Wifi Sensing Project Update - TCN Model

Name: Varun Bharti

Roll No.: 2022562

Advisor : Prof. Mukulika Maity

Date: 25th June 2024

Course: IP

INDRAPRASTHA INSTITUTE *of* INFORMATION TECHNOLOGY
DELHI

Why TCN

1. Our data is Time series data . For this , usage of RNNs have been traditional method . Issues with RNNs : -
   a. vanishing/exploding gradient problems
   b. lack of long-term memory (Models like LSTMs only have memory of 4 seconds)
2. Hence, to overcome this problem, we use the model architecture of Temporal Convolutional Network (TCN). TCN has main advantage of memory : -

   Since TCN model calculates the set of latent activitations and uploads it per layer , it works much faster than RNNs

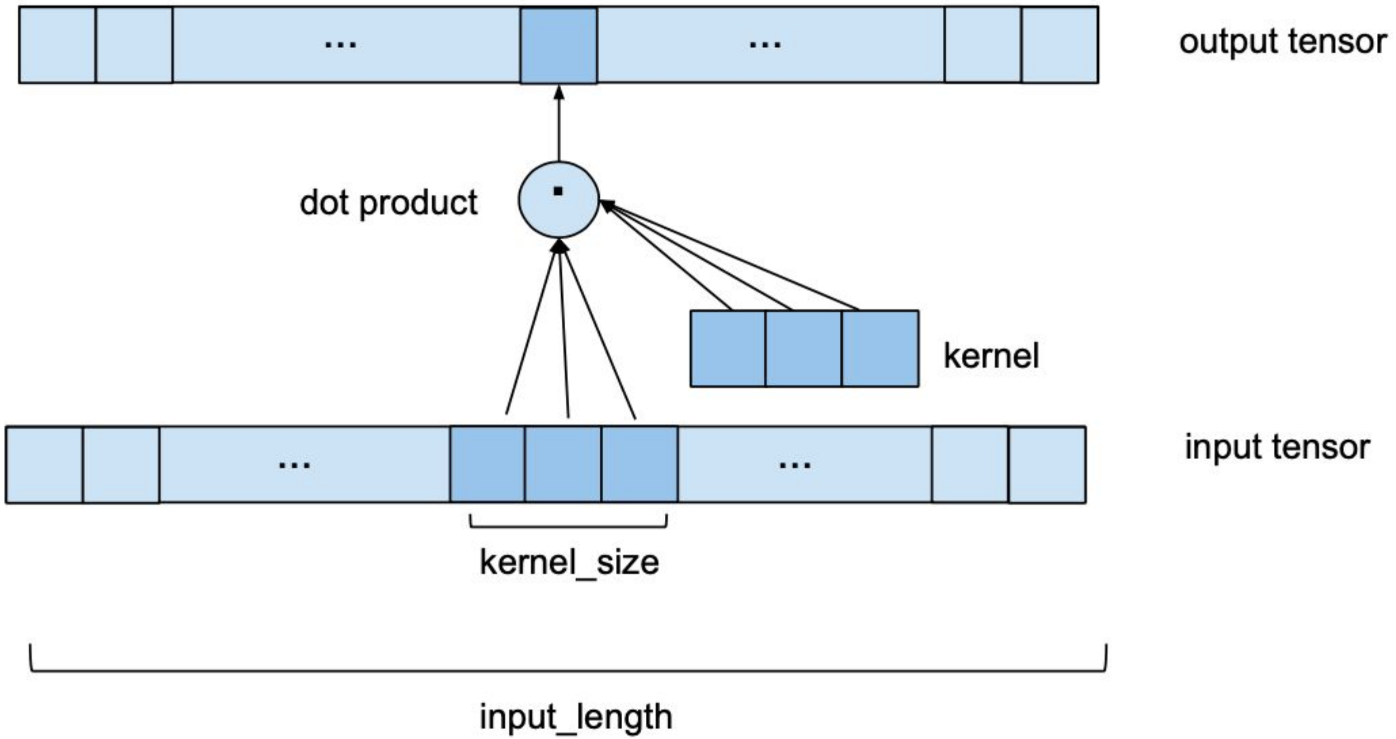
TCN model : Overview

1. TCNs are a type of convolutional neural network (CNN) that has been adapted for use with sequential data. In short , TCN can be viewed as : -

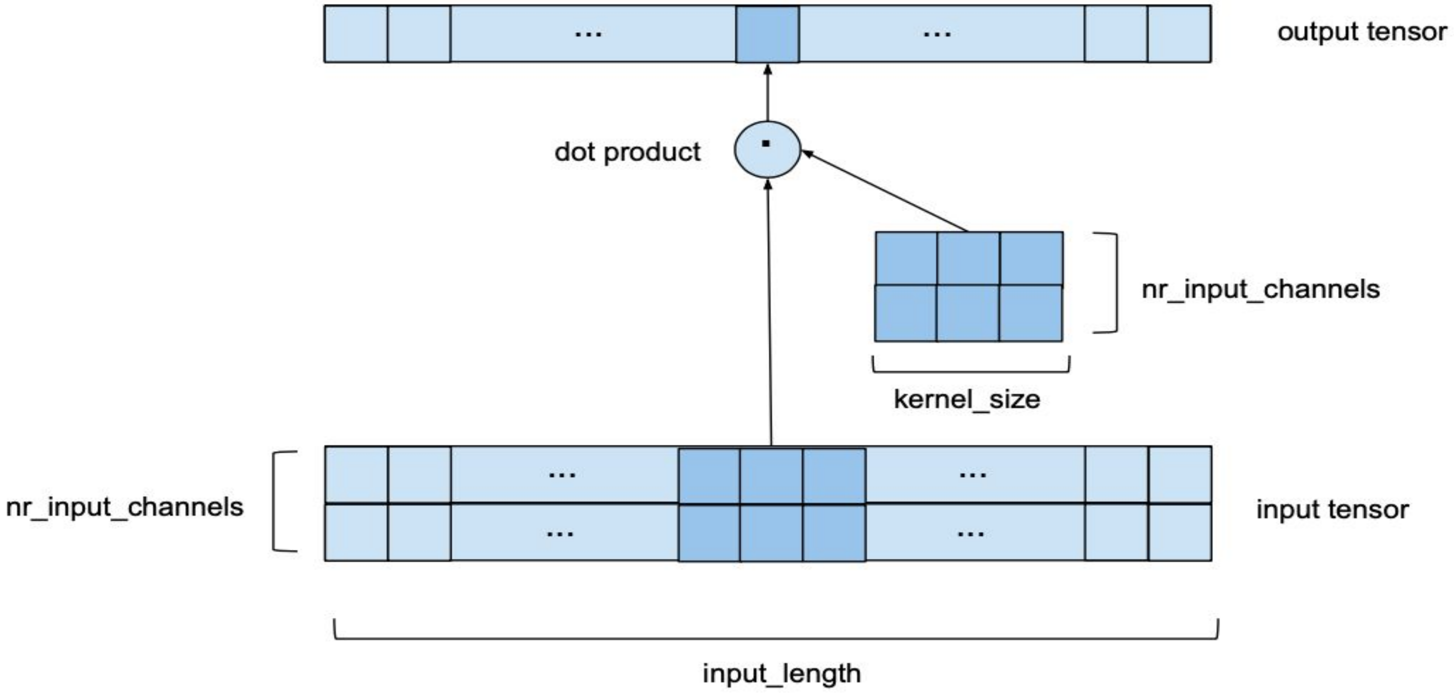
   TCN Model = Dilated (1-D FCN+ Casual Convolutionals)

2. Most Distinguishing features :
   a. No leakage of information from 'past' to 'present' (due to Casual Convolutionals)
   b. Input sequence and Output sequence are of same size (due to 1-D FCN)
   c. Large receptive field without actual increase in size (Dilation)
3. It takes as input a 3-D tensor and also outputs a 3-D tensor. The shape are : $\text{Shape}_{\text{Input/ Output}}$ = (batch_size, input_length, input_size)

TCN model : Overview



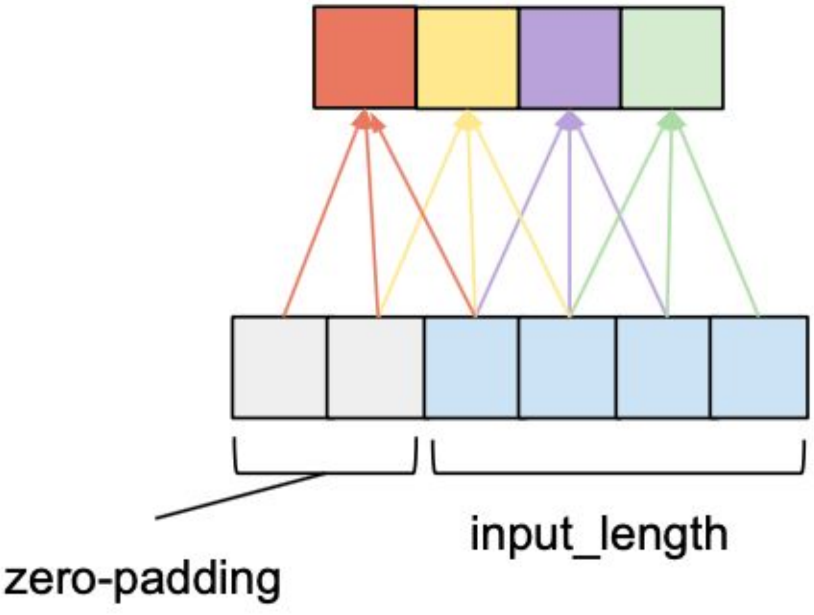
TCN model : Overview



1-D FCN Layer

1. This 1- Dimensional Fully Convolutional Layer is used to maintain same size of each layer as the input layer.
2. How ??

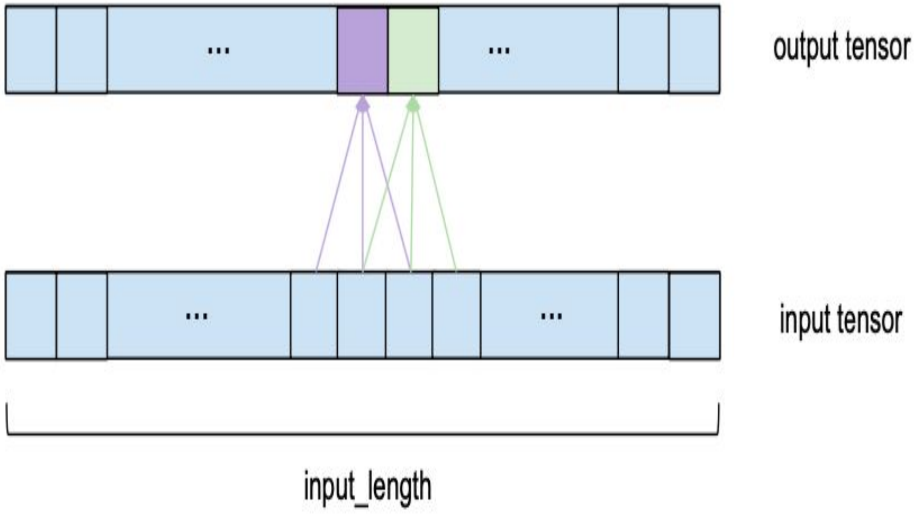
   In this architecture, each hidden layer is of same length due to addition of zero padding of length : $Size_{kernel} - 1$



Casual Convolutional

1. Used to maintain no leakage of Information i.e the unintended flow of information from future time steps to past time steps during the training process.
2. How ?

   Output at time = t is computed only using elements from that time + information from previous layers

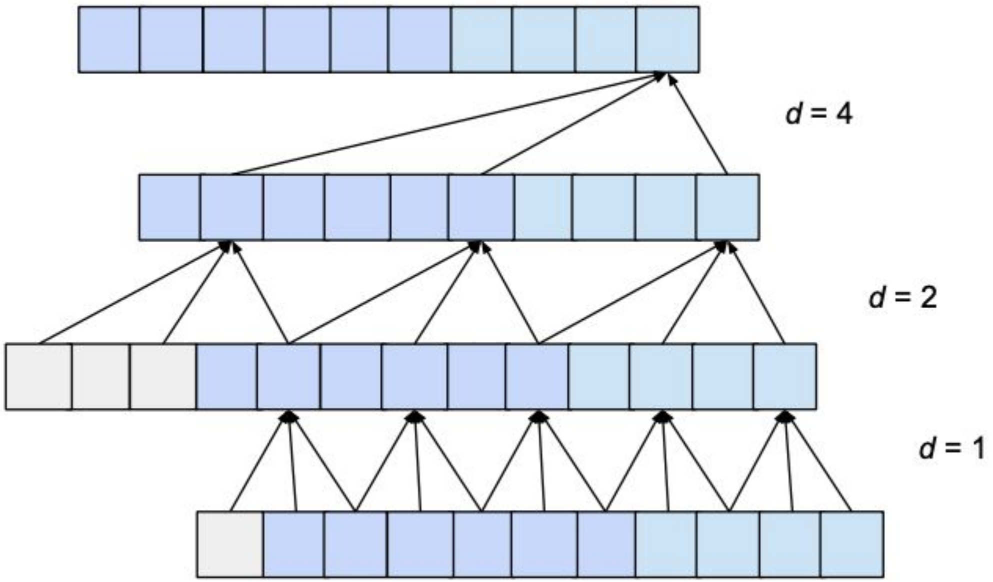


Dilation in TCN

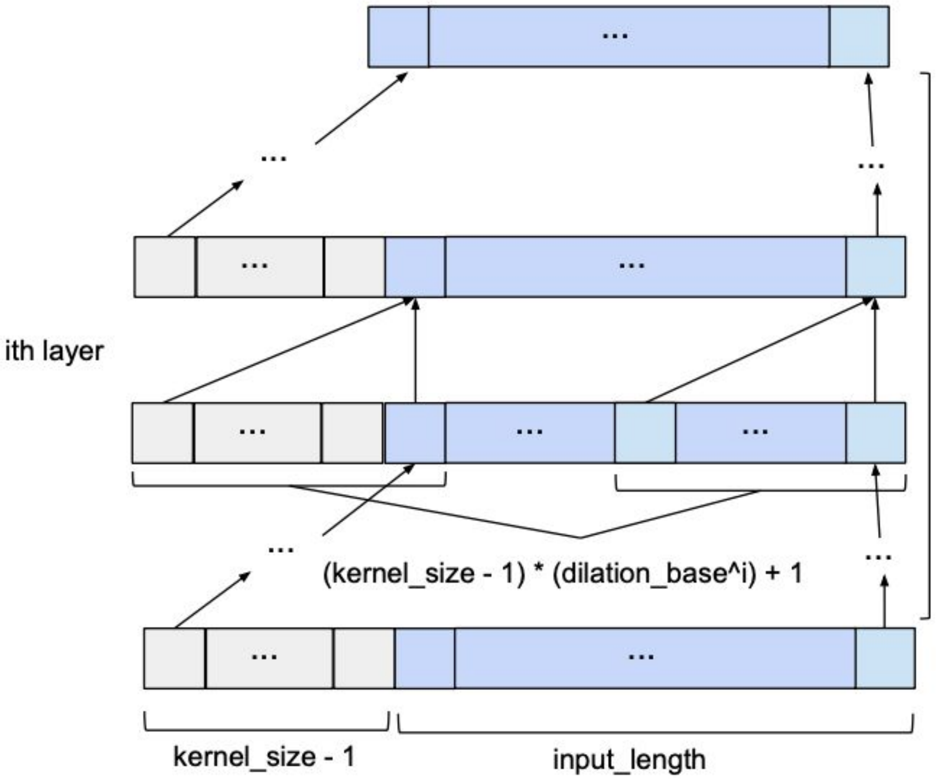
Why Dilation ?

In Simple Casual Convolutional , we look back at history as a fn which increase in size linear to the depth of the network. Hence dilation is used to increase the size of receptive field and capture larger history

Due to dilation , the size shows as exponential growth

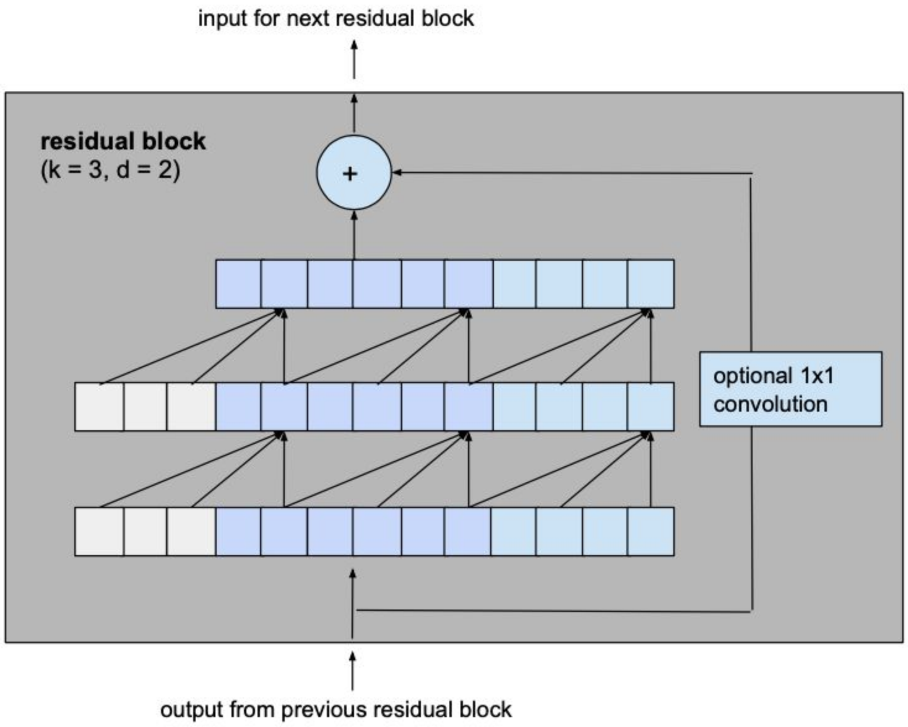


Basic TCN model Architecture



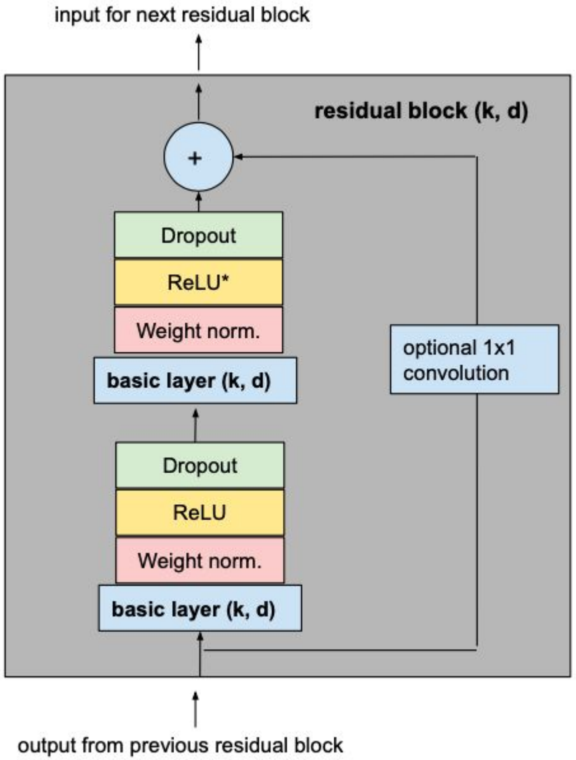
Residual Connection

1. This modification is done to solve the problem of 'vanishing and exploding gradient' in TCN model . It provides a bypass route for information to flow from layer (x) to output(o)
2. The output of the two convolutional layers will be added to the input of the residual block to produce the input for the next block.This change affects the calculus for the minimum number of required layers for full coverage

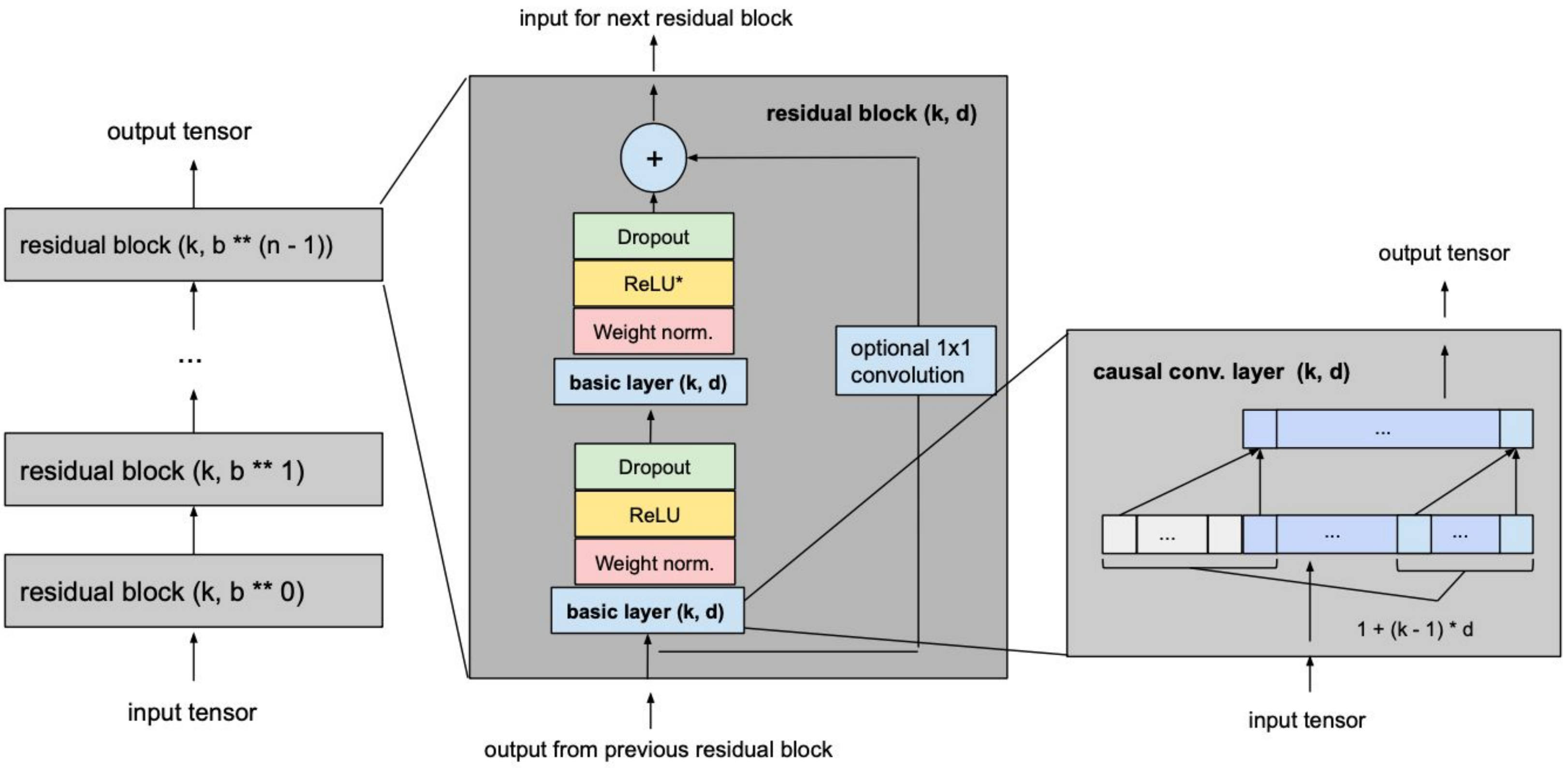


Other Modifications

1. Activation functions are need to be added on top of the convolutional layers to introduce non-linearities. ReLU activations are added to the residual blocks after both convolutional layers.
2. To normalize the input of hidden layers (which counteracts the exploding gradient problem among other things), weight normalization is applied to every convolutional layer.
3. In order to prevent overfitting, regularization is introduced via dropout after every convolutional layer in every residual block. The following figure shows the final residual block



Final TCN architecture



Resources and References

1. Paper on Basic TCN model : https://arxiv.org/pdf/1803.01271
2. Paper on TCN model for Action Segmentation : https://arxiv.org/pdf/1608.08242
3. Article Explaining the TCN model architecture : https://unit8.com/resources/temporal-convolutional-networks-and-forecasting/
4. Model Implementation using Pytorch : https://github.com/locuslab/TCN/blob/master/TCN/tcn.py
5. Model Implementation using Keras: https://github.com/philipperemy/keras-tcn

Thank You

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI