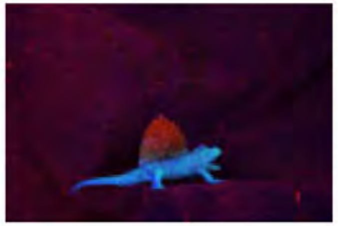


Visible light

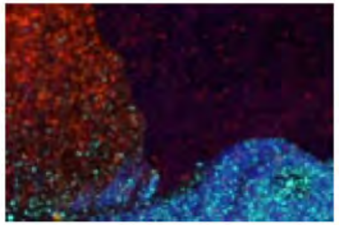
Visible light

Visible light (zoom)

UV light



UV light

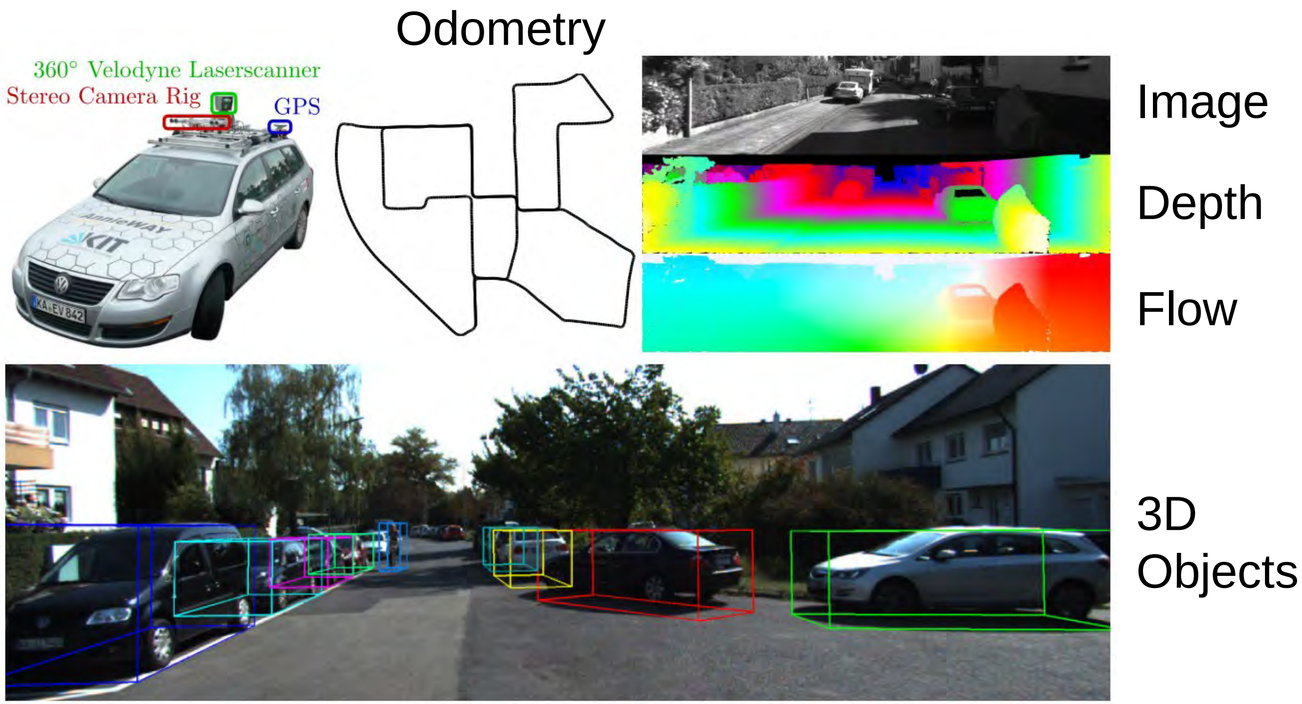


UV light (zoom)

A Database and Evaluation Methodology for Optical Flow, IJCV 2011

How to Make Optical Flow Datasets?

- KITTI
  ① Sensors: Cameras, Velodyne (LiDAR), GPS, IMU
  ② Collect data from sensors
  ③ Calibrate each data
  ④ Register 3D point clouds (with some manual matching)
  ⑤ Manually remove some ambiguous regions (windows, fences …)

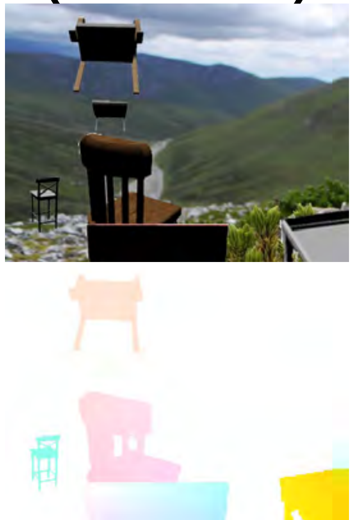


Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite, CVPR 2012

How to Make Optical Flow Datasets?

- **Real datasets are not enough (GT in low quality & low quantity)**
- Synthetic datasets
  - 👍 Infinitely many samples!
  - 👎 Lacks some realism…

FlyingChairs (ICCV 2015)



FlyingThings3D (CVPR 2017)

MPI Sintel (ECCV 2012)

Optical Flow Benchmarks (hidden test labels)

- MPI Sintel (ECCV2012)

- KITTI 2012 (CVPR 2012), KITTI 2015 (CVPR 2015)

Optical Flow Benchmarks (hidden test labels)

- MPI Sintel (ECCV2012)

- Spec
  - 1041 training pairs
  - 552 testing pairs
  - 1024x436 resolutions

- Focused on realistic effects
  - Motion blur, lighting effects, extreme camera movement ...

- Dense optical flow is provided
  - Rendered dataset!

Optical Flow Benchmarks (hidden test labels)

- KITTI 2012 (CVPR 2012),
  KITTI 2015 (CVPR 2015)

- Spec
  - 200 training pairs
  - 200 testing pairs
  - 1242x375 resolution

- Real-world driving data
  - Extreme shadows are the biggest challenge

- Sparse optical flow is provided
  - Real-world videos have non-matched pixels

SGVR Lab
KAIST

Optical Flow Benchmarks (hidden test labels)

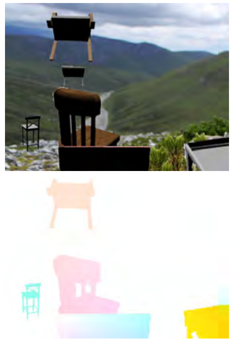
- MPI Sintel (ECCV2012)

- KITTI 2012 (CVPR 2012),
  KITTI 2015 (CVPR 2015)

Datasets for Optical Flow Estimation

- Datasets

Synthetic | Real

	FlyingChairs	**FlyingThings3D**	**Sintel**	**MiddleburyFlow**	**KITTI 2012**	
RGB						
						
Size	22K pairs	25K pairs	1K frames	0.1K frames	0.2K pairs	
Feature	2D Motion	3D motion	Realistic but not real	Dense	Not dense	Too small!

Unsupervised Optical Flow with Deep Feature Similarity

Unsupervised Learning of Optical Flow with Deep Feature Similarity

Woobin Im, Tae-Kyun Kim, and Sung-Eui Yoon

ECCV 2020

Horn-Schunk method for optical flow

- The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[\boxed{(I_x u + I_y v + I_t)}^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] \mathrm{d}x\mathrm{d}y$$

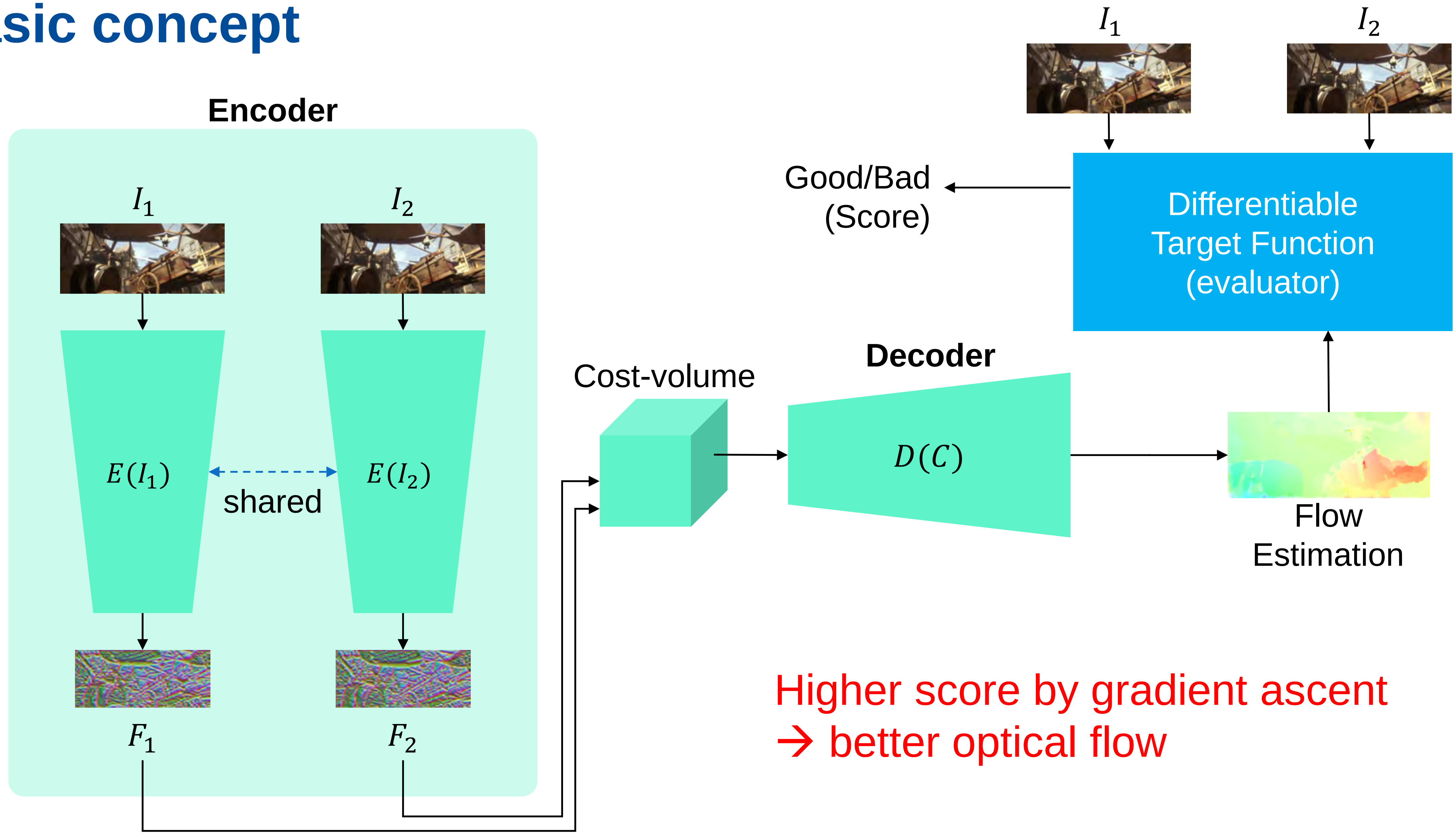
- The first part of the function is the brightness consistency.

Classical methods does not require GT,
but takes **few minutes / frame**

Can we learn **an end-to-end model**?

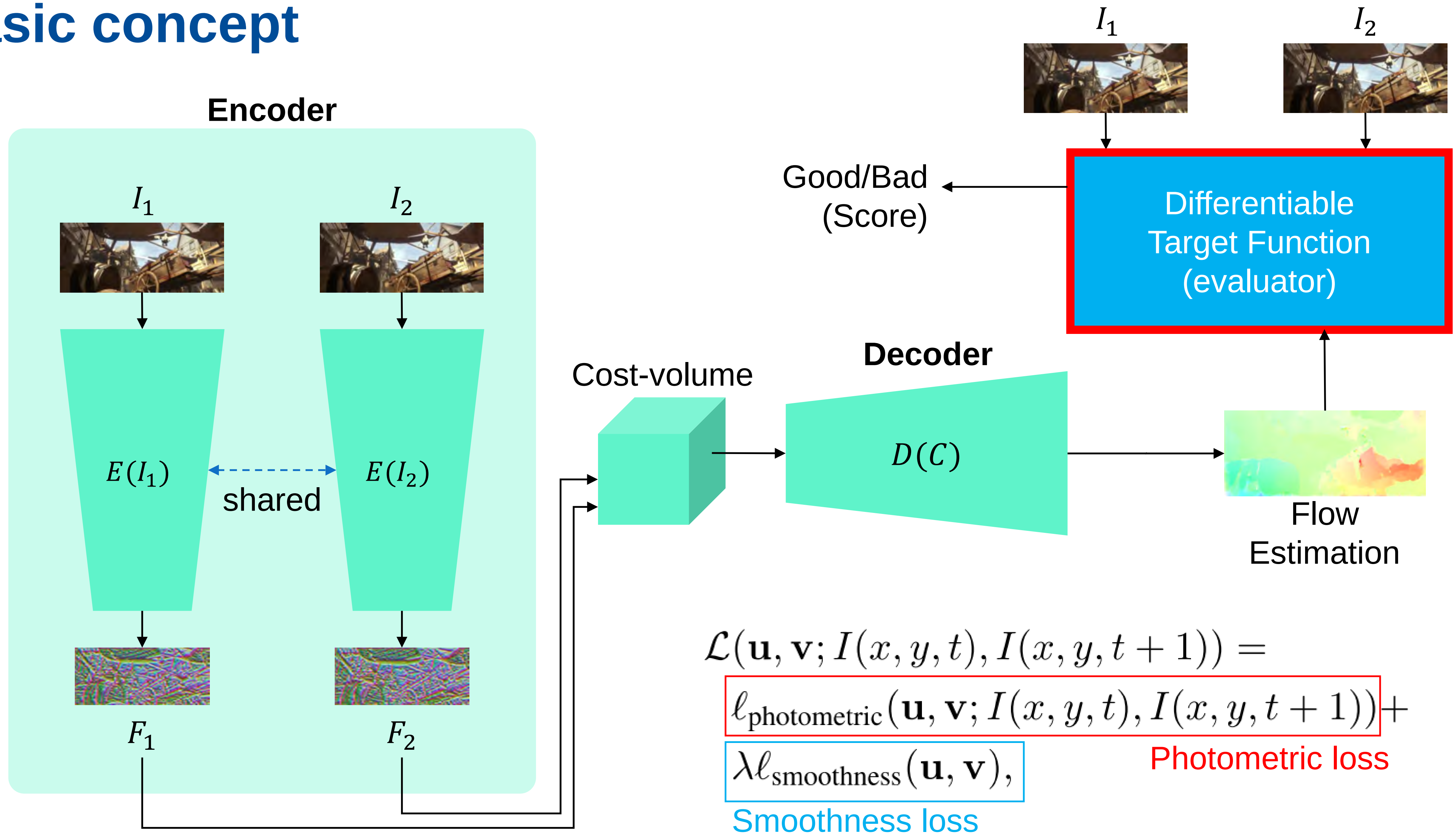
Credit: Juan Carlos Niebles and Ranjay Krishna @ Stanford Vision and Learning Lab

Basic concept



Higher score by gradient ascent
→ better optical flow

Basic concept



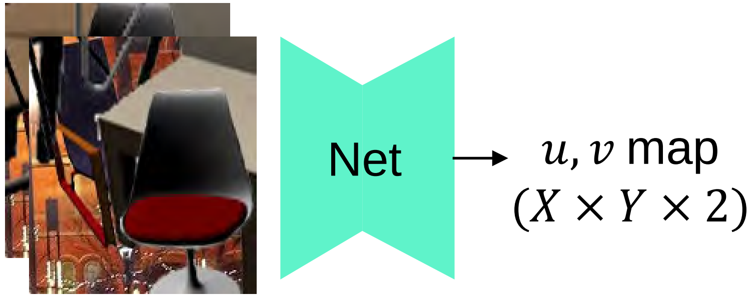
$$\mathcal{L}(\mathbf{u},\mathbf{v};I(x,y,t),I(x,y,t+1)) = \boxed{\ell_{\text{photometric}}(\mathbf{u},\mathbf{v};I(x,y,t),I(x,y,t+1))} + \boxed{\lambda\ell_{\text{smoothness}}(\mathbf{u},\mathbf{v})},$$

Photometric loss

Smoothness loss

Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness, ECCV workshop 2016

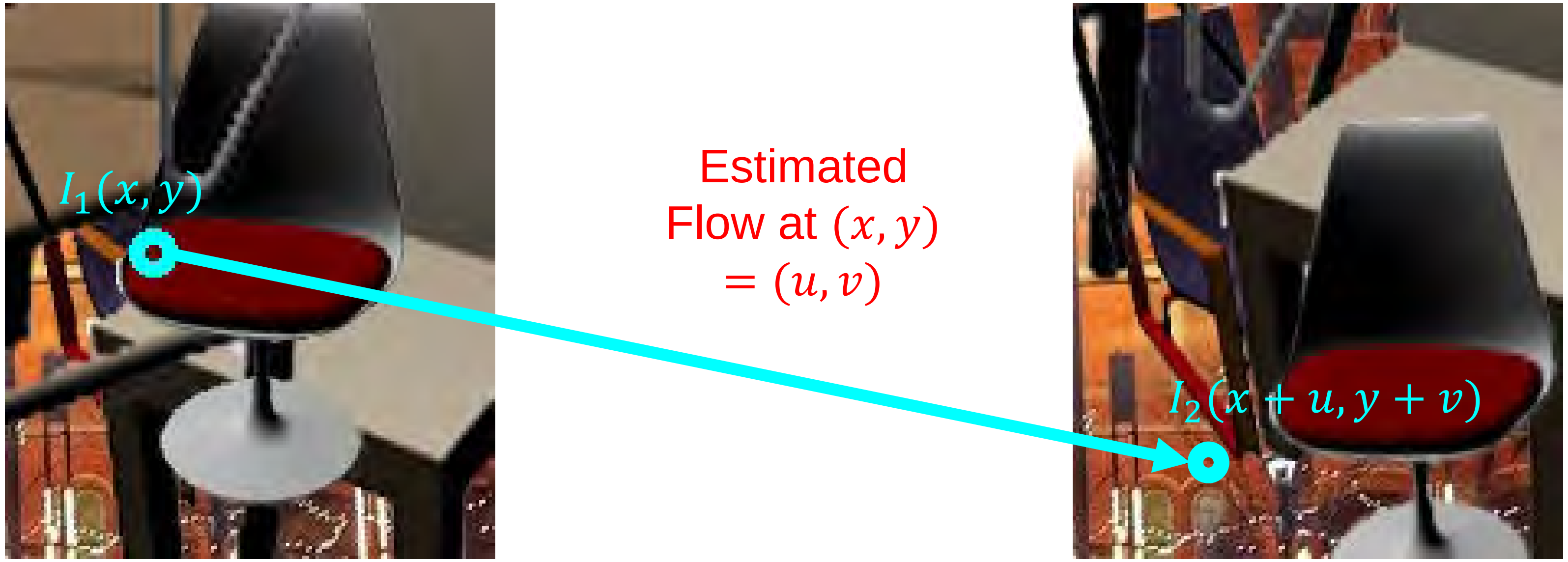
Photometric Consistency Loss



- Photometric consistency loss

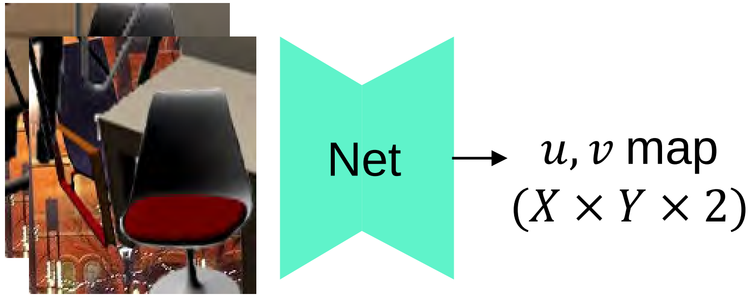
$$L_{photo} = \sum_{(x,y)\in\Omega} \|I_1(x,y) - I_2(x+u, y+v)\|_2^2$$

We can compute gradient w.r.t. (u, v) to obtain a better flow!



I_1 I_2

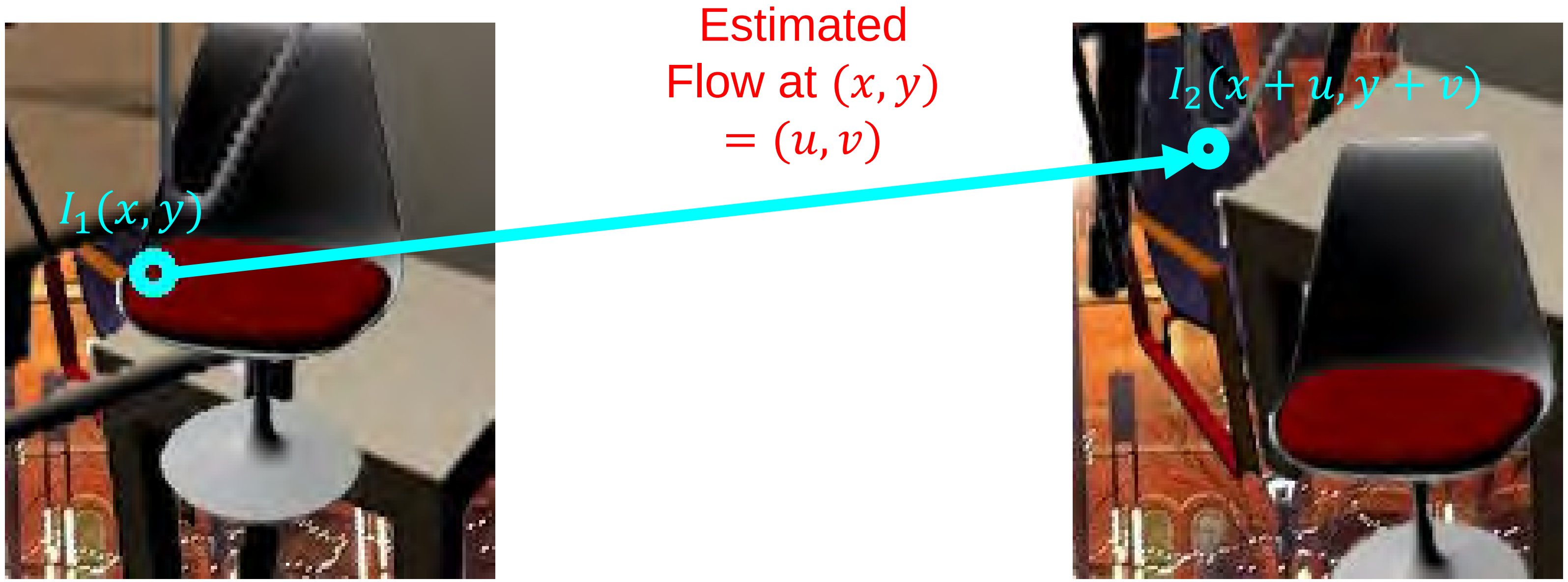
Photometric Consistency Loss



- Photometric consistency loss

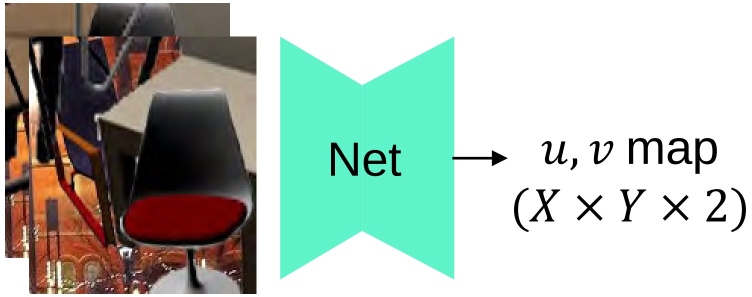
$$L_{photo} = \sum_{(x,y)\in\Omega} \|I_1(x,y) - I_2(x+u, y+v)\|_2^2$$

We can compute gradient w.r.t. (u, v) to obtain a better flow!



I_1 I_2

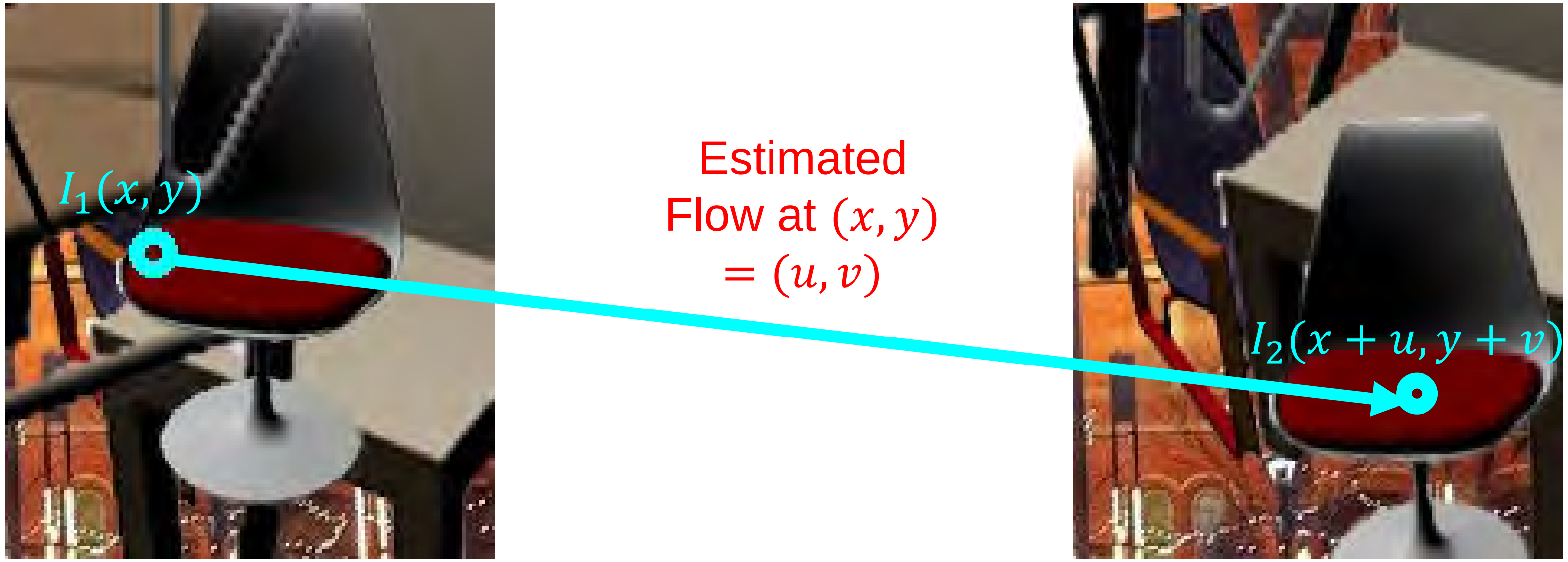
Photometric Consistency Loss



- Photometric consistency loss

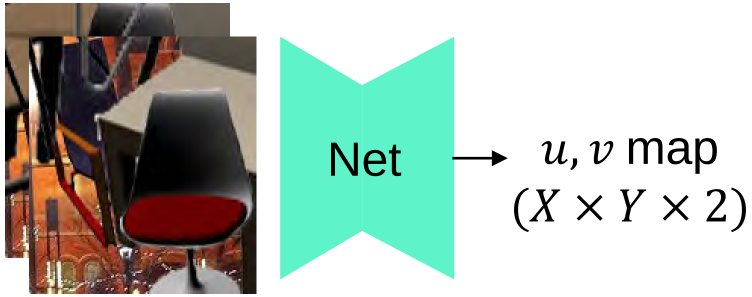
$$L_{photo} = \sum_{(x,y)\in\Omega} \|I_1(x, y) - I_2(x + u, y + v)\|_2^2$$

We can compute gradient w.r.t. (u, v) to obtain a better flow!



I_1 I_2

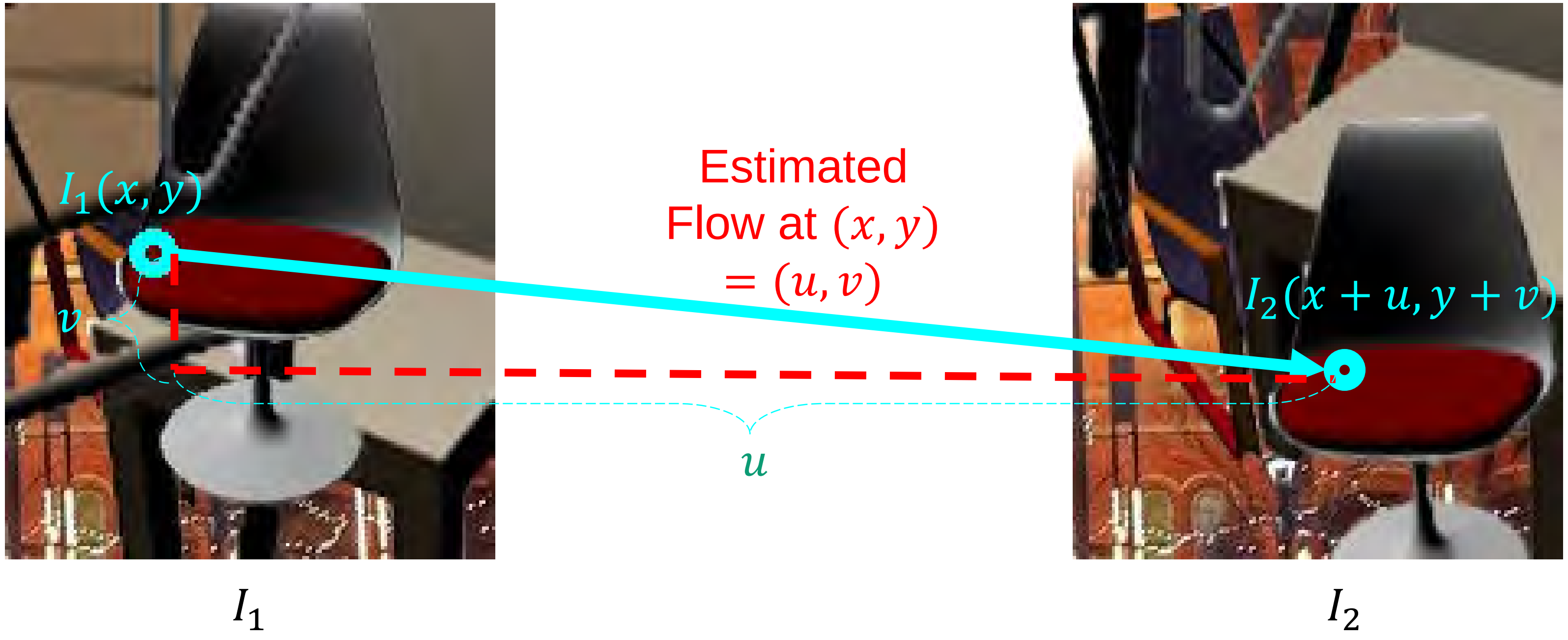
Photometric Consistency Loss



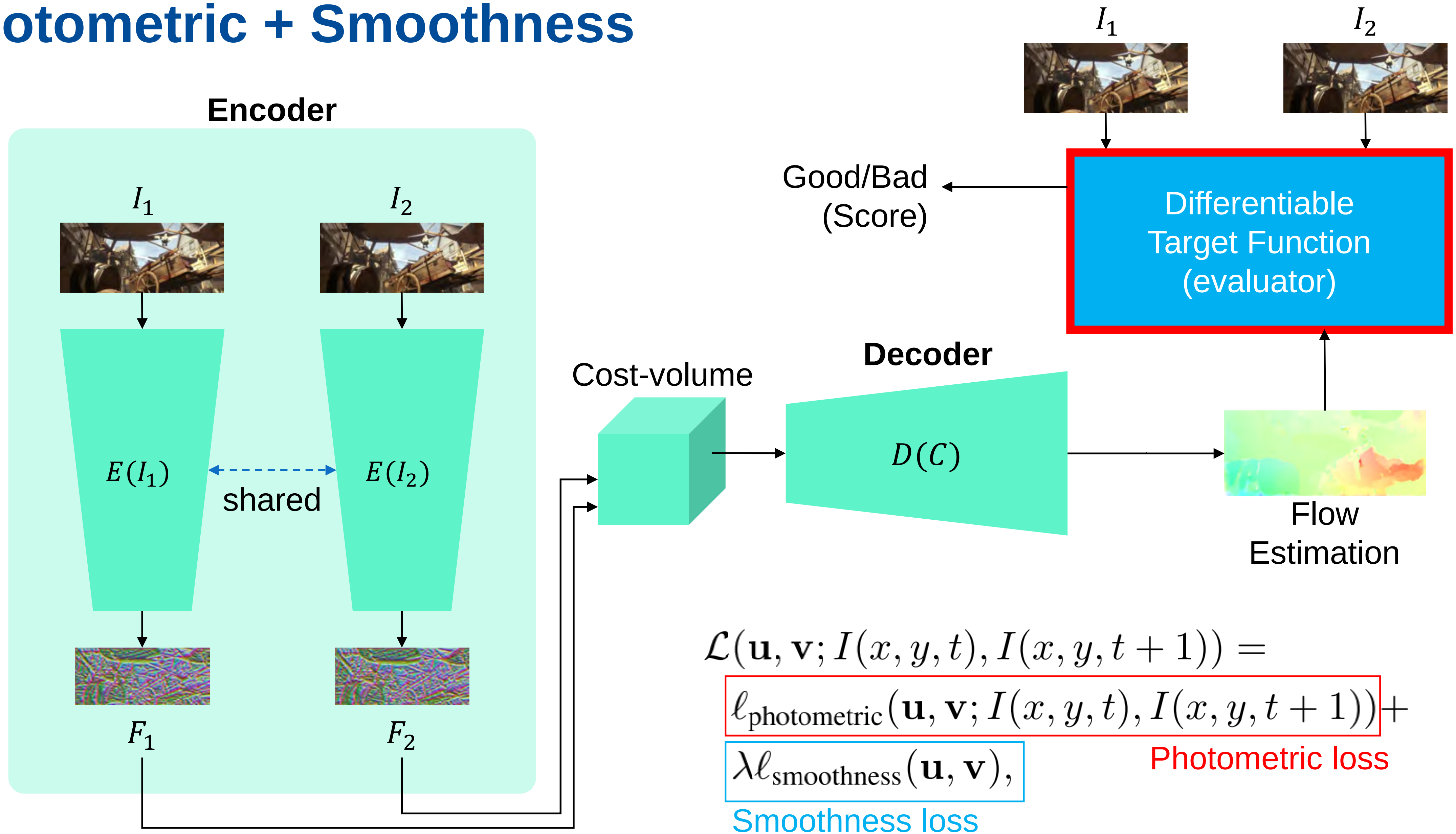
- Photometric consistency loss

$$L_{photo} = \sum_{(x,y)\in\Omega} \|I_1(x,y) - I_2(x+u, y+v)\|_2^2$$

We can compute gradient w.r.t. (u, v) to obtain a better flow!



Photometric + Smoothness



$$\mathcal{L}(\mathbf{u},\mathbf{v}; I(x,y,t), I(x,y,t+1)) = \boxed{\ell_{\text{photometric}}(\mathbf{u},\mathbf{v}; I(x,y,t), I(x,y,t+1))} + \boxed{\lambda \ell_{\text{smoothness}}(\mathbf{u},\mathbf{v})},$$

Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness, ECCV workshop 2016

SGVR Lab
KAIST

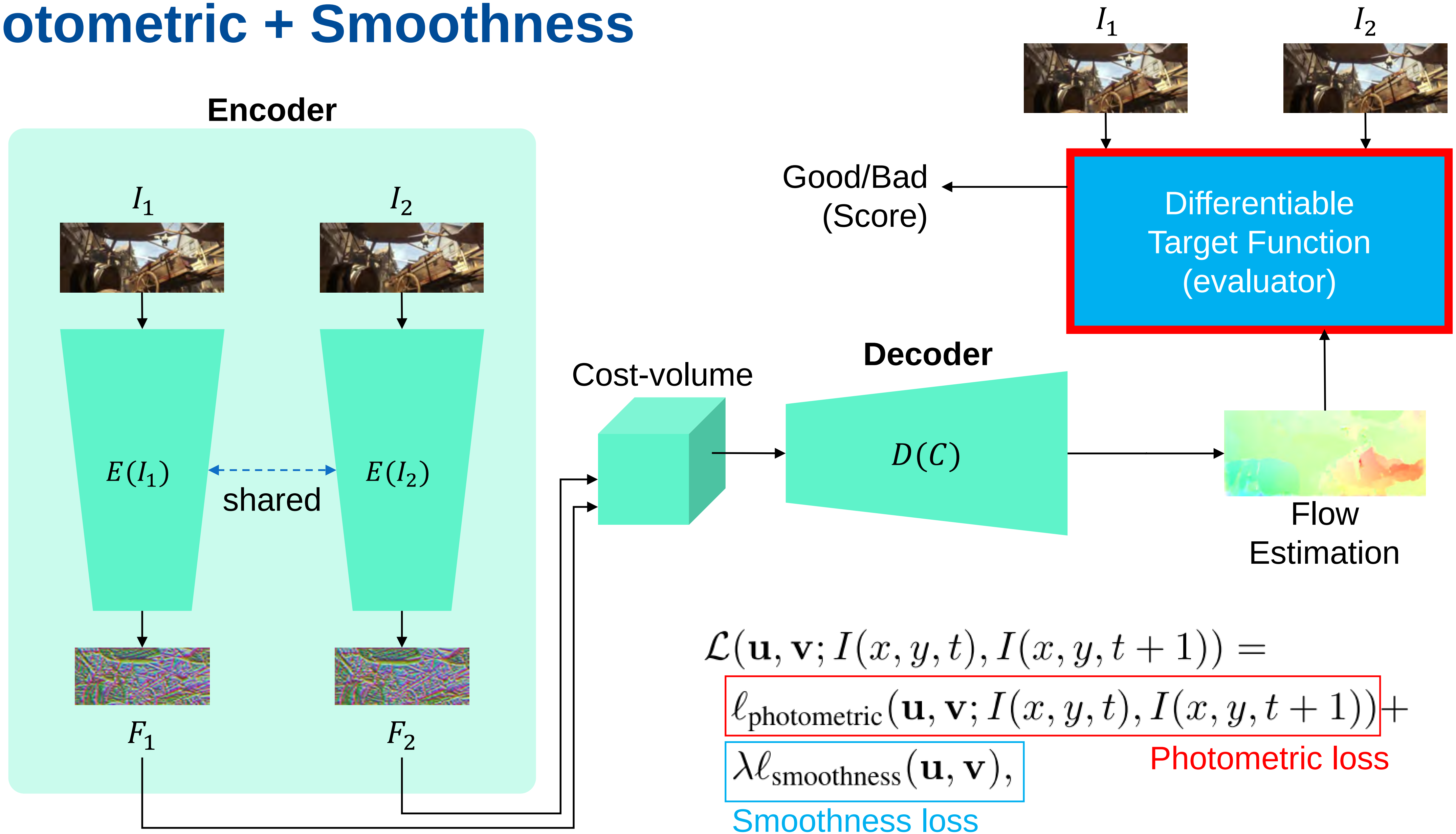
Smoothness constraint

- The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\boxed{\|\nabla u\|^2 + \|\nabla v\|^2} \right) \right] \mathrm{d}x\mathrm{d}y$$

- The second part is the smoothness constraint. It's trying to make sure that the changes between frames are small.

Photometric + Smoothness



$$\mathcal{L}(\mathbf{u}, \mathbf{v}; I(x, y, t), I(x, y, t+1)) =$$

$$\boxed{\ell_{\text{photometric}}(\mathbf{u}, \mathbf{v}; I(x, y, t), I(x, y, t+1))} +$$

$$\boxed{\lambda \ell_{\text{smoothness}}(\mathbf{u}, \mathbf{v})},$$

Photometric loss

Smoothness loss

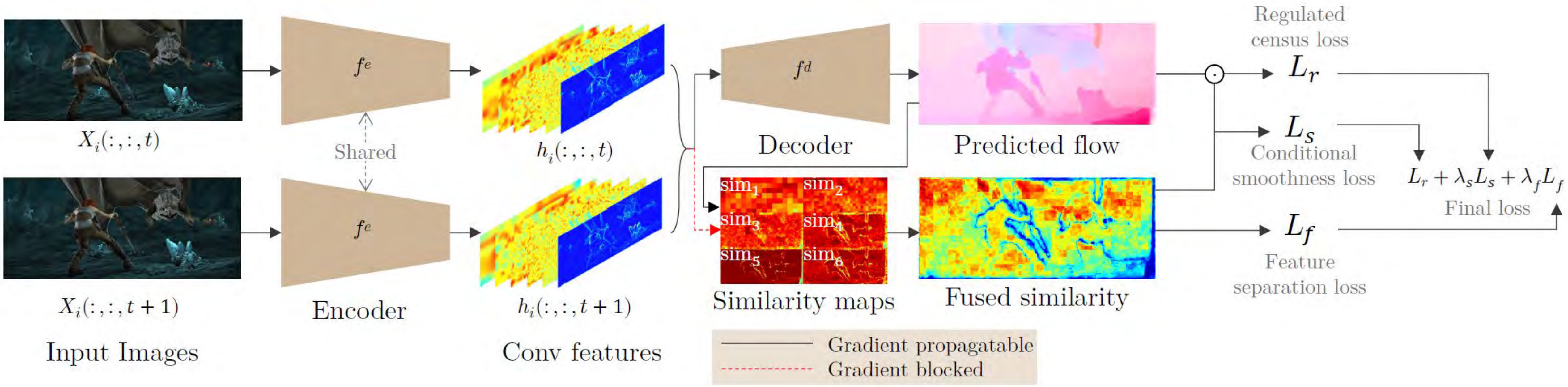
Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness, ECCV workshop 2016

Good/Bad (Score) ← Differentiable Target Function (evaluator)

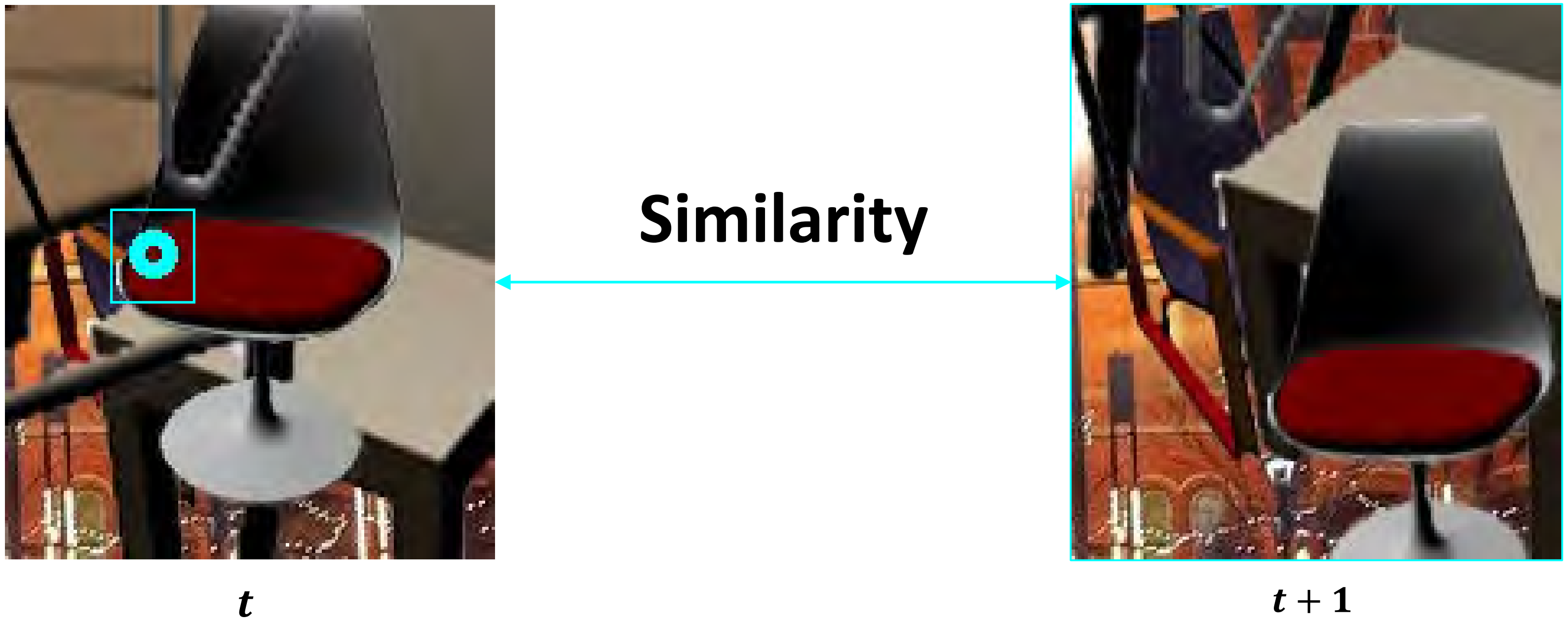
- As-is
  - Same as classical formulation $\left[(I_x u + I_y v + I_t)^2\right.$
- **To-be (ours)**
  - **Deep, self-supervised formulation**

Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020

- Why not use deep feature for optical flow learning?

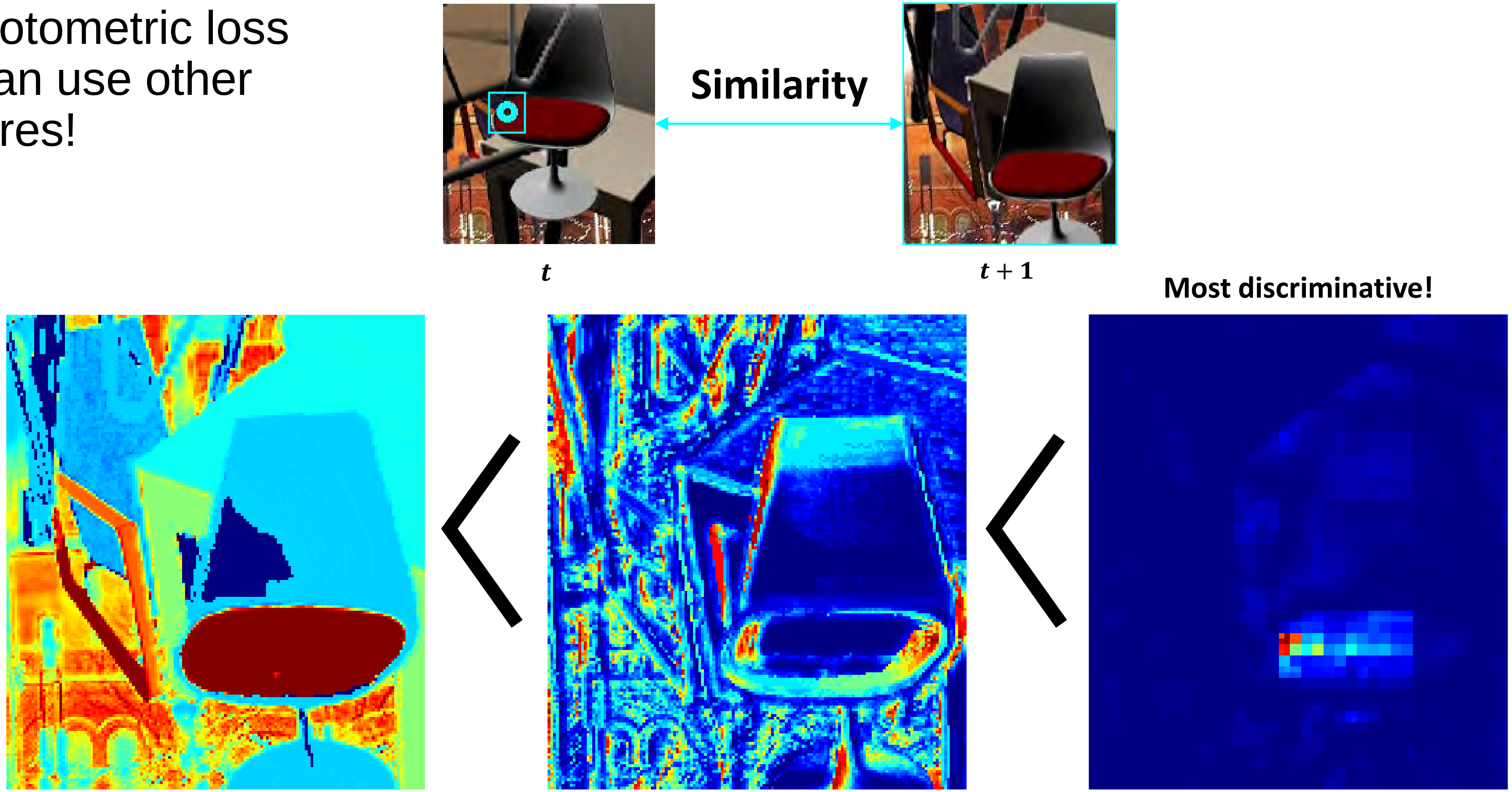


Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020



Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020

- In photometric loss we can use other features!

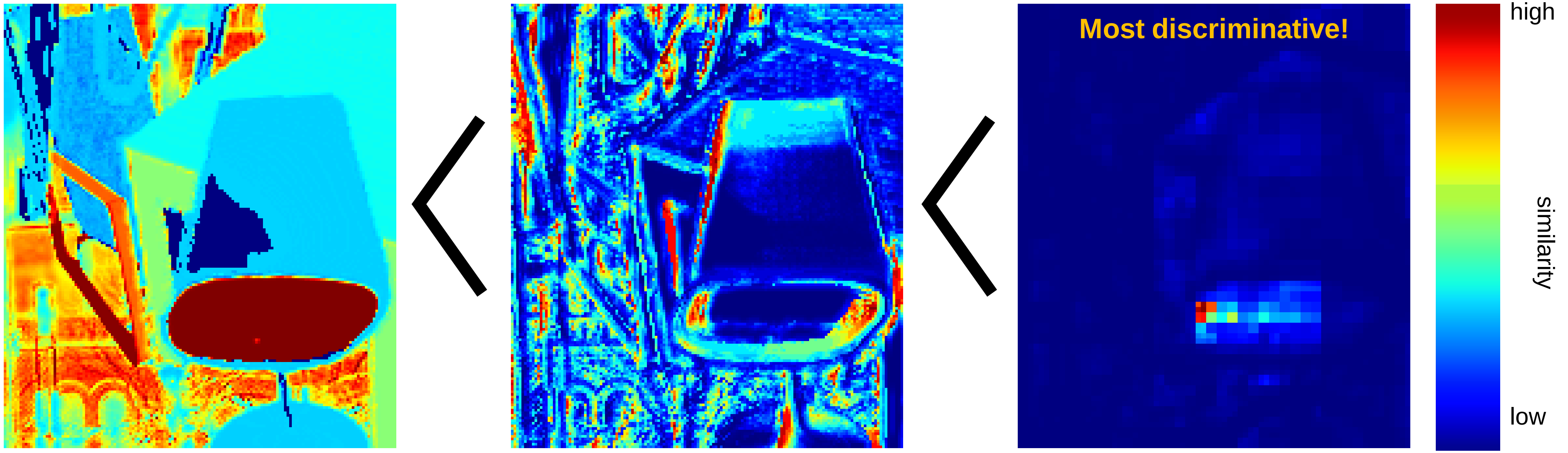


RGB

Census
(handcraft feature)

Deep feature

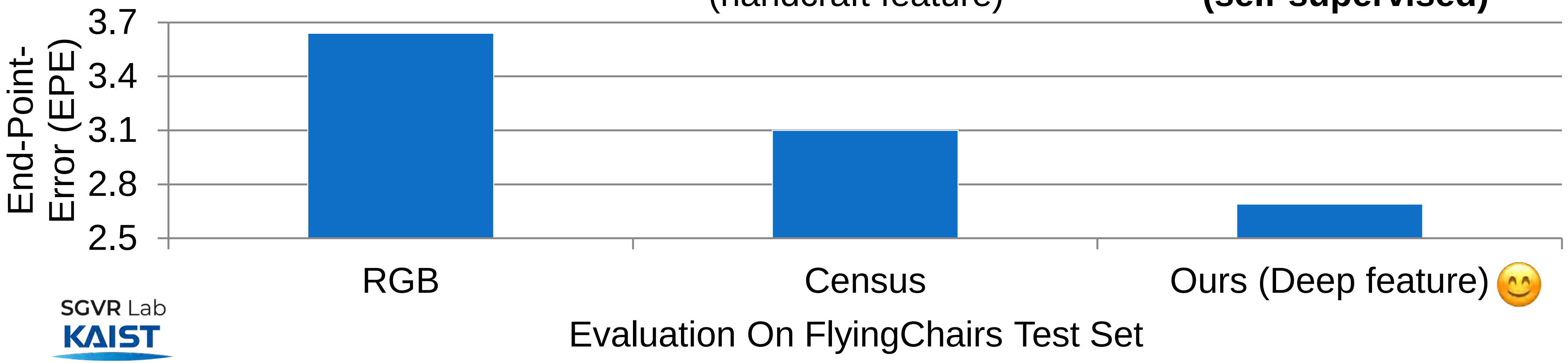
Which feature to use?



RGB

Census
(handcraft feature)

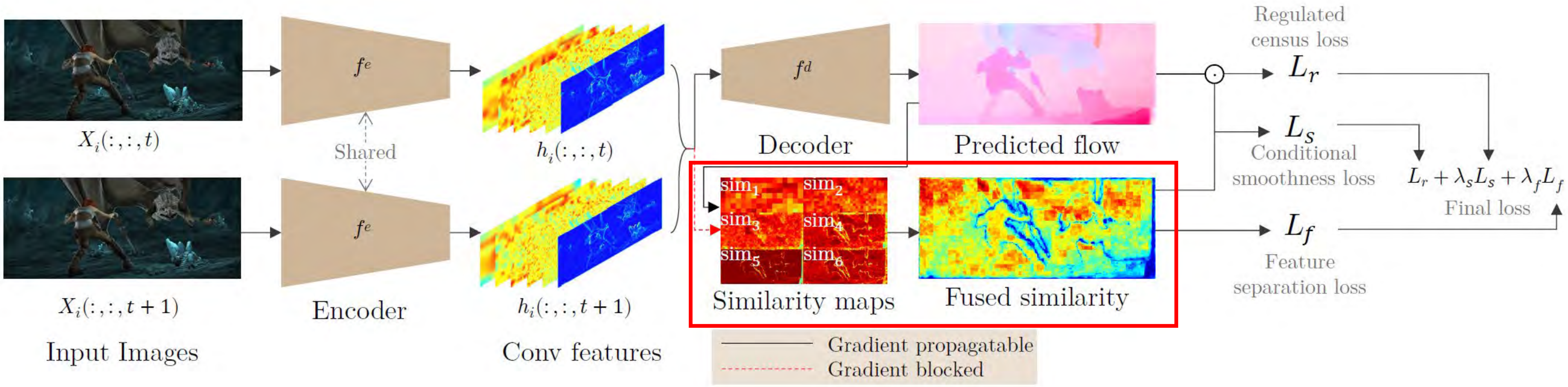
Deep feature
(self-supervised)



SGVR Lab
KAIST

Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020

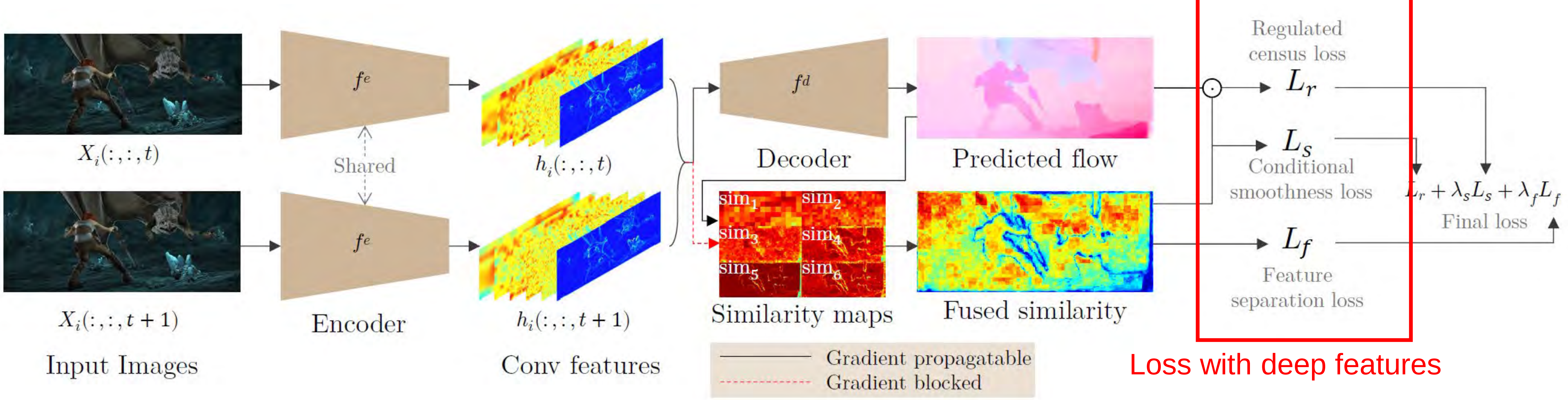
- Using deep feature for optical flow learning



Multi-layer feature fusion

Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020

- Why not use deep feature for optical flow learning?



Loss with deep features

Feature separation loss (L_f)

$$L_f = \frac{1}{N} \sum_{i}^{N} \sum_{(x,y,t)\in\Omega} -(\mathrm{sim}_f(x,y,t) - k)^2$$

- **Encourages** higher similarity for non-occluded ones
- **Discourages** higher similarity for occluded ones

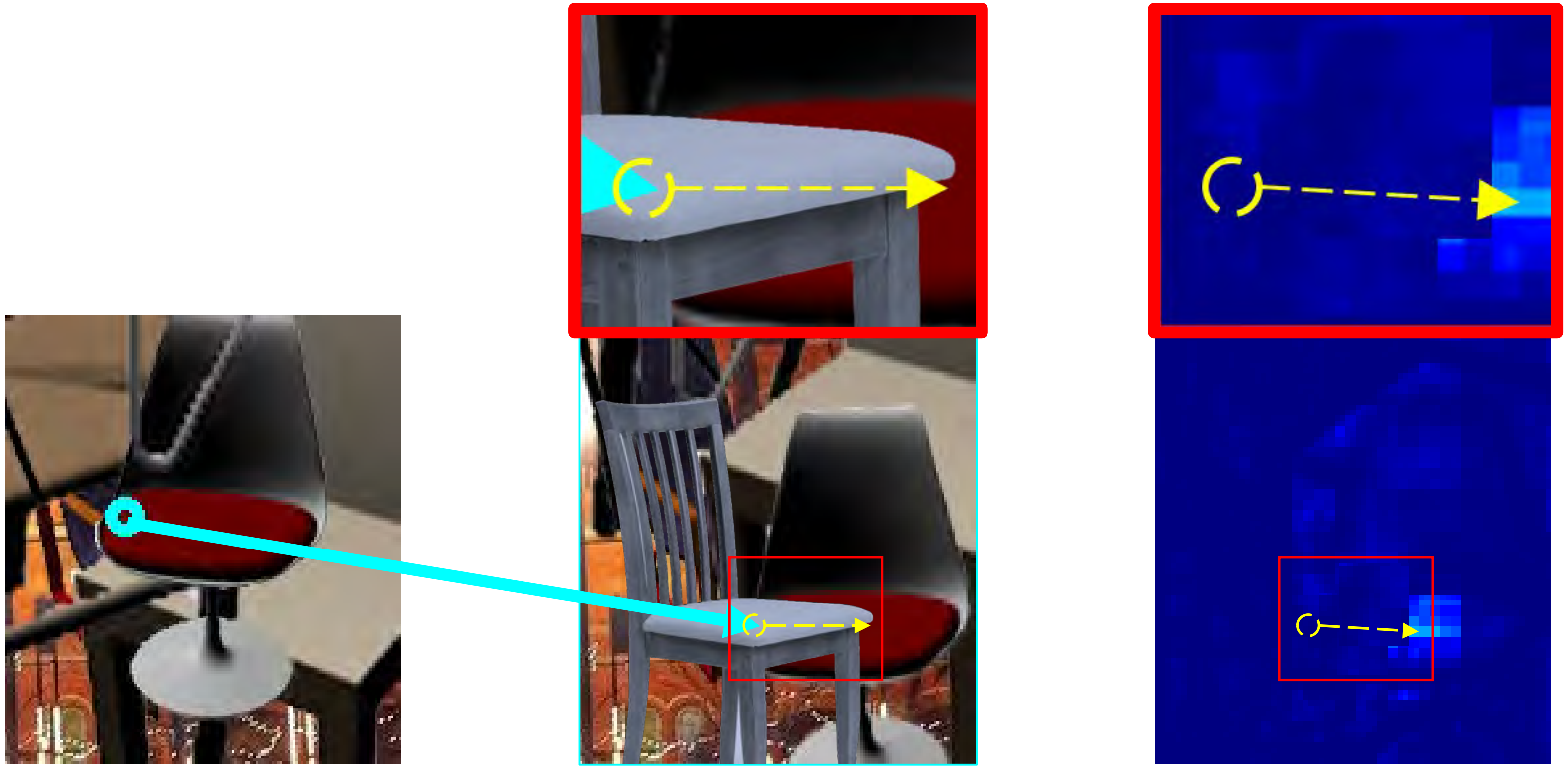
SGVR Lab KAIST

Deep Feature Separation Loss

Good match

Good match

Deep Feature Separation Loss



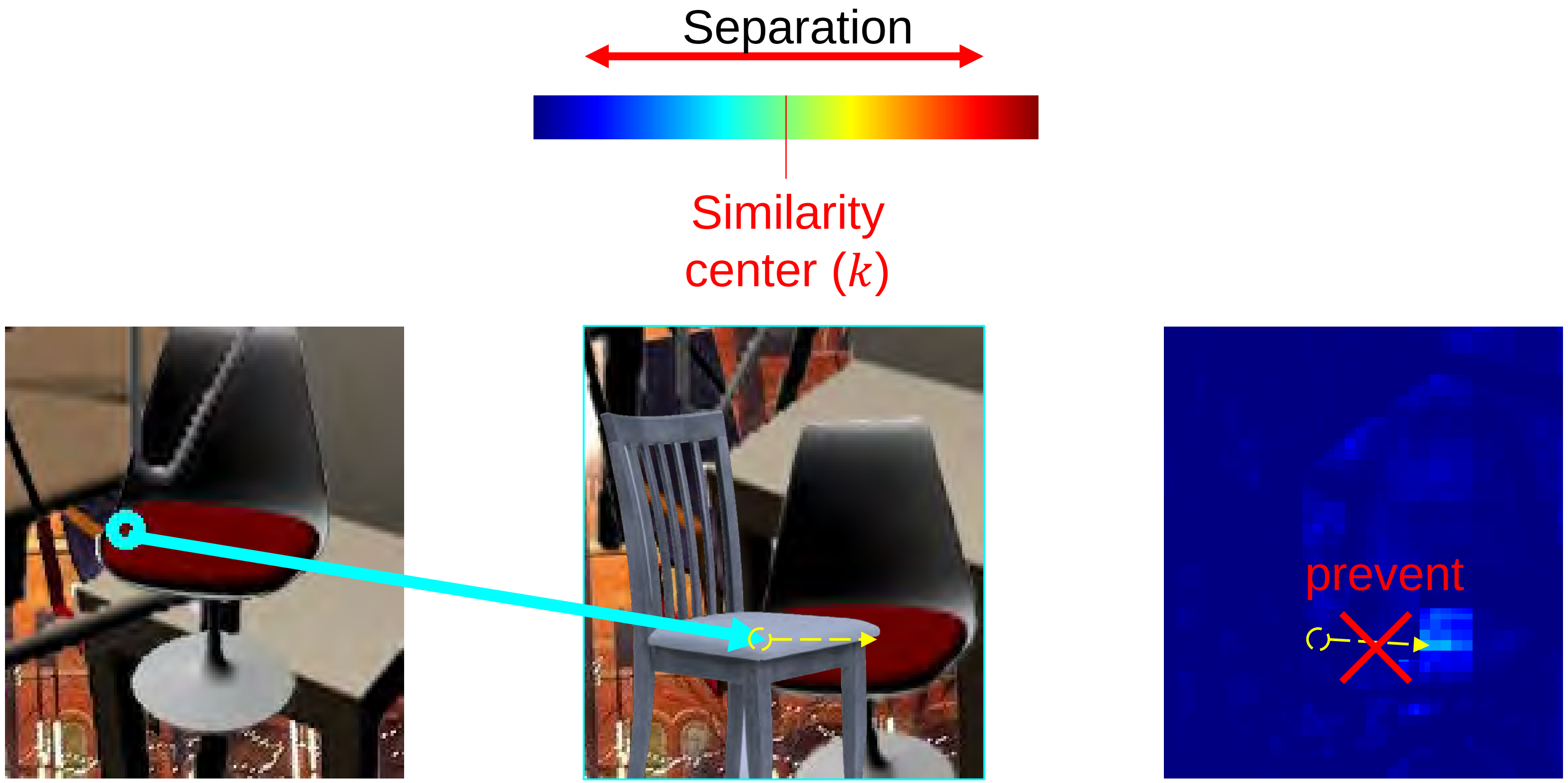
When occluded, maximizing similarity
results in a bad solution

Deep Feature Separation Loss

When occluded, maximizing similarity results in a bad solution

- Learning with photometric loss tends to make high-similarity solution
- **Deep feature separation loss** helps avoid this solution

Deep Feature Separation Loss



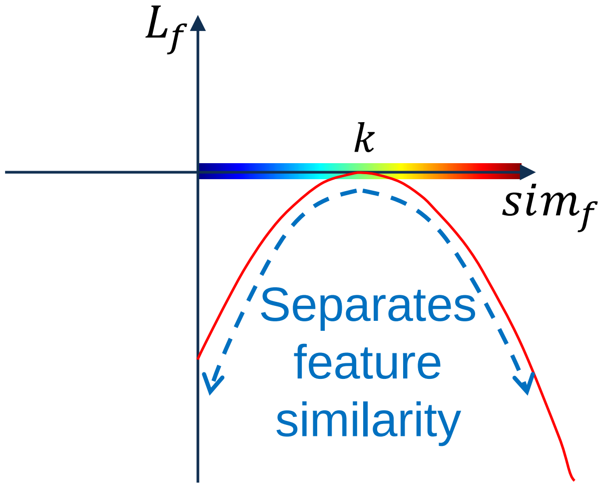
If $similarity < k$, minimize $similarity$
otherwise, maximize $similarity$

Deep Feature Separation Loss

$$L_f = \frac{1}{N}\sum_{i}^{N}\sum_{(x,y,t)\in\Omega} -(\mathrm{sim}_f(x,y,t) - k)^2$$

Similarity threshold

$$k = \frac{1}{2}(k_{noc} + k_{occ})$$

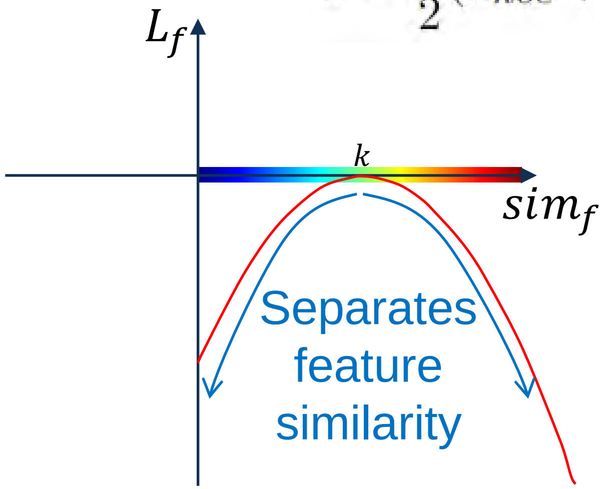


Deep Feature Separation Loss

$$L_f = \frac{1}{N} \sum_{i}^{N} \sum_{(x,y,t)\in\Omega} -(\mathrm{sim}_f(x,y,t) - k)^2$$

Similarity threshold

$$k = \frac{1}{2}(k_{noc} + k_{occ})$$



Related work (regularization for discriminative features)

- Guided Similarity Separation for Image Retrieval, NeurIPS 2019

$$\mathcal{L}(s_{ij}) = -\frac{\alpha}{2}(s_{ij} - \beta)^2$$

- Semi-supervised Learning by Entropy Minimization, NeurIPS 2004

$$C(\boldsymbol{\theta}, \lambda; \mathcal{L}_n) \quad = \quad L(\boldsymbol{\theta}; \mathcal{L}_n) - \lambda H_{\mathrm{emp}}(Y|X, Z; \mathcal{L}_n)$$

Final Loss Function

$$L = \boxed{L_r} + \lambda_f \boxed{L_f} + \lambda_s L_s$$

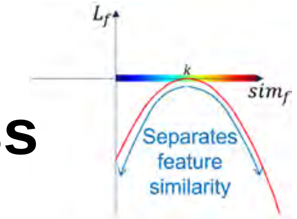
Smoothness loss

$$L_s = \frac{1}{N}\sum_{i}^{N}\sum_{(x,y,t)\in\Omega}(|\nabla u|^2 + |\nabla v|^2)M_l(x,y,t)$$

Deep Similarity-Aware Census Loss

$$L_r = \frac{1}{N}\sum_{i}^{N}\sum_{(x,y,t)\in\Omega}\underbrace{\Psi(\cdot)\hat{C}_i^o(x,y,t)}_{\text{Conventional loss}}\underbrace{\mathrm{sim}_f(x,y,t)}_{\text{Deep similarity}}$$

Deep Feature Separation Loss



$$L_f = \frac{1}{N}\sum_{i}^{N}\sum_{(x,y,t)\in\Omega} -(\underbrace{\mathrm{sim}_f(x,y,t)}_{\text{Deep similarity}} - k)^2$$

Unsupervised Learning of Optical Flow with Deep Feature Similarity, ECCV 2020

	FlyingChairs	Sintel Clean	Sintel Final
RGB	3.64	4.40	5.42
Census	2.93	3.15	3.86
Ours (deep)	**2.69**	**2.86**	**3.57**

Image GT flow SelFlow Ours
CVPR 2019

SGVR Lab
KAIST

Semi-Supervised Optical Flow by Flow Supervisor

Semi-Supervised Learning of Optical Flow by Flow Supervisor

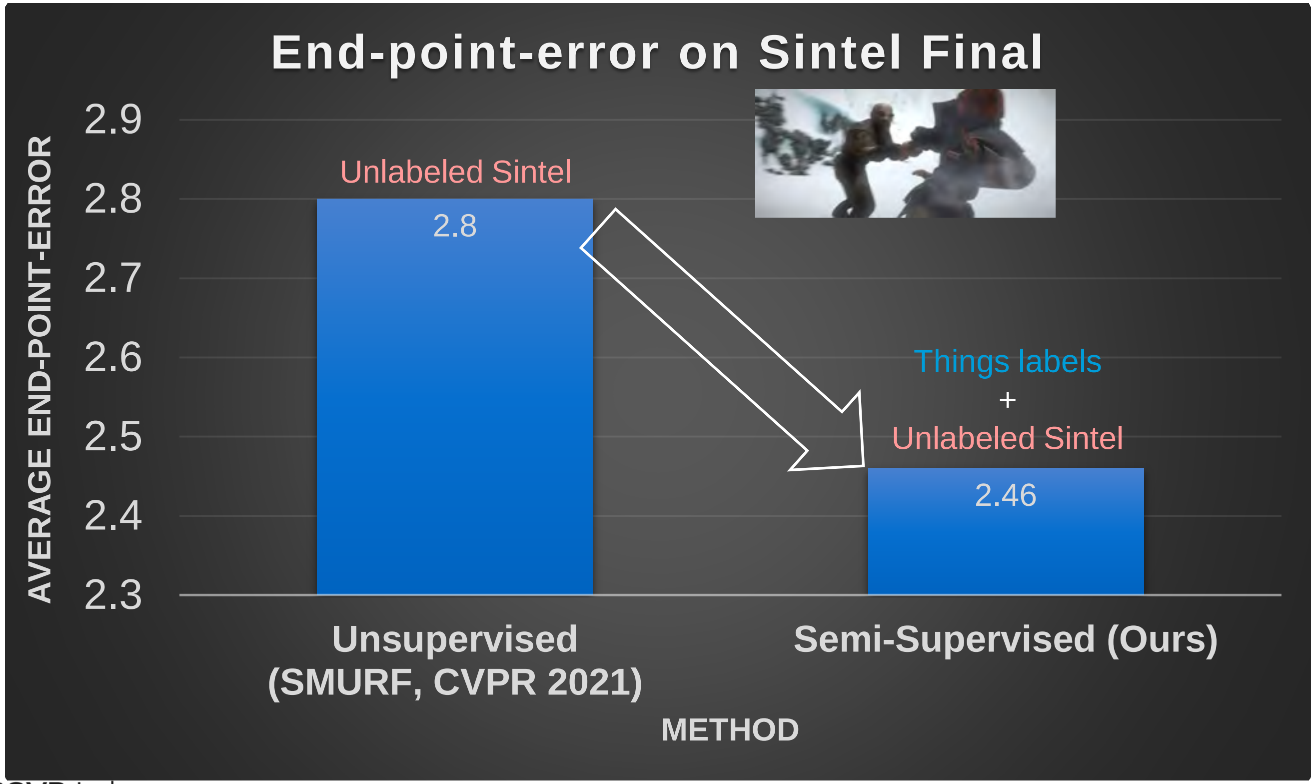
Woobin Im, Sebin Lee, and Sung-Eui Yoon

ECCV 2022

Semi-Supervised Optical Flow?

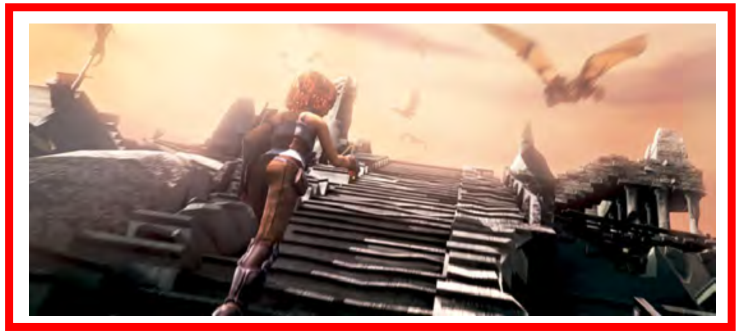
- Supervised methods **do not use unlabeled data**
- Unsupervised methods **do not use any label**

Semi-supervised learning method can improve by using synthetic labels



Things (labeled)

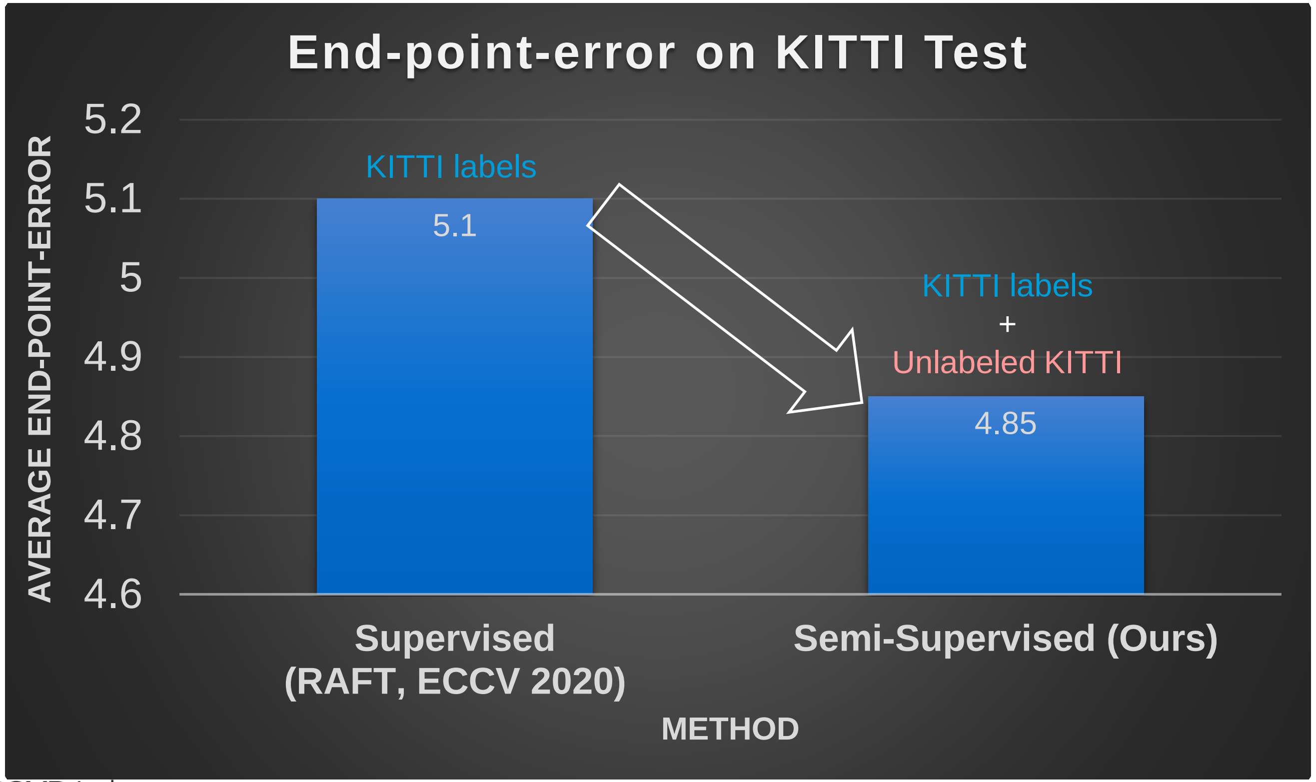
Sintel (unlabeled)



Semi-Supervised Optical Flow?

- Supervised methods **do not use unlabeled data**
- Unsupervised methods **do not use any label**

Semi-supervised learning method can improve by using additional labels



KITTI (labeled)

200 pairs labeled

KITTI (unlabeled)

4,200 pairs unlabeled

Naïve Approach

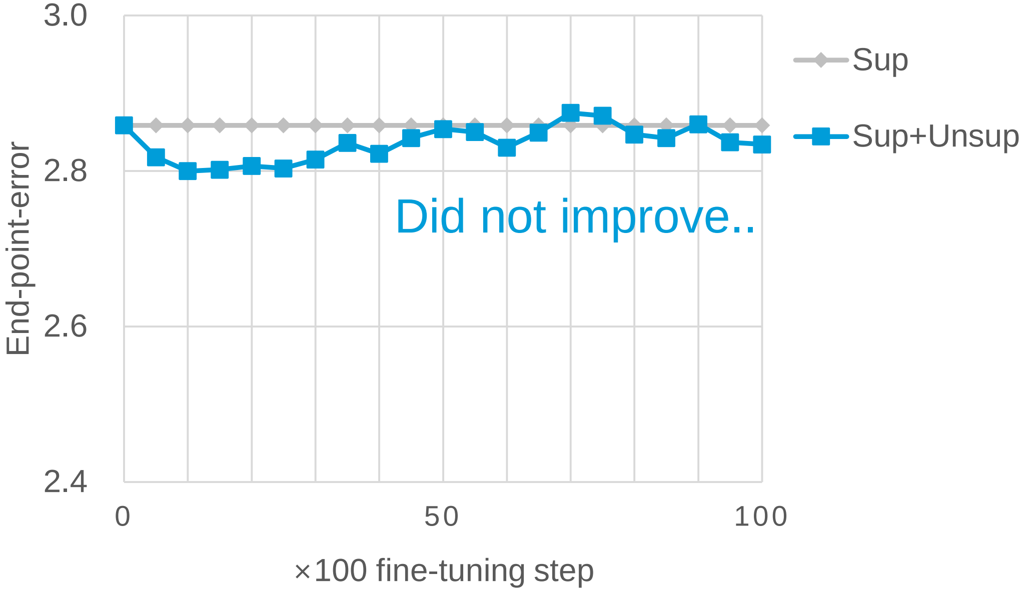
Things

Sintel (unlabeled)

$$Loss = \ell_{sup} + \lambda \ell_{unsup}$$

Pixel-wise loss
e.g. L1 loss

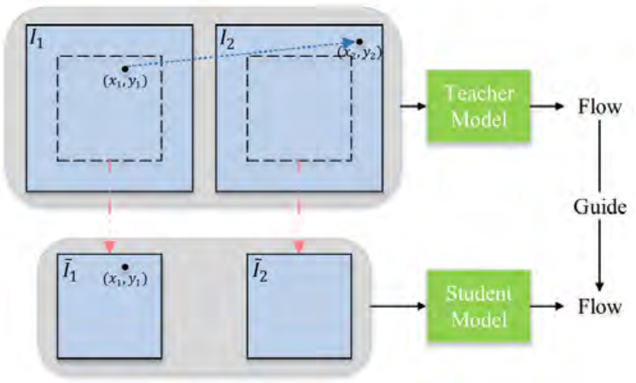
Photometric loss
(SMURF; CVPR 2021)



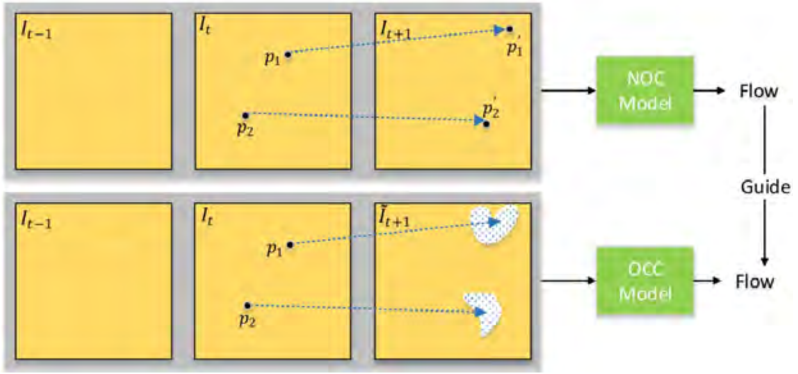
SGVR Lab
KAIST

Self-Supervision Loss for Optical Flow Learning

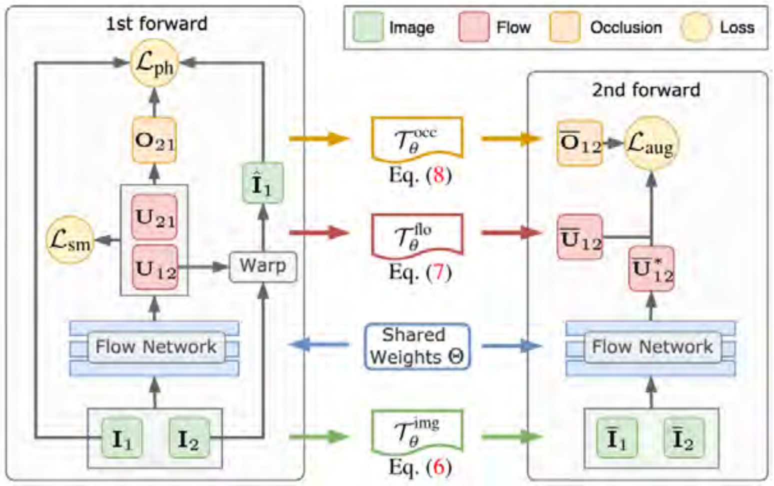
Self-supervision for optical flow



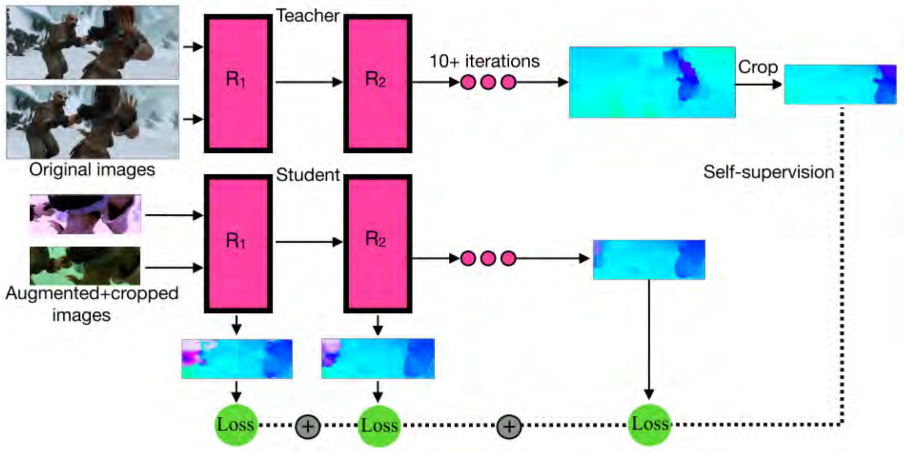
DDFlow (AAAI 2019)



SelFlow (CVPR 2019)



ARFlow (CVPR 2020)

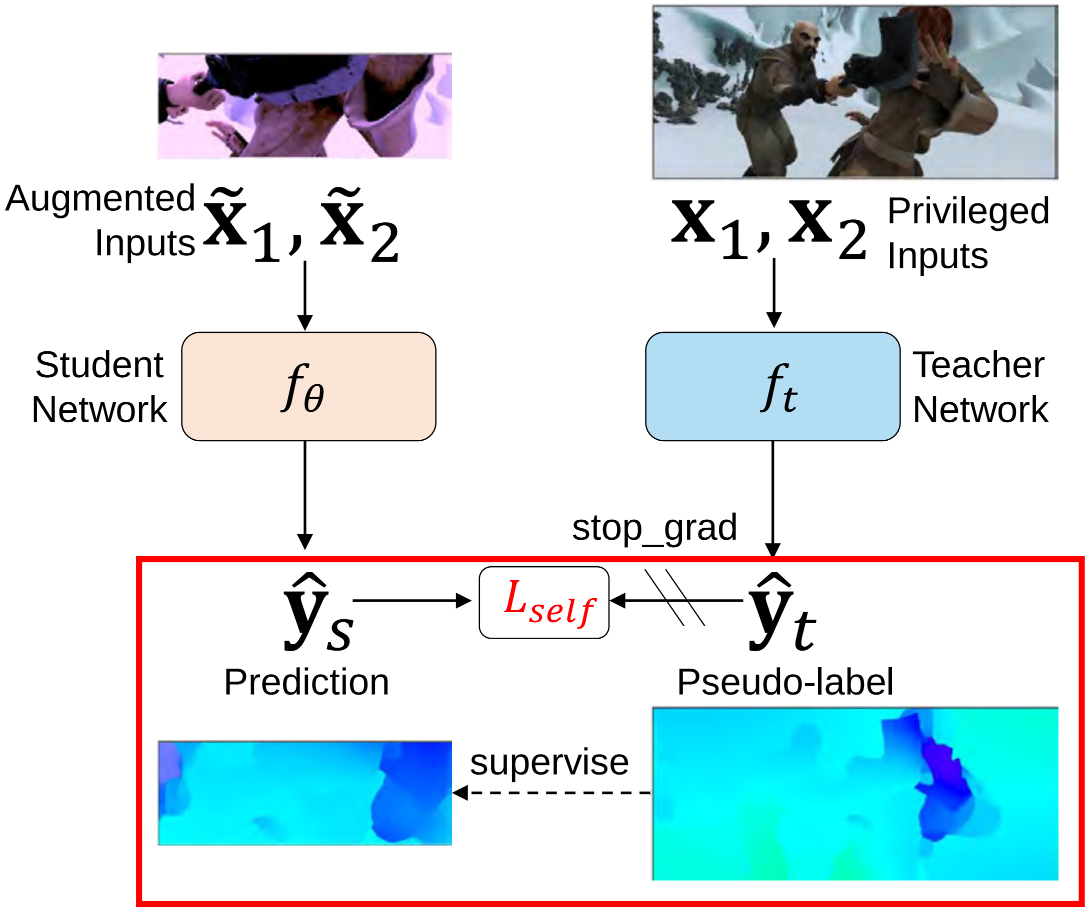


SMURF (CVPR 2022)

SGVR Lab
KAIST

Self-Supervision Loss for Optical Flow Learning

Self-supervision for optical flow



Privileged: No augmentation

Augmented:

color, cropping, erasing

i.e., privileged distillation
(Unifying distillation and privileged information, ICLR 2016)

Supervision with teacher output

Part of figure brought from SMURF (CVPR 2021)

Fixed Teacher Approach

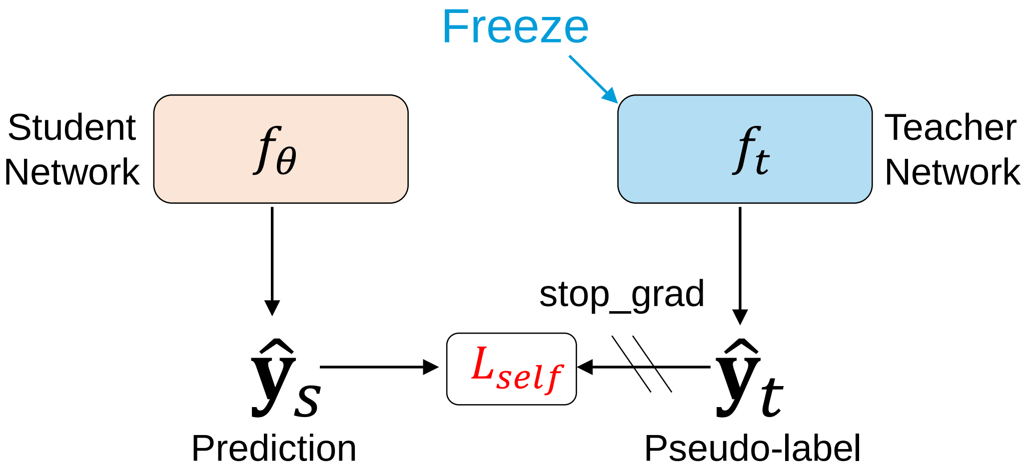
Things

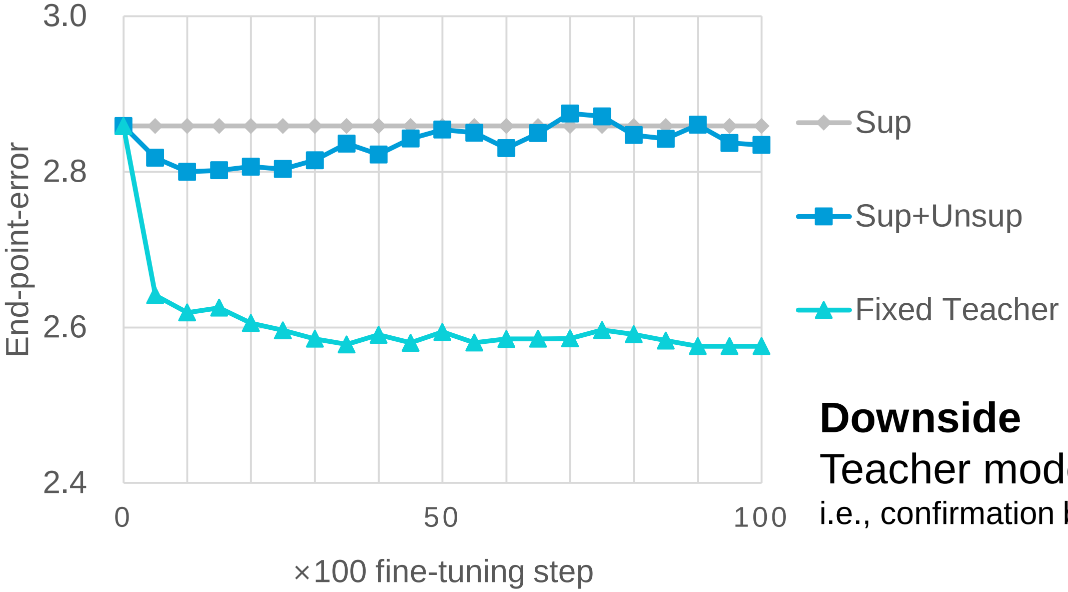
Sintel (unlabeled)

$$Loss = \ell_{sup} + \lambda \ell_{self}$$

Pixel-wise loss

Pixel-wise loss
(w/ pseudo-label)





Downside

Teacher model is not learned during training..

i.e., confirmation bias (MeanTeacher; NIPS 2017)

SGVR Lab KAIST

EMA (Moving Average) Approach

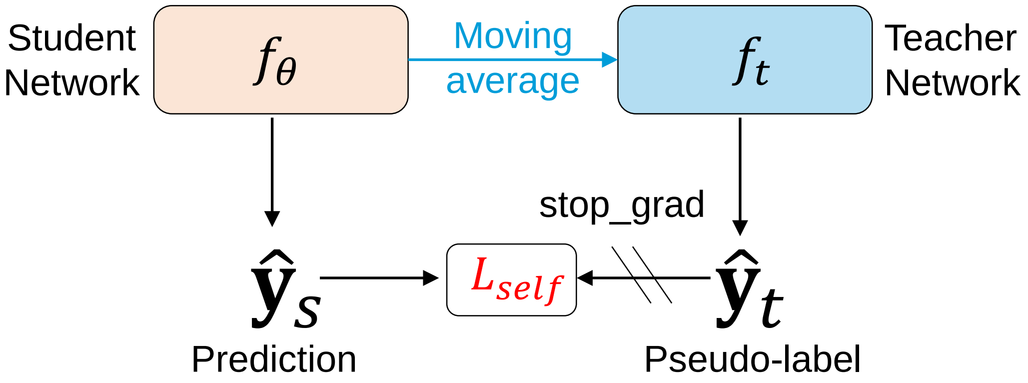
Things

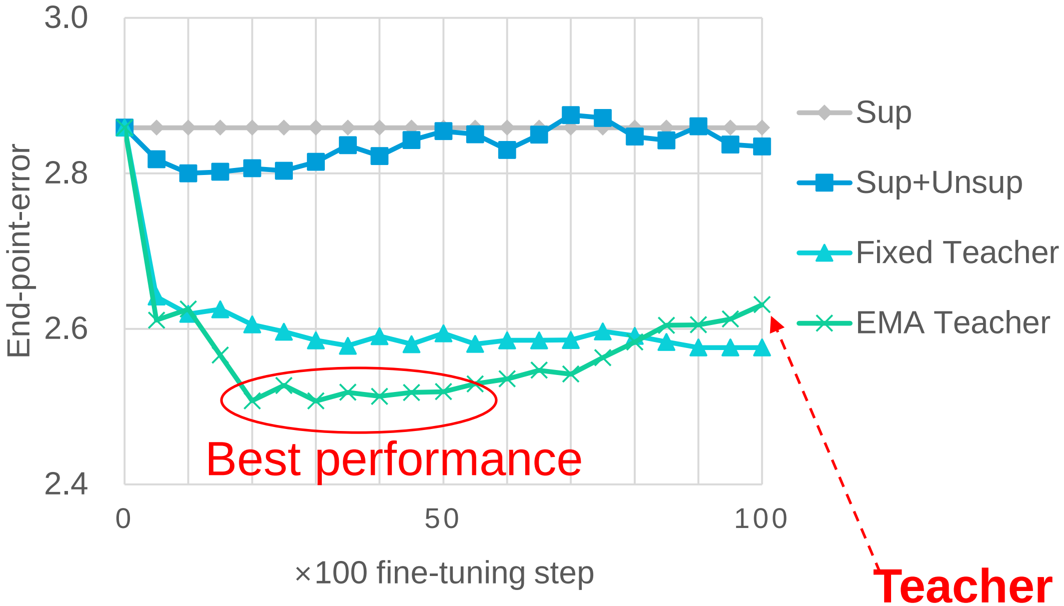
Sintel (unlabeled)

$$Loss = \ell_{sup} + \lambda \ell_{self}$$

Pixel-wise loss

Pixel-wise loss
(w/ pseudo-label)





Teacher is learned but unstable

FlowSupervisor (Ours)

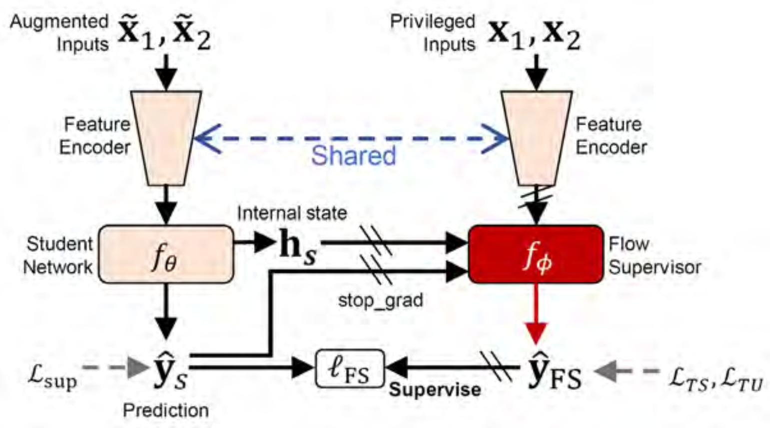
Things

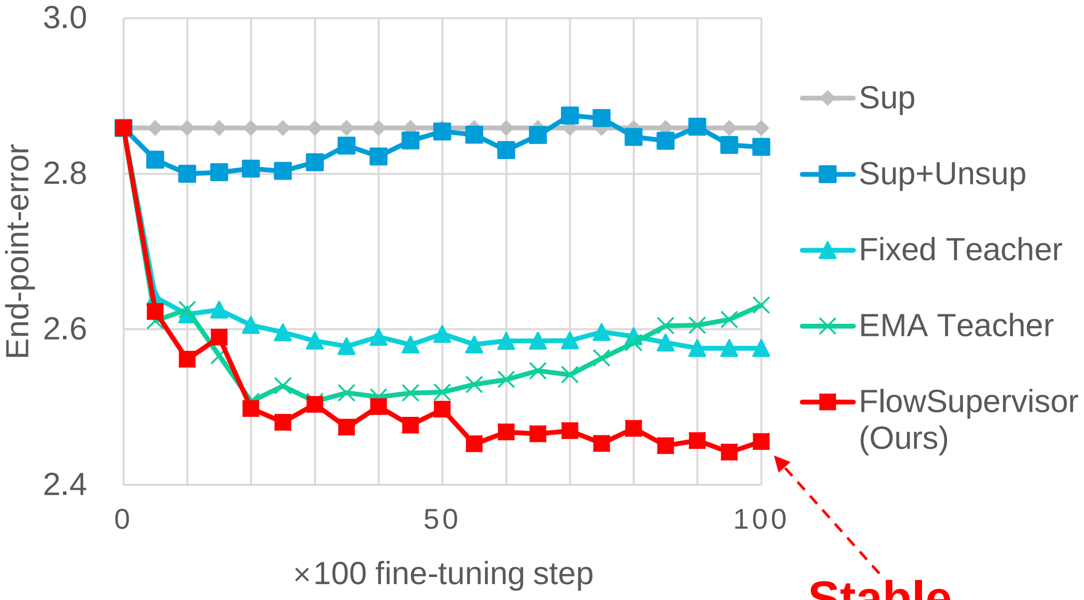
Sintel (unlabeled)

$$Loss = \ell_{sup} + \lambda \ell_{self}$$

Pixel-wise loss

Pixel-wise loss
(w/ pseudo-label)

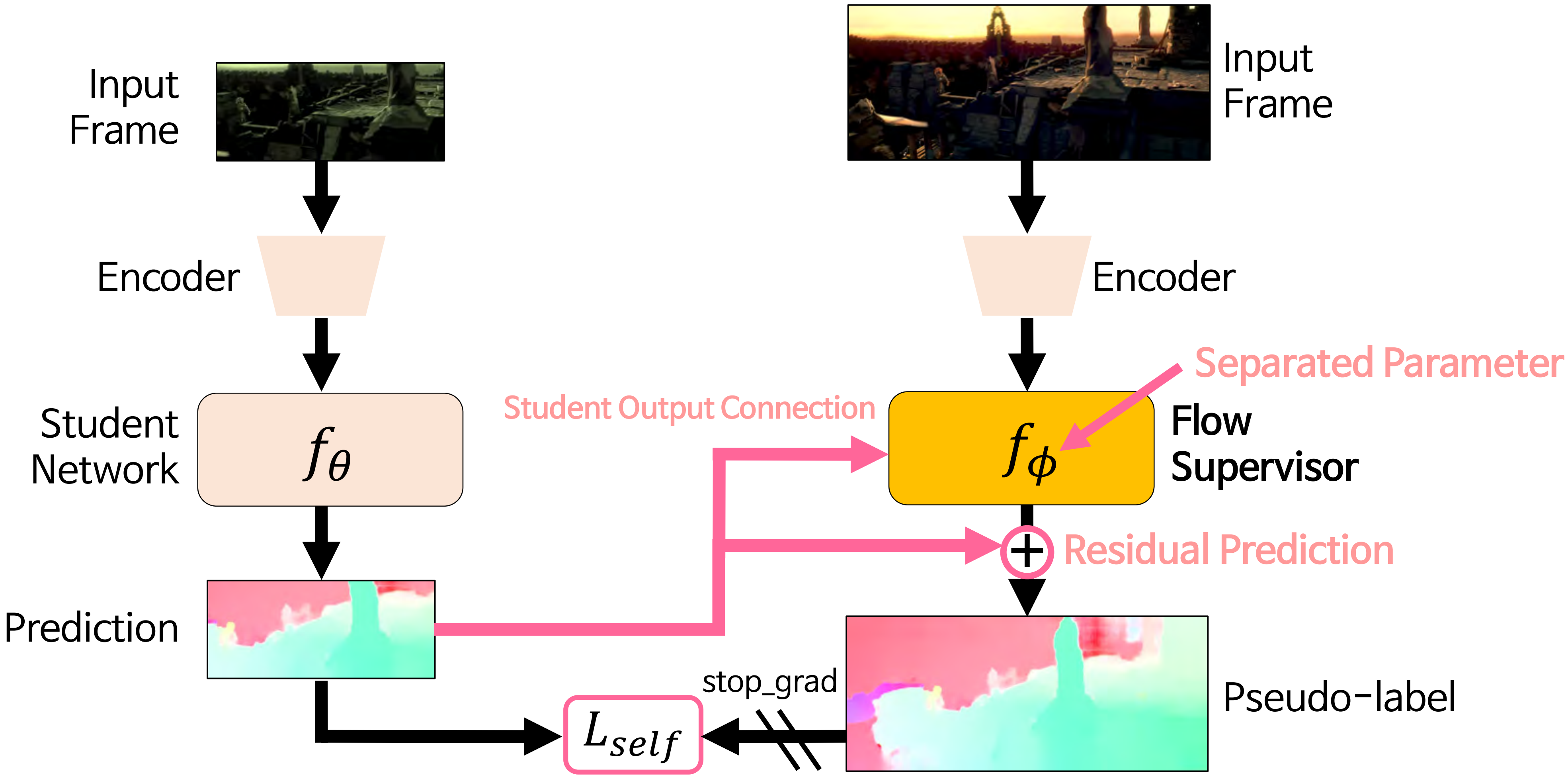




Stable
& best performance

SGVR Lab
KAIST

FlowSupervisor (Ours)



Comparison with Supervised Methods

W/ Label	W/O Label	Method	Sintel		KITTI	
			Clean	Final	EPE	Fl (%)
C+T	-	RAFT	1.46	2.80	5.79	18.8
C+T	S/K	**FlowSupervisor (RAFT)**	**1.30**	**2.46**	**3.35**	**11.12**

C: FlyingChairs

T: FlyingThings

S: Sintel

K: KITTI Multiview

Comparison with Supervised Methods

W/Label	W/O Label	Method	Sintel		KITTI	
			Clean	Final	EPE	Fl (%)
C+T	-	RAFT (ECCV 2020)	1.46	2.80	5.79	18.8
		GMA (CVPR 2021)	**1.30**	2.74	4.69	17.1
		SeparableFlow (CVPR 2021)	**1.30**	2.59	4.60	15.9
	S/K	**FlowSupervisor (RAFT)**	**1.30**	**2.46**	**3.35**	**11.12**

C: FlyingChairs

T: FlyingThings

S: Sintel

K: KITTI Multiview

Comparison with Supervised Methods

W/Label	W/O Label	Method	Sintel		KITTI	
			Clean	Final	EPE	Fl (%)
C+T	-	RAFT (ECCV 2020)	1.46	2.80	5.79	18.8
		GMA (CVPR 2021)	**1.30**	2.74	4.69	17.1
		SeparableFlow (CVPR 2021)	**1.30**	2.59	4.60	15.9
	S/K	**FlowSupervisor (RAFT)**	**1.30**	**2.46**	**3.35**	**11.12**
C+T+V	-	SeparableFlow (CVPR 2021)	-	-	2.60	7.74
	K	**FlowSupervisor (RAFT)**	-	-	**2.39**	**7.63**

C: FlyingChairs

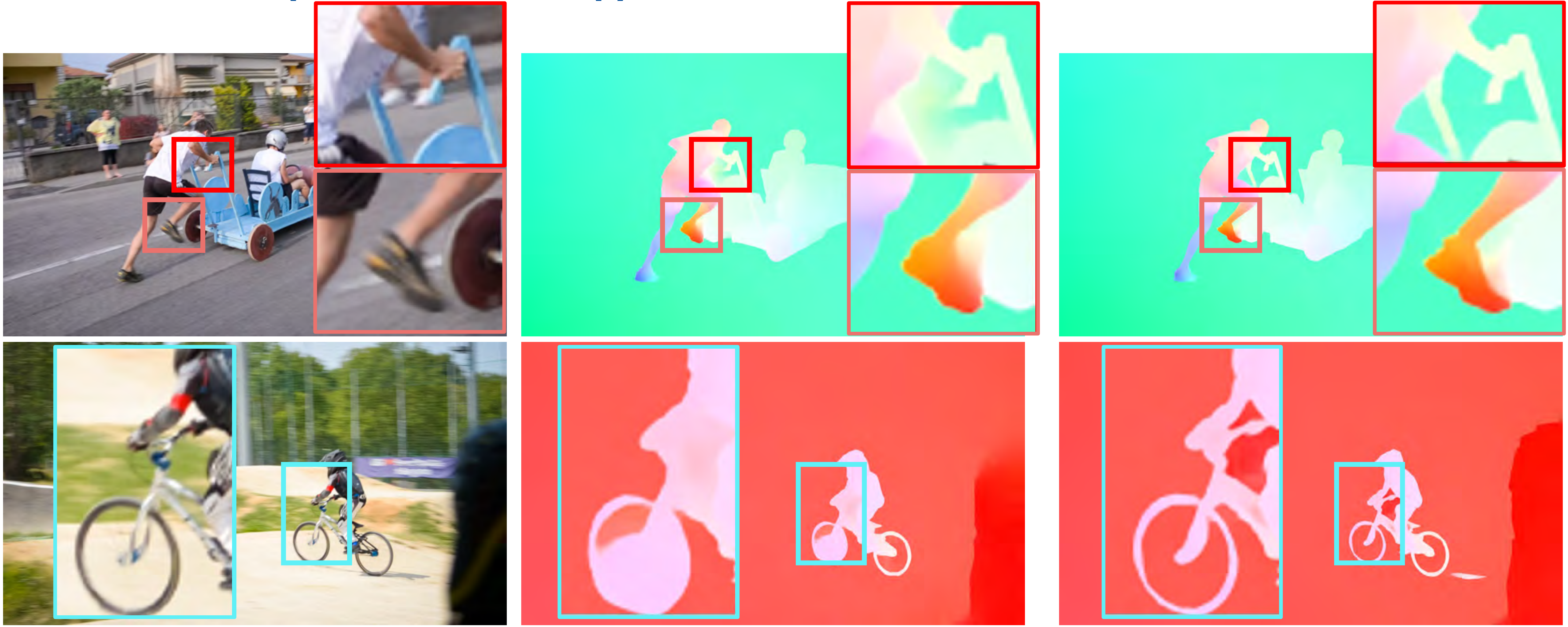
T: FlyingThings

S: Sintel

K: KITTI Multiview

V: Virtual KITTI

DAVIS dataset (real-world, 1080p)



Input

C+T+S+K+H
(Supervised-only)

C+T+S+K+H
(**Semi-supervised**)

DAVIS dataset (real-world, 1080p)

Input

C+T+S+K+H
(Supervised-only)

C+T+S+K+H
(**Semi-supervised**)

We’ve learned...

SGVR Lab
KAIST

- **What is optical flow?**
  - Pixel-level dense matching within a brief time frame.
- **Deep Optical Flow**
  - Fast, accurate optical flow
- **Unsupervised Deep Optical Flow**
  - Learn deep optical flow without ground truth
- **Semi-Supervised Optical Flow**
  - Use existing ground truth with free videos as training set

Q&A