Review optical flow papers

ZheweiMedia

1 7 papers in a row

1.1

Dosovitskiy, Alexey, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. **"Flownet: Learning optical flow with convolutional networks."** *In Proceedings of the IEEE International Conference on Computer Vision, pp. 2758-2766. 2015.*

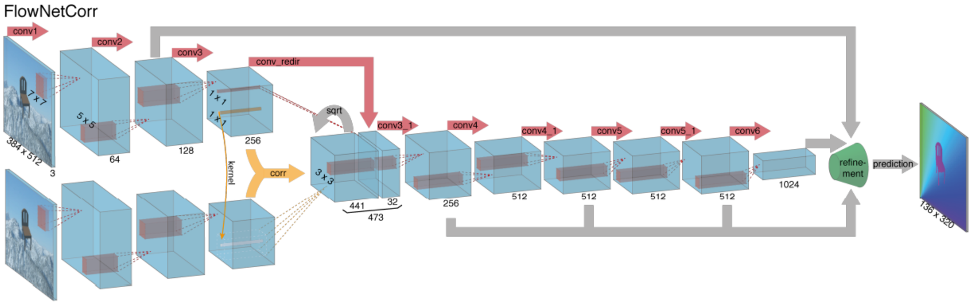


Figure 1: Whole Structure of Flownet

As shown in Fig. 1, two images go through two separate, yet identical branches to extract features, then a 'correlation layer' is applied, which is similiar to DeepMarching and can be done via convolution under CNN architecture, shown as Fig. 3.

The flow is generate on the highest feature, so refinement is needed, which is trivial in this paper and shown in Fig. 2.

Alternative upsampling methods from *Brox, Thomas, and Jitendra Malik. "Large displacement optical flow: descriptor matching in variational motion estimation." IEEE transactions on pattern analysis and machine intelligence 33, no. 3 (2011): 500-513.* are used, which improve the results.

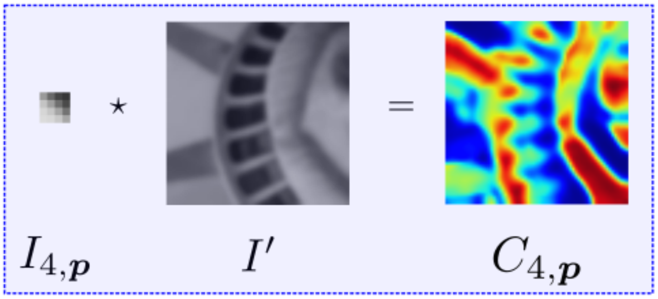


Figure 2: Correlation can be done as convolution

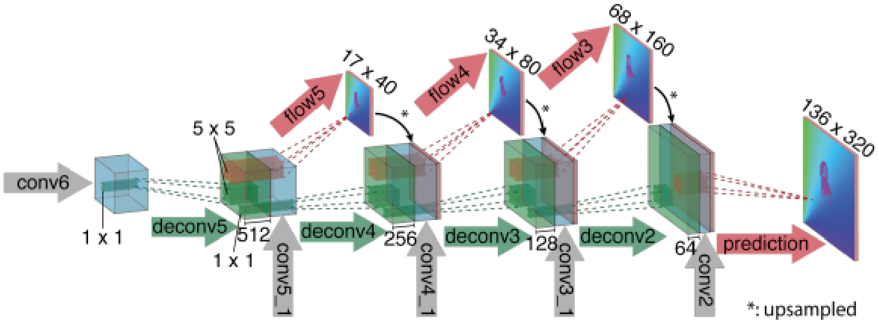


Figure 3: Correlation can be done as convolution

Data augmentation is also used.
Smooth function on the boundary is also used.

1.2

Ilg, Eddy, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. **"Flownet 2.0: Evolution of optical flow estimation with deep networks."** *arXiv preprint arXiv:1612.01925 (2016).*

The upper part of Fig. 4 is a residue network. Three contributions:
1. A learning schedule consisting of multiple datasets improves results significantly. Train on one dataset and fine-tune on anther achieves the best results.
2. Warping layer. Only stacking the network leads to over-fitting. This one

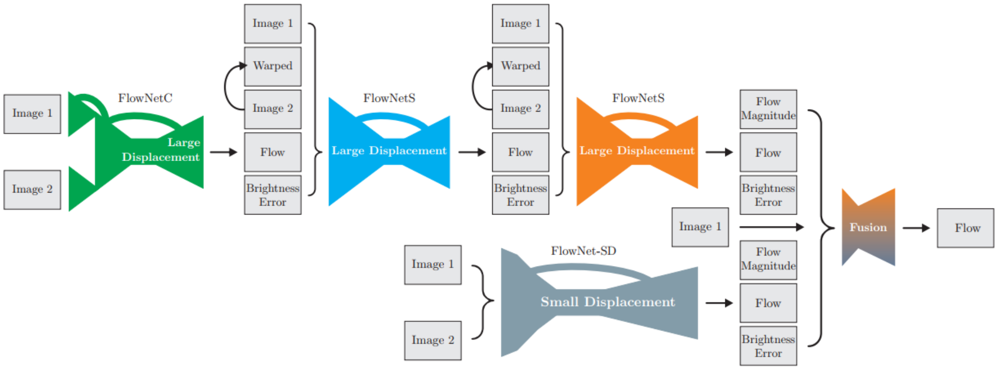


Figure 4: Flwonet2.0

is most useful part for us. The details of feedforward and backward is in supplementary material section 6. (compare with STN)
3. Small, subpixel motion, which is obtained by using small strides in convolution path and more deconvolutional layers in deconv path, which is shown in Fig. 4 lower part.

1.3

Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. **"Spatial transformer networks."** *In Advances in neural information processing systems, pp. 2017-2025. 2015.*

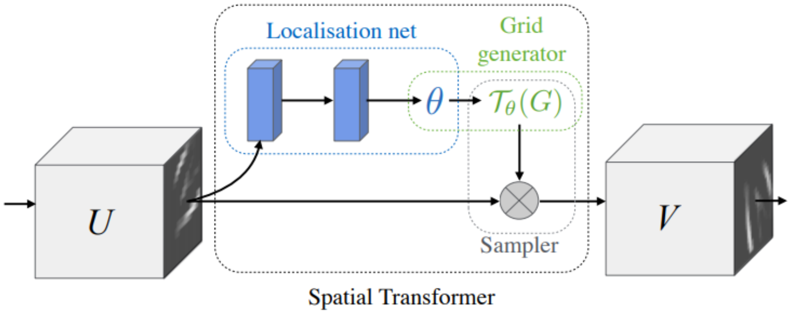


Figure 5: Spatial transformer networks

Unlike Flownet, spacial transformer networks learn parameters of transformation, such like affine transformation is 6-dimensional. This part is done by Localisation net. Then the transformation is used to transform grid. The transformed grid is applied on the input image **U** by Sampler (which is designed to produce differentiable sampling so the whole model can be trained end-to-end) to deliver the adjusted output **V**.

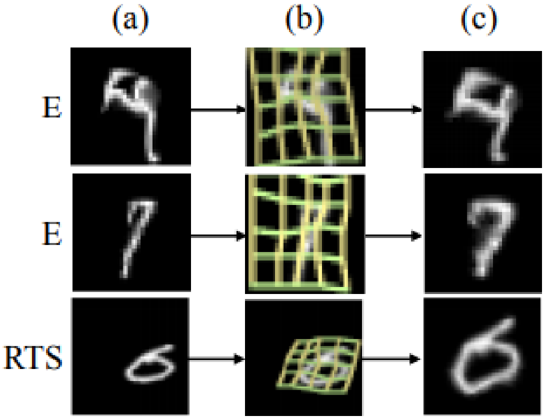


Figure 6: Spatial transformer networks

The Localisation net in Fig. 5 can produce Thin Plate Spline transformation as shown in Fig. 6

1.4

Ranjan, Anurag, and Michael J. Black. **"Optical flow estimation using a spatial pyramid network."** arXiv preprint arXiv:1611.00850 (2016).

One advantage: large motion is dealt with by pyramid.

Two problems in optic flow. One is long-range corelation and the other one is detailed. In spatial pyramid networks the first one is solved by pyramid, and the second one is tackled by deep learning. By using pyramid strategy for each module the flow is limited to a small level, then in this situation CNN filters are more meaningful than flownet.

The drawback of spatial pyramid networks or pyramid strategy is that it is not good at handle the small objects that move quickly. (can it be solved by using dilated CNN?)

In spatial pyramid networks the loss function is to minimize the residues between ground truth and the results the nework gives. So the G_k in Fig. 8

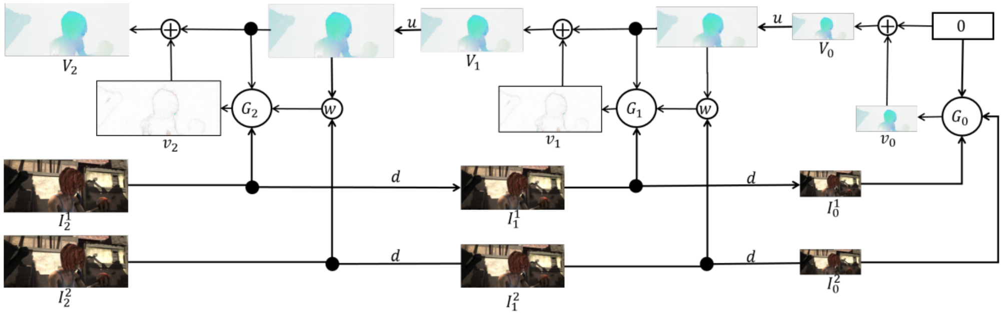


Figure 7: Spatial pyramid networks

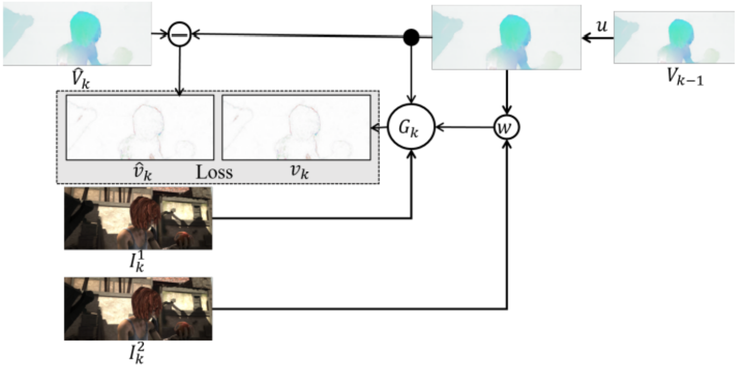


Figure 8: Module of spatial pyramid networks

produce residues of flow other than flow itself. The benefit of doing this way maybe because residue is small.

Each G_k is trained independently. (end-to-end should be better.) It is proved by `https://github.com/anuragranj/end2end-spynet`

Large CNN filter 7×7 on evert layer performs better than small ones.

1.5

Shan, Siyuan, Xiaoqing Guo, Wen Yan, Eric I. Chang, Yubo Fan, and Yan Xu. **"Unsupervised End-to-end Learning for Deformable Medical Image Registration."** *arXiv preprint arXiv:1711.08608 (2017).*

Flownet cannot handle medical image registration which is difficult to generate training data. So unsupervised learning is more promising.

The idea of unsupervised learning is the image after warping with the predicted transformation should be the same as target image. How to avoid predict arbitrary transformation? In STN, the transformation is predefined. The whole structure of the networks of this paper is show in Fig. 9.

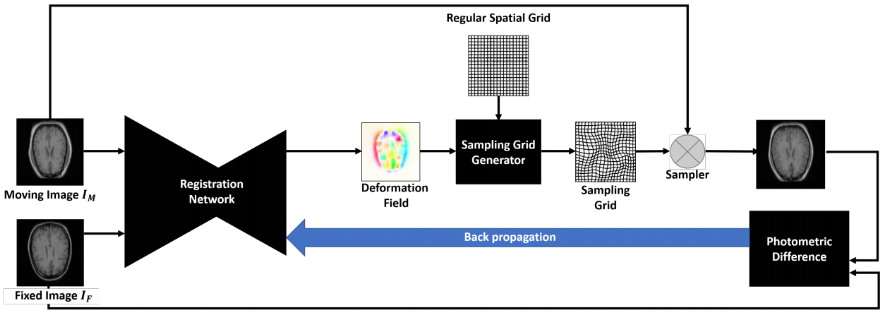


Figure 9: Unsupervised registration by STN

Compare Fig. 9 with Fig. 5, the structure are similiar except that in Fig. 9 the camparision between the warped image and the targe image is used to update the whole network. In details, in this paper the deformable field is not learned by the module Localization layer which is shown in Fig. 5, but in a vanilla CNN network shown in Fig. 10.

Some constrains such as deformation field smoothness loss is applied on the learned deformation field.

Seems like we cannot borrow dataset from this paper because all the dataset they use are about registration. (??)

The idea of this paper is very similiar to *Jason, J. Yu, Adam W. Harley, and Konstantinos G. Derpanis.* **"Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness."** *In European Conference on Computer Vision, pp. 3-10. Springer, Cham, 2016.* which use FlownetSample, and *Ren, Zhe, Junchi Yan, Bingbing Ni, Bin Liu, Xiaokang Yang, and Hongyuan Zha.* **"Unsupervised Deep Learning for Optical Flow Estimation."** *In AAAI, pp. 1495-1501. 2017.* which use STN.

1.6

de Vos, Bob D., Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Igum. **"End-to-end unsupervised deformable image registration with a convolutional neural network."** *In Deep Learning*

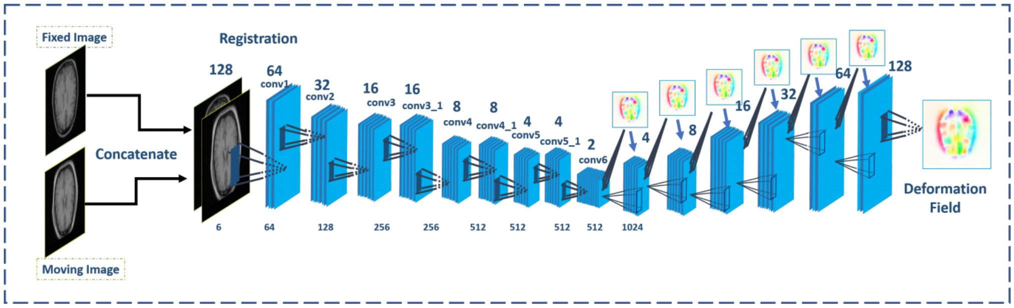


Figure 10: Deformation field learned in Unspervised network (No ground truth of deformation field, the filed is update via the final loss.)

in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 204-212. Springer, Cham, 2017.

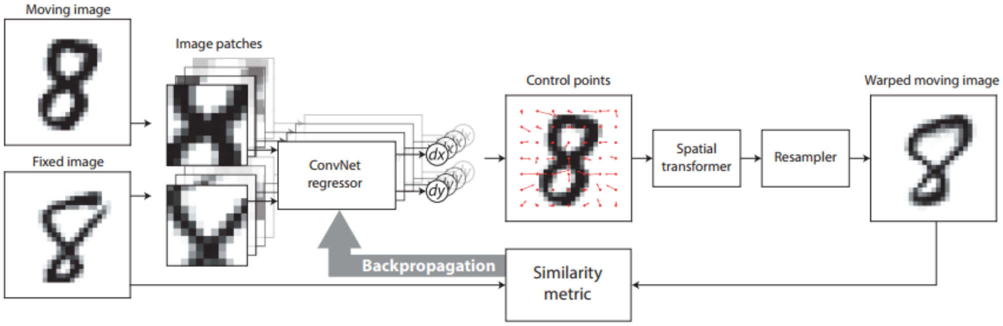
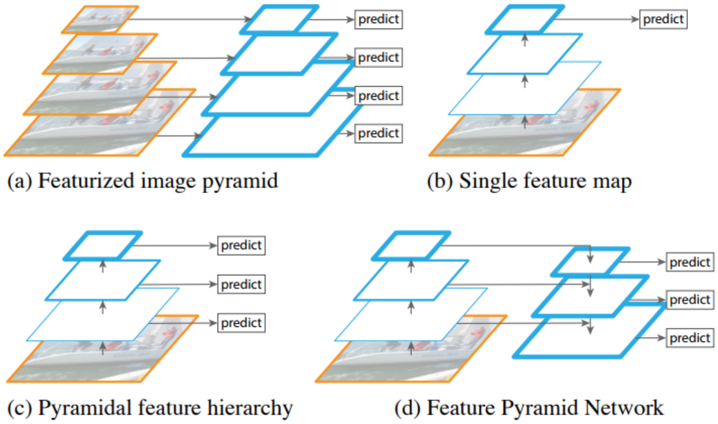


Figure 11: Deformable Image Registration

1.7

Lin, Tsung-Yi, Piotr Dollr, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. **"Feature pyramid networks for object detection."** *In CVPR, vol. 1, no. 2, p. 4. 2017.*



(a) Featurized image pyramid

(b) Single feature map

(c) Pyramidal feature hierarchy

(d) Feature Pyramid Network

Figure 12: Feature Pyramid Networks