LiteFlowNet360 - Revisiting Optical Flow Estimation in 360 Videos

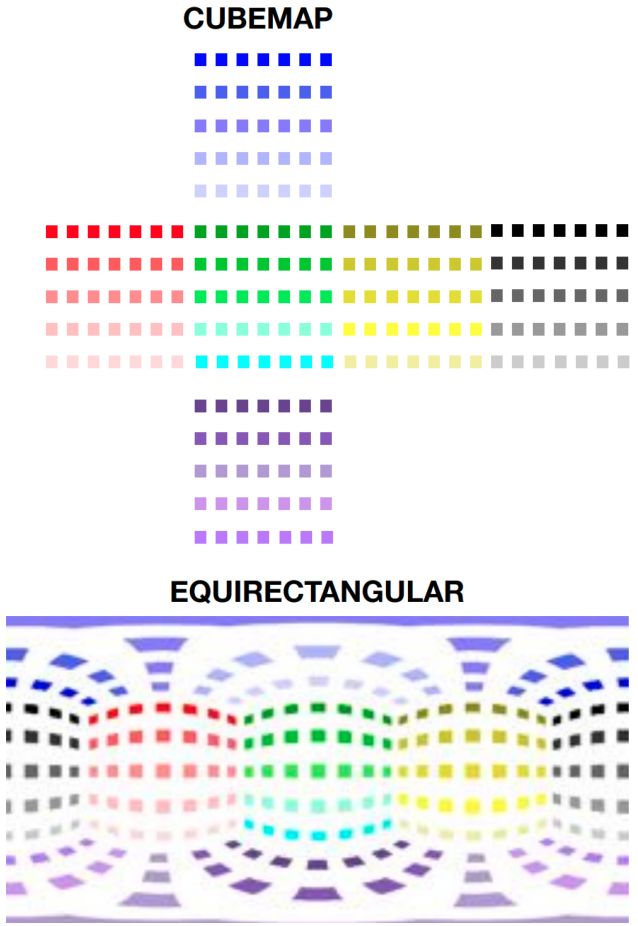
ICPR 2020 | PAPER ID : 2255
Texas State University
By
Keshav Bhandari, Ziliang Zong & Yan Yan
12/05/2020

MOTIVATION

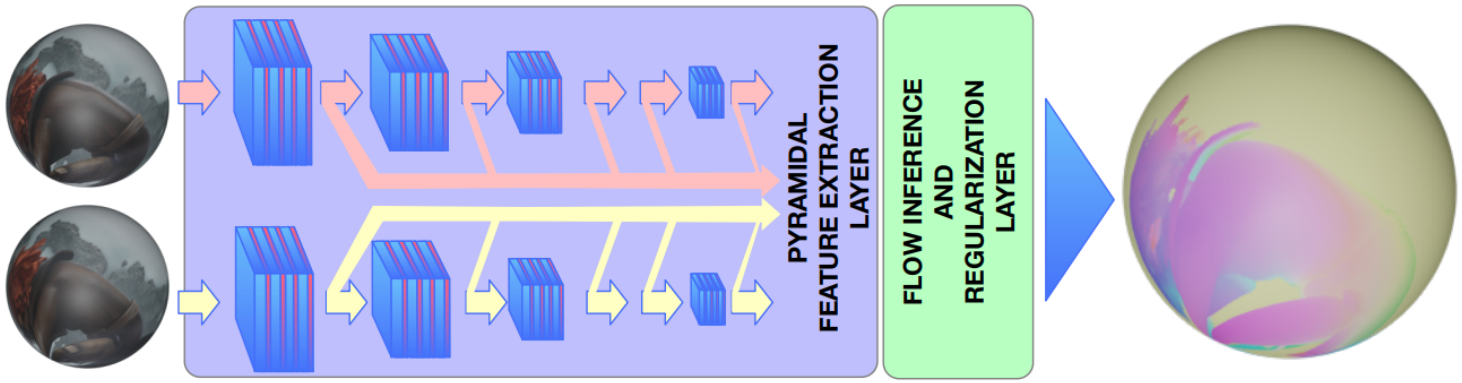
Are optical flow representation for normal and 360 videos are same?

How to exploit existing architecture to compute optical flow for 360 videos?

Do we have labelled data?

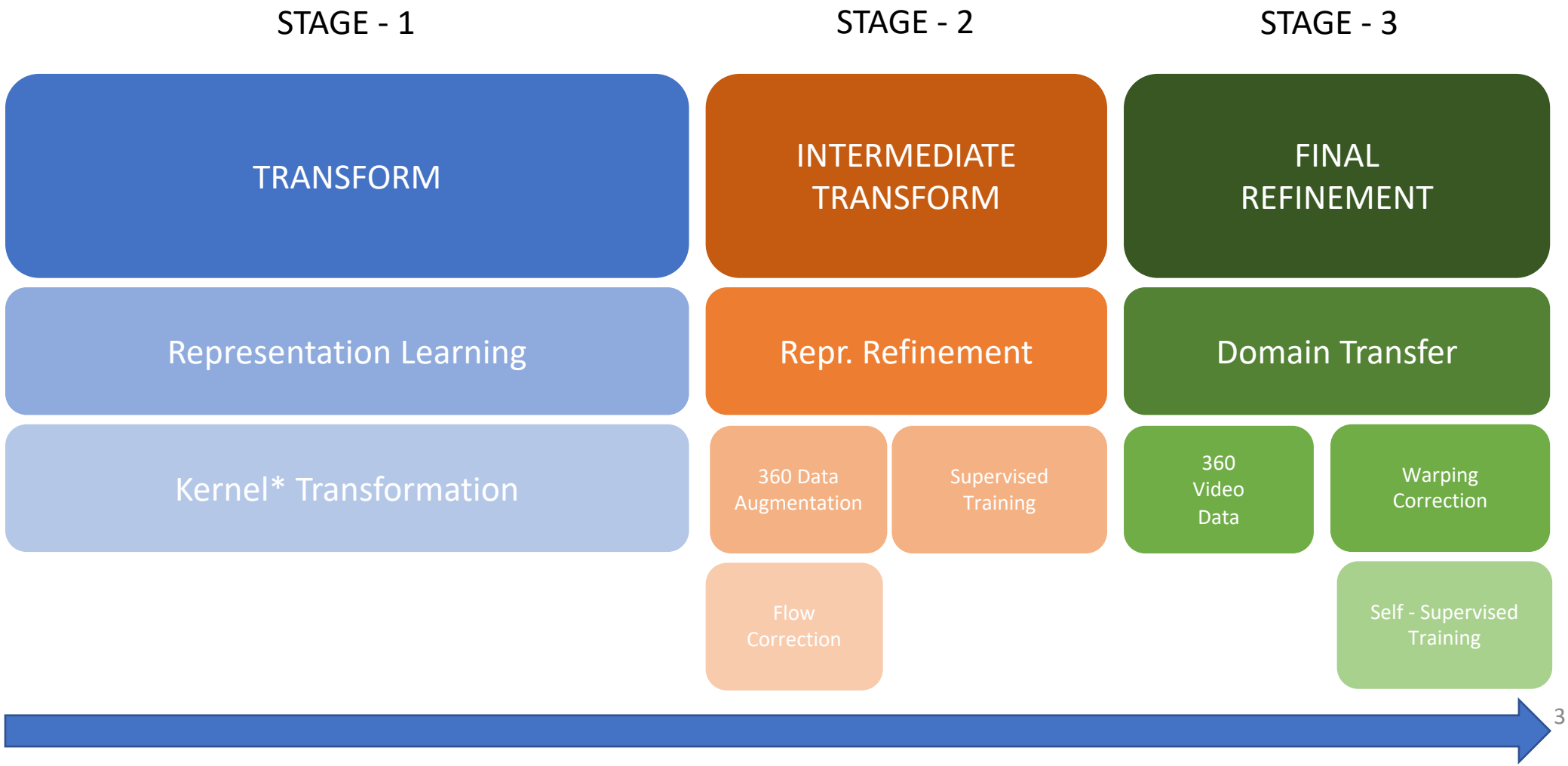


Architecture-Overview

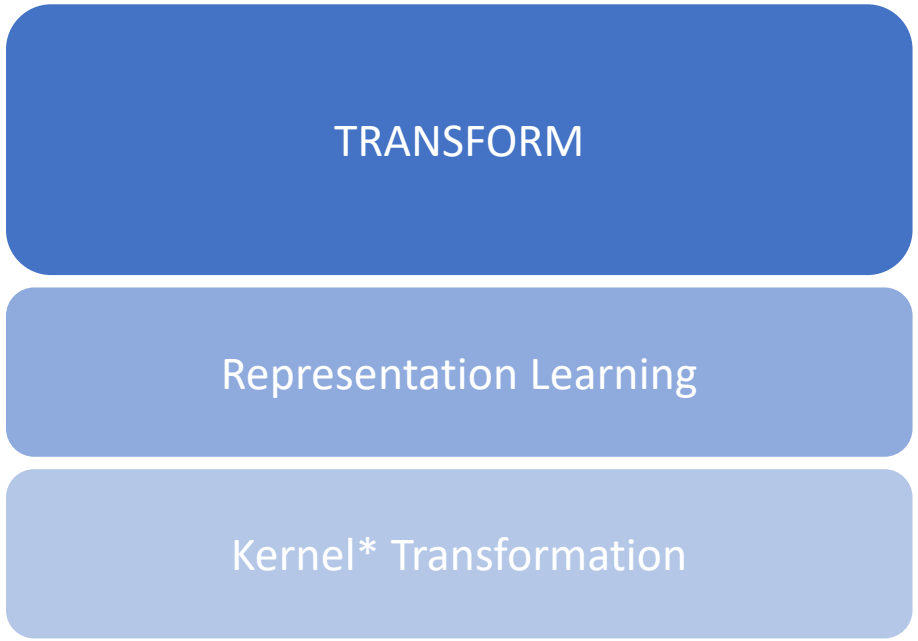


Architecture - ***LiteFlowNet360*** is inspired by Lite-FlowNet. Feature Extractions Layers are replaced with **"Spherical Convolution Learnable Layers"** using transformer network. We did flow inference by projecting the feature map to *tangential plane, rather than learning inference/regularization layer.*

Approach


STAGE - 1
STAGE - 2
STAGE - 3
TRANSFORM
INTERMEDIATE TRANSFORM
FINAL REFINEMENT
Representation Learning
Repr. Refinement
Domain Transfer
Kernel* Transformation
360 Data Augmentation
Supervised Training
360 Video Data
Warping Correction
Flow Correction
Self - Supervised Training

Representation Learning



$$F_k(X) = \begin{cases} F_0(X) & : k = 0 \\ (F_k \circ F_{k-1})(X) & : n > k > 0 \end{cases} \quad (1)$$

$$F'_k(X) = \begin{cases} F'_0(F_0(X), \Omega) & : k = 0 \\ (F'_k(F_k, \Omega) \circ F'_{k-1}(F_{k-1}, \Omega))(X) & : n > k > 0 \end{cases} \quad (2)$$

$$Y_k = F_k(X), Y'_k = F'_k(F_k(X), \Omega)$$

$$L_k = ||Y'_k - Y_k||^2$$

This is done in tangential planes, so multiple L_k

$$L'_k = \frac{1}{n_g} \sum_i^{n_g} L_k(\Omega(Y'^i_k), Y^i_k)$$

* Multiple variable size kernels are learned, these kernels are used to perform convolution in Spherical domain(that is in projected tangential planes)

Representation Refinement

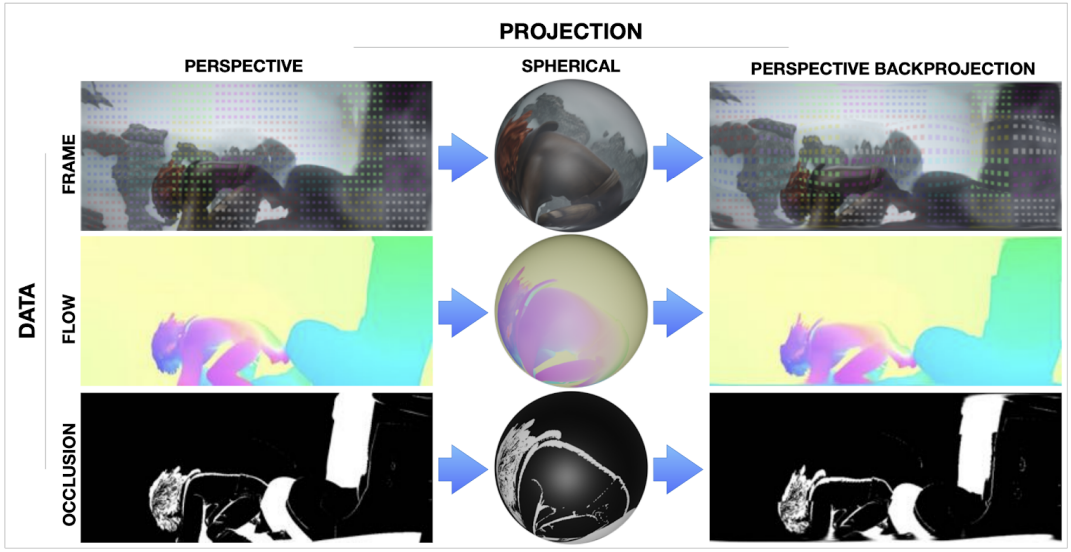
INTERMEDIATE TRANSFORM

Repr. Refinement

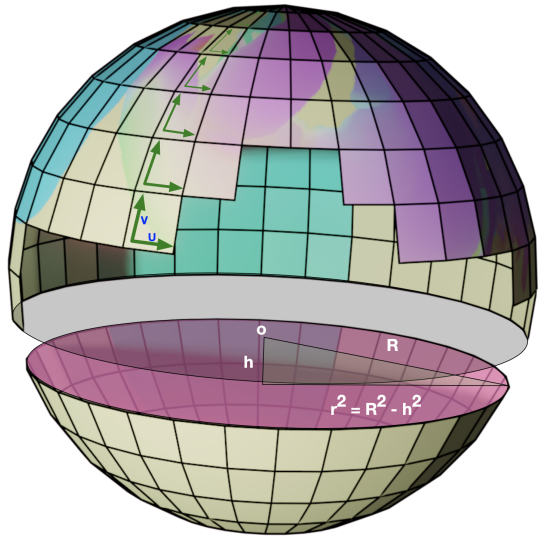
360 Data Augmentation

Supervised Training

Flow Correction



360 Data Augmentation



Flow Corrections

Domain Transfer

FINAL REFINEMENT

Domain Transfer

360 Video Data

Warping Correction

Self - Supervised Training

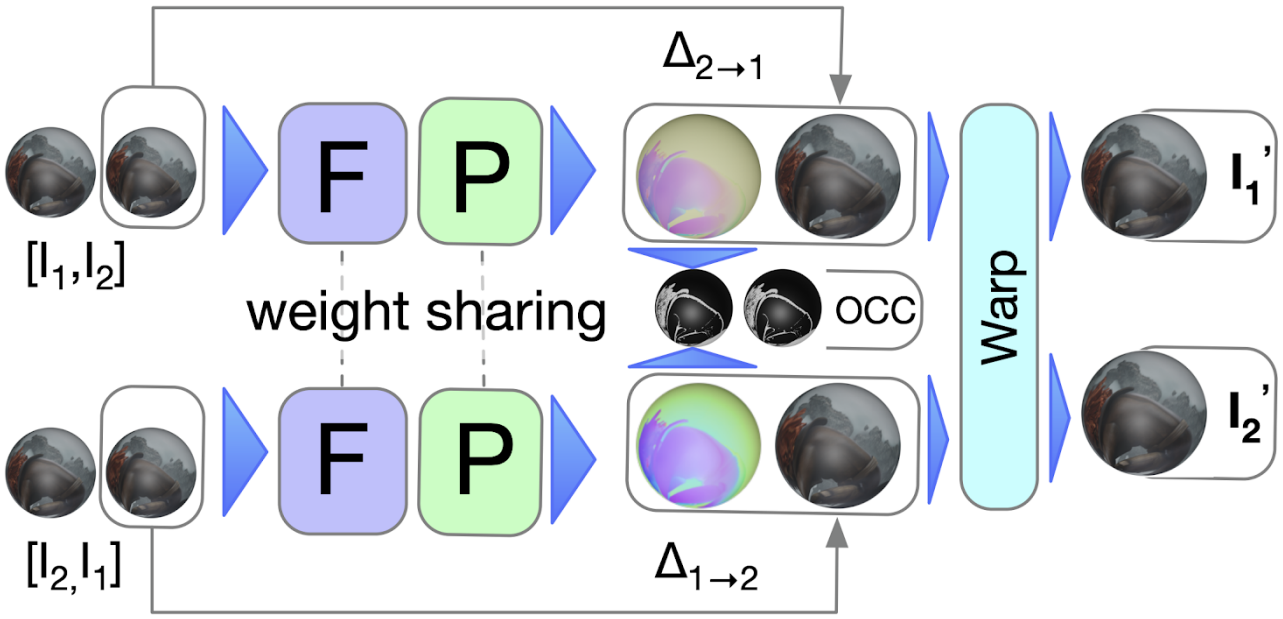


Fig. 5. Final Refinement process. Network from second stage is extended to have two parallel weight sharing architecture.

$$M_i = \begin{cases} 0 \\ 1 \quad if, |\Delta_{i\to j}| \leq \epsilon \end{cases}$$

$$\tilde{O}_{i\to j} = M_i \odot ((1 - M_j) + \tilde{O}_{j\to i})$$

$$L_p = \sum_{i,j} \frac{\sum \psi(I_i - I_i') \odot (1 - O_{i\to j})}{\sum 1 - O_{i\to j}}$$

*There is no boundary case for warping.
*Warping in 2D domain doesn't make sense
*Warping must be done in spherical domain.

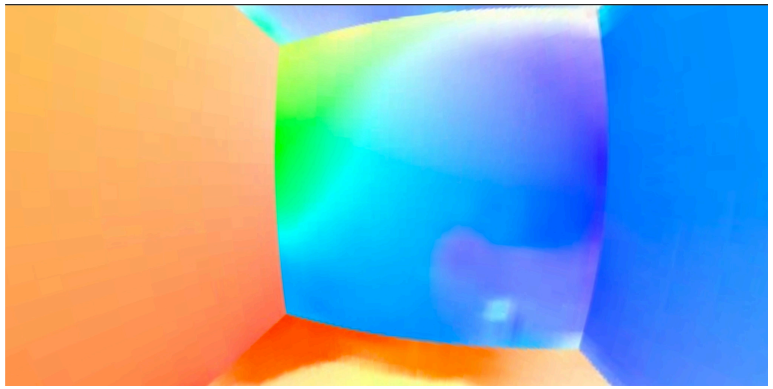
Quantitative Results

RESULTS ON SINTEL360 DATASET.

Model	Data	#Layers	EPE	L_p^*
LiteFlowNet[9]	Sintel360	0	$\sim$ 6.35	$\sim$ 1.30
LiteFlowNet +[14]	Sintel360	> 4	$\geq$ 17	$\geq$ 3.06
Ours, Stage-2	Sintel360	4	$\sim$ 6.35	$\sim$ 0.70
Ours, Final	Sintel360	4	$\sim$ 3.95	$\sim$ 0.60

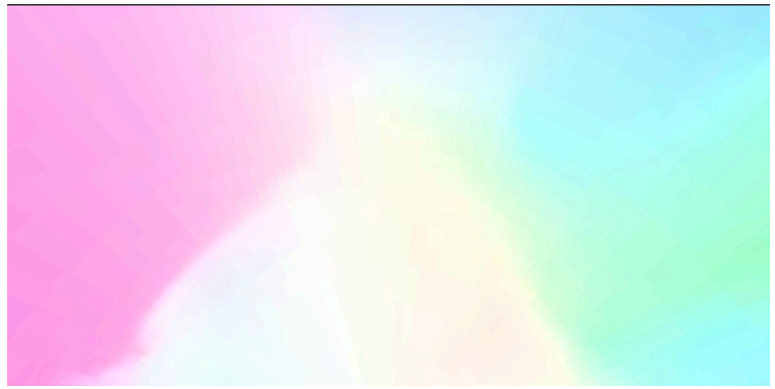
Qualitative Results

Conclusion



Flow Using Cube-maps only

Cube map using Lite-Flownet
Flow representations adapt to represent 360 properties
STAGE-2
(Equirectangular Projection)



OURS
(Equirectangular Projection)

<Conclusion>

LiteFlowNet360 is more of a domain adaption approach to estimate Optical flow for 360 Videos. The core idea is to light on fundamental considerations before exploiting off-the-shelf trained models.

<Thank You!!>

References

[1] Z. Zhang, Y. Xu, J. Yu, and S. Gao, "Saliency detection in 360 videos," in *Proceedings of the European Conference on Computer Vision (ECCV), 2018*.

[2] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *Proceedings of the 26th ACM international conference on Multimedia, 2018*.

[3] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360 sports videos," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017*.

[4] Y. Xu, Y. Dong, J. Wu, Z. Sun, Z. Shi, J. Yu, and S. Gao, "Gaze prediction in dynamic 360 immersive videos," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018*.

[5] B. K. Horn and B. G. Schunck, "Determining optical flow," in *Techniques and Applications of Image Understanding*. International Society for Optics and Photonics, 1981.

[6] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, 1981*.

[7] P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning Optical Flow with Convolutional Networks," *arXiv e-prints, 2015*.

[8] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition, 2017*.

[9] T.-W. Hui, X. Tang, and C. Change Loy, "LiteFlowNet: A Lightweight Convolutional Neural Network for Optical Flow Estimation," *arXiv e-prints, 2018*.

[10] Y.-C. Su and K. Grauman, "Learning spherical convolution for fast features from 360°imagery," in *Advances in Neural Information Processing Systems 30, 2017*.

[11] T. S. Cohen, M. Geiger, J. Koehler, and M. Welling, "Spherical CNNs," *arXiv e-prints, 2018*.

[12] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, "Learning SO(3) Equivariant Representations with Spherical CNNs," *arXiv e-prints, 2017*.

[13] B. Coors, A. P. Condurache, and A. Geiger, "Spherenet: Learning spherical representations for detection and classification in omnidirectional images," in *European Conference on Computer Vision (ECCV), 2018*.

[14] Y.-C. Su and K. Grauman, "Kernel Transformer Networks for Compact Spherical Convolution," *arXiv e-prints, 2018*.

[15] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning, 2009*.

[16] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision (ECCV), 2012*, A. Fitzgibbon et al. (Eds.), Ed.

[17] P. Liu, M. Lyu, I. King, and J. Xu, "Selflow: Self-supervised learning of optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019*.

[18] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Deep end2end voxel2voxel prediction," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2016*.

[19] A. Ahmadi and I. Patras, "Unsupervised convolutional neural networks for motion estimation," *arXiv e-prints, 2016*.

[20] J. Wulff and M. J. Black, "Efficient sparse-to-dense optical flow estimation using a learned basis and layers," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015*.

[21] J. J. Yu, A. W. Harley, and K. G. Derpanis, "Back to Basics: Unsupervised Learning of Optical Flow via Brightness Constancy and Motion Smoothness," *arXiv e-prints, 2016*.

[22] D. Teney and M. Hebert, "Learning to Extract Motion from Videos in Convolutional Neural Networks," *arXiv e-prints, 2016*.

[23] A. Ranjan and M. J. Black, "Optical Flow Estimation using a Spatial Pyramid Network," *arXiv e-prints, 2016*.

[24] D. Gadot and L. Wolf, "PatchBatch: a Batch Augmented Loss for Optical Flow," *arXiv e-prints, 2015*.

[25] C. Bailer, K. Varanasi, and D. Stricker, "CNN-based Patch Matching for Optical Flow with Thresholded Hinge Embedding Loss," *arXiv e-prints, 2016*.

[26] H.-Y. Tung, H.-W. Tung, E. Yumer, and K. Fragkiadaki, "Self-supervised learning of motion capture," in *Advances in Neural Information Processing Systems, 2017*.

[27] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, "Deep 360 Pilot: Learning a Deep Agent for Piloting through 360° Sports Video," *arXiv e-prints, 2017*.

[28] W.-S. Lai, Y. Huang, N. Joshi, C. Buehler, M.-H. Yang, and S. B. Kang, "Semantic-driven Generation of Hyperlapse from 360° Video," *arXiv e-prints, 2017*.

[29] W. Boomsma and J. Frellsen, "Spherical convolutions and their application in molecular modelling," in *Advances in Neural Information Processing Systems 30, 2017*.

[30] H.-T. Cheng, C.-H. Chao, J.-D. Dong, H.-K. Wen, T.-L. Liu, and M. Sun, "Cube Padding for Weakly-Supervised Saliency Prediction in 360° Videos," *arXiv e-prints, 2018*.

[31] R. Khasanova and P. Frossard, "Graph-Based Classification of Omnidirectional Images," *arXiv e-prints, 2017*.

[32] O. Shakernia, R. Vidal, and S. Sastry, "Omnidirectional egomotion estimation from back-projection flow," in *Conference on Computer Vision and Pattern Recognition Workshop, 2003*.

[33] A. Radgui, C. Demonceaux, E. M. Mouaddib, D. Aboutajdine, and M. Rziza, "An adapted lucas-kanade's method for optical flow estimation in catadioptric images, 2008."

[34] A. Radgui, C. Demonceaux, E. Mouaddib, M. Rziza, and D. Aboutajdine, "Optical flow estimation from multichannel spherical image decomposition," *Computer Vision and Image Understanding, 2011*.

[35] C. Kirisits, L. F. Lang, and O. Scherzer, "Decomposition of optical flow on the sphere," *GEM-International Journal on Geomathematics, 2014*.

[36] B. Alibouch, A. Radgui, M. Rziza, and D. Aboutajdine, "Optical flow estimation on omnidirectional images: an adapted phase based method," in *International Conference on Image and Signal Processing, 2012*. Springer.