References

[1] François Chollet. *Xception: Deep Learning with Depthwise Separable Convolutions*. 2016. arXiv: 1610.02357 [cs.CV].

[2] Alexis Conneau et al. "Very deep convolutional networks for text classification". In: *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*. 2017. ISBN: 9781510838604. DOI: 10.18653/v1/e17-1104.

[3] Jacob Devlin et al. "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*. 2019. ISBN: 9781950737130. arXiv: 1810.04805.

[4] Yoon Kim et al. "Character-Aware neural language models". In: *30th AAAI Conference on Artificial Intelligence, AAAI 2016*. 2016. ISBN: 9781577357605. arXiv: 1508.06615.

[5] Alex R. Kuefler. *Merging Recurrence and Inception-Like Convolution for Sentiment Analysis*. 2016. URL: https://cs224d.stanford.edu/reports/akuefler.pdf (visited on 06/17/2020).

[6] Girish Limaye, Manish Pandit, and Sawal Vinay. *BertNet: Combining BERT language representation with Attention and CNN for Reading Comprehension*. 2019. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15783457.pdf (visited on 06/17/2020).

[7] Min Lin, Qiang Chen, and Shuicheng Yan. "Network in network". In: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*. 2014.

[8] Xiaofei Ma et al. *Universal Text Representation from BERT: An Empirical Study*. 2019. arXiv: 1910.07973 [cs.CL].

[9] Matthew E. Peters et al. *Deep contextualized word representations*. 2018. arXiv: 1802.05365 [cs.CL].

[10] Pranav Rajpurkar et al. "SQuad: 100,000+ questions for machine comprehension of text". In: *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*. 2016. ISBN: 9781945626258.

[11] Prajit Ramachandran et al. *Stand-Alone Self-Attention in Vision Models*. 2019. arXiv: 1906.05909 [cs.CV].

[12] Danny Takeuchi and Kevin Tran. *Improving SQUAD 2.0 Performance using BERT + X*. 2019. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15737384.pdf (visited on 06/17/2020).

[13] Ian Tenney, Dipanjan Das, and Ellie Pavlick. "BERT rediscovers the classical NLP pipeline". In: *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*. 2020. ISBN: 9781950737482. DOI: 10.18653/v1/p19-1452. arXiv: 1905.05950.

[14] Ashish Vaswani et al. "Attention is all you need". In: *Advances in Neural Information Processing Systems*. 2017.