Fully-Trainable Deep Matching

James Thewlis
http://www.robots.ox.ac.uk/~jdt/
Shuai Zheng
http://www.robots.ox.ac.uk/~szheng/
Philip H. S. Torr
http://www.robots.ox.ac.uk/~tvg/
Andrea Vedaldi
http://www.robots.ox.ac.uk/~vedaldi/

Robotics Research Group
Department of Engineering Science
University of Oxford
Oxford, United Kingdom

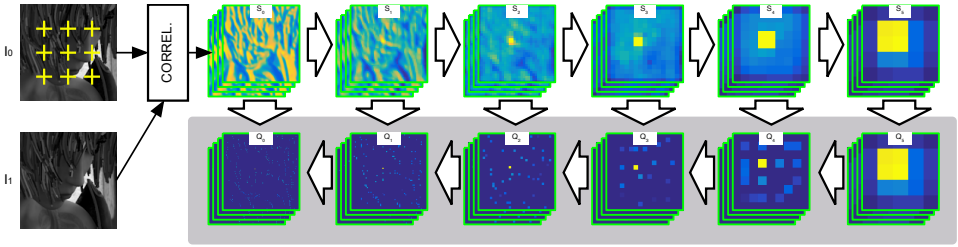

Figure 1: **Fully-Trainable Deep Matching.**

Deep Matching (DM) [3] is a popular method for establishing quasi-dense correspondences between images. An important application of DM is optical flow, where it is used for finding an initial set of image correspondences, which are then interpolated and refined by local optimisation. However, DM, as originally proposed, is *not* a deep neural network and cannot be trained end-to-end via backpropagation.

Our key contribution is to show that the full DM pipeline can be formulated as a Convolutional Neural Network (CNN) with a U-topology (Fig. 1). The fine-to-coarse stage of DM was already given as a CNN in [3]. Here, we complete the construction and show that the DM recursive decoding stage can be implemented by convolutional operators which reverse the ones used in the fine-to-coarse stage. The architecture can be trained using backpropagation, for which we propose a structured-output loss.

We demonstrate the utility of the approach by improving the performance of DM, replacing the HOG image features with the first layers of a pretrained classification CNN. We then train it end-to-end on an image matching task using synthetic data, showing that the accuracy of the matches improves (Table. 1). We further use [2] to interpolate the matches and give dense optical flow, which also sees an improvement.

We compare to FlowNet [1], an alternative end-to-end trainable CNN architecture for image matching, noting the significant advantage of the DM architecture that is further improved by the ability to train from data.

Method	Training	Acc@10 (matches)	EPE (flow)
FlowNet [1]	Flying Chairs	—	4.76
DM-HOG	—	89.39%	3.72
DM-CNN	Pretrained	89.48%	3.63
DM-CNN	Flying Chairs	**90.03%**	**3.50**

Table 1: **Sintel Final training set performance.** DM variants evaluated on Sintel Final training set. EPE (endpoint-error, in pixels) gives mean distance between predicted and correct flow vectors. Acc@10 [3] assigns each pixel a nearby match, measuring the proportion correct within 10px, hence illustrating the quality of the quasi-dense matches.

[1] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazırbaş, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *IEEE ICCV*, 2015.

[2] Jérôme Revaud, Philippe Weinzaepfel, Zaïd Harchaoui, and Cordelia Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *IEEE CVPR*, 2015.

[3] Jérôme Revaud, Philippe Weinzaepfel, Zaïd Harchaoui, and Cordelia Schmid. Deepmatching: Hierarchical deformable dense matching. *IJCV*, 2015.