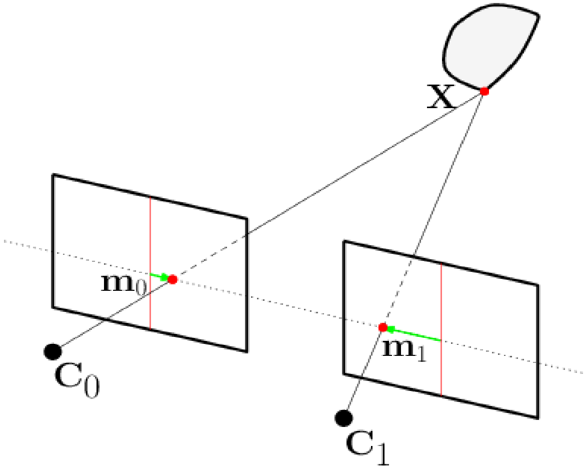
Dense correspondences estimation

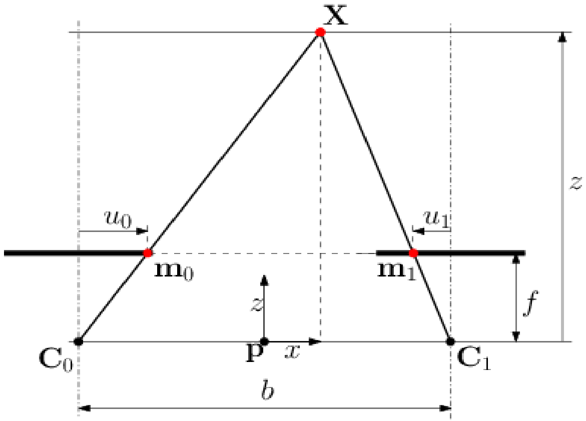
optical flow, disparity & scene flow

Michal Neoral

Disparity (Stereo)

- **Problem**:
  - Densely find correspondences between **two frames captured at same time** (two cameras)
  - 1D problem with calibrated cameras - estimate difference along epipolar line = disparity
- **Motivation**:
  - observed scene 3D reconstructure





- Direct estimation of depth **z** from disparity **d** baseline **b** and focal length **f**:

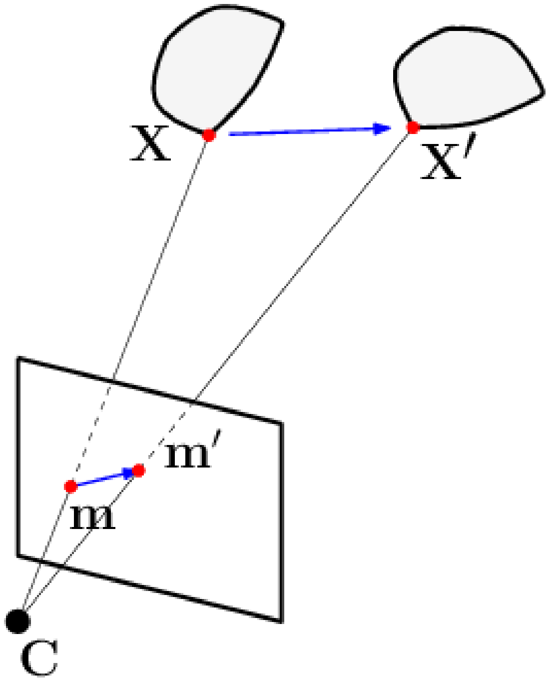
$$z = \frac{bf}{d}$$

- More detailed in subject TDV next semester

Optical flow

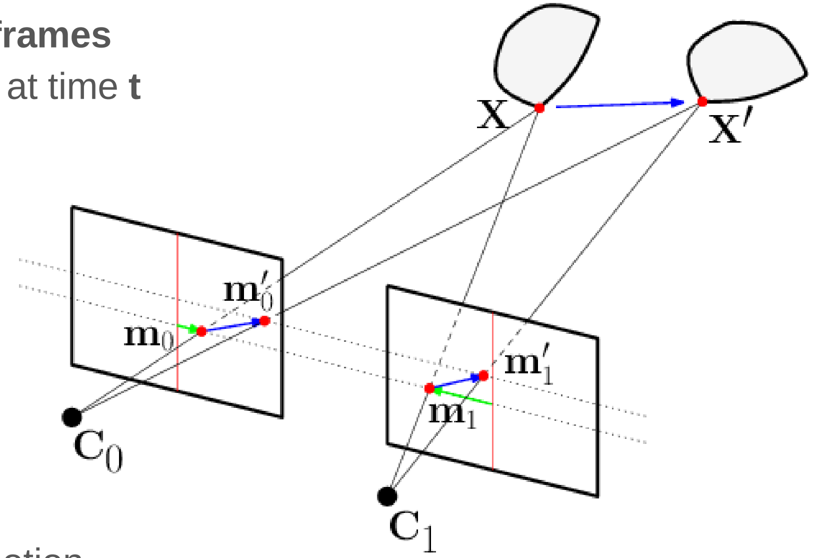
- **Problem**:
  - densely find correspondences between **two frames captured at different time** (frames from video at time **t** and **t'**)
  - 2D problem - displacement **p = (du, dv)**

- **Motivation**:
  - low-level motion cues in observed scene
  - robotics, autonomous driving, video processing
  - frame rate-up conversion, motion segmentation and estimation, structure from motion



Scene flow

- **Problem**:
  - Densely estimate correspondences from **two stereo-frames captured at different time** (frames from stereo-video at time **t** and **t'**) - 3D motion vector for each visible 3D point

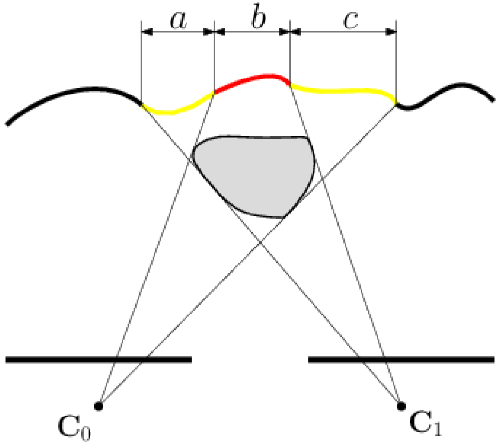


- **Motivation**:
  - estimate motion in 3D directly and more precisely in comparison with OF
  - robotics, autonomous driving
  - motion segmentation and estimation, structure from motion

Challanges - why is problem so hard?

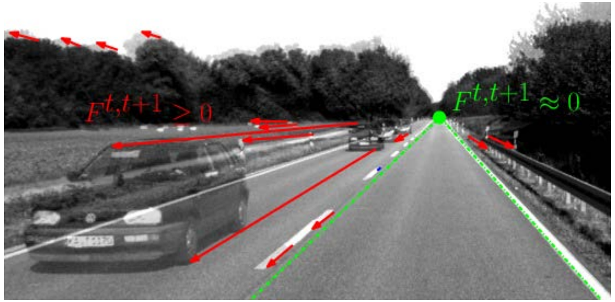
Large displacements

Occlusions

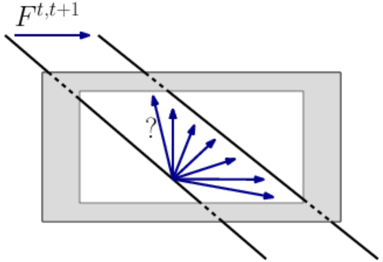


Illumination conditions

Focus of expansion problem



Aperture problem



Repetitive patterns

Optical Flow

Standard methods

- **Input -** two consecutive images from different time steps
- **Output -** 2D vector field - estimated displacement for each pixel
- **Standard approaches - variational**:
  - based on Horn-Schunck [**1**]
  - **global** method for **dense** OF (Lucas-Kanade [**2**] is local and sparse)
  - energy minimisation (optimisation) of

$$E = \iint D\left(\mathbf{I}_2(\mathbf{x} + \partial\mathbf{x}), \mathbf{I}_1(\mathbf{x})\right) + S(\nabla u(\mathbf{x}), \nabla v(\mathbf{x}))\ d\mathbf{x}$$

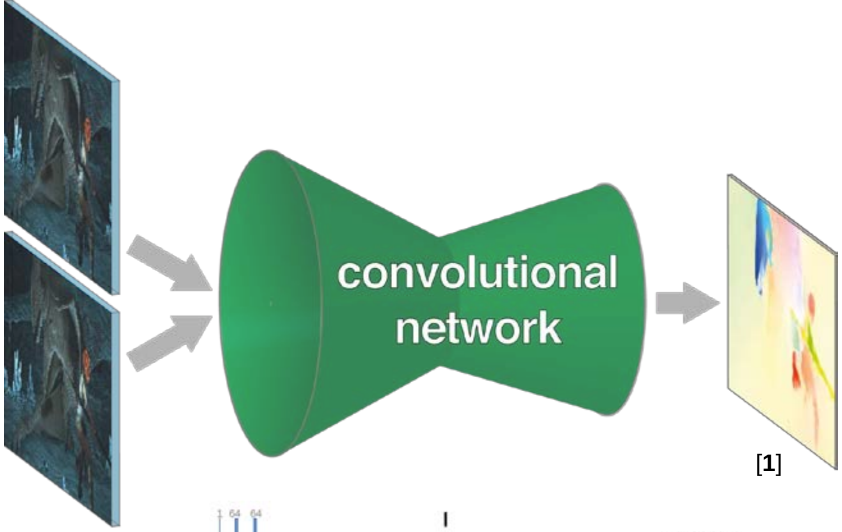
where data term *D* describes how close are pixels **x** from reference image $\mathbf{I}_1$ and pixels **x + dx** from $\mathbf{I}_2$ in term of appearance, smoothness term *S* describes how similar is flow in neighborhood pixels

- Other approaches:
  - Discrete energy MRF, Patch based (super-pixels), Region growing [Cech-CVPR-2011]

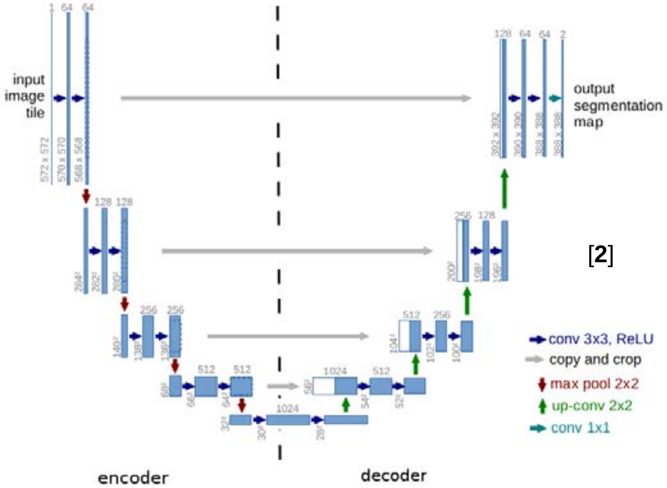
[**1**] Horn, B. K. & Schunck, B. G. Determining optical flow **1981**
[**2**] Lucas, B. D.; Kanade, T. & others An iterative image registration technique with an application to stereo vision **1981**
[**3**] Butler, D. J.; Wulff, J.; Stanley, G. B. & Black, M. J. A. Fitzgibbon et al. (Eds.) (Ed.) A naturalistic open source movie for optical flow evaluation ECCV **2012**

CNN methods

- Fully convolutional
- Most common architecture - U-Net [**2**]
  - Skip connections between encoder-decoder
- In general without max-pooling, batch normalisation or dropout
- Supervised versions:
  - small number of real-world OF ground-truth KITTI, HD1C
  - learning on huge number of CG images FlyingChairs, FlyingThings, Sintel, VirtualKITTI, ...



[**1**]



[**2**]

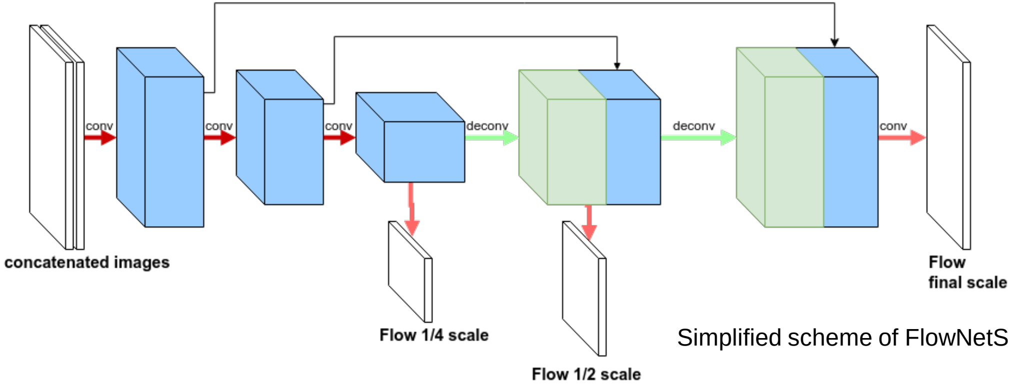
[**1**] Dosovitskiy, A.; Fischer, P.; Ilg, E.; Häusser, P.; Hazibas;, C.; Golkov, V.; van der Smagt, P.; Cremers, D. & Brox, T. FlowNet: Learning Optical Flow with Convolutional Networks, *ICCV* **2015**
[**2**] Ronneberger, O.; Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation MICCAI **2015**

FlowNet

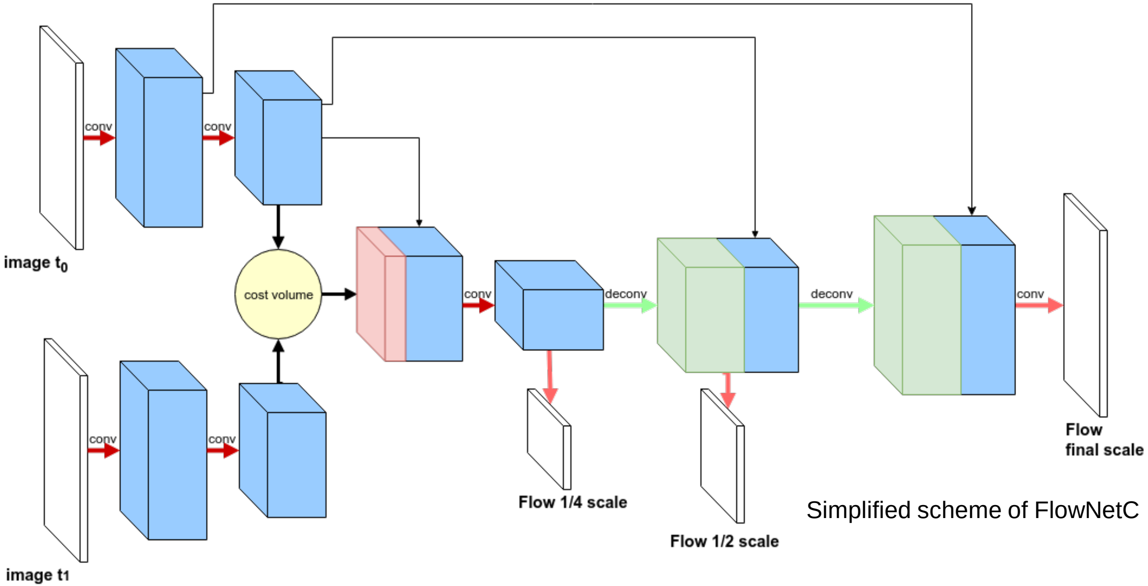
- FlowNet is first end-to-end CNN method for OF
- supervised method
- Training on CGI FlyingChairs, then fine tuned on KITTI or Sintel
- Two versions - simple and correlation

$$\mathbf{cv}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{N}\mathbf{x}_1^{\mathrm{T}} \cdot \mathbf{x}_2$$

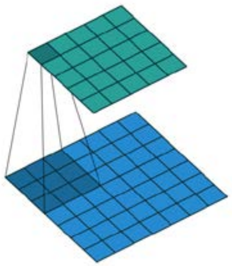
- Not achieved state-of-the-art accuracy, but very fast (50 FPS)



Simplified scheme of FlowNetS



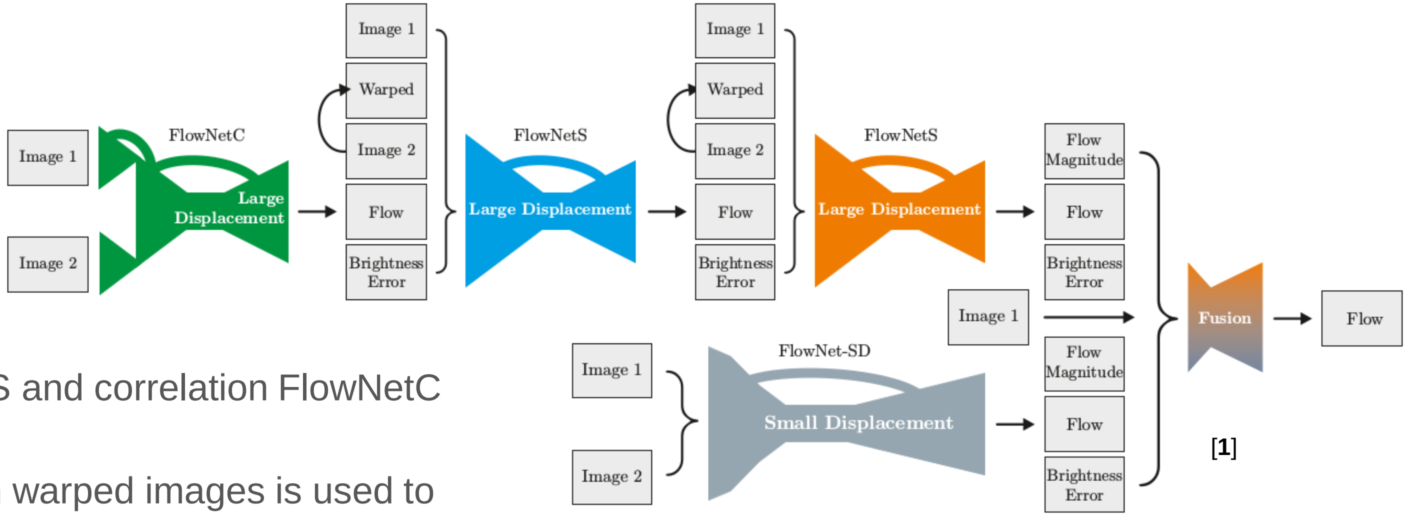
Simplified scheme of FlowNetC



Deconvolution [2]
input, output

[1] Dosovitskiy, A.; Fischer, P.; Ilg, E.; Häusser, P.; Hazibas;, C.; Golkov, V.; van der Smagt, P.; Cremers, D. & Brox, T. FlowNet: Learning Optical Flow with Convolutional Networks, *ICCV* **2015**
[2] Dumoulin, V. & Visin, F. A guide to convolution arithmetic for deep learning arXiv preprint arXiv:1603.07285, 2016

FlowNet 2.0

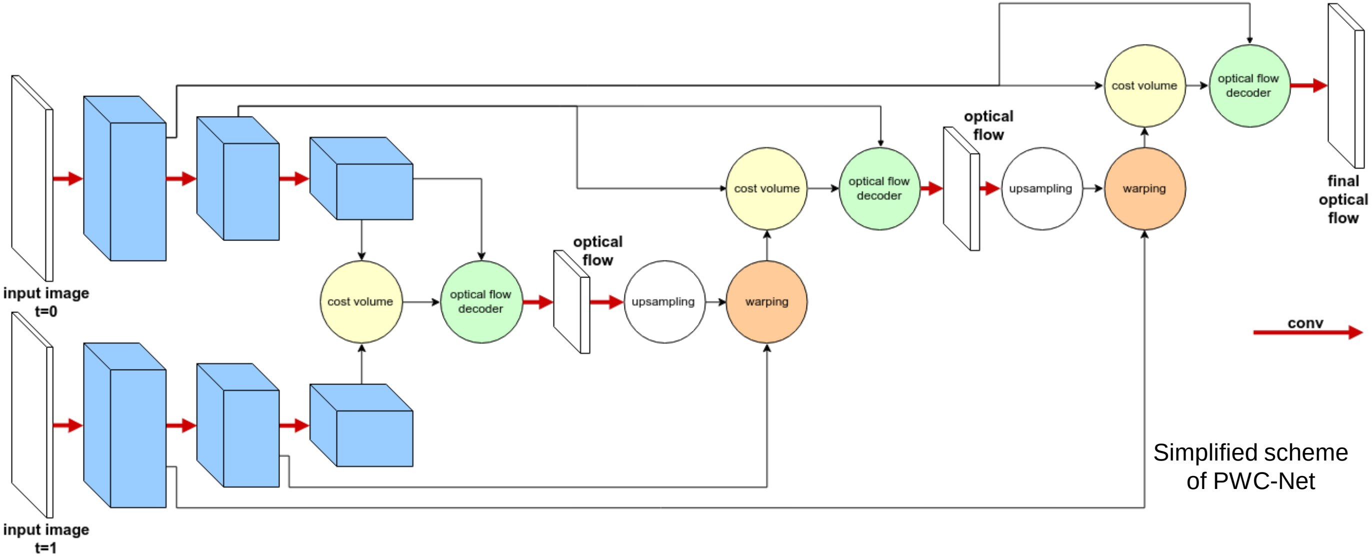


[1]

- several simple FlowNetS and correlation FlowNetC stacked in one network
- uses brightness error on warped images is used to correct estimated OF in iterative part of a network
- specialised part for small displacements
- comparable accuracy with SOTA standard approaches, but several times faster (15 FPS)

[**1**] Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A. & Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks CoRR, **2016**

PWC-Net



Simplified scheme of PWC-Net

- **pyramid**, **warping**, **cost**
- current state-of-the-art
- 30 FPS, small number of parameters (up to 9 millions)
- classical “**coarse-to-fine**” principle of dense estimation within CNN
  - features **pyramid** are extracted from both input images with shared weights features extractor
  - estimated optical flow from coarser level is upsampled and used for **warping** features to reference view
  - **cost** volume layer is computed as correlation of reference view features and warped features
  - optical flow is estimated using convnet from cost layer and reference features
  - used for each scale

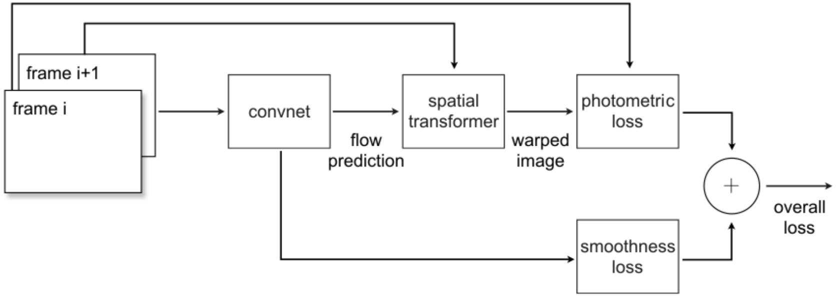
[**1**] Sun, D.; Yang, X.; Liu, M.-Y. & Kautz, J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume *CVPR* **2018**

Self-supervised flow

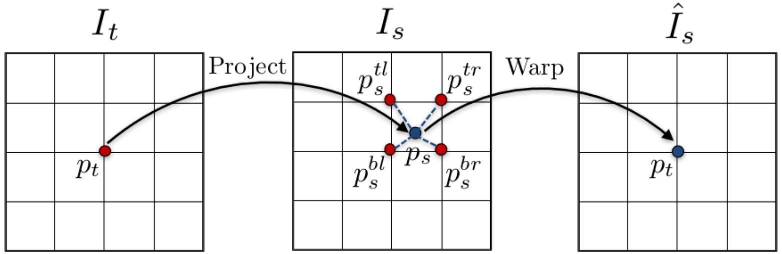
- needs only monocular video sequence with **no GT**
- changes in loss, architecture remains the same
  - estimated flow ***F*** between source and target image is used to warp (back-project) source to target
  - bilinear interpolation - differentiable
- loss - analogical to standard variational methods

$$\mathcal{L} = \mathcal{L}_{\mathrm{p}}\left(I_s, \hat{I}_s\right) + \alpha \mathcal{L}_{\mathrm{s}}\left(\hat{F}\right)$$

  - $\mathbf{L_p}$ photometric loss - minimise difference between source and target
  - $\mathbf{L_s}$ smoothness loss - avoid huge artificial steps between neighbor pixels



Scheme of self-supervised loss learning [**1**]



Bilinear interpolation [**2**]

[**3**]

target I_t

source I_s

estimated flow

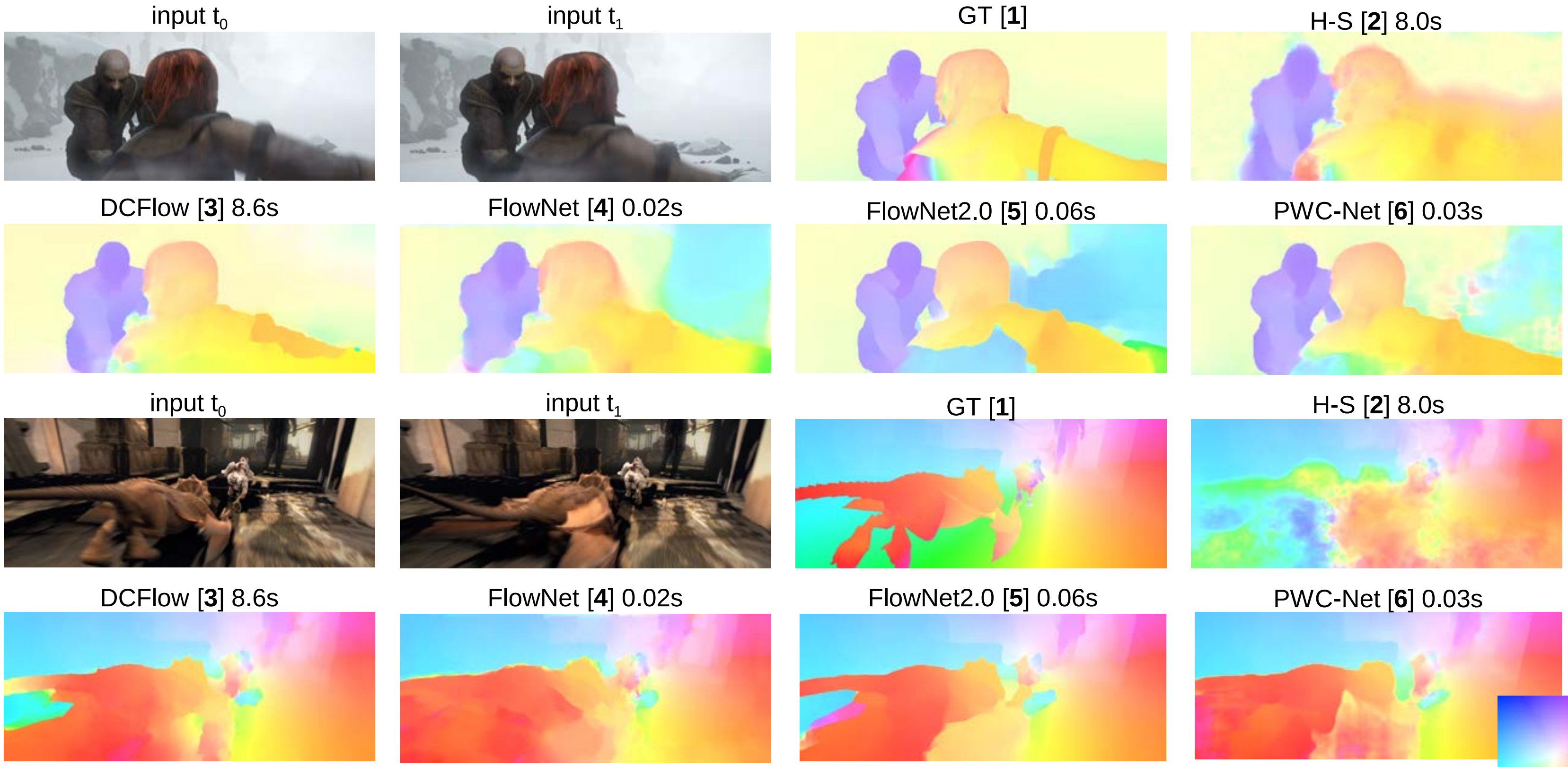
warped image

photometric loss

smoothness loss

[**1**] Yu, J. J.; Harley, A. W. & Derpanis, K. G. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness ECCV **2016**
[**2**] Zhou, T.; Brown, M.; Snavely, N. & Lowe, D. G. Unsupervised learning of depth and ego-motion from video arXiv **2017**
[**3**] Butler, D. J.; Wulff, J.; Stanley, G. B. & Black, M. J. A. Fitzgibbon et al. (Eds.) (Ed.) A naturalistic open source movie for optical flow evaluation ECCV **2012**

Comparison of OF algorithms



[**1**] Butler, D. J.; Wulff, J.; Stanley, G. B. & Black, M. J. A. Fitzgibbon et al. (Eds.) (Ed.) A naturalistic open source movie for optical flow evaluation ECCV **2012**
[**2**] Horn, B. K. & Schunck, B. G. Determining optical flow **1981**
[**3**] Xu, J.; Ranftl, R. & Koltun, V. Accurate optical flow via direct cost volume processing CVPR **2017**
[**4**] Dosovitskiy, A.; Fischer, P.; Ilg, E.; Häusser, P.; Hazibas;, C.; Golkov, V.; van der Smagt, P.; Cremers, D. & Brox, T. FlowNet: Learning Optical Flow with Convolutional Networks, *ICCV* **2015**
[**5**] Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A. & Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks CoRR, **2016**
[**6**] Sun, D.; Yang, X.; Liu, M.-Y. & Kautz, J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume *CVPR* **2018**

Current trends in CNN optical flow estimation

- incorporating knowledge from 40 years of computer vision into CNN
  - better architecture choices
  - PWC-Net as example
- multi-task learning
  - (even loosely) related tasks learning together achieve better results than learning separately
  - Semantic and Flow, Segmentation and Flow
- temporal consistency
  - using more than two consecutive images or using flow estimated in previous frames