REGRESSION ANALYSIS FOR BOUNDARY VALUE PROBLEMS IN SOLID MECHANICS USING RESIDUAL CONVOLUTIONAL NEURAL NETWORKS

Vivek Chavan

MSc. Computer-Aided Conception and Production in Mechanical Engineering
Matriculation Number: 404543

This paper explores the use of Convolutional Neural Networks for the regression analysis of Boundary Value Problems in structural mechanics. In recent years, deep Convolutional Neural Networks have proven to be immensely successful at approximating complex functions in the field of computer vision applications. The current problem is perceived as one such application, where the aim is to predict standard FE simulation outputs using available inputs. This paper presents the use of 1D and 2D convolutions along sequential and spatial dimensions, respectively, for comparing the performance and accuracy of regression. For the 2D convolutions, two different networks were tested: a plain network and a Residual network (making use of skip connections) The paper also discusses how ‘going deep’ with neural networks is not necessarily the most optimal strategy for improving its performance. We used a well-tuned Recurrent network as the benchmark for comparing the convolutional networks.

KEYWORDS: Supervised learning, 1D and 2D convolutions, Skip Connection, residual block, LSTM.

INTRODUCTION: In recent years, there has been a growing interest in implementing machine learning to mechanical engineering applications, esp. those dealing with computer simulations that require massive computational resources. The argument in favor of machine learning is that the solver (Computer-Aided Engineering-CAE, Finite Element Analysis-FEA) does not play an active role in solving the given problem, but instead solves the given Boundary Value Problem (BVP) based on the FEA code. The FEA code is impervious to the parameters of a given problem (materials, geometry, applied loads) such that if the same problem were submitted to the solver multiple times, it would take the solver the same amount of time to solve it each time. Thus, the patent conclusion is that the machine experience gained during the computation is lost. It is this aspect of simulation sciences that machine learning seeks to explore and exploit. The aim is to train machine learning models on the training data that comprises the results of standard CAE assessments and use the model to predict structure behavior under varying loads with a reasonable degree of accuracy without having to rerun the CAE assessment on the test set.

Koeppe et al. [1] demonstrated the use of supervised learning to train models to create meta-elements (intelligent elements) that can accurately compute the non-linear response of a mechanical system at a structural level. The current paper deals with a single intelligent element in the form of a square continuum with elastoplastic material behavior subjected to deformation. The resultant forces are then predicted using a neural network model and are compared with the results from FEA.

Computer vision applications often require the model to extract sophisticated features from the input data and approximate very intricate features, which is why deep neural networks are optimal. However, a substantial drawback to training deep networks is vanishing and exploding gradients. He et al. [2] proposed a novel solution to this problem: reformulating the layers as learning residual functions with reference to the layer inputs (i.e., the outputs of the previous layers) instead of learning unreferenced functions. With this approach, they were able to cut down the ImageNet test set error to 3.57%. The current BVP can be approximated as a similar application, and models developed for computer vision applications can be implemented with our data. This study aims to model the neural network using Convolutional Neural Networks (CNNs). CNNs such as AlexNet and ResNet yield accurate results when it comes to classification problems (e.g., ImageNet Challenge) [2]. In the current BVP, however, the aim is to accurately predict the output values, which necessitates a regression analysis.

METHODS: The work presented in this paper was carried out using the Tensorflow GPU and Keras libraries. The data from FE simulations is structured as a 5D tensor (Samples (N), Sequence (S), X-Axis (X), Y-Axis (Y), Features (F)) and was imported in a Tensorflow native format (.tf). The entire dataset comprises 66513 samples, each of which represents a 2D 9x9 grid (X and Y axis) of an elastoplastic material on which displacements are applied, which induces in stresses in the plate along with displacements along the entire grid and forces at its boundary. For developing the Neural Network model, we split the dataset into a training set (70%), test set (15%), and validation set (15%). Furthermore, each sample was edited to contain only the first 25 sequences, since most of the further sequences (75-100) are empty. Thus, we can train the model to recognize essential features in the training dataset rather than excessively focusing on the empty values. For the current BVP, the displacements and the stresses act as the input features and the forces as output features. All these values were normalized using its mean and the standard deviation along all the other axes. The resultant values mostly lie between ±3.

When training the model using a CNN, the convolutions can be carried out along the spatial axes or the sequence axes. For the convolutions along the sequence axis, the spatial axes were flattened, which resulted in a 3D tensor, which serves as an input for 1D convolutions. For convolutions along the spatial axes, we implemented 2D convolutions (which require 4D tensor as an input) in conjunction with time distributed layers.

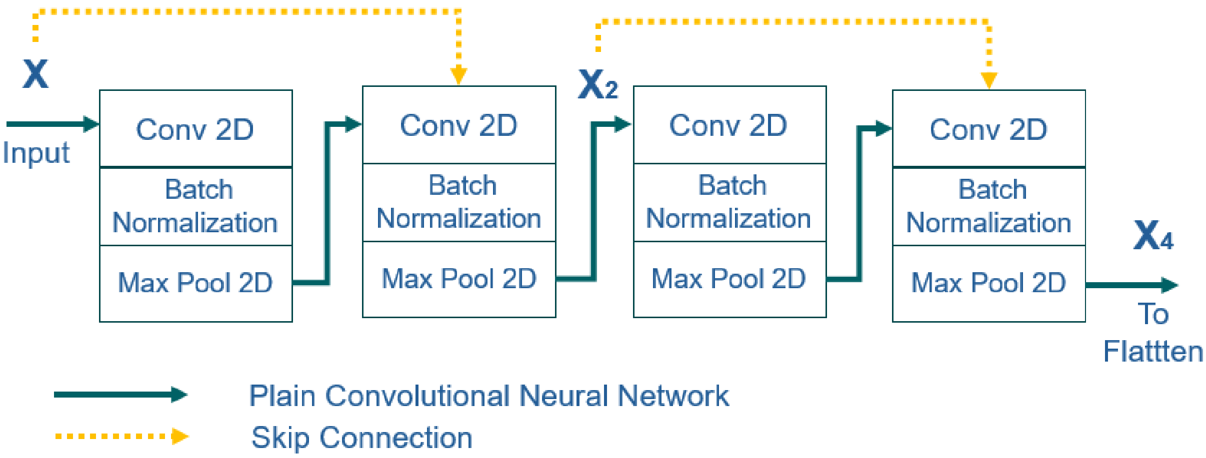

Figure 1 Schematic representation of the Convolutional Neural Network with skip connections

The relevant parameters for tuning the hyperparameters were the number of layers, learning rate, batch size, number of filters used for 2D convolution, kernel size, stride, max-pooling layers. Grid search was implemented for different sets of hyperparameter combinations (32 different networks in total). The tuned CNN had the following hyperparameter values: Number of layers=4 convolutional layers, each followed by batch normalization and max-pooling; learning rate=0.001; batch size=16; Kernel sizes= (2,2) and (1,1): Moreover, two skip connections were added to the plain network as shown in Figure 1.

Mean Absolute Error (MAE) was the loss function, and the models were trained over 80 epochs. We also used the Symmetric Mean Absolute Percentage Error (SMAPE) as a metric for comparing the performance. To compare the performance of these convolutional networks, an already tuned Recurrent Neural Network was used with LSTM layers [3]. The plain CNN had a total of 102,636 parameters; the Residual Network had 55,780 parameters, whereas the Recurrent network had 1,553,634 parameters.

RESULTS: The 1D CNN failed to learn the essential features of the training data, and its output is irrelevant to the problem. The results from 2D CNNs are applicable and are discussed further. With respect to the training, validation, and testing losses for the networks, the Recurrent network performs much better than the 2D CNNs, as shown in Figure 2. The Residual Network performs slightly better than the plain CNN. The SMAPE values were: Plain CNN=0.13%; Residual Network=0.12%; Recurrent Network=0.08%. However, the Residual Network proved to be the best in terms of its computational cost.

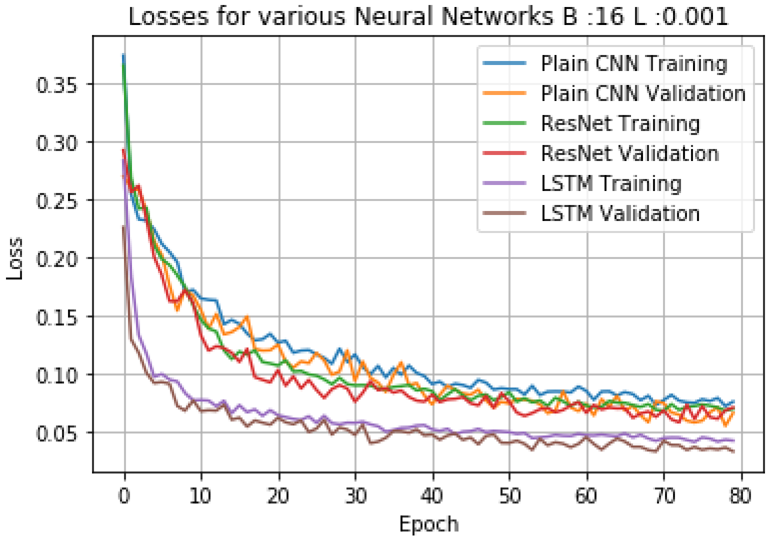

Figure 2 Losses for (a)Plain CNN (b)Residual Network and (c)Recurrent Network-LSTM

The output values (forces) of the three neural networks were accurate and comparable to each other. Figure 3 shows the plot of force over multiple samples (1 sample=25 sequences on the X-axis). Slight under prediction may is a common trend among the samples. We can attribute it to the presence of dummy data (zero value) in some sequences. Using sequential masking may have been a good option. However, TensorflowGPU does not allow the use of masking layers in their networks (needs verification).

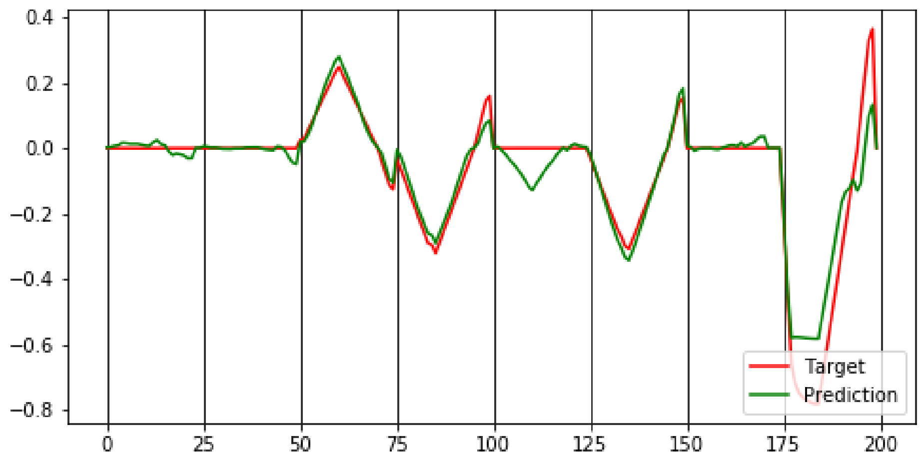

Figure 3 Prediction of the forces induced at one corner of the grid over multiple samples using CNN with skip connections

DISCUSSION: The study aims to explore the use of Convolutional Neural Networks for the regression analysis of Boundary Value Problems in structural mechanics, using the stresses and the displacements as the input features and the boundary forces as the predicted output features. 1D CNN failed to produce meaningful results. This approach requires further scrutiny and debugging before it is dismissed as incorrect or unsuitable for the task. With regards to the 2D CNNs, it is clear that the Residual Network performs slightly better than the plain CNN over all the metrics. The Residual Networks failed to capitalize on their biggest strength- designing deeper networks. However, since the spatial dimensions of the structure are too small (9x9), adding more layers to the network would mean that we would need to manipulate the hyperparameters (kernel size, stride) against the optimally tuned values. Hence, although adding more layers should not inherently worsen the performance of the model, the author found that the deeper networks showed a slight increase in loss and SMAPE values.

Moreover, we can conjecture that deeper neural networks are useful in computer vision applications where there is no theoretical understanding of the relation between the input features and the output features [4]. However, for the boundary value problems in structural mechanics, the relation between the input features and the output features is well understood, and the order of the variables in these functions is relatively low; as a result of which, exploring deep neural networks is not necessary. Hence, using skip connections and residual blocks to construct a deep neural network is not of much use for this problem. Koeppe et al. [1] were able to accurately model the problem with a fully connected feedforward neural network.

He et al. [2] also asserted that increasing the number of layers in a neural network is not necessarily the ideal way to optimize the performance of a neural network application. Perhaps it would be worth exploring other avenues for improving the prediction of this neural network, such as using sequential masking, increasing the grid size of the spatial dimensions, and exploring Conv-LSTM networks [5].

CONCLUSION: We conclude that the addition of skip connections to a Convolutional Neural Network can improve its performance. However, it is necessary to balance deep learning with other aspects of model tuning to get an optimal result. The addition of skip connections did not significantly improve the performance of the model for the current BVP. We can also infer that Recurrent networks are ideally suited for the current Boundary Value Problem.

REFERENCES:

[1] Koeppe, Arnd, Franz Bamer, and Bernd Markert. "An Intelligent Meta-Element for Linear Elastic Continua." Pamm 18, no. 1 (2018). https://doi.org/10.1002/pamm.201800283.

[2] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. https://doi.org/10.1109/cvpr.2016.90.

[3] Koeppe, Arnd, Mundt, Marion, "Computational Intelligence in Engineering: Lecture notes", RWTHmoodle. Accessed February 19, 2020. https://moodle.rwth-aachen.de/mod/resource/view.php?id=136686.

[4] Ruiz, Pablo. "Understanding and Visualizing ResNets." Medium. Towards Data Science, April 23, 2019. https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8.

[5] Tan, Chao, Xin Feng, Jianwu Long, and Li Geng. "FORECAST-CLSTM: A New Convolutional LSTM Network for Cloudage Nowcasting." *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018. https://doi.org/10.1109/vcip.2018.8698733.

ACKNOWLEDGMENT: The author is deeply grateful to Univ.-Prof. Dr.-Ing. Bernd Markert for facilitating the course (Computational Intelligence in Engineering) and Dipl.-Ing. Arnd Koeppe and MSc. Marion Mundt for their valuable suggestions and feedback, as well as for making the topic very engaging for the students. The author would also like to thank Li Chi Shing and Ahmet Küpeli for their continued support while working on the projects for this course.