References

[Hill '79] Banu Musa brothers (9th century). The book of ingenious devices (Kitab al-hiyal). Translated by D. R. Hill (1979), Springer, p. 44, ISBN 90-277-0833-9.

[Wiener '48] N. Wiener (1948). Time, communication, and the nervous system. Teleological mechanisms. Annals of the N.Y. Acad. Sci. 50 (4): 197-219.

[Fuegi '03] Fuegi, J., & Francis, J. (2003). Lovelace & Babbage and the creation of the 1843'notes'. *IEEE Annals of the History of Computing*, *25*(4), 16-26.

[Kagermann '11] Kagermann, H., Lukas, W. D., & Wahlster, W. (2011). Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution. *VDI nachrichten*, *13*(1), 2-3.

[Hey '09] Hey, T., Tansley, S., & Tolle, K. M. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery (Vol. 1). Redmond, WA: Microsoft Research.

[Mitchell '97] Mitchell, T.M. (1997) Machine Learning. McGraw Hill.

[Géron '19] Géron, A. (2019). Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd Edition, O'Reilly.

[Bayes '63] Bayes, T. (1763). LII. An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, (53), 370-418.

[Provost '13] Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.".

[Russell '21] Russell, S., Norvig, P. (2021). Artificial intelligence: A Modern Approach (4th Edition). Prentice Hall.

[Chapmann '17] Chapmann, J. (2017) Machine Learning: Fundamental Algorithms for Supervised and Unsupervised Learning With Real-World Applications.

[Knuth '97] Knuth, D. E. (1997). The Art of Computer Programming (Vol. 1). Pearson Education.

[Bhargava '16] Bhargava, A. (2016). Grokking Algorithms: An illustrated guide for programmers and other curious people. Manning Publications Co.

[Han '11] Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. 3rd Edition. Elsevier.

[Allen '83] Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832-843.

[Schneider '19] Schneider, S., Baevski, A., Collobert, R., & Auli, M. (2019). wav2vec: Unsupervised pre-training for speech recognition. arXiv preprint arXiv:1904.05862.

[Wolpert '97] Wolpert, D.H., Macready, W.G. (1997), "No Free Lunch Theorems for Optimization", *IEEE Transactions on Evolutionary Computation* 1, 67

[Topol '19] Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.

[Zheng '18] Zheng, A., & Casari, A. (2018). Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. " O'Reilly Media, Inc.".

[Deisenroth '19] Deisenroth, M.P, Faisal, A. and Ong C.S. (2019) Mathematics for Machine Learning. Cambridge University Press.

[Roweis '00] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, *290* (5500), 2323-2326.

[Saul '00] Saul, L. K., & Roweis, S. T. (2000). An introduction to locally linear embedding. unpublished. Available at: http://www.cs.toronto.edu/~ roweis/lle/publications. html.

[Van der Maaten '08] Van der Maaten, L.J.P.; Hinton, G.E. (2008). Visualizing Data Using t-SNE. *Journal of Machine Learning Research*. 9: 2579–2605.

[McInnes '18] McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

[Barnes '86] Barnes, J. & Hut, P. (1986). A hierarchical O(N log N) force-calculation algorithm. *Nature*. 324 (4): 446–449

[Harrison '19] Harrison M. (2019) Machine Learning Pocket Reference: Working with Structured Data in Python. *O'reilly Media.*

[Fisher '36]Fisher, R. (1936). Linear discriminant analysis. Ann. Eugenics, 7, 179.

[Griffiths '09] Griffin Dawn (2009) *Head First Statistics*. Publisher: O'Reilly.

[Javaheri '21] Javaheri, T., Homayounfar, M., Amoozgar, Z., Reiazi, R., Homayounieh, F., Abbas, E., ... & Rawassizadeh, R. (2021). CovidCTNet: an open-source deep learning approach to diagnose covid-19 using small cohort of CT images. *NPJ digital medicine*, *4*(1), 1-10.

[Rawassizadeh '19] Rawassizadeh, R., Dobbins, C., Akbari, M. & Pazzani, M. (2019) Indexing Multivariate Data through Spatio -Temporal Event Detection and Clustering. *Sensors*, 19 (3), 448.

[Rawassizadeh '13] Rawassizadeh, R., Tomitsch, M., Wac, K., & Tjoa, A. M. (2013). UbiqLog: a generic mobile phone-based life-log framework. *Personal and ubiquitous computing*, *17*(4), 621-637.

[Tufte '01] Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (Vol. 2). Cheshire, CT: Graphics press.

[Abela '13] Abela, A. (2013) *Advanced Presentations by Design: Creating Communication that Drives Action*. Pfeiffer, 2nd edition.

[Knaflic '15] Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

[Wilke '19] Wilke, C. O. (2019). *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media.

[Rawassizadeh '15] Rawassizadeh, R., Momeni, E., Dobbins, C., Mirza-Babaei, P., & Rahnamoun, R. (2015). Lesson learned from collecting quantified self information via mobile and wearable devices. Journal of Sensor and Actuator Networks, 4(4), 315-335.

[Rumsey '15] Rumsey, D. J. (2015). U Can: statistics for dummies. John Wiley & Sons.

[Rumsey '09] Rumsey, D. J. (2009). Statistics II for Dummies. Wiley Publishing, Inc.

[Boslaugh '12] Boslaugh, S. (2012). Statistics in a nutshell: A desktop quick reference. " O'Reilly Media, Inc.".

[Skiena '17] Skiena, S. S. (2017). The Data Science Design Manual. Springer.

[Neyman '37] Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *236*(767), 333-380.

[Chandola '09] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 15.

[Aggarwal '01] Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *International conference on database theory* (pp. 420-434). Springer, Berlin, Heidelberg.

[Goodfellow '16] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). Cambridge: MIT press.

[Shapiro '65] Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*. 52 (3–4): 591–611.

[Bonferroni '36] Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936

[Sullivan '12] Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the P-value is not Enough. *Journal of graduate medical education*, *4*(3), 279-282.

[Cohen '92] Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155 - 159.

[Cliff '93] Cliff, Norman (1993). "Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions". *Psychological Bulletin*. 114 (3): 494-509.

[Stigler '92] Stigler, S. M. (1989). "Francis Galton's Account of the Invention of Correlation". *Statistical Science*. 4 (2), 73–79.

[Spearman '04] Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, *15*(1), 72-101.

[Shannon '49] Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell system technical journal*, *28*(4), 656-715.

[Bishop '06] Bishop, C.M. (2006) Pattern Recognition and Machine Learning. Springer.

[Kullback '51] Kullback, S., Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*. 22 (1): 79–86.

[Bruce '17] Bruce, P., & Bruce, A. (2017). Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly Media, Inc.

[Dunn '61] Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*. 56 (293), 52–64

[Aggrawal '15] Aggrawal, C. C. (2015). *Data Mining: The Textbook*. Springer.

[Rao '00] Rao, K. R., & Yip, P. C. (2000). *The Transform and Data Compression Handbook*. CRC press.

[Pighizzini '01] Pighizzini, G. (2001). How hard is computing the edit distance?. *Information and Computation*, *165*(1), 1-13.

[Ratanamahatana '05] Ratanamahatana, C. A., & Keogh, E. (2005). Three Myths about Dynamic Time Warping Data Mining. In *Proceedings of the 2005 SIAM International Conference on Data Mining* (pp. 506-510). Society for Industrial and Applied Mathematics.

[Xu '05] Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on neural networks*, *16*(3), 645-678.

[Ester '96] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD* (Vol. 96, No. 34, pp. 226-231).

[Ankerst '99] Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. In *ACM Sigmod record* (Vol. 28, No. 2, pp. 49-60). ACM.

[Sibson '73] Sibson, R. (1973). SLINK: An Optimally Efficient Algorithm for the Single-Link Cluster Method. *The Computer Journal*. British Computer Society. 16 (1): 30–34.

[Defays '77] Defays, D. (1977). An Efficient Algorithm for a Complete Link Method. *The Computer Journal*. British Computer Society. 20 (4): 364–366.

[Kaufman '09] Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to cluster analysis (Vol. 344). John Wiley & Sons.

[Larose '15] Larose, D. T. (2015). Data Mining and Predictive Analytics. John Wiley & Sons.

[Manning '10] Manning, C., Raghavan, P., & Schütze, H. (2010). Introduction to information retrieval. *Natural Language Engineering*, *16*(1), 100-103.

[Kurzweil '06] Kurzweil, R. (2006). The Singularity is Near: When Humans Transcend Biology. Gerald Duckworth & Co.

[Deza '09] Deza, M. M., & Deza, E. (2009). Encyclopedia of distances. In *Encyclopedia of distances* (pp. 1-583). Springer, Berlin, Heidelberg.

[Zhang '96] Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. In *ACM Sigmod Record* (Vol. 25, No. 2, pp. 103-114).

[Guha '98] Guha, S., Rastogi, R., & Shim, K. (1998). CURE: An Efficient Clustering Algorithm for Large Databases. In ACM Sigmod Record (Vol. 27, No. 2, pp. 73-84).

[Edgeworth '1908] Edgeworth, F. Y. (1908). "On the / errors of frequency-constants". *Journal of the Royal Statistical Society*. **71** (3): 499–512. doi:10.2307/2339293.

[Duda '73] Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. *A Wiley-Interscience Publication, New York: Wiley, 1973*.

[Dempster '77] Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. 39 (1): 1–38.

[Yule '36] Yule, G. U., Filon, L. N. G. (1936). "Karl Pearson. 1857-1936". *Obituary Notices of Fellows of the Royal Society*. 2 (5): 72

[Zadeh '96] Zadeh, L. A. (1996). Fuzzy sets. In Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh (pp. 394-432).

[Rawassizadeh '19] Rawassizadeh, R., Dobbins, C., Akbari, M., & Pazzani, M. (2019). Indexing multivariate mobile data through spatio-temporal event detection and clustering. *Sensors*, *19*(3), 448.

[Leskovec '14] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Mining of Massive Datasets. *Cambridge University Press*.

[Rousseeuw '87] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

[Dunn '74] Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, *4*(1), 95-104.

[Davies '79] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.

[Caliński '74] Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1-27.

[Sedgewick '13] Sedgewick, R., & Flajolet, P. (2013). An introduction to the analysis of algorithms. Pearson Education India.

[McDowell '16] McDowell, G. L. (2016). Cracking the Coding Interview: 189 Programming Questions and Solutions. CareerCup, LLC.

[Comer '79] Comer, D. (1979). Ubiquitous B-tree. *ACM Computing Surveys (CSUR)*, *11*(2), 121-137.

[Zaki '14] Zaki, M. J., Meira Jr, W., & Meira J., W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.

[Sedgewick '11] Sedgewick, R., & Wayne, K. (2011). *Algorithms, 4th Edition*. Addison-Wesley Professional.

[Dean '08] Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

[Coremen '09] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms, 3rd Edition*. MIT press.

[Ziv '78] Ziv, J., & Lempel, A. (1978). Compression of Individual Sequences via Variable-Rate Coding. *IEEE transactions on Information Theory*, *24*(5), 530-536.

[Welch '84] Welch, T. A. (1984). Technique for high-performance data compression. *Computer*, (52).

[Agrawal '94] Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In *Proceedings of 20th International Conference of Very Large Data Bases, VLDB* (Vol. 1215, pp. 487-499).

[Han '00] Han, J., Pei, J., & Yin, Y. (2000). Mining Frequent Patterns Without Candidate Generation. In *ACM Sigmod Record* (Vol. 29, No. 2, pp. 1-12).

[Zaki '97] Zaki, M. J., Parthasarathy, S., Ogihara, M., & Li, W. (1997). New Algorithms for Fast Discovery of Association Rules. In *Knowledge Discovery and Data Mining* (Vol. 97, pp. 283-286).

[Srikant '96] Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In *International Conference on Extending Database Technology* (pp. 1-17).

[Zaki '01] Zaki, M. J. (2001). SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning, 42(1-2), 31-60.

[Han '00] Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U., & Hsu, M. C. (2000). FreeSpan: Frequent Pattern-Projected Sequential Pattern Mining. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 355-359).

[Han '05] Han, J., Pei, J., & Yan, X. (2005). Sequential pattern mining by pattern-growth: Principles and extensions. In *Foundations and Advances in Data Mining* (pp. 183-220). Springer, Berlin, Heidelberg.

[Aggarwal '14] Aggarwal, C. C., & Han, J. (Eds.). (2014). *Frequent Pattern Mining*. Springer.

[Pei '04] Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. C. (2004). Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Transactions on Knowledge & Data Engineering*, (11), 1424-1440.

[Gonzalez '17] Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2017). *Digital Image Processing, 4th Edition*. Pearson-Prentice-Hall.

[Jurafsky '18] Jurafsky, D. and H. Martin, J. (2018) Speech and Language Processing (3rd edition draft) online: https://web.stanford.edu/~jurafsky/slp3/

[Ozdemir '18] Ozdemir, S., & Susarla, D. (2018). Feature Engineering Made Easy: Identify unique features from your dataset in order to build powerful machine learning systems. Packt Publishing Ltd.

[Chandrashekar '14] Chandrashekar, G., & Sahin, F. (2014). A Survey on Feature Selection Methods. *Computers & Electrical Engineering*, *40*(1), 16-28.

[Quinlan '86] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

[Tibshirani '96] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

[Rawassizadeh '15] Rawassizadeh, R., Price, B. A., & Petre, M. (2015). Wearables: Has the Age of Smartwatches Finally Arrived?. *Communications of the ACM*, *58*(1), 45-47.

[Box '64] Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society*, Series B, 26, 211-252.

[Ng '18] Ng, A. (2018). Machine Learning Yearning. *URL: http://www.mlyearning.org/(96)*.

[Mikolov '13] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

[Krishnamoorthy '13] Krishnamoorthy, A., & Menon, D. (2013). Matrix Inversion Using Cholesky Decomposition. In *2013 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* (pp. 70-72).

[Cai '08] Cai, D., He, X., & Han, J. (2008). Training Linear Discriminant Analysis in Linear Time. In *2008 IEEE 24th International Conference on Data Engineering* (pp. 209-217).

[Bellman '61] Bellman, R. (1961). Curse of dimensionality. *Adaptive control processes: a guided tour*. Princeton University Press, 3, 2.

[Deerwester '90] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391-407.

[Blei '03] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

[Harris '88] Harris, C. G., & Stephens, M. (1988). A Combined Corner and Edge Detector. In *Alvey vision conference*. Vol. 15, No. 50, (pp. 147-151).

[Matas '04] Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust Wide-baseline Stereo from Maximally Stable Extremal Regions. *Image and Vision Computing*, *22*(10), 761-767.

[Smith '97] Smith, S. W. (1997) *The Scientists & Engineer's Guide to Digital Signal Processing*, California Technical Pub. San Diego.

[Page '54] Page, E. S. (1954). Continuous Inspection Scheme. *Biometrika*. 41 (1/2): 100–115.

[Aminikhanghahi '17] Aminikhanghahi, S., & Cook, D. J. (2017). A Survey of Methods for Time Series Change Point Detection. *Knowledge and information systems*, *51*(2), 339-367.

[Dalal '05] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *International Conference on Computer Vision & Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886-893).

[Lowe '99] Lowe, D. G. (1999). Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision*, (Vol. 99, No. 2, pp. 1150-1157).

[Kornilov '18] Kornilov, A. S., & Safonov, I. V. (2018). An overview of watershed algorithm implementations in open source libraries. *Journal of Imaging*, *4*(10), 123.

[Heil '89] Heil, Christopher E., and David F. Walnut. Continuous and discrete wavelet transforms. *SIAM review* 31.4 (1989): 628-666.

[Haar '10] Haar, Alfréd (1910), Zur Theorie der orthogonalen Funktionensysteme, *Mathematische Annalen*, 69(3): 331–371.

[Keogh '01] Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. (2001). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, *3*(3), 263-286.

[Lin '03] Jessica, L., Keogh, E., Lonardi, S., and Chiu, B. (2003). A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pp. 2-11.

[Cholesky '10] Cholesky, A.L. (1910) On Numerical Solutions of Systems of Linear Equations. Available: http://bibnum.education.fr/mathematiques/algebre/sur-la-resolution-numerique-des-systemes-d-equations-lineaires

[Sra '06] Sra, S., & Dhillon, I. S. (2006). Generalized Nonnegative Matrix Approximations with Bregman divergences. In *Advances in Neural Information Processing Systems* (pp. 283-290).

[Golub '71] Golub, G. H., & Reinsch, C. (1971). Singular Value Decomposition and Least Squares Solutions. In *Linear Algebra*(pp. 134-151). Springer, Berlin, Heidelberg.

[Rehman '16] Ur Rehman, M. H., Liew, C. S., Abbas, A., Jayaraman, P. P., Wah, T. Y., & Khan, S. U. (2016). Big data reduction methods: a survey. *Data Science and Engineering*, *1*(4), 265-284.

[Kolda '09] Kolda, Tamara G., and Brett W. Bader. "Tensor decompositions and applications." *SIAM review* 51.3 (2009): 455-500.

[Hitchcock '27] Hitchcock, L. (1927) The expression of a tensor or a polyadic as a sum of products, Journal of Math and Physics, 6, pp. 164–189.

[Harshman '70] Harshman, R. A. (1970) Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics.* pp. 1-84.
[Carroll '70] Carroll, D. and Chang, J. J. (1970) Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of "Eckart-Young" Decomposition,

Psychometrika, 35, pp. 283–319.

[Håstad '90] Håstad, J. (1990). Tensor rank is NP-complete. *Journal of Algorithms*, *11*(4), 644-654.

[Tucker '66] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279-311.

[Kroonenberg '80] Kroonenberg, P. M., & De Leeuw, J. (1980). Principal Component Analysis of Three-mode Data by Means of Alternating Least Squares Algorithms. *Psychometrika*, *45*(1), 69-97.

[Oseledets '11] Oseledets, I. V. (2011). Tensor-Train Decomposition. *SIAM Journal on Scientific Computing*, *33*(5), 2295-2317.

[James '13] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York: springer.

[Domencich '75] Domencich, T. A., & McFadden, D. (1975). Urban travel demand-a behavioral analysis.

[Wald '45] Wald, A. (1945). Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, *16*(2), 117-186.

[Akaike '74] Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

[Schwarz '78] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

[Grus '19] Grus, J. (2019) Data Science from Scratch: First Principles with Python. *O'Reilly Media*.

[Domingos '12] Domingos, P. M. (2012). A few useful things to know about machine learning. *Communication of ACM 55*(10), 78-87.

[Thikhonov '43] Thikhonov, A. H. (1943). Об устойчивости обратных задач. In *Докл. Proceedings of the USSR Academy of Sciences* (Vol. 39, No. 5, pp. 195-198).

[Santosa '86] Santosa, F., Symes, W.W. (1986). Linear Inversion of Band-Limited Reflection Seismograms". *SIAM Journal on Scientific and Statistical Computing*. SIAM. 7 (4): 1307–1330.

[Tibshirani '96] Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society*. Series B (methodological). Wiley. 58 (1): 267–88.

[Efron '04] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407-499.

[Zou '05] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.

[Breiman '95] Breiman, L. (1995). Better Subset Regression using the Nonnegative Garrote. *Technometrics*, *37*(4), 373-384.

[Xiong '10] Xiong, S. (2010). Some Notes on the Nonnegative Garrote. *Technometrics*,*52*(3), 349-361.

[Robbins '51] Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400-407.

[Burkov '19] Burkov, A. (2019). The Hundred-Page Machine Learning Book. Andriy Burkov.

[Kochenderfer '19] Kochenderfer, M. J., & Wheeler, T. A. (2019). *Algorithms for optimization*. MIT Press.

[Kingma '14] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[Brownlee '16] Brownlee, J. (2016). Master Machine Learning Algorithms: discover how they work and implement them from scratch. Machine Learning Mastery.

[Cortes '95] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine learning*, *20*(3), 273-297.

[Bordes '05] Bordes, A., Ertekin, S., Weston, J., & Bottou, L. (2005). Fast kernel classifiers with online and active learning. *Journal of Machine Learning Research*, *6*(Sep), 1579-1619.

[Scholkopf '01] Scholkopf, B., & Smola, A. J. (2001). Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

[Kass '80] Kass, Gordon V.; An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics, Vol. 29, No. 2 (1980), pp. 119–127

[Quinlan '94] Quinlan, J. R. (1994). C4. 5: Programs for Machine Learning. *Elsevier*

[Breimanetal '84] Breimanetal, L. (1984). Classification and Regression Trees. Wadsworth, Belmont.

[Breiman '96] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.

[Schapire '90] Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197-227.

[Wolpert '92] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259.

[Breiman '05] Breiman, L. (2001). Random Forests. *Machine learning*, *45*(1), 5-32.

[Freund '95] Freund, Y., & Schapire, R. E. (1995). A Desicion-Theoretic Generalization of On-line Learning and an Application to Boosting. In *European conference on computational learning theory* (pp. 23-37). Springer, Berlin, Heidelberg.

[Friedman '01] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of statistics*, 1189-1232.

[Si '17] Si, S., Zhang, H., Keerthi, S. S., Mahajan, D., Dhillon, I. S., & Hsieh, C. J. (2017). Gradient boosted decision trees for high dimensional sparse output. In International conference on machine learning (pp. 3182-3190). PMLR.

[Mason '99] Mason, L., Baxter, J., Bartlett, P., & Frean, M. (1999, May). Boosting algorithms as gradient descent in function space. NIPS.

[Chen '15] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). XGBoost: extreme gradient boosting. *R package version 0.4-2*, *1*(4).

[Ke '17] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30,* 3146-3154.

[Prokhorenkova '18] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *31*.

[Boehmke '20] Boehmke, B., & Greenwell, B. M. (2019). Hands-on machine learning with R. CRC Press.

[Keshavarz '21] Keshavarz, H., Abadeh, M. S., & Rawassizadeh, R. (2021). SEFR: A fast linear-time classifier for ultra-low power devices. *arXiv preprint arXiv:2006.04620*.

[McCulloch '43] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133.

[Rosenblatt '58] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

[Minsky '69] Minsky, M., & Papert, S. A. (1969). Perceptrons: An introduction to computational geometry. MIT press.

[Ciresan '11] Ciresan, D. C. , Meier, U., Masci, J.. Gambardella, L. M.. Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification.
(2011) *International Joint Conference on Artificial Intelligence (IJCAI-2011)*

[Krizhevsky '12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[Aggarwal '18] Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*.

[Hofstadter '79] Hofstadter, D. R. (1979). Gödel, Escher, Bach: an eternal golden braid (Vol. 13). New York: Basic books.

[Rawassizadeh '19] Rawassizadeh, R., Sen, T., Kim, S. J., Meurisch, C., Keshavarz, H., Mühlhäuser, M., & Pazzani, M. (2019). Manifestation of virtual assistants and robots into daily life: Vision and challenges. *CCF Transactions on Pervasive Computing and Interaction*, 1-12.

[Chollet '18] Chollet, F. (2017). Deep Learning with Python, *Manning Publications CO*.

[Nair '10] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *In ICML*.

[Maas '13] Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

[Hellinger '09] Hellinger, E. (1909), Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen, *Journal für die reine und angewandte Mathematik* (in German), 136: 210–271

[Bharath '99] Bharath, B., & Borkar, V. S. (1999). Stochastic approximation algorithms: Overview and recent trends. *Sadhana*, *24*(4-5), 425-452.

[Polyak '64] Polyak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, *4*(5), 1-17.

[Nesterov '83] Nesterov, Y. A method of solving a convex programming problem with convergence rate O(1/sqr(k)). *Soviet Mathematics Doklady,* 27:372–376, 1983.

[Sutskever '13] Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning* (pp. 1139-1147).

[Duchi '11] Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, *12*(7).

[Zeiler '12] Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

[Hinton '12] Hinton, G., Srivastava, N., & Swersky, K. (2012). RMSprop: Divide the Gradient by a Running Average of its Recent Magnitude. *Neural Networks for Machine Learning, Lecture 6e*.

[Riedmiller '93] Riedmiller, M., & Braun, H. (1993). A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP algorithm. In *IEEE International Conference on Neural Networks* (pp. 586-591).

[Schmidt '20] Schmidt, R. M., Schneider, F., & Hennig, P. (2020). Descending through a Crowded Valley--Benchmarking Deep Learning Optimizers. arXiv preprint arXiv:2007.01547.

[Chen '19] Chen, J., Kyrillidis, A. Decaying Momentum Helps Neural Network Training. arXiv preprint arXiv:1910.04952, 2019.

[Zhang '17] Zhang, J., & Mitliagkas, I. (2017). Yellowfin and the art of momentum tuning. *arXiv preprint arXiv:1706.03471*.

[Schneider '19] Schneider, Frank, Lukas Balles, and Philipp Hennig. (2019) DeepOBS: A Deep Learning Optimizer Benchmark Suite. *International Conference on Learning Representations*.

[Kelley '60] Kelley, H. J. (1960). Gradient theory of optimal flight paths. *Ars Journal*, *30*(10), 947-954.

[Bryson '62] Bryson, A. E. (1961, April). A gradient method for optimizing multi-stage allocation processes. In *Proc. Harvard Univ. Symposium on digital computers and their applications* (Vol. 72).

[Rumelhart '86] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533-536.

[Glorot '10] Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256).

[He '15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

[Srivastava '14] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

[Gal '16] Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059).

[Konečný '16] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*.

[Warden '19] Warden, P., & Situnayake, D. (2019). TinyML: Machine learning with tensorflow lite on Arduino and ultra-low-power microcontrollers. O'Reilly Media, Inc.

[Huhel '59] Hubel, DH; Wiesel, T.N. (1959). "Receptive fields of single neurones in the cat's striate cortex". *J. Physiol*. 148 (3): 574–91.

[Fukushima '80] Fukushima, K. (1980). "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position" (PDF). *Biological Cybernetics*. **36** (4): 193–202.

[LeCun '89] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel (1989) Backpropagation Applied to Handwritten Zip Code Recognition.

[Deng '09] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on Computer Vision and Pattern Recognition* (pp. 248-255).

[Brown '20] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Agarwal, S. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

[Hochreiter '97] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[Greff '16] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.

[Chung '14] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[Karpathy '15] Karpathy, A. (2015). The Unreasonable Effectiveness of Recurrent Neural Networks. *Andrej Karpathy blog*, *21*, 23.

[Dumoulin '16] Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*.

[Chen '17] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[Ramesh '22] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*.

[Foster '19] Foster, D. (2019). Generative deep learning: teaching machines to paint, write, compose, and play. O'Reilly Media.

[Kohonen '82] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59-69.

[Tian '14] Tian, J., Azarian, M. H., & Pecht, M. (2014). Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. In *Proceedings of the European Conference of the Prognostics and Health Management Society* (pp. 1-9). Citeseer.

[LeCun '06] LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., & Huang, F. (2006). A tutorial on energy-based learning. *Predicting structured data*, *1*(0).

[Ackley '85] Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive science*, *9*(1), 147-169.

[Gibbs '02] Gibbs, J. W. (1902). *Elementary principles in statistical mechanics*: developed with especial reference to the rational foundations of thermodynamics. C. Scribner's Sons.

[Hinton '02] Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, *14*(8), 1771-1800.

[Hinton '95] Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The" wake-sleep" algorithm for unsupervised neural networks. *Science*, *268*(5214), 1158-1161.

[Wang '17] Wang, H., & Raj, B. (2017). On the Origin of Deep Learning. *arXiv preprint arXiv:1702.07800*.

[Hopfield '82] Hopfield, J. J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

[Bengio '09] Bengio, Y. (2009). Learning Deep Architectures for AI. Now Publishers Inc.

[Kingma '13] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Makhzani '13] Makhzani, A., & Frey, B. (2013). K-sparse autoencoders. *arXiv preprint arXiv:1312.5663*.

[Ng '11] Ng, A. (2011). Sparse autoencoder. CS294A Lecture notes, 72 (2011), 1-19.

[Vincent '10] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research, 11(12).

[Rifai '11] Rifai, S., Vincent, P., Muller, X., Glorot, X., & Bengio, Y. (2011). Contractive auto-encoders: Explicit invariance during feature extraction. In *International Conference of Machine Learning*.

[Bengio '07] Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2006). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, *19*.

[Hinton '06] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504-507.

[Foster '19] Foster, D. (2019). Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. O'Reilly Media.

[Langr '19] Langr, J., & Bok, V. (2019). GANs in action: deep learning with generative adversarial networks. Manning Publications.

[Ramesh '21] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092.

[Ronneberger '15] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241).

[Goodfellow '14] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

[Niemitalo '11] Niemitalo, O. (2010). A method for training artificial neural networks to generate missing data within a variable context. *Blog post, Internet Archive*. https://web.archive.org/web/201203122111546/http://yehar.com:80/blog/?p=167

[Li '13] Li, W., Gauci, M., & Groß, R. (2013). A coevolutionary approach to learn animal behavior through controlled interaction. I*n Proceedings of the 15th annual conference on Genetic and Evolutionary Computation* (pp. 223-230).

[Karras '19] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4401-4410).

[Salimans '16] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. arXiv preprint arXiv:1606.03498.

[Heusel '17] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500.

[Szegedy '16] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[Mirza '14] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[Rüschendorf '85] Rüschendorf, L. (1985). The Wasserstein distance and approximation theorems. Probability Theory and Related Fields, 70(1), 117-129.

[Gulrajani '17] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028.

[Isola '17] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1125-1134).

[Zhu '17] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

[Karras '20] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8110-8119).

[Huang '17] Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1501-1510).

[Karras '17] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

[Karras '21] Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021). Alias-Free Generative Adversarial Networks. *arXiv preprint arXiv:2106.12423*.

[Brownlee '19] Brownlee, J. (2019). Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image Translation. Machine Learning Mastery.

[Zakharov '19] Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9459-9468).

[Bengio '13] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828.

[Le-Khac '20] Le-Khac, P. H., Healy, G., & Smeaton, A. F. (2020). Contrastive representation learning: A framework and review. IEEE Access, 8, 193907-193934.

[Bell '10] Bell, G., & Gemmell, J. (2010). Your life, uploaded: The digital way to better memory, health, and productivity. Penguin.

[Chen '20 A] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

[Chen '20 B] Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. E. (2020). Big self-supervised models are strong semi-supervised learners. Advances in neural information processing systems, 33, 22243-22255.

[He '20] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9729-9738).

[Chen '20 C] Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297.

[Chopra '05] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 539-546).

[Schroff '15] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

[Radford '21] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In International Conference on Machine Learning (pp. 8748-8763). PMLR.

[Crowson '22] Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., & Raff, E. (2022). Vqgan-clip: Open domain image generation and editing with natural language guidance. *arXiv preprint arXiv:2204.08583*.

[Van Den Oord '17] Van Den Oord, A., & Vinyals, O. (2017). Neural discrete representation learning. *Advances in neural information processing systems*, *30*.

[Ramesh '21] Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. (2021) "Zero-shot text-to-image generation." In International Conference on Machine Learning, pp. 8821-8831. PMLR.

[Maddison '16] Maddison, C. J., Mnih, A., & Teh, Y. W. (2016). The concrete distribution: A continuous relaxation of discrete random variables. arXiv preprint arXiv:1611.00712.

[Jang '16] Jang, E., Gu, S., & Poole, B. (2016). Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144.

[Ramesh '22] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.

[Nichol '21] Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., ... & Chen, M. (2021). Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint arXiv:2112.10741.

[Rombach '22] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).

[Zhang '21] Zhang, K., Liang, J., Van Gool, L., & Timofte, R. (2021). Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4791-4800).

[Saharia '22] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., ... & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv preprint arXiv:2205.11487.

[Yu '22] Yu, J., Xu, Y., Koh, J. Y., Luong, T., Baid, G., Wang, Z., ... & Wu, Y. (2022). Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. arXiv preprint arXiv:2206.10789.

[Dhariwal '21] Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in Neural Information Processing Systems, 34, 8780-8794.

[Yu '21] Yu, J., Li, X., Koh, J. Y., Zhang, H., Pang, R., Qin, J., ... & Wu, Y. (2021). Vector-quantized image modeling with improved VQGAN. arXiv preprint arXiv:2110.04627.

[Vaswani '17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[Sutskever '14] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

[Cho '14] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[Bahdanau '14] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[Luong '15] Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

[Peters '18] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[Howard '18] Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

[Lewis '19] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

[Hendrycks '16] Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415.

[Liu '19] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[Lee '20] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234-1240.

[Lan '19] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

[Sanh '19] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

[Lample '19] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. arXiv preprint arXiv:1901.07291.

[Raffel '20] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21, 1-67.

[Radford '18] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[Zhu '15] Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international conference on computer vision (pp. 19-27).

[Liu '18] Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, L., & Shazeer, N. (2018). Generating wikipedia by summarizing long sequences. arXiv preprint arXiv:1801.10198.

[Radford '19] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

[Larochelle '08] Larochelle, H., Erhan, D., & Bengio, Y. (2008). Zero-data learning of new tasks. In AAAI (Vol. 1, No. 2, p. 3).

[Chang '08] Chang, M. W., Ratinov, L. A., Roth, D., & Srikumar, V. (2008). Importance of Semantic Representation: Dataless Classification. In Aaai (Vol. 2, pp. 830-835).

[Brown '20] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

[Child '19] Child, R., Gray, S., Radford, A., & Sutskever, I. (2019). Generating long sequences with sparse transformers. *arXiv preprint arXiv:1904.10509*.

[LeCun '89] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

[Devlin '18] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[Jelinek '77] Jelinek, F., Mercer, R. L., Bahl, L. R., & Baker, J. K. (1977). Perplexity—a measure of the difficulty of speech recognition tasks. The Journal of the Acoustical Society of America, 62(S1), S63-S63.

[Morris '04] Morris, A. C., Maier, V., & Green, P. (2004). From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition. In Eighth International Conference on Spoken Language Processing.

[Lin '04] Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).

[Banerjee '05] Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization (pp. 65-72).

[Papineni '02] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

[Wu '16] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

[Napoles '15] Napoles, C., Sakaguchi, K., Post, M., & Tetreault, J. (2015). Ground truth for grammatical error correction metrics. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 588-593).

[Sellam '20] Sellam, T., Das, D., & Parikh, A. P. (2020). BLEURT: Learning robust metrics for text generation. arXiv preprint arXiv:2004.04696.

[Simonyan '14] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[Hubel '59] Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology, 148(3), 574-591.

[Szegedy '15] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[Zhang '21] Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into deep learning. https://d2l.ai.

[He '16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[Loffe '15] Loffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). ICML.

[Ulyanov '17] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2017). Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

[Ba '16] Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. arXiv preprint arXiv:1607.06450.

[Wu '18] Wu, Y., & He, K. (2018). Group normalization. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[Li '17] Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2017). Visualizing the loss landscape of neural nets. arXiv preprint arXiv:1712.09913.

[Szegedy '17] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

[Dosovitskiy '20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[Sutton '18] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

[Page '99] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

[Kelinberg '99] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604-632.

[Blondel '08] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

[Traag '19] Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, *9*(1), 1-12.

[Traag '11] Traag, V. A., Van Dooren, P., & Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, *84*(1), 016114.

[Girshick '15] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[Uijlings '13] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2), 154-171.

[Van de Sande '11] Van de Sande, K. E., Uijlings, J. R., Gevers, T., & Smeulders, A. W. (2011). Segmentation as selective search for object recognition. In 2011 international conference on computer vision (pp. 1879-1886). IEEE.

[Girshick '15] Girshick, R. (2015). Fast R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[Ren '16] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99.

[Liu '16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

[Redmon '16] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[Redmon '17] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

[Farhadi '18] Farhadi, A., & Redmon, J. (2018). Yolov3: An incremental improvement. In Computer Vision and Pattern Recognition (pp. 1804-2767). Berlin/Heidelberg, Germany: Springer.

[Redmon '17] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

[Lin '17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

[Ge '21] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YoloX: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[Bochkovskiy '20] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[Jocher '20] Jocher, G. (2020) https://github.com/ultralytics/yolov5

[Luo '16] Luo, W., Li, Y., Urtasun, R., & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems (pp. 4905-4913).

[Long '15] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[He '17] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[Jaderberg '15] Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Spatial transformer networks. Advances in neural information processing systems, 28, 2017-2025.

[Chen '14] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

[Chen '17 A] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

[Chen '17 B] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[Chen '18] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

[Chollet '17] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[Kling '21] Kling, M. U. (2021). *Quality Land*. Éditions Actes Sud.

[Masi '18] Masi, I., Wu, Y., Hassner, T., & Natarajan, P. (2018). Deep face recognition: A survey. In 2018 31st IEEE SIBGRAPI conference on graphics, patterns and images (SIBGRAPI) (pp. 471-478).

[Taigman '14] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

[Sun '14] Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1898).

[Sun '15] Sun, Y., Wang, X., & Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2892-2900).

[Schroff '15] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).