Vision and Perception

Deep Learning for Video: Motion estimation

SAPIENZA
UNIVERSITÀ DI ROMA

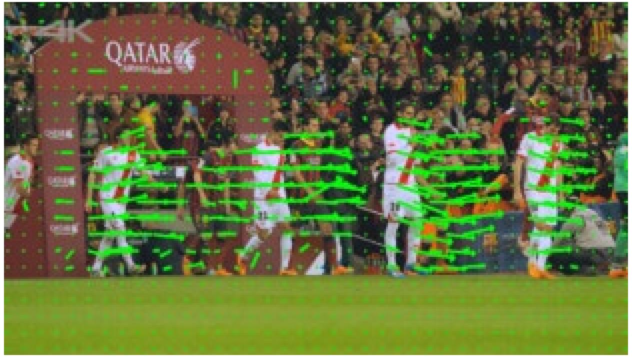
References

- Basic reading: Szeliski, Chapter 9.3

Optical Flow

- Motion estimation in video
- "Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image." [1]
- For subsequent frames, determine displacement vector for each pixel

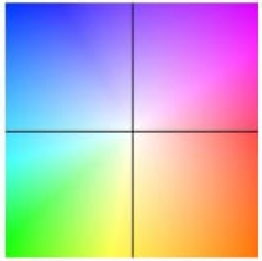
- Possible applications: visual odometry, autonomous driving, semantic segmentation…
→ Whenever motion conveys useful information

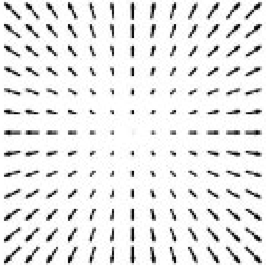


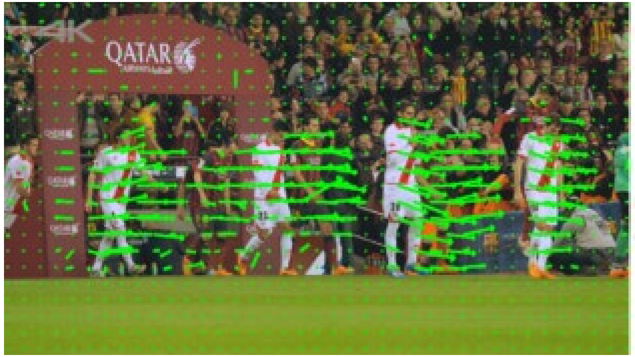
[1] Horn, Berthold KP, And Brian G. Schunck. "Determining Optical Flow." Artificial Intelligence 17.1-3 (1981): 185-203. https://devblogs.nvidia.com/an-introduction-to-the-nvidia-optical-flow-sdk/, retrieved 18.11.2019

Color code visualization

- Colour code for visualisation:







Baghaie, Ahmadreza, Roshan D’Souza, and Zeyun Yu. "Dense descriptors for optical flow estimation: a comparative study." Journal of Imaging 3.1 (2017)
https://devblogs.nvidia.com/an-introduction-to-the-nvidia-optical-flow-sdk/

Optical flow with deep neural networks

- “Classical” approaches: complex optimization problems, computationally expensive
→ Not suitable for real-time applications

- First DNN approaches: trade-off between accuracy and size of the model

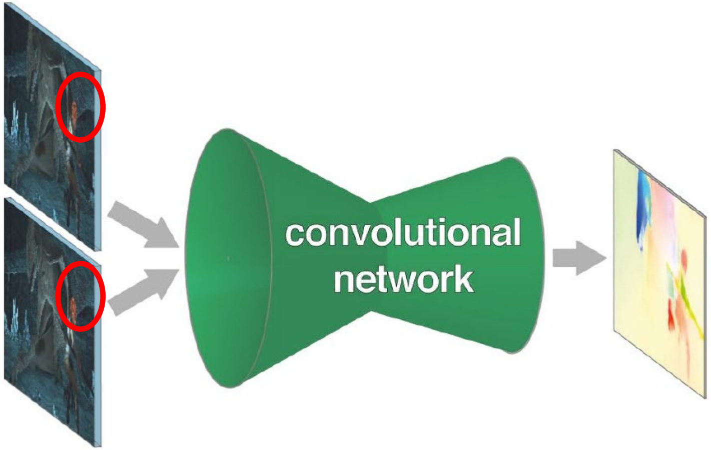
FlowNet

End to end supervised learning for optical flow estimation:

- encoder/decoder architecture

The idea: given two consecutive frames need to match features at different location in input images

The network has to find if an object in the first image is at another location in the second images

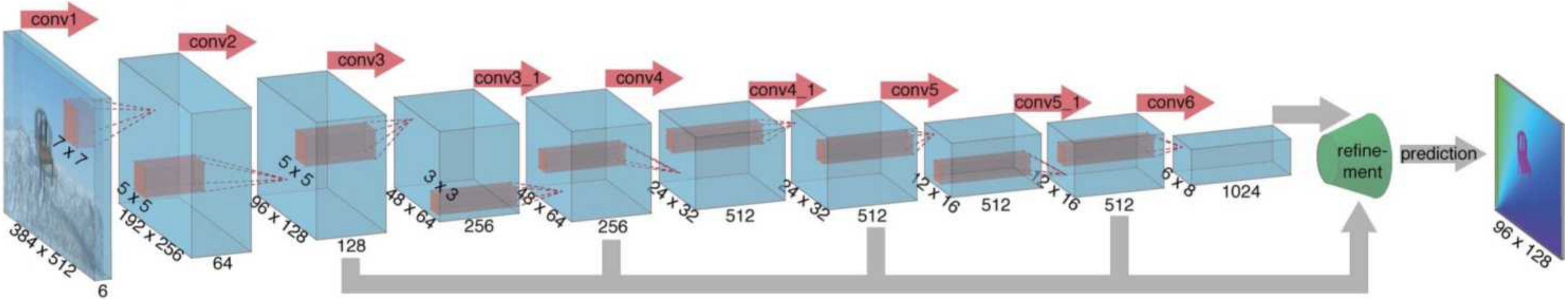


Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van derSmagt, P., Cremers, D. and Brox, T., FlowNet: Learning Optical Flow With Convolutional Networks. ICCV 2015

FlowNet (encoder)

Option A: stack both input images together and feed them through a generic network.

FlowNetSimple

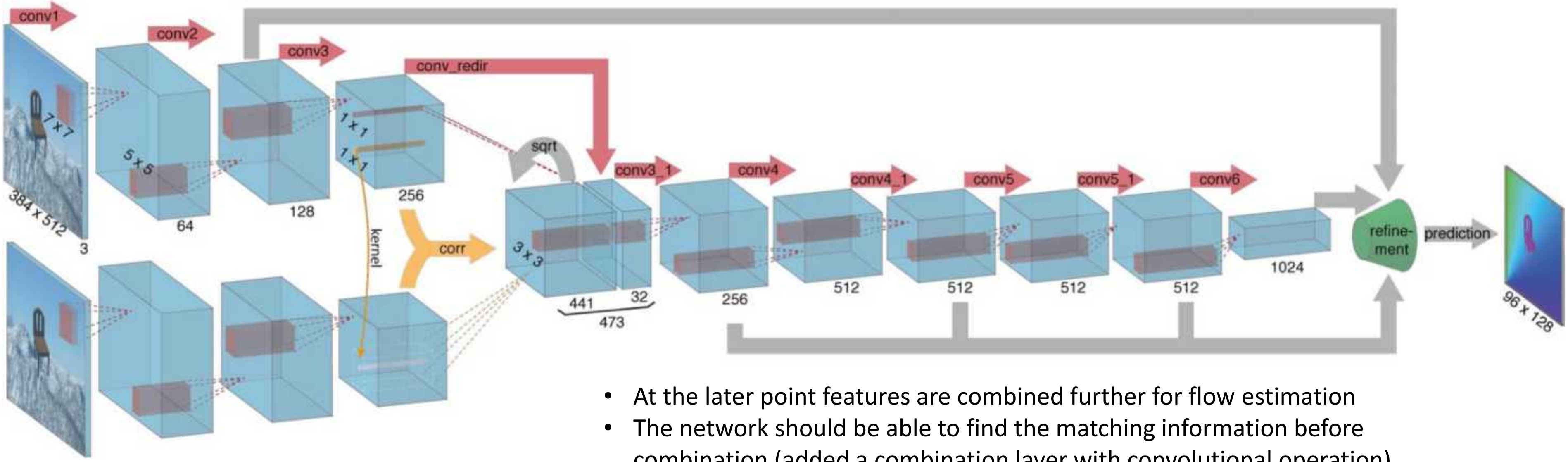


Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van derSmagt, P., Cremers, D. and Brox, T., FlowNet: Learning Optical Flow With Convolutional Networks. ICCV 2015

FlowNet (encoder)

Option B: create two separate, yet identical processing streams for the two images and combine them at a later stage.

FlowNetCorr



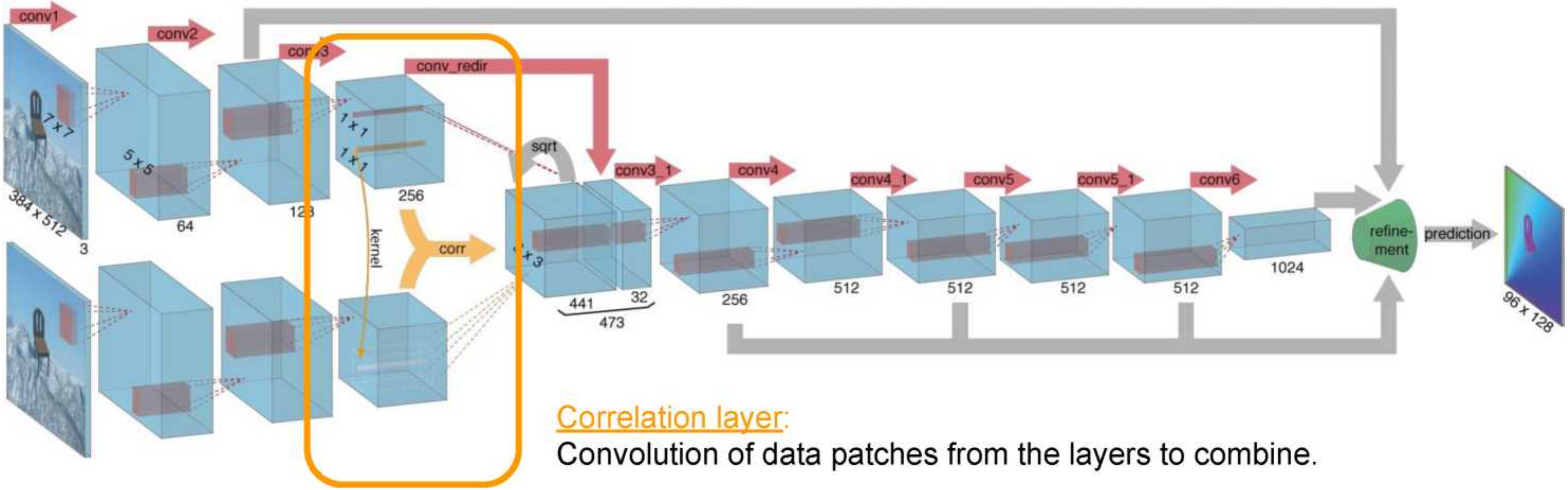
- At the later point features are combined further for flow estimation
- The network should be able to find the matching information before combination (added a combination layer with convolutional operation)
- The features are skip connected to the refinement layer

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van derSmagt, P., Cremers, D. and Brox, T., FlowNet: Learning Optical Flow With Convolutional Networks. ICCV 2015

FlowNet (encoder)

Option B: create two separate, yet identical processing streams for the two images and combine them at a later stage.

FlowNetCorr



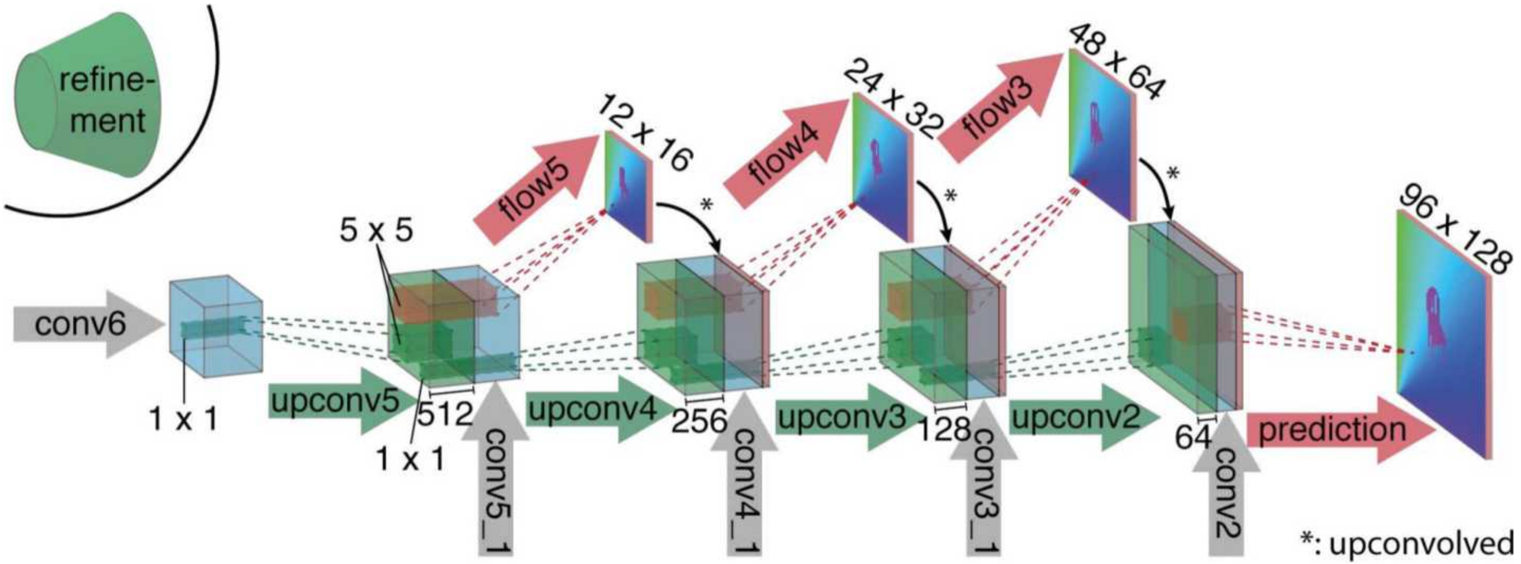
Correlation layer:
Convolution of data patches from the layers to combine.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van derSmagt, P., Cremers, D. and Brox, T., FlowNet: Learning Optical Flow With Convolutional Networks. ICCV 2015

FlowNet (decoder)

Upconvolutional layers: Unpooling features maps + convolution.

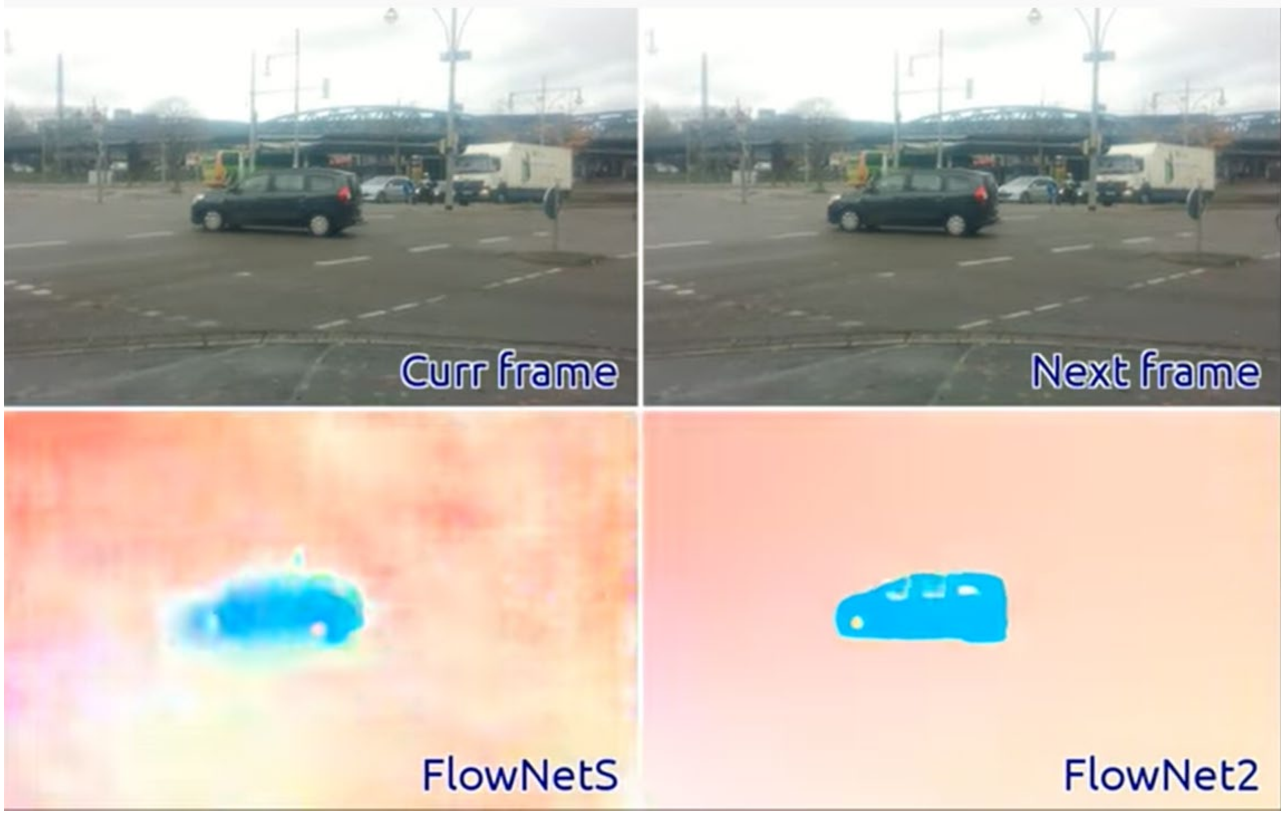
Upconvolutioned feature maps are concatenated with the corresponding map from the contractive part.



- Features map are scaled to the original resolution
- Upconvolutional layer (unpooling and convolution) doubling the resolution
- Concatenate the ‘upconvolution’ results with the features from the ‘contractive’ part of the network
- Bilinear upsampling (variational upsampling) in the last stage

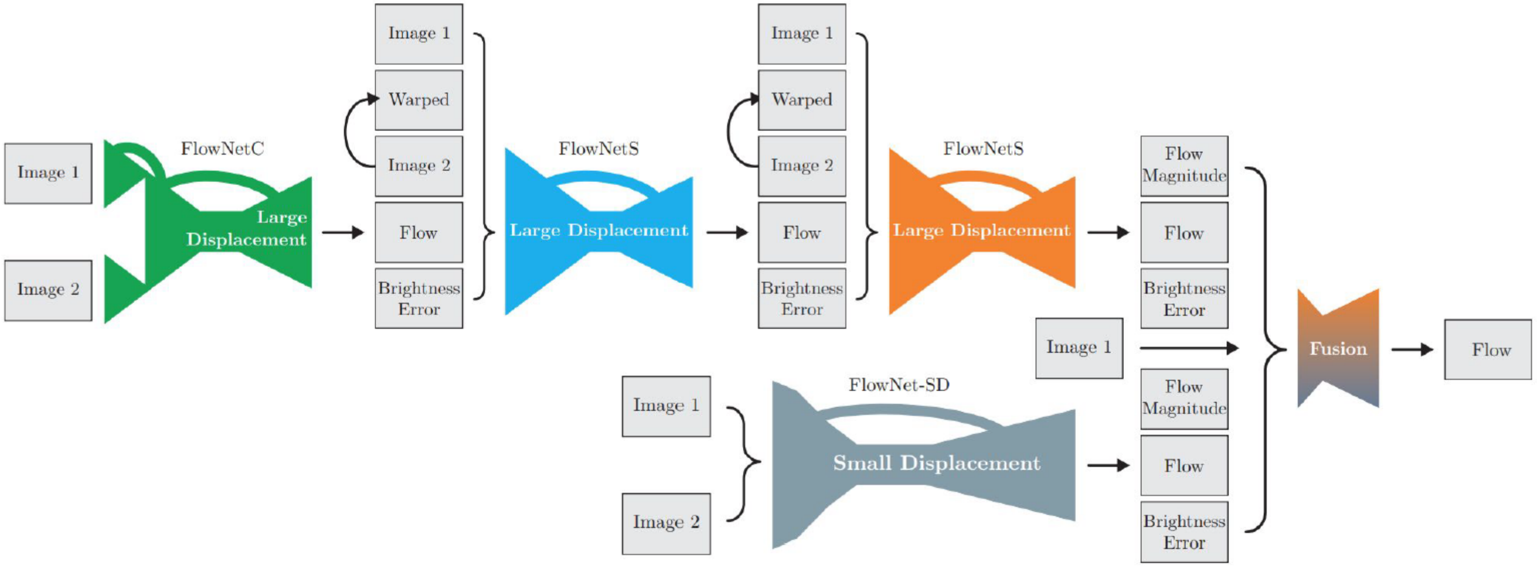
Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van derSmagt, P., Cremers, D. and Brox, T., FlowNet: Learning Optical Flow With Convolutional Networks. ICCV 2015

FlowNet 2.0



Ilg, Eddy, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. "Flownet2.0: Evolution of optical flow estimation with deep networks." CVPR 2017.

Optical Flow: FlowNet 2.0



Ilg, Eddy, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. "Flownet2.0: Evolution of optical flow estimation with deep networks." CVPR 2017. [code]

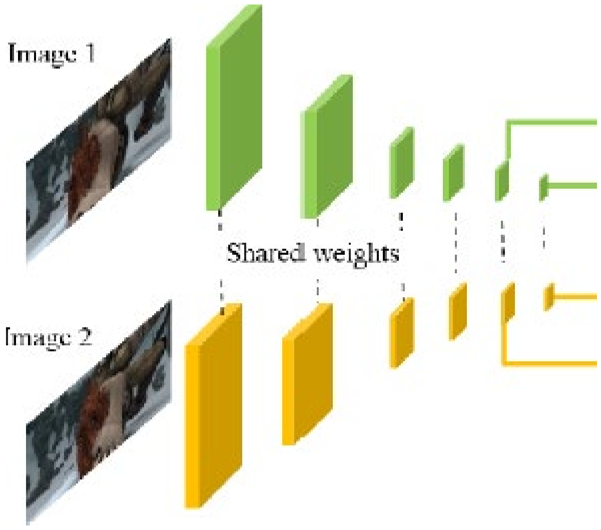
PWC-Net (Pyramid, Warping, Cost volume)

1. Feature extraction from input images with feature pyramid, i.e. convolutional layers
    - Reduction of spatial resolution

2. Optical flow estimation for every level of feature pyramid
    - Start with last convolutional layer, finish on input level
    - Warping and cost volume used in optical flow estimation

Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

PWC-Net



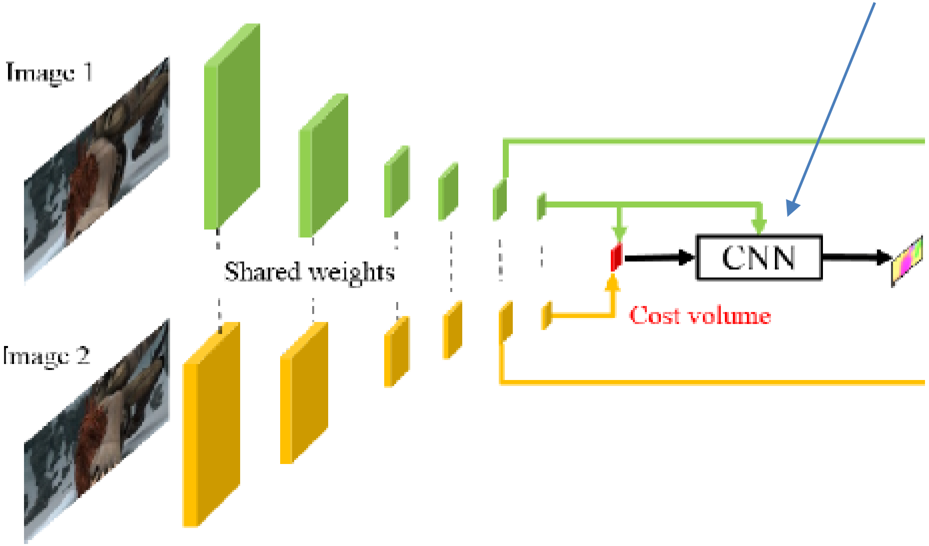
1. Compute cost volume: find most similar pixel in features w.r.t other image

- We can use normalized cross correlation to get the cost volume of a pixel
- Invariance to color change, not in scale so pyramidal processing
- Learning the features pyramid at lower scale we can learn global context to do the matching

Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

PWC-Net

Flow estimator



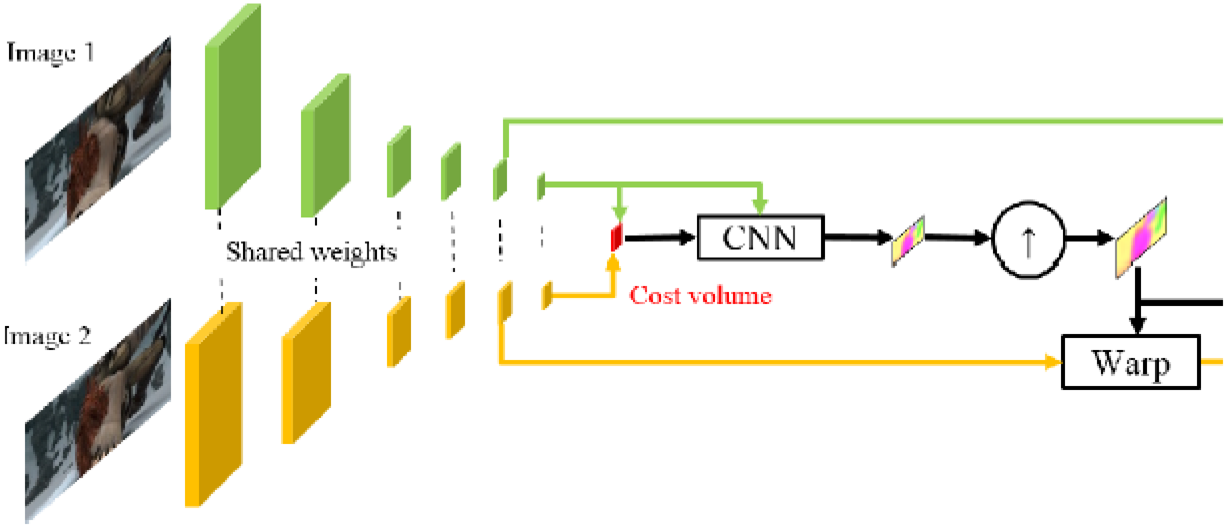
2. Optical flow estimation:

- cost volume
- features of the first image (image 1)

→Output OF for lowest level

Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

PWC-Net



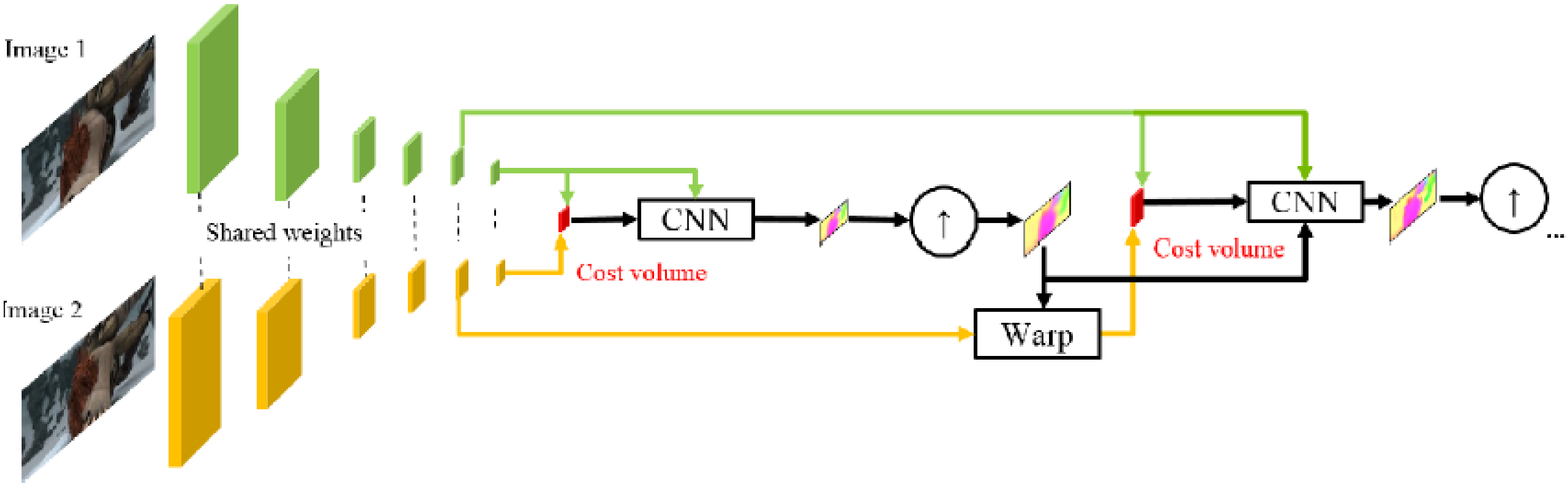
1. Upsample OF to match spatial dimensions
2. Warp the features of the 2nd image towards the features of the 1st image (using the estimated flow field)

Why warping?

- Second image becomes more similar to first image
- Pixel displacement becomes smaller

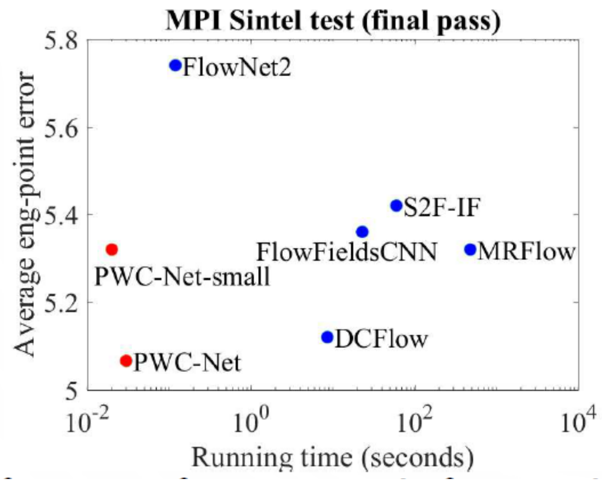
Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

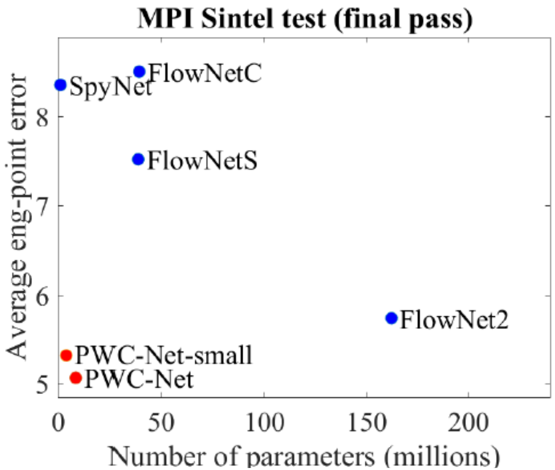
PWC-Net



Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

PWC-Net results





- State-of-the-art accuracy with end-to-end training
- Inference fast enough for real-time application
- PWC-Net-small for mobile applications

Code publicly available: https://github.com/NVlabs/PWC-Net

Sun, Deqing, et al. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
Sun, Deqing, et al. "Models Matter, So Does Training: An Empirical Study of CNNs for Optical Flow Estimation." arXiv preprint arXiv:1809.05571 (2018).

Acknowledgement: some slides and material from Bernt Schiele, Mario Fritz, Michael Black, Bill Freeman, Fei-Fei, Justin Johnson, Serena Yeung, R. Szelisky, Fabio Galasso, Ioannis Gkioulekas, Noah Snavely, Abe Davis, Kris Kitani, Xavier Giró-i-Nieto