



## REFERENCES

- [1] A. J. Hawkins, “Tesla relied on too many robots to build the model 3, elon musk says,” *The Verge*, April 2018. [Online]. Available: <https://www.theverge.com/2018/4/13/17234296/tesla-model-3-robots-production-hell-elon-musk>
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’12, 2012, pp. 1097–1105.
- [3] D. C. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3642–3649.
- [4] C. Lawrence, “How much has amazon invested in automation?” *Amazon Picking Challenge*, August 2011. [Online]. Available: <https://amazonpickingchallenge.org/how-much-has-amazon-invested-in-automation/>
- [5] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Alet, N. C. Daffe, R. Holladay, I. Morona, P. Q. Nair, D. Green, I. Taylor, W. Liu, T. Funkhouser, and A. Rodriguez, “Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2018.
- [6] C. Eppner, S. Höfer, R. Jonschkowski, R. Martín-Martín, A. Sieverling, V. Wall, and O. Brock, “Four aspects of building robotic systems: lessons from the amazon picking challenge 2015,” *Autonomous Robots*, vol. 42, no. 7, pp. 1459–1475, Oct 2018. [Online]. Available: <https://doi.org/10.1007/s10514-018-9761-2>
- [7] S. Caldera, A. Rassau, and D. Chai, “Review of deep learning methods in robotic grasp detection,” *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 57, 2018. [Online]. Available: <https://doi.org/10.3390/mti2030057>
- [8] D. Zhao and S. Li, “A 3d image processing method for manufacturing process automation,” *Computers in Industry*, vol. 56, pp. 975–985, 2005.
- [9] A. C. Romea, D. Berenson, S. Srinivasa, and D. I. Ferguson, “Object recognition and full pose registration from a single image for robotic manipulation,” May 2009.
- [10] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, “Rigid 3d geometry matching for grasping of known objects in cluttered scenes,” *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 538–553, 2012. [Online]. Available: <https://doi.org/10.1177/0278364911436019>



- [11] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. K. Marks, and R. Chellappa, “Fast object localization and pose estimation in heavy clutter for robotic bin picking,” *The International Journal of Robotics Research*, vol. 31, no. 8, pp. 951–973, 2012. [Online]. Available: <https://doi.org/10.1177/0278364911436018>
- [12] C.-Y. Tsai, K.-J. Hsu, and H. Nisar, “Efficient model-based object pose estimation based on multi-template tracking and pnp algorithms,” *Algorithms*, vol. 11, no. 8, p. 122, 2018. [Online]. Available: <https://doi.org/10.3390/a11080122>
- [13] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal, “Learning of grasp selection based on shape-templates,” *Autonomous Robots*, vol. 36, no. 1, pp. 51–65, Jan 2014. [Online]. Available: <https://doi.org/10.1007/s10514-013-9366-8>
- [14] M. Gualtieri, A. Ten Pas, K. Saenko, and R. Platt, “High precision grasp pose detection in dense clutter,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 598–605.
- [15] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, “Grasp pose detection in point clouds,” *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1455–1473, 2017. [Online]. Available: <https://doi.org/10.1177/0278364917735594>
- [16] J. Wei, H. Liu, G. Yan, and F. Sun, “Robotic grasping recognition using multi-modal deep extreme learning machine,” *Multidimensional Systems and Signal Processing*, vol. 28, no. 3, pp. 817–833, Jul 2017. [Online]. Available: <https://doi.org/10.1007/s11045-016-0389-0>
- [17] D. Guo, F. Sun, H. Liu, T. Kong, B. Fang, and N. Xi, “A hybrid deep architecture for robotic grasp detection,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 1609–1614.
- [18] J. Bohg, A. Morales, T. Asfour, and D. Kragic, “Data-driven grasp synthesis—a survey,” *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, April 2014.
- [19] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, “Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes,” *CoRR*, vol. abs/1711.00199, 2017. [Online]. Available: <http://arxiv.org/abs/1711.00199>
- [20] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Y. Goldberg, “Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics,” *ArXiv*, vol. abs/1703.09312, 2017.
- [21] J. Mahler and K. Goldberg, “Learning deep policies for robot bin picking by simulating robust grasping sequences,” in *Proceedings of the 1st Annual Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, S. Levine, V. Vanhoucke, and



- K. Goldberg, Eds., vol. 78. PMLR, 13–15 Nov 2017, pp. 515–524. [Online]. Available: <http://proceedings.mlr.press/v78/mahler17a.html>
- [22] T. Do, M. Cai, T. Pham, and I. D. Reid, “Deep-6dpose: Recovering 6d object pose from a single RGB image,” *CoRR*, vol. abs/1802.10367, 2018. [Online]. Available: <http://arxiv.org/abs/1802.10367>
- [23] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, “Deep object pose estimation for semantic robotic grasping of household objects,” in *Conference on Robot Learning (CoRL)*, 2018. [Online]. Available: <https://arxiv.org/abs/1809.10790>
- [24] M. Danielczuk, M. Matl, S. Gupta, A. Li, A. Lee, J. Mahler, and K. Goldberg, “Segmenting unknown 3d objects from real depth images using mask r-cnn trained on synthetic data,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2019.
- [25] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng, “Robotic grasping of novel objects,” in *Advances in Neural Information Processing Systems 19*. MIT Press, 2007, pp. 1209–1216. [Online]. Available: <http://papers.nips.cc/paper/3010-robotic-grasping-of-novel-objects.pdf>
- [26] S. Kumra and C. Kanan, “Robotic grasp detection using deep convolutional neural networks,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 769–776.
- [27] E. Johns, S. Leutenegger, and A. J. Davison, “Deep learning a grasp function for grasping under gripper pose uncertainty,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2016, pp. 4461–4468.
- [28] Q. Lu, K. Chenna, B. Sundaralingam, and T. Hermans, “Planning multi-fingered grasps as probabilistic inference in a learned deep network,” *CoRR*, vol. abs/1804.03289, 2018. [Online]. Available: <http://arxiv.org/abs/1804.03289>
- [29] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [31] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: <http://arxiv.org/abs/1905.11946>



- [32] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [33] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
- [34] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “Bisenet: Bilateral segmentation network for real-time semantic segmentation,” *CoRR*, vol. abs/1808.00897, 2018. [Online]. Available: <http://arxiv.org/abs/1808.00897>
- [35] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” *CoRR*, vol. abs/1611.05431, 2016. [Online]. Available: <http://arxiv.org/abs/1611.05431>
- [36] D. Mahajan, R. B. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. van der Maaten, “Exploring the limits of weakly supervised pretraining,” *CoRR*, vol. abs/1805.00932, 2018. [Online]. Available: <http://arxiv.org/abs/1805.00932>
- [37] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, “Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution,” *CoRR*, vol. abs/1904.05049, 2019. [Online]. Available: <http://arxiv.org/abs/1904.05049>
- [38] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou, “Fixing the train-test resolution discrepancy,” *CoRR*, vol. abs/1906.06423, 2019. [Online]. Available: <http://arxiv.org/abs/1906.06423>
- [39] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” *CoRR*, vol. abs/1707.07012, 2017. [Online]. Available: <http://arxiv.org/abs/1707.07012>
- [40] M. Tan, B. Chen, R. Pang, V. Vasudevan, and Q. V. Le, “Mnasnet: Platform-aware neural architecture search for mobile,” *CoRR*, vol. abs/1807.11626, 2018. [Online]. Available: <http://arxiv.org/abs/1807.11626>
- [41] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [42] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [43] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.



- [44] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: A structure for efficient numerical computation," *Computing in Science Engineering*, vol. 13, no. 2, pp. 22–30, March 2011.
- [45] A. B. Jung, K. Wada, J. Crall, S. Tanaka, J. Graving, S. Yadav, J. Banerjee, G. Vecsei, A. Kraft, J. Borovec, C. Vallentin, S. Zhydenko, K. Pfeiffer, B. Cook, I. Fernández, W. Chi-Hung, A. Ayala-Acevedo, R. Meudec, M. Laporte *et al.*, "imgaug," <https://github.com/aleju/imgaug>, 2019.
- [46] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.
- [47] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015. [Online]. Available: <https://doi.org/10.1177/0278364914549607>
- [48] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker Jr, A. Rodriguez, and J. Xiao, "Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge," in *ICRA*, 2016.
- [49] E. Matsumoto, M. Saito, A. Kume, and J. Tan, "End-to-end learning of object grasp poses in the amazon robotics challenge," in *Warehouse Picking Automation Workshop (WPAW), ICRA*, 2017.