

REFERENCES

- [1] Koetsier, J. (2021) There are now 8.9 million mobile apps, and China is 40% of mobile app spending, Forbes. Available at: <https://www.forbes.com/sites/johnkoetsier/2020/02/28/there-are-now-89-million-mobile-apps-and-china-is-40-of-mobile-app-spending/?sh=6623021d21dd> (Accessed: 12 March 2024).
- [2] Total number of websites (no date) Total number of Websites - Internet Live Stats. Available at: <https://www.internetlivestats.com/total-number-of-websites/> (Accessed: 12 March 2024).
- [3] T. A. Nguyen and C. Csallner, "Reverse Engineering Mobile Application User Interfaces with REMAUI (T)," 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2015, pp. 248-259, doi: 10.1109/ASE.2015.32.
- [4] K. Moran, C. Bernal-Cárdenas, M. Curcio, R. Bonett and D. Poshyvanyk, "Machine Learning-Based Prototyping of Graphical User Interfaces for Mobile Apps," in IEEE Transactions on Software Engineering, vol. 46, no. 2, pp. 196-221, 5 June. 2018, doi: 10.1109/TSE.2018.2844788.
- [5] Android apps on Google Play Google. Available at: https://play.google.com/store/games?hl=en_US&gl=US (Accessed: 12 March 2024).
- [6] Xu, Y., Bo, L., Sun, X., Li, B., Jiang, J. and Zhou, W., 2021. image2emmet: Automatic code generation from web user interface image. Journal of Software: Evolution and Process, 33(8).
- [7] Beltramelli, T.: pix2code: generating code from a graphical user interface screenshot. In: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 1–6 (2018).
- [8] D. Baulé, C. Wangenheim, A. Wangenheim and J. Hauck, "Recent Progress in Automated Code Generation from GUI Images Using Machine Learning Techniques", JUCS - Journal of Universal Computer Science, vol. 26, no. 9, pp. 1095-1127, 2020. Available: 10.3897/jucs.2020.058.
- [9] B. Aşiroğlu et al., "Automatic HTML Code Generation from Mock-Up Images Using Machine Learning Techniques," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 2019, pp. 1-4, doi: 10.1109/EBBT.2019.8741736.
- [10] W. Yang, M. R. Prasad, and T. Xie, "A grey-box approach for automated gui-model generation of mobile applications," in Proceedings of the 16th International Conference on Fundamental Approaches to Software Engineering, ser. FASE'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 250–265.
- [11] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on

Computer Vision and Pattern Recognition, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[12] Wang Ling, Edward Grefenstette, Karl Moritz Hermann, Tomáš Kočisk`y, Andrew Senior, FuminWang, and Phil Blunsom. 2016. Latent predictor networks for code generation. arXiv preprint arXiv:1603.06744 (2016).

[13] Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: A neural image caption generator." 2015. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156-3164.

[14] Xu, Kelvin, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. "Show, attend and tell: Neural image caption generation with visual attention." In International conference on machine learning, pp. 2048-2057. PMLR,

[15] Anderson, Peter, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. "Bottom-up and top-down attention for image captioning and visual question answering." 2018. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6077-6086.

[16] Download python Python.org. Available at: <https://www.python.org/downloads/> (Accessed: 12 March 2024).

[17] What is python used for? A beginner's guide Coursera. Available at: <https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python> (Accessed: 12 March 2024).

[18] PYTORCH documentation - PyTorch 2.2 documentation. Available at: <https://pytorch.org/docs/stable/index.html> (Accessed: 12 March 2024).

[19] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A., 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.

[20] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521, no. 7553 (2015): 436-444.

[21] Guo, Yanming, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. "Deep learning for visual understanding: A review." Neurocomputing 187 (2016): 27-48.

[22] Buduma, Nithin, Nikhil Buduma, and Joe Papa. Fundamentals of deep learning. " O'Reilly Media, Inc.", 2022.

[23] Gonzalez, Rafael C. Digital image processing. Pearson education India, 2009.

[24] Khan, Abdullah Ayub, Asif Ali Laghari, and Shafique Ahmed Awan. "Machine learning in computer vision: a review." EAI Endorsed Transactions on Scalable Information Systems 8, no. 32 (2021): e4-e4.

- [25] Proakis, John G. Digital signal processing: principles, algorithms, and applications, 4/E. Pearson Education India, 2007.
- [26] Christopher, D. Manning, and Schütze Hinrich. "Foundations of statistical natural language processing." (1999).
- [27] NLTK. Available at: <https://www.nltk.org/> (Accessed: 12 March 2024).
- [28] What is natural language processing? IBM. Available at: <https://www.ibm.com/topics/natural-language-processing> (Accessed: 12 March 2024).
- [29] Wang, Dong, and Thomas Fang Zheng. "Transfer learning for speech and language processing." In 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1225-1237. IEEE, 2015.
- [30] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [31] Graves, Alex, and Alex Graves. "Long short-term memory." Supervised sequence labelling with recurrent neural networks (2012): 37-45.
- [32] Zhao, Nannan. "Research on Web GUI Image Recognition System Based on Improved Convolutional Neural Network." Journal of Web Engineering 21, no. 5 (2022): 1727-1747.
- [33] Alzubi, Jafar A., Rachna Jain, Preeti Nagrath, Suresh Satapathy, Soham Taneja, and Paras Gupta. "Deep image captioning using an ensemble of CNN and LSTM based deep neural networks." Journal of Intelligent & Fuzzy Systems 40, no. 4 (2021): 5761-5769.
- [34] Sharma, Himanshu, Manmohan Agrahari, Sujeet Kumar Singh, Mohd Firoj, and Ravi Kumar Mishra. "Image captioning: a comprehensive survey." In 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), pp. 325-328. IEEE, 2020.
- [35] Feng, Yang, Lin Ma, Wei Liu, and Jiebo Luo. "Unsupervised image captioning." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4125-4134. 2019.
- [36] Herdade, Simao, Armin Kappeler, Kofi Boakye, and Joao Soares. "Image captioning: Transforming objects into words." Advances in neural information processing systems 32 (2019).
- [37] Reiter, Ehud. "A structured review of the validity of BLEU." Computational Linguistics 44, no. 3 (2018): 393-401.
- [38] Callison-Burch, Chris, Miles Osborne, and Philipp Koehn. "Re-evaluating the role of BLEU in machine translation research." In 11th conference of the european chapter of the association for computational linguistics, pp. 249-256. 2006.

- [39] Pal, Kuntal Kumar, and K. S. Sudeep. "Preprocessing for image classification by convolutional neural networks." In 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1778-1781. IEEE, 2016.
- [40] Vijayarani, S., Ms J. Ilamathi, and Ms Nithya. "Preprocessing techniques for text mining-an overview." *International Journal of Computer Science & Communication Networks* 5, no. 1 (2015): 7-16.
- [41] Kadhim, Ammar Ismael. "An evaluation of preprocessing techniques for text classification." *International Journal of Computer Science and Information Security (IJCSIS)* 16, no. 6 (2018): 22-32.
- [42] Blum, Avrim. "Machine learning theory." *Carnegie Melon University, School of Computer Science* 26 (2007).
- [43] Alzubi, Jafar, Anand Nayyar, and Akshi Kumar. "Machine learning from theory to algorithms: an overview." In *Journal of physics: conference series*, vol. 1142, p. 012012. IOP Publishing, 2018.
- [44] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science* 349, no. 6245 (2015): 255-260.
- [45] Simeone, Osvaldo. "A very brief introduction to machine learning with applications to communication systems." *IEEE Transactions on Cognitive Communications and Networking* 4, no. 4 (2018): 648-664.
- [46] L'heureux, Alexandra, Katarina Grolinger, Hany F. Elyamany, and Miriam AM Capretz. "Machine learning with big data: Challenges and approaches." *Ieee Access* 5 (2017): 7776-7797.
- [47] Sun, Shiliang, Zehui Cao, Han Zhu, and Jing Zhao. "A survey of optimization methods from a machine learning perspective." *IEEE transactions on cybernetics* 50, no. 8 (2019): 3668-3681.
- [48] Deka, Biplab, Zifeng Huang, Chad Franzen, Joshua Hirschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. "Rico: A mobile app dataset for building data-driven design applications." In *Proceedings of the 30th annual ACM symposium on user interface software and technology*, pp. 845-854. 2017.
- [49] Wu, Jason, Siyan Wang, Siman Shen, Yi-Hao Peng, Jeffrey Nichols, and Jeffrey P. Bigham. "Webui: A dataset for enhancing visual ui understanding with web semantics." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1-14. 2023.
- [50] Muñoz, Andrés, and Daniel Borrajo. "Computer User Interface Understanding. A New Dataset and a Learning Framework." *arXiv preprint arXiv:2403.10170* (2024).
- [51] Feng, Weixi, Wanrong Zhu, Tsu-jui Fu, Varun Jampani, Arjun Akula, Xuehai He, Sugato Basu, Xin Eric Wang, and William Yang Wang. "Layoutgpt:

Compositional visual planning and generation with large language models." *Advances in Neural Information Processing Systems* 36 (2024).

[52] Wang, Bryan, Gang Li, and Yang Li. "Enabling conversational interaction with mobile ui using large language models." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1-17. 2023.

[53] Ko, Hyung-Kwon, Hyeon Jeon, Gwanmo Park, Dae Hyun Kim, Nam Wook Kim, Juho Kim, and Jinwook Seo. "Natural language dataset generation framework for visualizations powered by large language models." *arXiv preprint arXiv:2309.10245* (2023).

[54] Wu, Jason, Kashyap Todi, Joannes Chan, Brad A. Myers, and Ben Lafreniere. "FrameKit: A Tool for Authoring Adaptive UIs Using Keyframes." In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pp. 660-674. 2024.

[55] Salvi, Giuseppe. "Web UI code generation: a transformer-based model applied to real-world screenshots." PhD diss., Politecnico di Torino, 2023.

[56] Jiang, Yue, Yuwen Lu, Clara Kliman-Silver, Christof Lutteroth, Toby Jia-Jun Li, Jeffrey Nichols, and Wolfgang Stuerzlinger. "Computational Methodologies for Understanding, Automating, and Evaluating User Interfaces." In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA'24)*. 2024.

[57] McCann, Michael T., Kyong Hwan Jin, and Michael Unser. "Convolutional neural networks for inverse problems in imaging: A review." *IEEE Signal Processing Magazine* 34, no. 6 (2017): 85-95.

[58] Li, Zewen, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. "A survey of convolutional neural networks: analysis, applications, and prospects." *IEEE transactions on neural networks and learning systems* 33, no. 12 (2021): 6999-7019.

[59] Aloysius, Neena, and M. Geetha. "A review on deep convolutional neural networks." In *2017 international conference on communication and signal processing (ICCSP)*, pp. 0588-0592. IEEE, 2017.

[60] Ciancetta, Fabrizio, Giovanni Bucci, Edoardo Fiorucci, Simone Mari, and Andrea Fioravanti. "A new convolutional neural network-based system for NILM applications." *IEEE Transactions on Instrumentation and Measurement* 70 (2020): 1-12.

[61] Tu, Fengbin, Shouyi Yin, Peng Ouyang, Shibin Tang, Leibo Liu, and Shaojun Wei. "Deep convolutional neural network architecture with reconfigurable computation patterns." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 25, no. 8 (2017): 2220-2233.

[62] Le Callet, Patrick, Christian Viard-Gaudin, and Dominique Barba. "A convolutional neural network approach for objective video quality assessment." *IEEE transactions on neural networks* 17, no. 5 (2006): 1316-1327.

- [63] Nebauer, Claus. "Evaluation of convolutional neural networks for visual recognition." *IEEE transactions on neural networks* 9, no. 4 (1998): 685-696.
- [64] Pustokhin, Denis A., Irina V. Pustokhina, Phuoc Nguyen Dinh, Son Van Phan, Gia Nhu Nguyen, Gyanendra Prasad Joshi, and Shankar K. "An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19." *Journal of Applied Statistics* 50, no. 3 (2023): 477-494.
- [65] Zakaria, Neili, Fezari Mohamed, Redjati Abdelghani, and Kenneth Sundaraj. "Three resnet deep learning architectures applied in pulmonary pathologies classification." In *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, pp. 1-8. IEEE, 2021.
- [66] Demir, Ahmet, Feyza Yilmaz, and Onur Kose. "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3." In *2019 medical technologies congress (TIPTEKNO)*, pp. 1-4. IEEE, 2019.
- [67] Kiliç, Şafak, Iman Askerzade, and Yilmaz Kaya. "Using ResNet transfer deep learning methods in person identification according to physical actions." *IEEE Access* 8 (2020): 220364-220373.
- [68] Mascarenhas, Sheldon, and Mukul Agarwal. "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification." In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, vol. 1, pp. 96-99. IEEE, 2021.
- [69] Zakaria, Neili, Fezari Mohamed, Redjati Abdelghani, and Kenneth Sundaraj. "VGG16, ResNet-50, and GoogLeNet deep learning architecture for breathing sound classification: a comparative study." In *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, pp. 1-6. IEEE, 2021.
- [70] Ye, Mu, Ni Ruiwen, Zhang Chang, Gong He, Hu Tianli, Li Shijun, Sun Yu, Zhang Tong, and Guo Ying. "A lightweight model of VGG-16 for remote sensing image classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021): 6916-6922.
- [71] Shaha, Manali, and Meenakshi Pawar. "Transfer learning for image classification." In *2018 second international conference on electronics, communication and aerospace technology (ICECA)*, pp. 656-660. IEEE, 2018.
- [72] Naveen, P., and B. Diwan. "Pre-trained VGG-16 with CNN Architecture to classify X-Rays images into Normal or Pneumonia." In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 102-105. IEEE, 2021.
- [73] Wang, Zhumei, Xing Su, and Zhiming Ding. "Long-term traffic prediction based on lstm encoder-decoder architecture." *IEEE Transactions on Intelligent Transportation Systems* 22, no. 10 (2020): 6561-6571.
- [74] Gers, Felix A., and E. Schmidhuber. "LSTM recurrent networks learn simple context-free and context-sensitive languages." *IEEE transactions on neural networks* 12, no. 6 (2001): 1333-1340.

- [75] Wang, Minsi, Li Song, Xiaokang Yang, and Chuanfei Luo. "A parallel-fusion RNN-LSTM architecture for image caption generation." In 2016 IEEE international conference on image processing (ICIP), pp. 4448-4452. IEEE, 2016.
- [76] Castro, Roberto, Israel Pineda, Wansu Lim, and Manuel Eugenio Morocho-Cayamcela. "Deep learning approaches based on transformer architectures for image captioning tasks." IEEE Access 10 (2022): 33679-33694.
- [77] Ghandi, Taraneh, Hamidreza Pourreza, and Hamidreza Mahyar. "Deep learning approaches on image captioning: A review." ACM Computing Surveys 56, no. 3 (2023): 1-39.
- [78] Sharma, Himanshu, Manmohan Agrahari, Sujeet Kumar Singh, Mohd Firoj, and Ravi Kumar Mishra. "Image captioning: a comprehensive survey." In 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), pp. 325-328. IEEE, 2020.
- [79] Cui, Yin, Guandao Yang, Andreas Veit, Xun Huang, and Serge Belongie. "Learning to evaluate image captioning." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5804-5812. 2018.
- [80] He, Xiaodong, and Li Deng. "Deep learning for image-to-text generation: A technical overview." IEEE Signal Processing Magazine 34, no. 6 (2017): 109-116.
- [81] Wang, Cheng, Haojin Yang, Christian Bartz, and Christoph Meinel. "Image captioning with deep bidirectional LSTMs." In Proceedings of the 24th ACM international conference on Multimedia, pp. 988-997. 2016.
- [82] Aneja, Jyoti, Aditya Deshpande, and Alexander G. Schwing. "Convolutional image captioning." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5561-5570. 2018.
- [83] Lai, Taotao, Riqing Chen, Changcai Yang, Qiming Li, Hamido Fujita, Alireza Sadri, and Hanzi Wang. "Efficient robust model fitting for multistructure data using global greedy search." IEEE Transactions on Cybernetics 50, no. 7 (2019): 3294-3306.
- [84] Chickering, David Maxwell. "Optimal structure identification with greedy search." Journal of machine learning research 3, no. Nov (2002): 507-554.
- [85] Wang, Xiao, Zhe Chen, Bo Jiang, Jin Tang, Bin Luo, and Dacheng Tao. "Beyond greedy search: tracking by multi-agent reinforcement learning-based beam search." IEEE Transactions on Image Processing 31 (2022): 6239-6254.
- [86] Vijayakumar, Ashwin K., Michael Cogswell, Ramprasath R. Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. "Diverse beam search: Decoding diverse solutions from neural sequence models." arXiv preprint arXiv:1610.02424 (2016).
- [87] Chen, Yun, Victor OK Li, Kyunghyun Cho, and Samuel R. Bowman. "A stable and effective learning strategy for trainable greedy decoding." arXiv preprint arXiv:1804.07915 (2018).

- [88] Post, Matt. "A call for clarity in reporting BLEU scores." arXiv preprint arXiv:1804.08771 (2018).
- [89] Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "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. 2002.
- [90] Yang, Muyun, Junguo Zhu, Jufeng Li, Lixin Wang, Haoliang Qi, Sheng Li, and Liu Daxin. "Extending BLEU evaluation method with linguistic weight." In 2008 The 9th International Conference for Young Computer Scientists, pp. 1683-1688. IEEE, 2008.