Here we have a look at the input, embedding and prediction networks that constitute almost all the deep learning models. The thing to notice here is that the first and the last part of the model (i.e. input and prediction networks) essentially have the same functional form for all the models.

The input maps a sequence of L amino acids (1-D) to a L x 20 binary array (2-D). The columns of Lx20 are one-hot encoding of the input amino acid. For example, A is represented are [1,0,0, . . .] and similarly, ADE is also represented as a set of one hot encoded vectors

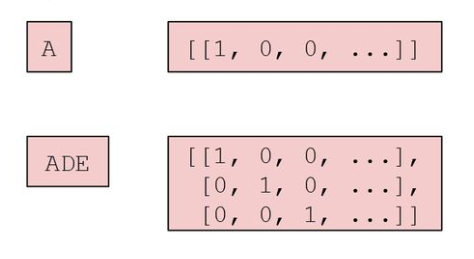


Figure - One hot Encoding of input vector

The encoded one hot vector is then padded to the length of the longest sequence in the batch with all-zero vectors appended on the right. For example, the vectors above are padded with zero like mentioned below

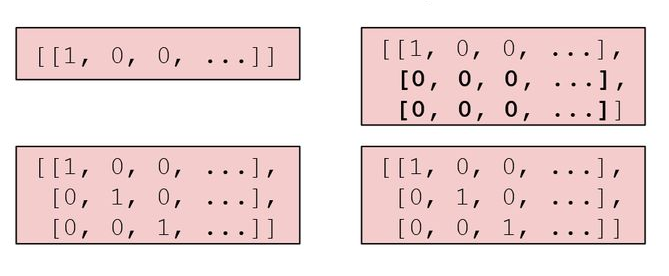


Figure - Encoded vectors with padded zeroes

The next network, that is the embedding network maps the Lx20 array (containing the padded one hot encoded vectors) to a bigger LxF array which contains embedding for each and every sequence. Here F is a tunable parameter (the default value of F for ProtCNN is 1100). For example, the padded one hot encoded vectors in the Figure 2 are embedded as below

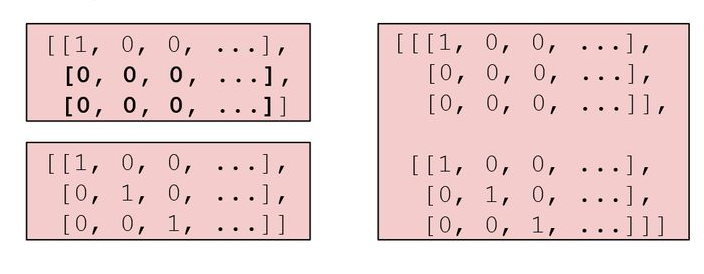


Figure - Batch of embbeded one-hot encoded vectors

For embedding purposes, ProtCNN uses residues of convolutional neural network (ResNets), which means that each layer feeds into the next layer and directly into the layers about 2–3 hops away. For a better understanding refer the image of a single residual block in Figure 4.

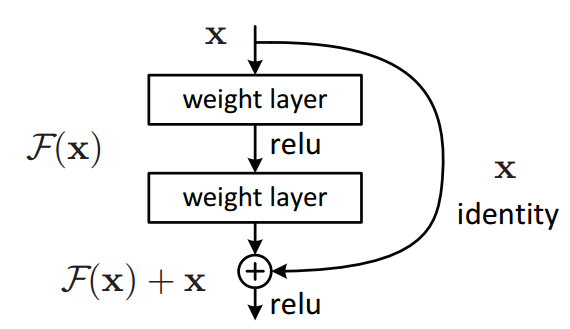


Figure - Single Residual Block

For this project we have used ResNets with dilated convolutions, the architecture of which is depicted in Figure 5. Dilated convolutions are just a convolution applied on the input but with defined gaps or applied after being dilated (See Figure 6)

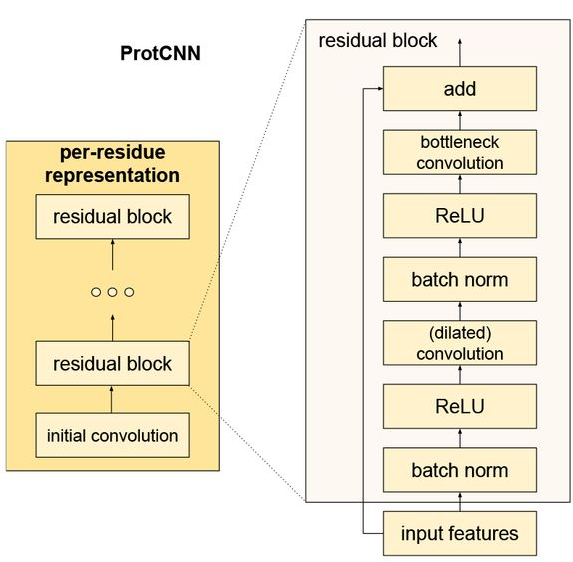


Figure - ResNet architecture used in the project

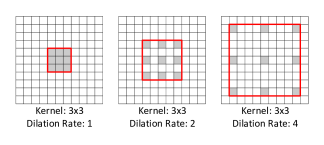


Figure - Dilated Convolutions

All in all, to summarize and give a short overview of model and the entire functioning of the model we can have a look at the Figure 7.

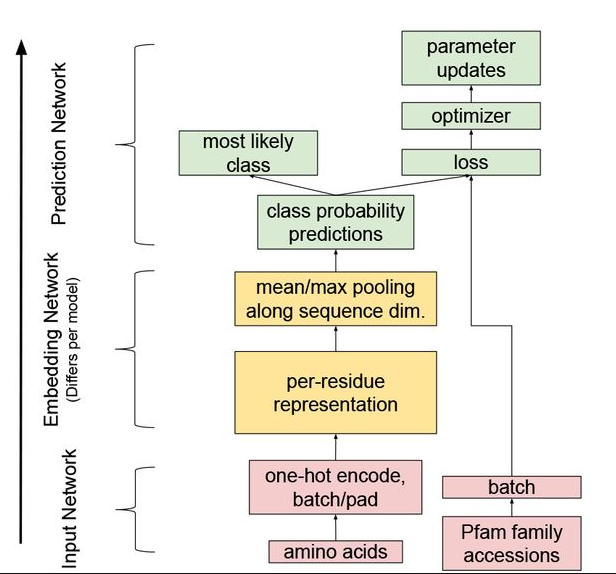


Figure - Overview