IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Study on Self-Supervised Multi-Frame with Full-Image Warping

[1]Ms.Rajashree R. Shinde, [2]Mr. Devidas S. Thosar

[1,2]Research Scholar

[1]Electronics & Telecommunication Engineering,

[1]Dr. A.P.J. Abdul Kalam University,Indore, Indore

Abstract: Self Supervised multi-frame with Full Image Warping is a strategy for finding out with regards to solo optical stream that works on the cutting edge of all benchmarks from 36% to 40% (contrasted with the past best UFlow technique) and some even outperforms administered approaches like PWC-Net and FlowNet2. Our strategy for compositional improvements from the administered optical stream, otherwise known as the RAFT model, coordinates novel thoughts for solo picking up including loss of arrangement mindful self-management, procedure for taking care of off-outline movement, and a learning approach. Viably from multi-outline video information with just two casings needed for inductions.

Index Terms **-Self-supervised learning, Multi-frame, Full-Image Warping.**

I. INTRODUCTION

Unsupervised learning is a promising direction to address this issue as it allows training optical flow models from unlabeled videos of any domain. The unsupervised approach works by combining ideas from classical methods and supervised-learning – training the same neural networks as in supervised approaches but optimizing them with objectives such as smoothness and photometric similarity from classical methods. Unlike those classical methods, unsupervised approaches perform optimization not per image pair but jointly for the entire training set. Since unsupervised optical flow takes inspiration from classical and supervised learning methods, we can make substantial progress by properly combining novel ideas with insights from these two directions. In this paper, we do exactly that and make the following three contributions:

1. We integrate the current best supervised model, RAFT [29] with unsupervised learning and perform key changes to the loss functions and data augmentation to properly regularize this model for unsupervised learning.

2. We perform unsupervised learning on image crops while using the full image to compute unsupervised losses. This technique, which we refer to as full-image warping, improves flow quality near image boundaries.

3. We leverage a classical method for multi-frame flow refinement [19] to generate better labels for selfsupervision from multi-frame input. This technique improves performance especially in occluded regions without requiring more than two frames for inference.

Our method Self-Teaching Multi-frame Unsupervised RAFT with Full-Image Warping (SMURF) combines these three contributions and improves the state of the art (SOTA) in unsupervised optical flow in all major benchmarks, i.e. it reduces errors by 40 / 36 / 39 % in the Sintel Clean / Sintel Final / KITTI 2015 benchmarks relative to the prior SOTA set by UFlow [12]. These improvements also reduce the gap to supervised approaches, as SMURF is the first unsupervised optical flow method that outperforms supervised FlowNet2 [9] and PWC-Net [28] on all benchmarks.

II. RELATED WORK

of five years. The time series monthly data is collected on stock prices for sample firmsand relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

Optical flow was first studied in psychology to describe motion in the visual field [6]. Later, the rise of computing in the 1980s led to analytical techniques to estimate optical flow from images [8, 18]. These techniques introduced photometric consistency and smoothness assumptions. These early methods do not perform any learning but instead solve a system of equations to find flow vectors that minimize the objective function for a given image pair. Follow-up work continued to improve flow accuracy, e.g. through better optimization techniques and occlusion reasoning [2, 27]. Machine learning has helped to improve results substantially. First approaches used supervised convolutional neural networks that had relatively little flow-specific structure [5, 9]. Others introduced additional inductive biases from classical approaches such as coarse-to-fine search [23, 28, 37]. The current best network architecture is the Recurrent All-Pairs Field Transforms (RAFT) model. It follows classical work that breaks with the coarse-to-fine assumption [4, 26, 36] and computes the cost volume between all pairs of pixels and uses that information to iteratively refine the flow field [29]. All supervised methods rely heavily on synthetic labeled data for training. While producing excellent results in the supervised setting, RAFT has not previously been used in unsupervised learning. Unsupervised approaches appeared after supervised methods and showed that even without labels, deep learning can greatly outperform classical flow methods [10, 11, 12, 14, 16, 17, 21, 24, 25, 33, 35, 38, 39, 42, 43]. Besides being more accurate than classical methods, learned methods are also faster at inference because all optimization occurs during training instead of during inference time.1 A recent study [12] performed an extensive comparison of the many proposed advances in unsupervised flow estimation and amalgamated these different works into a state of the art method called UFlow. We take this work as a starting point and build on the techniques suggested there such as range-map based occlusion reasoning [35], the Census loss for photometric consistency [21, 40], edge aware smoothness [30], and self supervision [16, 17]. We build on and substantially extend this prior work by enabling the RAFT model to work in an unsupervised learning setting through changes in the loss function and data augmentation. We also propose full-image warping to make the photometric loss useful for pixels that leave the (cropped) image plane. And we utilize a flow refinement technique [19] to leverage multi-frame input during training to self-generate improved labels for self-supervision.

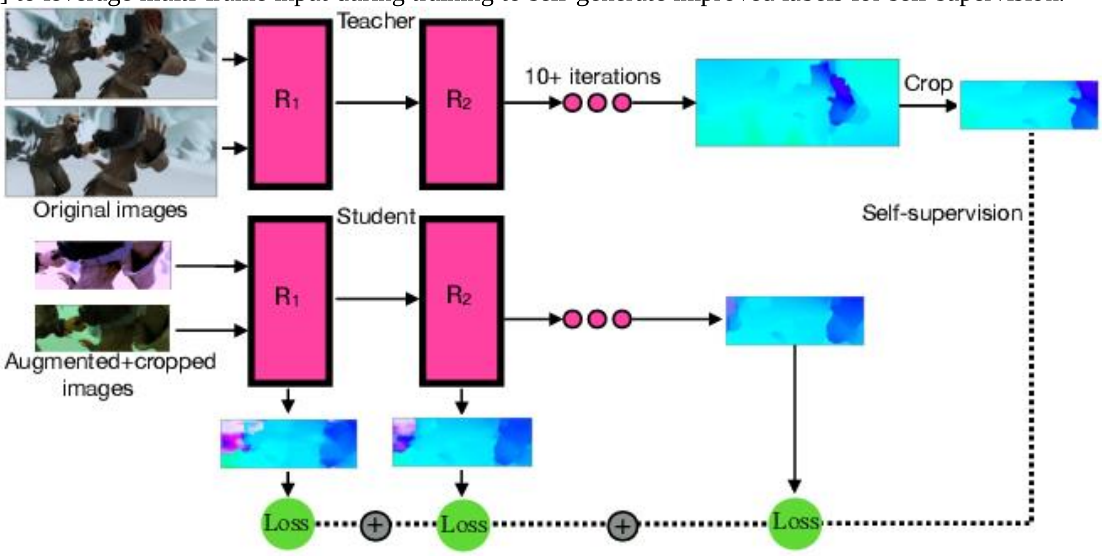

Figure 1. Self-supervision with sequence loss and augmentation.

We use a single model as both "student" and "teacher". As the teacher, we apply the model on full non-augmented images. As student, the model only sees a cropped and augmented version of the same images. The final output of the teacher is then cropped and used to supervise the predictions at all iterations of the student (to which the smoothness and photometric losses are applied as well). The advantages of this self-supervision method are threefold: (1) the model learns to ignore photometric augmentations (2) the model learns to make better predictions at the borders and in occluded areas of the image, and (3) early iterations of the recurrent model learn from the output at the final iteration.

III. SELF-TEACHING MULTI FRAME WITH FULL IMAGE WARPING

3.1 Full Image Warping

The photometric loss, which is essential for unsupervised optical flow estimation, is generally limited to flow vectors that stay inside the image frame because vectors that point outside of the frame have no pixels to compare their photometric appearance to. We address this limitation by computing the flow field from a cropped version of the images I1 and I2 while referencing the full, uncropped image I2 when warping it with the estimated flow V1 before computing the photometric loss (see Figure 2). As we also no longer mark these flow vectors that move outside the image frame as occluded, they now provide the model with a learning signal. We use full-image warping for all datasets except Flying Chairs, where we found that cropping the already small images hurt performance.

3.2 Muti-Frame Self supervision

Finally, we propose to leverage multi-frame information for self-supervision to generate better labels in occluded areas, inspired by work that used a similar technique for inference [19]. For multi-frame self-supervision, we take a frame t and compute the forward flow to the next frame ($t \rightarrow t+1$) and the backward flow to the previous frame ($t \rightarrow t-1$). We then use the backward flow to predict the forward flow through a tiny learned inversion model and use that prediction to inpaint areas that were occluded in the original forward flow but were not occluded in the backward flow – which is why the estimate from the backward flow is more accurate (see Figure 3). The tiny model for the backward-forward inversion consists of three layers of 3×3 convolutions with [16, 16, 2] channels that are applied on the backward flow and the image coordinates normalized to [−1,1]. The model is re-initialized and trained per frame using the non-occluded forward flow as supervision, after which the in-painted flow field is stored and used for selfsupervision. We apply multi-frame self-supervision only at the final stage of training. Importantly, we use multiple frames only during training and not for inference.

3.3 Extensive Data Augmentation

To regularize RAFT, we use the same augmentation as supervised RAFT [RAFT] which is much stronger than what has typically been used in unsupervised optical flow, except for the recent ARFlow [liu2020learning]. We randomly vary hue, brightness, saturation, stretching, scaling, random cropping, random flipping left/right and up/down, and we apply a random eraser augmentation that removes random parts of each image. All augmentations are applied to the model inputs, but not to the images used to compute the photometric and smoothness losses. The self-generated labels for self-supervision are computed from un-augmented images, which has the benefit of training the model to ignore these augmentations.

IV. CONCLUSION

We have presented SMURF, an effective method for unsupervised learning of optical flow that reduces the gap to supervised approaches and shows excellent generalization across datasets and even to "zero-shot" depth estimation. SMURF brings key improvements, most importantly (1) enabling the RAFT architecture to work in an unsupervised setting via modifications to the unsupervised losses and data augmentation, (2) full-image warping for learning to predict out of frame motion, and (3) multi-frame self-supervision for improved flow estimates in occluded regions. We believe that these contributions are a step towards making unsupervised optical flow truly practical, so that optical flow models trained on unlabeled videos can provide high quality pixel-matching in domains without labelled data.

V. ACKNOWLEDGMENT

We express our heartiest thanks to our known and unknown well-wishers for their unreserved cooperation, encouragement and suggestions. Last but not the least, we would like to thanks to all our teachers, and all our friends who helped us with the ever daunting task of gathering information for the topic.

REFERENCES

[1] Austin Stone, Daniel Maurer, Alper Ayvaci, Anelia Angelova, Rico Jonschkowski, SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping, CVPR 2021.
[2] Filippo Aleotti, Fabio Tosi, Li Zhang, Matteo Poggi, and Stefano Mattoccia. Reversing the cycle: Self-supervised deep stereo through enhanced monocular distillation. In ECCV, 2020.
[3] Thomas Brox, Andres Bruhn, Nils Papenberg, and Joachim ´ Weickert. High accuracy optical flow estimation based on a theory for warping. In ECCV, 2004.
[4] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In ECCV, 2012.
[5] Qifeng Chen and Vladlen Koltun. Full flow: Optical flow estimation by global optimization over regular grids. In CVPR, 2016.
[6] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazırbas¸, Vladimir Golkov, Patrick van der ¨ Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In ICCV, 2015.
[7] H. Hirschmuller. Stereo processing by semi-global match- ¨ ing and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328–341, February 2008.
[8] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. AI, 1981.
[9] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In CVPR, 2017.
[10] Woobin Im, Tae-Kyun Kim, and Sung-Eui Yoon. Unsupervised learning of optical flow with deep feature similarity. In ECCV, 2020.
[11] Joel Janai, Fatma Guney, Anurag Ranjan, Michael J. Black, ¨ and Andreas Geiger. Unsupervised learning of multi-frame optical flow with occlusions. In ECCV, 2018.
[12] Rico Jonschkowski, Austin Stone, Jonathan T Barron, Ariel Gordon, Kurt Konolige, and Anelia Angelova. What matters in unsupervised optical flow. In ECCV, 2020.
[13] Diederick P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In ICLR, 2015.
[14] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In CVPR, 2020.
[15] Pengpeng Liu, Irwin King, Michael Lyu, and Jia Xu. Flow2Stereo: Effective self-supervised learning of optical flow and stereo matching. In CVPR, 2020.
[16] Pengpeng Liu, Irwin King, Michael R. Lyu, and Jia Xu. DDFlow: Learning optical flow with unlabeled data distillation. In AAAI, 2019.
[17] Pengpeng Liu, Michael R. Lyu, Irwin King, and Jia Xu. Selflow: Self-supervised learning of optical flow. CVPR, 2019.
[18] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. DARPA Image Understanding Workshop, 1981.
[19] Daniel Maurer and Andres Bruhn. Proflow: Learning to pre- ´ dict optical flow. In BMVC, 2018.
[20] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In CVPR, 2016.
[21] Simon Meister, Junhwa Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In AAAI, 2018.
[22] Moritz Menze, Christian Heipke, and Andreas Geiger. Joint 3d estimation of vehicles and scene flow. ISPRS Workshop on Image Sequence Analysis, 2015.
[23] Anurag Ranjan and Michael J. Black. Optical flow estimation using a spatial pyramid network. In CVPR, 2017.
[24] Anurag Ranjan, Varun Jampani, Lukas Balles, Kihwan Kim, Deqing Sun, Jonas Wulff, and Michael J. Black. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In CVPR, 2019.
[25] Zhe Ren, Junchi Yan, Bingbing Ni, Bin Liu, Xiaokang Yang, and Hongyuan Zha. Unsupervised deep learning for optical flow estimation. In AAAI, 2017.
[26] Frank Steinbrucker, Thomas Pock, and Daniel Cremers. ¨ Large displacement optical flow computation without warping. In ICCV, 2009.
[27] Deqing Sun, Stefan Roth, and Michael J. Black. Secrets of optical flow estimation and their principles. In CVPR, 2010.
[28] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In CVPR, 2018.
[29] Zachary Teed and Jia Deng. Raft: Recurrent all pairs field transforms for optical flow. In ECCV, 2020.
[30] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In ICCV, 1998.

[31] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. arXiv:1607.08022, 2016.
[32] James J. Gibson. The Perception of the Visual World. Houghton Mifflin, 1950.
[33] Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. In CVPR, 1999.
[34] Yang Wang, Peng Wang, Zhenheng Yang, Chenxu Luo, Yi Yang, and Wei Xu. Unos: Unified unsupervised optical-flow and stereo-depth estimation by watching videos. In CVPR, 2019.
[35] Yang Wang, Peng Wang, Zhenheng Yang, Chenxu Luo, Yi Yang, and Wei Xu. Unos: Unified unsupervised optical-flow and stereo-depth estimation by watching videos. In CVPR, 2019.
[36] Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang, and Wei Xu. Occlusion aware unsupervised learning of optical flow. In CVPR, 2018.
[37] Jia Xu, Rene Ranftl, and Vladlen Koltun. Accurate optical ´ flow via direct cost volume processing. In CVPR, 2017.
[38] Gengshan Yang and Deva Ramanan. Volumetric correspondence networks for optical flow. In NeurIPS, 2019.
[39] Zhichao Yin and Jianping Shi. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In CVPR, 2018.
[40] 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 ECCV Workshop, 2016.
[41] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In ECCV, 1994.
[42] Yiran Zhong, Yuchao Dai, and Hongdong Li. Selfsupervised learning for stereo matching with self-improving ability. arXiv:1709.00930, 2017.
[43] Yiran Zhong, Pan Ji, Jianyuan Wang, Yuchao Dai, and Hongdong Li. Unsupervised deep epipolar flow for stationary or dynamic scenes. In CVPR, 2019.
[44] Yuliang Zou, Zelun Luo, and Jia-Bin Huang. DF-Net: Unsupervised joint learning of depth and flow using cross-task consistency. In ECCV, 2018.