# Explanation for the TODO tasks within the script

TODO: In the following lines we add a (trainable) positional encoding to our representation. Why is this needed? Can you think of another similar task where the positional encoding would not be necessary?

Positional encoding is needed because the positions of X and Y play a pivotal role. For every random sequence generated the positions of X and Y are fixed therefore positional encoding needs to be introduced to cater the aspect of position.

Positional encoding may not be necessary in some kind of a computer vision task where the counting of objects is the objective. It wouldn't matter where the object appears in the image as long as it is being counted for.

TODO: Summarize the idea of attention in a few sentences. What are Q, K and V?

Attention is a cognitive process of selectively concentrating on one or a few things while ignoring others therefore it is a mechanism in deep learning which attempts to implement the same action of selectively concentrating on a few relevant things, while ignoring others.

To calculate attention a query (Q) and keys (K) are used. The attention network compares the query (Q) with the keys (K) and get scores/weights for the values (V). Each score/weight is the relevancy between the query and each key. And networks reweigh the values with the scores/weights, and take the summation of the reweighted values.