Reference Table:

For the theory I used the “attention is all you need paper”

|  |  |  |  |
| --- | --- | --- | --- |
| Nr | Reference | What value I made over the reference | Referenced code |
| 1 | https://github.com/aladdinpersson/Machine-Learning-Collection/blob/master/ML/Pytorch/more\_advanced/transformer\_from\_scratch/transformer\_from\_scratch.py | I added a classification head in the encoder and reduced sum | Line 27-109  Line 121-134 |