REFERENCES

1. Mart´ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Ma1hieu Devin, and others. 2016. Tensorfiow: Large-scale machine learning on heterogeneous dis- tributed systems. *arXiv preprint arXiv:1603.04467* (2016).
2. Martin Arjovsky, Soumith Chintala, and Lé on Bo1ou. 2017. Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017).
3. James R Bergen, Peter J Burt, Rajesh Hingorani, and Shmuel Peleg. 1990. Com- puting two motions from three frames. In *Computer Vision, 1990. Proceedings, Lird International Conference on*. IEEE, 27–32.
4. ftomas Brox, André s Bruhn, Nils Papenberg, and Joachim Weickert. 2004. High accuracy optical fiow estimation based on a theory for warping. *Computer Vision-ECCV 2004* (2004), 25–36.
5. Bert De Brabandere, Xu Jia, Tinne Tuytelaars, and Luc Van Gool. 2016. Dynamic filter networks. In *Neural Information Processing Systems (NIPS)*.
6. Emily L Denton, Soumith Chintala, Rob Fergus, and others. 2015. Deep Gen- erative Image Models using a Laplacian Pyramid of Adversarial Networks. In *Advances in neural information processing systems*. 1486–1494.
7. Chelsea Finn, Ian Goodfellow, and Sergey Levine. 2016. Unsupervised learning for physical interaction through video prediction. In *Advances In Neural Information Processing Systems*. 64–72.
8. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
9. Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. 2015. DRAW: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623* (2015).

10. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity map- pings in deep residual networks. In *European Conference on Computer Vision*. Springer, 630–645.
11. Minh Hoai and Fernando De la Torre. 2014. Max-margin early event detectors. *International Journal of Computer Vision* 107, 2 (2014), 191–202.
12. Max Jaderberg, Karen Simonyan, Andrew Zisserman, and others. 2015. Spatial transformer networks. In *Advances in Neural Information Processing Systems*. 2017–2025.
13. Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 2013. 3D convolutional neural networks for human action recognition. *IEEE transactions on paMern analysis and machine intelligence* 35, 1 (2013), 221–231.
14. Nal Kalchbrenner, Aaron van den Oord, Karen Simonyan, Ivo Danihelka, Oriol Vinyals, Alex Graves, and Koray Kavukcuoglu. 2016. Video pixel networks. *arXiv preprint arXiv:1610.00527* (2016).
15. Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
16. Kris M Kitani, Brian D Ziebart, James Andrew Bagnell, and Martial Hebert. 2012. Activity forecasting. In *European Conference on Computer Vision*. Springer, 201–214.
17. Tian Lan, Tsung-Chuan Chen, and Silvio Savarese. 2014. A hierarchical represen- tation for future action prediction. In *European Conference on Computer Vision*. Springer, 689–704.
18. Ziwei Liu, Raymond Yeh, Xiaoou Tang, Yiming Liu, and Aseem Agarwala. 2017. Video Frame Synthesis using Deep Voxel Flow. *arXiv preprint arXiv:1702.02463* (2017).
19. Dhruv Mahajan, Fu-Chung Huang, Wojciech Matusik, Ravi Ramamoorthi, and Peter Belhumeur. 2009. Moving gradients: a path-based method for plausible image interpolation. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 42.
20. Michael Mathieu, Camille Couprie, and Yann LeCun. 2015. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440* (2015).

21. Anish Mi1al, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No- reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21, 12 (2012), 4695–4708.
22. Silvia L Pintea, Jan C van Gemert, and Arnold WM Smeulders. 2014. Déja vu. In *European Conference on Computer Vision*. Springer, 172–187.
23. Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representa- tion learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
24. MarcAurelio Ranzato, Arthur Szlam, Joan Bruna, Michael Mathieu, Ronan Col- lobert, and Sumit Chopra. 2014. Video (language) modeling: a baseline for generative models of natural videos. *arXiv preprint arXiv:1412.6604* (2014).
25. Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. 2012. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402* (2012).
26. Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. 2015. Unsuper- vised Learning of Video Representations using LSTMs.. In *ICML*. 843–852.
27. Ilya Sutskever, Oriol Vinyals, and <OC V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.
28. ftiow Keng Tan, Choong Seng Boon, and Yoshinori Suzuki. 2006. Intra prediction by template matching. In *Image Processing, 2006 IEEE International Conference on*. IEEE, 1693–1696.
29. Lucas fteis, Aäron van den Oord, and Ma1hias Bethge. 2015. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844* (2015).
30. Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. 2015. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 4489–4497.
31. Joost van Amersfoort, Anitha Kannan, Marc'Aurelio Ranzato, Arthur Szlam, Du Tran, and Soumith Chintala. 2017. Transformation-based models of video sequences. *arXiv preprint arXiv:1701.08435* (2017).
32. Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol.

2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. ACM, 1096–1103.

33. Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. 2016. Generating videos with scene dynamics. In *Advances In Neural Information Processing Systems*. 613– 621.
34. Jacob Walker, Carl Doersch, Abhinav Gupta, and Martial Hebert. 2016. An uncertain future: Forecasting from static images using variational autoencoders.
35. Jacob Walker, Abhinav Gupta, and Martial Hebert. 2014. Patch to the future: Unsupervised visual prediction. In *Proceedings of the IEEE Conference on Computer Vision and PaMern Recognition*. 3302–3309.
36. John YA Wang and Edward H Adelson. 1993. Layered representation for motion analysis. In *Computer Vision and PaMern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*. IEEE, 361–366.
37. Zhou Wang and Alan C Bovik. 2009. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine* 26, 1 (2009), 98–117.
38. Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
39. Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Deep3d: Fully automatic 2D- to-3D video conversion with deep convolutional neural networks. In *European Conference on Computer Vision*. Springer, 842–857.
40. Jianwen Xie, Song-Chun Zhu, and Ying Nian Wu. 2016. Synthesizing Dynamic Textures and Sounds by Spatial-Temporal Generative ConvNet. *arXiv preprint arXiv:1606.00972* (2016).
41. Tianfan Xue, Jiajun Wu, Katherine Bouman, and Bill Freeman. 2016. Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks. In *Advances in Neural Information Processing Systems*. 91–99.
42. Xinchen Yan, Jimei Yang, Kihyuk Sohn, and Honglak Lee. 2016. A1ribute2image: Conditional image generation from visual a1ributes. In *European Conference on Computer Vision*. Springer, 776–791.