

References

- [1] Alexis Conneau et al. “Very deep convolutional networks for text classification”. In: (2017). DOI: [10.18653/v1/e17-1104](https://doi.org/10.18653/v1/e17-1104).
- [2] Jacob Devlin et al. “BERT: Pre-training of deep bidirectional transformers for language understanding”. In: (2019). arXiv: 1810.04805.
- [3] Yoon Kim et al. “Character-Aware neural language models”. In: (2016). arXiv: 1508.06615.
- [4] 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).
- [5] 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).
- [6] Min Lin, Qiang Chen, and Shuicheng Yan. “Network in network”. In: (2014).
- [7] Pranav Rajpurkar et al. “SQuad: 100,000+ questions for machine comprehension of text”. In: (2016).
- [8] 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).
- [9] Ashish Vaswani et al. “Attention is all you need”. In: (2017). ISSN: 10495258.