[1] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." arXiv, Feb. 15, 2014. doi: 10.48550/arXiv.1402.3722.

[2] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN." arXiv, Dec. 06, 2017. doi: 10.48550/arXiv.1701.07875.

[3] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised Visual Representation Learning by Context Prediction." arXiv, Jan. 16, 2016. doi: 10.48550/arXiv.1505.05192.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." arXiv, Jan. 07, 2016. doi: 10.48550/arXiv.1511.06434.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." arXiv, Aug. 24, 2020. doi: 10.48550/arXiv.1703.10593.

[6] J. Rocca, "Understanding Variational Autoencoders (VAEs)," *Medium*, Mar. 21, 2021. https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73 (accessed Oct. 03, 2022).

[7] P. Shree, "The Journey of Open AI GPT models," *Walmart Global Tech Blog*, Nov. 10, 2020. https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2 (accessed Oct. 04, 2022).

[8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," p. 38.

[9] R. Zhang, P. Isola, and A. A. Efros, "Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction." arXiv, Apr. 20, 2017. doi: 10.48550/arXiv.1611.09842.

[10] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-Attention Generative Adversarial Networks." arXiv, Jun. 14, 2019. doi: 10.48550/arXiv.1805.08318.

[11] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing Flows for Probabilistic Modeling and Inference." arXiv, Apr. 08, 2021. doi: 10.48550/arXiv.1912.02762.

[12] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear Independent Components Estimation." arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1410.8516.

[13] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, "Neural Discrete Representation Learning." arXiv, May 30, 2018. Accessed: Oct. 03, 2022. [Online]. Available: http://arxiv.org/abs/1711.00937

[14] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners." arXiv, Dec. 19, 2021. doi: 10.48550/arXiv.2111.06377.

[15] Y. LeCun, "Learning Invariant Feature Hierarchies," in *Computer Vision – ECCV 2012. Workshops and Demonstrations*, vol. 7583, A. Fusiello, V. Murino, and R. Cucchiara, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 496–505. doi: 10.1007/978-3-642-33863-2_51.

[16] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in Translation: Contextualized Word Vectors." arXiv, Jun. 20, 2018. doi: 10.48550/arXiv.1708.00107.

[17] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis." arXiv, Feb. 25, 2019. doi: 10.48550/arXiv.1809.11096.

[18] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets." arXiv, Jun. 11, 2016. doi: 10.48550/arXiv.1606.03657.

[19] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," p. 12.

[20] "Implicit Generation and Generalization Methods for Energy-Based Models," *OpenAI*, Mar. 21, 2019. https://openai.com/blog/energy-based-models/ (accessed Oct. 03, 2022).

[21] Y. Du and I. Mordatch, "Implicit Generation and Generalization in Energy-Based Models." arXiv, Jun. 29, 2020. doi: 10.48550/arXiv.1903.08689.

[22] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543. doi: 10.3115/v1/D14-1162.

[23] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[24] I. J. Goodfellow *et al.*, "Generative Adversarial Networks." arXiv, Jun. 10, 2014. doi: 10.48550/arXiv.1406.2661.

[25] A. Graves, "Generating Sequences With Recurrent Neural Networks." arXiv, Jun. 05, 2014. Accessed: Oct. 04, 2022. [Online]. Available: http://arxiv.org/abs/1308.0850

[26] A. Razavi, A. van den Oord, and O. Vinyals, "Generating Diverse High-Fidelity Images with VQ-VAE-2." arXiv, Jun. 02, 2019. doi: 10.48550/arXiv.1906.00446.

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space." arXiv, Sep. 06, 2013. doi: 10.48550/arXiv.1301.3781.

[28] A. Pentland, "E i g e d c e s for Recognition," p. 16.

[29] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion Models in Vision: A Survey." arXiv, Sep. 10, 2022. doi: 10.48550/arXiv.2209.04747.

[30] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP." arXiv, Feb. 27, 2017. doi: 10.48550/arXiv.1605.08803.

[31] M. E. Peters *et al.*, "Deep contextualized word representations." arXiv, Mar. 22, 2018. Accessed: Oct. 04, 2022. [Online]. Available: http://arxiv.org/abs/1802.05365

[32] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive Learning for Unpaired Image-to-Image Translation." arXiv, Aug. 20, 2020. doi: 10.48550/arXiv.2007.15651.

[33] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting." arXiv, Nov. 21, 2016. doi: 10.48550/arXiv.1604.07379.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.

[35] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT Pre-Training of Image Transformers." arXiv, Sep. 03, 2022. Accessed: Oct. 05, 2022. [Online]. Available: http://arxiv.org/abs/2106.08254

[36] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT Pre-Training of Image Transformers." arXiv, Sep. 03, 2022. doi: 10.48550/arXiv.2106.08254.

[37] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes." arXiv, May 01, 2014. doi: 10.48550/arXiv.1312.6114.

[38] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data," in *Proceedings of the 35th International Conference on Machine Learning*, Jul. 2018, pp. 195–204. Accessed: Oct. 03, 2022. [Online]. Available: https://proceedings.mlr.press/v80/almahairi18a.html

[39] A. Vaswani *et al.*, "Attention Is All You Need." arXiv, Dec. 05, 2017. doi: 10.48550/arXiv.1706.03762.

[40] D. P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," *FNT in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019, doi: 10.1561/2200000056.

[41] D. P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," *FNT in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019, doi: 10.1561/2200000056.

[42] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, Jun. 03, 2021. doi: 10.48550/arXiv.2010.11929.