

Neural Scene Representation and Rendering

Related Work

Owing to space constraints we have only been able to cite a small sample of references in the main manuscript. Below we provide a more comprehensive list of related works. Please email aeslami@google.com or danilor@google.com to make additions to this list.

1. Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703.
2. Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B*, 200(1140), 269-294.
3. Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2), 115.
4. Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive psychology*, 21(2), 233-282.
5. Becker, S., & Hinton, G. E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355(6356), 161.
6. Faugeras, O. (1993). Three-dimensional computer vision: a geometric viewpoint. MIT press.
7. Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural computation*, 7(5), 889-904.
8. Williams, C. K., Revow, M., & Hinton, G. E. (1997). Instantiating deformable models with a neural net. *Computer vision and image understanding*, 68(1), 120-126.
9. Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (pp. 189-198).
10. Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*(pp. 187-194).

11. Stevens, M. R., & Beveridge, J. R. (2000). *Integrating graphics and vision for object recognition* (Vol. 589). Springer Science & Business Media.
12. Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635.
13. Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. *Probabilistic models of the brain*, 13-36.
14. Ramamoorthi, R. (2002). A signal-processing framework for forward and inverse rendering. Stanford University.
15. Schaffalitzky, F., & Zisserman, A. (2002). Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In European conference on computer vision (pp. 414-431). Springer, Berlin, Heidelberg.
16. Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision*. Cambridge university press.
17. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., & Koch, R. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3), 207-232.
18. Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual review of psychology*, 55.
19. Williams, C. K., & Titsias, M. K. (2004). Greedy learning of multiple objects in images using robust statistics and factorial learning. *Neural Computation*, 16(5), 1039-1062.
20. LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2, pp. II-104).
21. Brown, M., & Lowe, D. G. (2005). Unsupervised 3D object recognition and reconstruction in unordered datasets. In *Fifth International Conference on 3-D Digital Imaging and Modeling, 2005* (pp. 56-63).
22. Cao, L., Liu, J., & Tang, X. (2005). 3D object reconstruction from a single 2D line drawing without hidden lines. In *Tenth IEEE International Conference on Computer Vision* (pp. 272-277).
23. Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis?. *Trends in cognitive sciences*, 10(7), 301-308.

24. Itti, L., & Baldi, P. F. (2006). Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems* (pp. 547-554).
25. Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7), 299-306.
26. Snavely, N., Seitz, S. M., & Szeliski, R. (2008). Modeling the world from internet photo collections. *International journal of computer vision*, 80(2), 189-210.
27. Saxena, A., Sun, M., & Ng, A. Y. (2009). Make3D: Learning 3D scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5), 824-840.
28. Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3), 230-247.
29. Salakhutdinov, R., & Larochelle, H. (2010). Efficient learning of deep Boltzmann machines. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. 693-700).
30. Memisevic, R., & Hinton, G. E. (2010). Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural computation*, 22(6), 1473-1492.
31. Liebelt, J., & Schmid, C. (2010). Multi-view object class detection with a 3D geometric model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010 (pp. 1688-1695).
32. Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005), 845-851.
33. Kivinen, J. J., & Williams, C. K. (2011). Transformation equivariant Boltzmann machines. In *International Conference on Artificial Neural Networks* (pp. 1-9). Springer, Berlin, Heidelberg.
34. Geiger, A., Lauer, M., & Urtasun, R. (2011). A generative model for 3D urban scene understanding from movable platforms. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011 (pp. 1945-1952).
35. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., ... & Fitzgibbon, A. (2011). KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 559-568).

36. Hinton, G. E., Krizhevsky, A., & Wang, S. D. (2011). Transforming auto-encoders. In *International Conference on Artificial Neural Networks* (pp. 44-51).
37. DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415-434.
38. Eslami, S., & Williams, C. (2012). A generative model for parts-based object segmentation. In *Advances in Neural Information Processing Systems* (pp. 100-107).
39. Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, 76(4), 677-694.
40. Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., & Poggio, T. (2013). Unsupervised learning of invariant representations in hierarchical architectures. *arXiv preprint arXiv:1311.4158*.
41. Zia, M. Z., Stark, M., Schiele, B., & Schindler, K. (2013). Detailed 3D representations for object recognition and modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(11), 2608-2623.
42. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
43. Cohen, T. S., & Welling, M. (2014). Transformation properties of learned visual representations. *arXiv preprint arXiv:1412.7659*.
44. Jahangiri, E., Vidal, R., Younes, L., & Geman, D. (2014). Object-Level Generative Models for 3D Scene Understanding. In *SUNw: Scene Understanding Workshop*.
45. Chen, C. Y., & Grauman, K. (2014). Inferring unseen views of people. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2003-2010).
46. Guzman-Rivera, A., Kohli, P., Glocker, B., Shotton, J., Sharp, T., Fitzgibbon, A., & Izadi, S. (2014). Multi-output learning for camera relocalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1114-1121).
47. Kulkarni, T. D., Mansinghka, V. K., Kohli, P., & Tenenbaum, J. B. (2014). Inverse graphics with probabilistic CAD models. *arXiv preprint arXiv:1407.1339*.
48. Kulkarni, T., Yildirim, I., Kohli, P., Freiwald, W., & Tenenbaum, J. B. (2014). Deep generative vision as approximate bayesian computation. In *NIPS 2014 ABC Workshop*.

49. Loper, M. M., & Black, M. J. (2014). OpenDR: An approximate differentiable renderer. In *European Conference on Computer Vision* (pp. 154-169).
50. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
51. Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*.
52. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3D shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1912-1920).
53. Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D. (2015). DRAW: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*.
54. Agrawal, P., Carreira, J., & Malik, J. (2015). Learning to see by moving. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 37-45).
55. Jampani, V., Eslami, S. A., Tarlow, D., Kohli, P., & Winn, J. (2015). Consensus message passing for layered graphical models. In *Artificial Intelligence and Statistics* (pp. 425-433).
56. Jampani, V., Nowozin, S., Loper, M., & Gehler, P. V. (2015). The informed sampler: A discriminative approach to bayesian inference in generative computer vision models. *Computer Vision and Image Understanding*, 136, 32-44.
57. Jayaraman, D., & Grauman, K. (2015). Learning image representations tied to ego-motion. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1413-1421).
58. Zhang, Y., Xu, W., Tong, Y., & Zhou, K. (2015). Online structure analysis for real-time indoor scene reconstruction. *ACM Transactions on Graphics (TOG)*, 34(5), 159.
59. Zuffi, S., & Black, M. J. (2015). The stitched puppet: A graphical model of 3D human shape and pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3537-3546).

60. Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. (2015). Multi-view convolutional neural networks for 3D shape recognition. In Proceedings of the IEEE international conference on computer vision (pp. 945-953).
61. Kulkarni, T. D., Kohli, P., Tenenbaum, J. B., & Mansinghka, V. (2015). Picture: A probabilistic programming language for scene perception. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4390-4399).
62. Kulkarni, T. D., Whitney, W. F., Kohli, P., & Tenenbaum, J. (2015). Deep convolutional inverse graphics network. In *Advances in Neural Information Processing Systems* (pp. 2539-2547).
63. Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332-1338.
64. Huang, H., Kalogerakis, E., & Marlin, B. (2015). Analysis and synthesis of 3D shape families via deep-learned generative models of surfaces. In *Computer Graphics Forum* (Vol. 34, No. 5, pp. 25-38).
65. Yildirim, I., Kulkarni, T. D., Freiwald, W. A., & Tenenbaum, J. B. (2015). Efficient and robust analysis-by-synthesis in vision: A computational framework, behavioral tests, and modeling neuronal representations. In Annual conference of the cognitive science society.
66. Choy, C. B., Stark, M., Corbett-Davies, S., & Savarese, S. (2015). Enriching object detection with 2D-3D registration and continuous viewpoint estimation. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on* (pp. 2512-2520).
67. Wu, J., Zhang, C., Xue, T., Freeman, B., & Tenenbaum, J. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *Advances in Neural Information Processing Systems* (pp. 82-90).
68. Girdhar, R., Fouhey, D. F., Rodriguez, M., & Gupta, A. (2016). Learning a predictable and generative vector representation for objects. In *European Conference on Computer Vision* (pp. 484-499).
69. Flynn, J., Neulander, I., Philbin, J., & Snavely, N. (2016). DeepStereo: Learning to predict new views from the world's imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5515-5524).
70. Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. (2016). 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In *European Conference on Computer Vision* (pp. 628-644).

71. Fouhey, D. F., Gupta, A., & Zisserman, A. (2016). 3D shape attributes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1516-1524).
72. Gregor, K., Besse, F., Rezende, D. J., Danihelka, I., & Wierstra, D. (2016). Towards conceptual compression. In *Advances In Neural Information Processing Systems* (pp. 3549-3557).
73. van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O., & Graves, A. (2016). Conditional image generation with pixelcnn decoders. In *Advances in Neural Information Processing Systems* (pp. 4790-4798).
74. Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2016). Density estimation using Real NVP. arXiv preprint arXiv:1605.08803.
75. Tatarchenko, M., Dosovitskiy, A., & Brox, T. (2016). Multi-view 3D models from single images with a convolutional network. In *European Conference on Computer Vision* (pp. 322-337).
76. Zamir, A. R., Wekel, T., Agrawal, P., Wei, C., Malik, J., & Savarese, S. (2016). Generic 3D representation via pose estimation and matching. In *European Conference on Computer Vision* (pp. 535-553).
77. Eslami, S. A., Heess, N., Weber, T., Tassa, Y., Szepesvari, D., & Hinton, G. E. (2016). Attend, infer, repeat: Fast scene understanding with generative models. In *Advances in Neural Information Processing Systems* (pp. 3225-3233).
78. Brock, A., Lim, T., Ritchie, J. M., & Weston, N. (2016). Generative and discriminative voxel modeling with convolutional neural networks. arXiv preprint arXiv:1608.04236.
79. Yan, X., Yang, J., Yumer, E., Guo, Y., & Lee, H. (2016). Perspective transformer nets: Learning single-view 3D object reconstruction without 3D supervision. In *Advances in Neural Information Processing Systems* (pp. 1696-1704).
80. Malik, J., Arbeláez, P., Carreira, J., Fragkiadaki, K., Girshick, R., Gkioxari, G., ... & Tulsiani, S. (2016). The three R's of computer vision: Recognition, reconstruction and reorganization. *Pattern Recognition Letters*, 72, 4-14.
81. Zhou, T., Tulsiani, S., Sun, W., Malik, J., & Efros, A. A. (2016). View synthesis by appearance flow. In *European conference on computer vision* (pp. 286-301).
82. Moreno, P., Williams, C. K., Nash, C., & Kohli, P. (2016). Overcoming occlusion with inverse graphics. In *European Conference on Computer Vision* (pp. 170-185).

83. Zhang, Y., Bai, M., Kohli, P., Izadi, S., & Xiao, J. (2016). Deepcontext: Context-encoding neural pathways for 3d holistic scene understanding. arXiv preprint arXiv:1603.04922.
84. Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience*, 17(3), 173.
85. Rezende, D. J., Eslami, S. A., Mohamed, S., Battaglia, P., Jaderberg, M., & Heess, N. (2016). Unsupervised learning of 3D structure from images. In *Advances In Neural Information Processing Systems* (pp. 4996-5004).
86. Yan, X., Khansari, M., Bai, Y., Hsu, J., Pathak, A., Gupta, A., ... & Lee, H. (2017). Learning grasping interaction with geometry-aware 3D representations. *arXiv preprint arXiv:1708.07303*.
87. Dosovitskiy, A., Springenberg, J. T., Tatarchenko, M., & Brox, T. (2017). Learning to generate chairs, tables and cars with convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(4), 692-705.
88. Zhao, B., Wu, X., Cheng, Z. Q., Liu, H., Jie, Z., & Feng, J. (2017). Multi-view image generation from a single-view. arXiv preprint arXiv:1704.04886.
89. Jayaraman, D., & Grauman, K. (2017). Learning to look around. arXiv preprint arXiv:1709.00507.
90. Jayaraman, D., Gao, R., Grauman K. (2017). ShapeCodes: Self-Supervised Feature Learning by Lifting Views to Viewgrids. arXiv preprint arXiv:1709.00505.
91. Jayaraman, D., & Grauman, K. (2017). Learning image representations tied to egomotion from unlabeled video. *International Journal of Computer Vision*, 125(1-3), 136-161.
92. Siddharth, N., Paige, B., Van de Meent, J. W., Desmaison, A., Wood, F., Goodman, N. D., Kohli, P. & Torr, P. H. (2017). Learning disentangled representations with semi-supervised deep generative models. ArXiv e-prints.
93. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
94. Yildirim, I., Janner, M., Belledonne, M., Wallraven, C., Freiwald, W., & Tenenbaum, J. B. (2017). Causal and compositional generative models in online perception. In *at 39th Annual Conference of the Cognitive Science Society*.

95. Godard, C., Mac Aodha, O., & Brostow, G. J. (2017). Unsupervised monocular depth estimation with left-right consistency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2, No. 6, p. 7).
96. Ji, D., Kwon, J., McFarland, M., & Savarese, S. (2017). Deep view morphing. Technical report.
97. Lassner, C., Pons-Moll, G., & Gehler, P. V. (2017). A generative model of people in clothing. arXiv preprint arXiv:1705.04098
98. Achlioptas, P., Diamanti, O., Mitliagkas, I., & Guibas, L. (2017). Learning Representations and Generative Models for 3D Point Clouds. arXiv preprint arXiv:1707.02392.
99. Smith, E., & Meger, D. (2017). Improved adversarial systems for 3D object generation and reconstruction. arXiv preprint arXiv:1707.09557.
100. Soltani, A. A., Huang, H., Wu, J., Kulkarni, T. D., & Tenenbaum, J. B. (2017). Synthesizing 3D shapes via modeling multi-view depth maps and silhouettes with deep generative networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1511-1519).
101. Lin, C. H., Kong, C., & Lucey, S. (2017). Learning efficient point cloud generation for dense 3D object reconstruction. arXiv preprint arXiv:1706.07036.
102. Pronobis, A., & Rao, R. P. (2017). Learning deep generative spatial models for mobile robots. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on* (pp. 755-762).
103. Umetani, N. (2017). Exploring generative 3D shapes using autoencoder networks. In *SIGGRAPH Asia 2017 Technical Briefs* (p. 24). ACM.
104. Zhou, T., Brown, M., Snavely, N., & Lowe, D. G. (2017). Unsupervised learning of depth and ego-motion from video. In *CVPR* (Vol. 2, No. 6, p. 7).
105. Kato, H., Ushiku, Y., & Harada, T. (2017). Neural 3D Mesh Renderer. arXiv preprint arXiv:1711.07566.
106. Li, J., Xu, K., Chaudhuri, S., Yumer, E., Zhang, H., & Guibas, L. (2017). Grass: Generative recursive autoencoders for shape structures. *ACM Transactions on Graphics (TOG)*, 36(4), 52.
107. Han, X., Gao, C., & Yu, Y. (2017). DeepSketch2Face: a deep learning based sketching system for 3D face and caricature modeling. *ACM Transactions on Graphics (TOG)*, 36(4), 126.

108. Chen, Q., & Koltun, V. (2017). Photographic image synthesis with cascaded refinement networks. In *IEEE International Conference on Computer Vision (ICCV)* (Vol. 1).
109. Pontes, J. K., Kong, C., Eriksson, A., Fookes, C., Sridharan, S., & Lucey, S. (2017). Compact model representation for 3D reconstruction. *arXiv preprint arXiv:1707.07360*.
110. Liu, Y., Gupta, A., Abbeel, P., & Levine, S. (2017). Imitation from observation: Learning to imitate behaviors from raw video via context translation. *arXiv preprint arXiv:1707.03374*.
111. Park, E., Yang, J., Yumer, E., Ceylan, D., & Berg, A. C. (2017). Transformation-grounded image generation network for novel 3d view synthesis. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 702-711).
112. Jiang, C., Zhu, Y., Qi, S., Huang, S., Lin, J., Guo, X., ... & Zhu, S. C. (2017). Configurable, Photorealistic Image Rendering and Ground Truth Synthesis by Sampling Stochastic Grammars Representing Indoor Scenes. *arXiv preprint arXiv:1704.00112*.
113. Wu, J., Tenenbaum, J. B., & Kohli, P. (2017). Neural scene de-rendering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2).
114. Liu, C., Wu, J., Kohli, P., & Furukawa, Y. (2017). Raster-to-Vector: Revisiting Floorplan Transformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2195-2203).
115. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances in Neural Information Processing Systems* (pp. 3859-3869).
116. Romaszko, L., Williams, C. K., Moreno, P., & Kohli, P. (2017). Vision-as-inverse-graphics: Obtaining a rich 3D explanation of a scene from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (pp. 851-859).
117. Nash, C., & Williams, C. K. (2017). The shape variational autoencoder: A deep generative model of part-segmented 3D objects. In *Computer Graphics Forum* (Vol. 36, No. 5, pp. 1-12).
118. Häne, C., Tulsiani, S., & Malik, J. (2017). Hierarchical surface prediction for 3D object reconstruction. *arXiv preprint arXiv:1704.00710*.
119. Tan, Q., Gao, L., Lai, Y. K., & Xia, S. (2017). Variational Autoencoders for Deforming 3D Mesh Models. *arXiv preprint arXiv:1709.04307*.

120. Greff, K., van Steenkiste, S., & Schmidhuber, J. (2017). Neural expectation maximization. In *Advances in Neural Information Processing Systems* (pp. 6694-6704).
121. Lin, C. H., Yumer, E., Wang, O., Shechtman, E., & Lucey, S. (2018). ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing. *arXiv preprint arXiv:1803.01837*.
122. Wayne, G., Hung, C. C., Amos, D., Mirza, M., Ahuja, A., Grabska-Barwinska, A., ... & Gemici, M. (2018). Unsupervised Predictive Memory in a Goal-Directed Agent. *arXiv preprint arXiv:1803.10760*.
123. Ha, D., & Schmidhuber J., "World Models" (2018). <https://doi.org/10.5281/zenodo.1207631>
124. Esser, P., Sutter, E., & Ommer, B. (2018). A Variational U-Net for Conditional Appearance and Shape Generation. *arXiv preprint arXiv:1804.04694*.
125. de Bem, R., Ghosh, A., Ajanthan, T., Miksik, O., Siddharth, N., & Torr, P. H. (2018). DGPose: Disentangled Semi-supervised Deep Generative Models for Human Body Analysis. *arXiv preprint arXiv:1804.06364*.
126. Stutz, D., & Geiger, A. (2018). Learning 3D Shape Completion from Laser Scan Data with Weak Supervision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
127. Ranjan, A., Jampani, V., Kim, K., Sun, D., Wulff, J., & Black, M. J. (2018). Adversarial Collaboration: Joint Unsupervised Learning of Depth, Camera Motion, Optical Flow and Motion Segmentation. *arXiv preprint arXiv:1805.09806*.
128. Liu, M., He, X., & Salzmann, M. (2018). Geometry-aware Deep Network for Single-Image Novel View Synthesis. *arXiv preprint arXiv:1804.06008*.
129. Tulsiani, S., Efros, A. A., & Malik, J. (2018). Multi-view Consistency as Supervisory Signal for Learning Shape and Pose Prediction. *arXiv preprint arXiv:1801.03910*.
130. Kanazawa, A., Tulsiani, S., Efros, A. A., & Malik, J. (2018). Learning Category-Specific Mesh Reconstruction from Image Collections. *arXiv preprint arXiv:1803.07549*.
131. van Steenkiste, S., Chang, M., Greff, K., & Schmidhuber, J. (2018). Relational Neural Expectation Maximization: Unsupervised Discovery of Objects and their Interactions. *arXiv preprint arXiv:1802.10353*.

132. Mirchev, A., Kayalibay, B., van der Smagt, P., & Bayer, J. (2018). Approximate Bayesian inference in spatial environments. arXiv preprint arXiv:1805.07206.

