TestConX 中国 China™

Virtual Archive

October 26 – 29, 2021
Virtual Event

www.testconx.org

With Thanks to Our Sponsors!

Honored

Distinguished

Exhibitor

Application of Deep Learning on Shmoo Data Result Prediction and Analysis

Chao Zhou, Teradyne

TestConX 中国 China

Virtual ▪ October 26-29, 2021

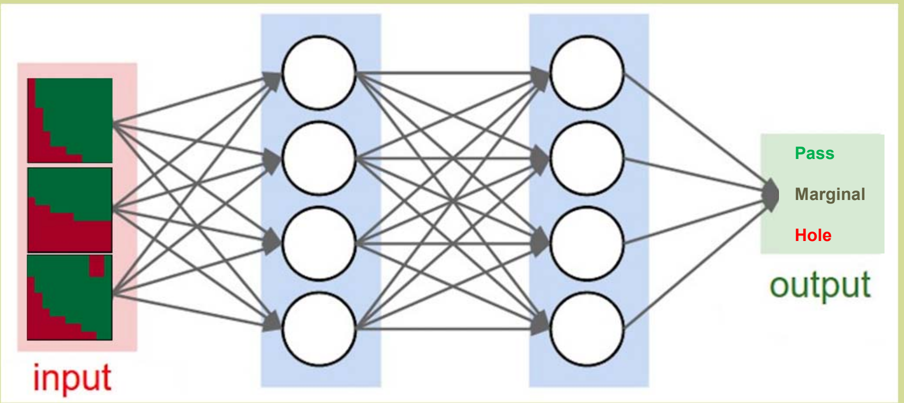
TERADYNE

Contents

- Introduction
  - Why did I write this application?
  - Machine learning VS. deep learning
  - What functional blocks included in this application?
  - How is the training data set composed?
  - Advantages & Limitations of SHM Detect Tool
- How SHM Detect Tool composed
  - Shmoo training data set
  - Convolutional neural network structure of this tool
  - Optimize test accuracy
  - Fighting with overfitting
  - Examine the network
- Result & Future Work
- Conclusion

Why was this tool written?

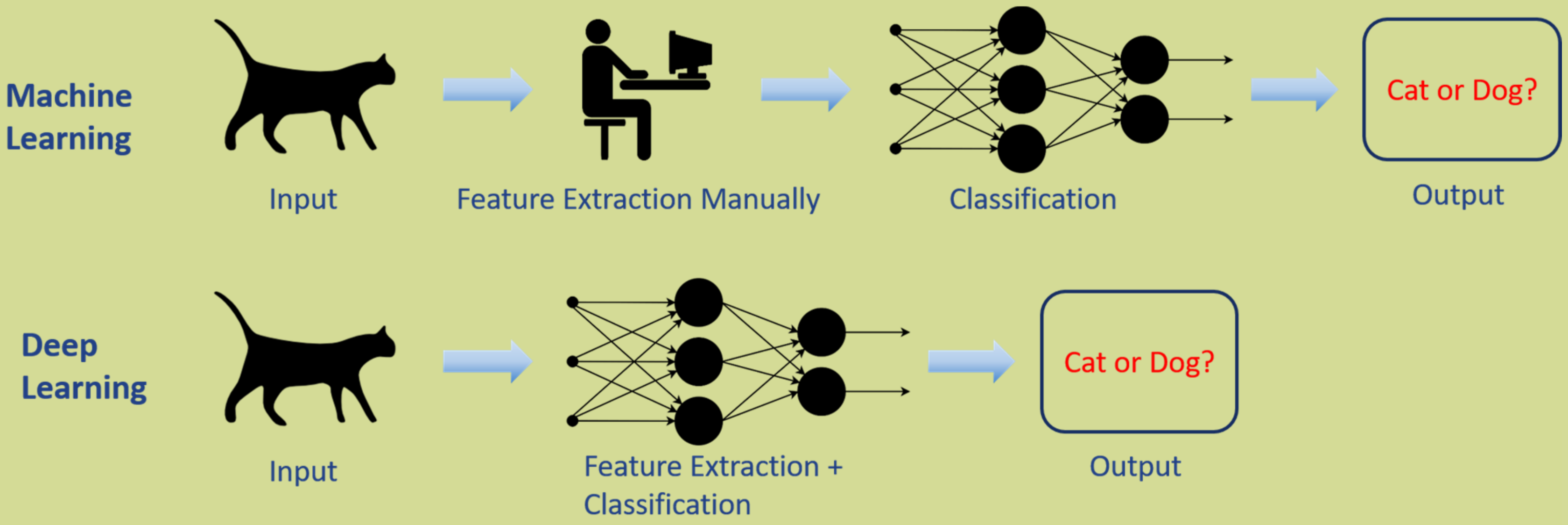
- With this tool, large amount of Shmoo tests result can be analyzed automatically
- Try to apply deep learning(DL) technique to predict/ the Shmoo result, like voltage wall, marginal and fail hole
- “Shmoo Detect Tool” is developed based on Pytorch library, which is a **Python machine learning package** based on Torch used for applications such as computer vision and natural language processing
- Source code link: Taishan Gitlab URL: SHM_Detect_Tool



Hidden Layers

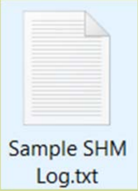
Traditional Machine Learning VS. Deep Learning

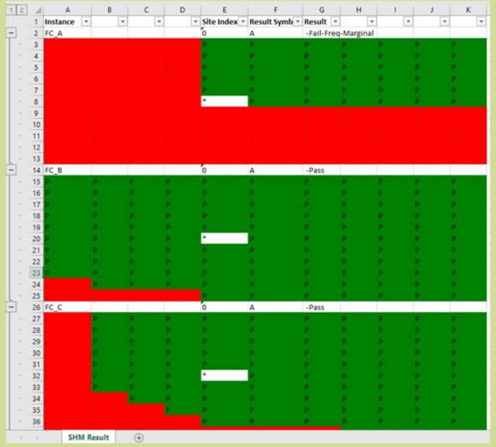
- Difference between traditional machine learning and deep learning



What functional blocks included in this application ?

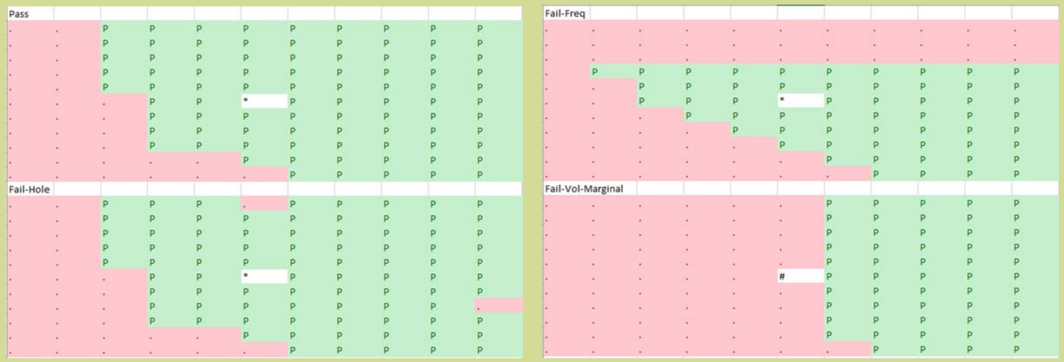
- Training CNN(Convolutional Neural Network): this button is used to train a new model
- Load SHM Log: this is to convert the Shmoo data log to a CSV file
- Analyse SHM Log: do the detection and prediction work and output an XLSX file report
- Exam CNN: this button is used to optimze the CNN structure in debug processing
- 2 types of results are available:
  - Simple one: only Pass/Fail flags are used
  - Detail one: 4 labels are provided to show detail results (Vol, Freq, Marginal, Hole)





How is the training data set composed?

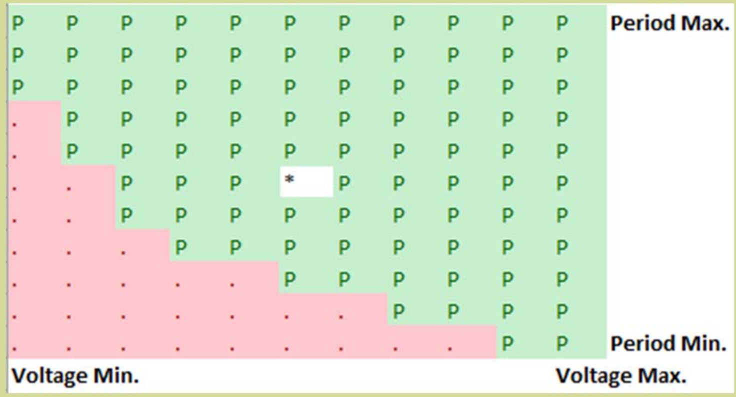
- Training Data structure: Each training data composed by two parts, result and 11x11 Shmoo plot
- X axis: voltage VS Y axis: period
- 6 labels:
  - Pass: good shmoo plot
  - Fail: bad shmoo plot
  - Vol-Wall: number of pass points changes dramatically on X-axis
  - Freq-Wall: number of pass points changes dramatically on Y-axis
  - Marginal: fail test point appears near central
  - Hole: fail test point surrounded by pass ones



2021

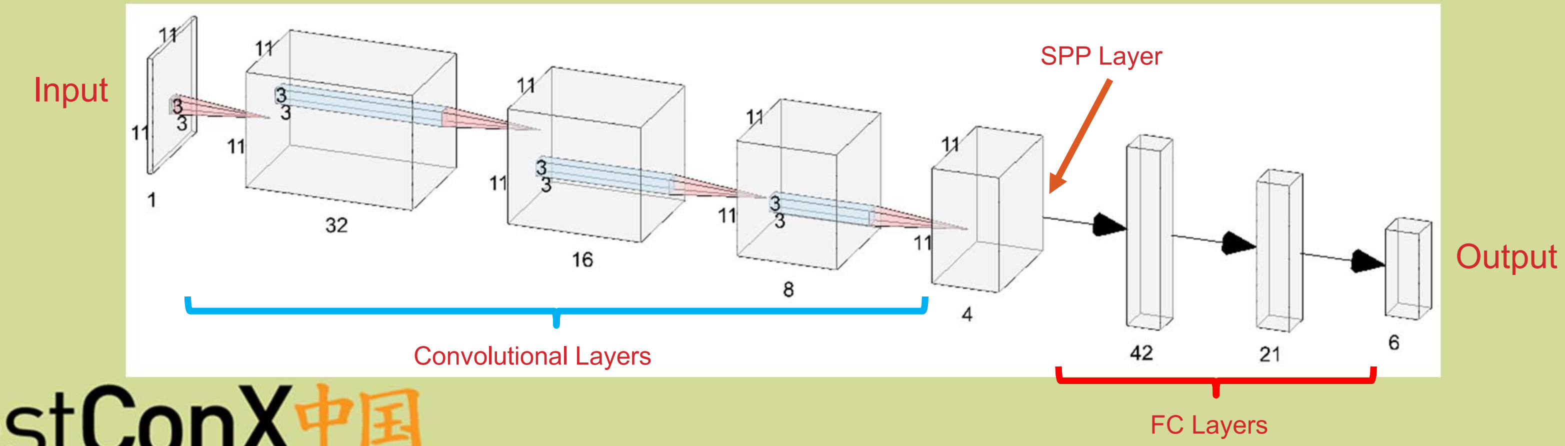
Advantages & Limitations of SHM Detect Tool

- Advantages of Shmoo Detect Tool
  - High test accuracy built-in trained model
  - User can train their own model
  - Variable shape of Shmoo can be set as input
- Limitations of Shmoo Detect Tool
  - Only support Shmoo with Period as Y-axis and Voltage as X-axis



CNN Structure of Shmoo Detect Tool

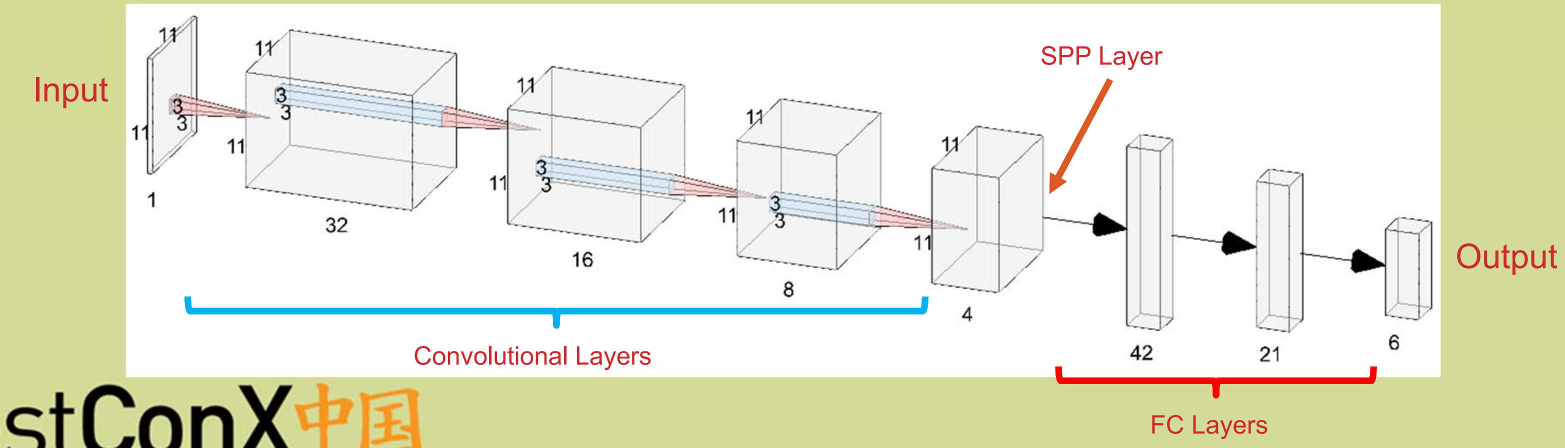
- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Then, there are 4 convolutional layers to extract the features of Shmoo
  - 1 Spatial Pyramid Pooling(SPP) layer is applied after convolutional layers to process variable size of input
  - There are 3 fully-connected(FC) layers to detect the result of Shmoo



CNN Structure of Shmoo Detect Tool

- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Each Shmoo test point is converted to one input pixel, so all the pixels are sensitive to the prediction result
  - Since pooling layer could reduce the input information of Shmoo plot, the pooling layers are not employed in my CNN network

```
tensor([[[[1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
          [0., 0., 0., 0., 1., 1., 1., 1., 1., 1., 1.]]]])
```

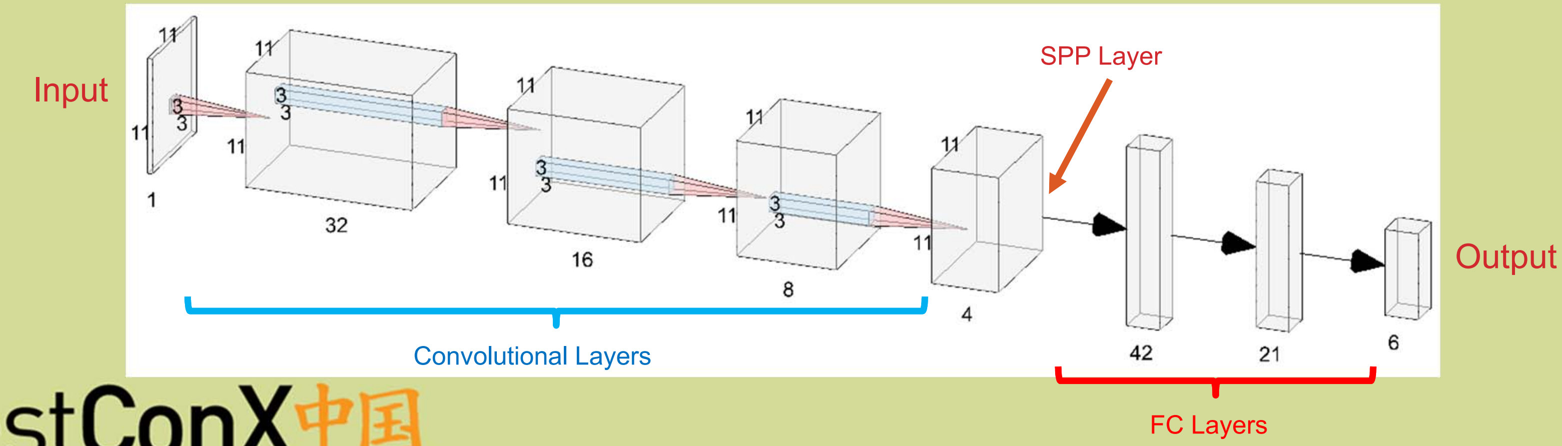


Application of Deep Learning on Shmoo Data Analysis

TestConX China 2021

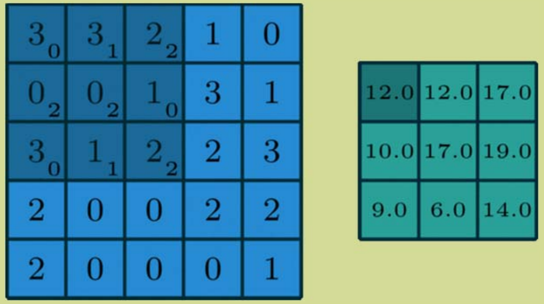
CNN Structure of Shmoo Detect Tool

- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Then, there are 4 convolutional layers to extract the features of Shmoo



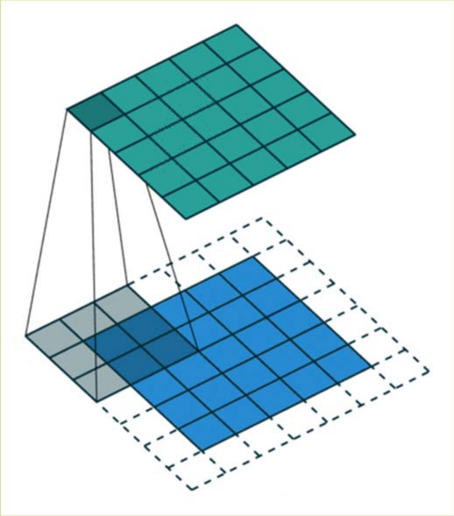
CNN Structure of Shmoo Detect Tool

- Convolutional layer
  - A convolutional layer contains a set of kernels whose parameters need to be learned
  - Convolutional layer to extract the features
  - Kernel is slid across the width and height of the input
  - Each kernel is convolved with the input volume to compute a feature map
  - The convolution kernel shape of all layers in the network is 3x3 in this application



Note: Image retrieved from Google Image

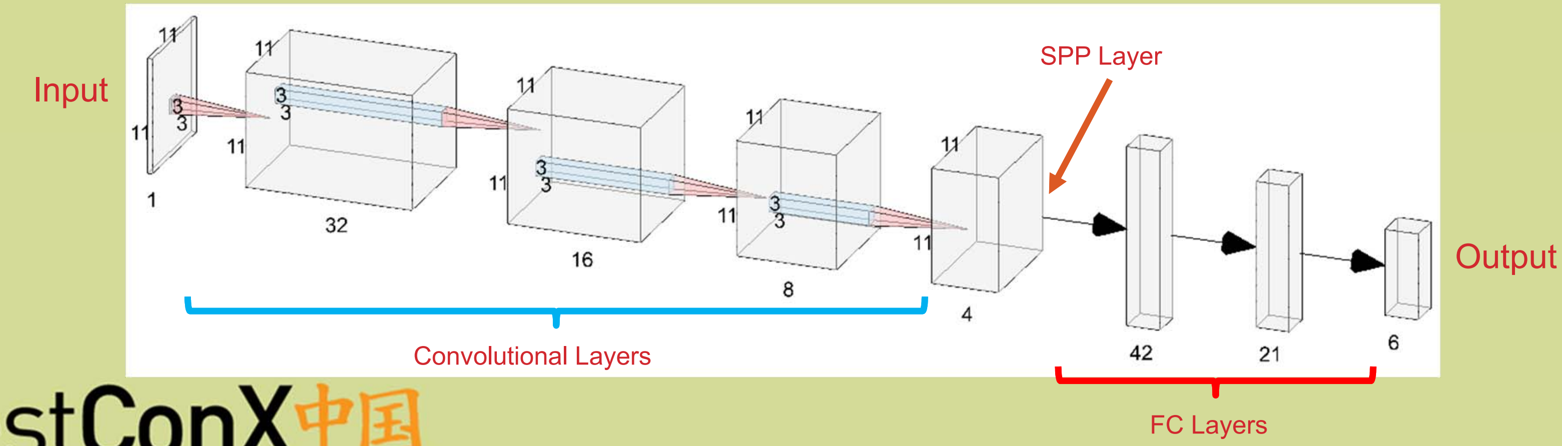
- Padding
  - This is used to preserve pixels information of input image edge
  - It can used to maintain the pixels size of on the perimeter of image
  - The values of the extra pixels is set to zero typically
  - The padding size is set to 1 in this application
  - Work with convolutional kernel to keep the shape of convolutional layer's output



Note: Image retrieved from Google Image

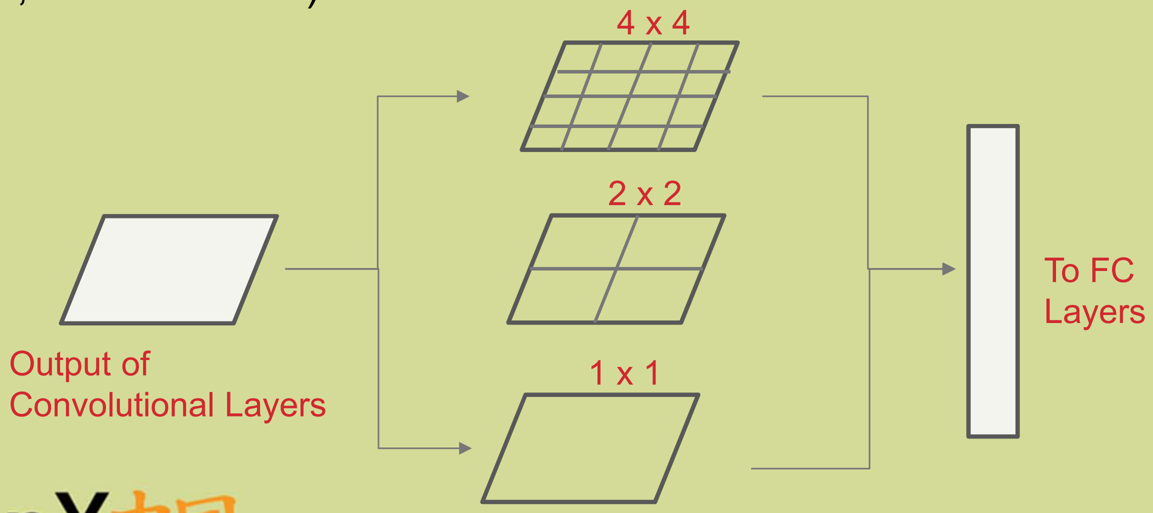
CNN Structure of Shmoo Detect Tool

- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Then, there are 4 convolutional layers to extract the features of Shmoo
  - 1 Spatial Pyramid Pooling(SPP) layer is applied after convolutional layers to process variable size of input



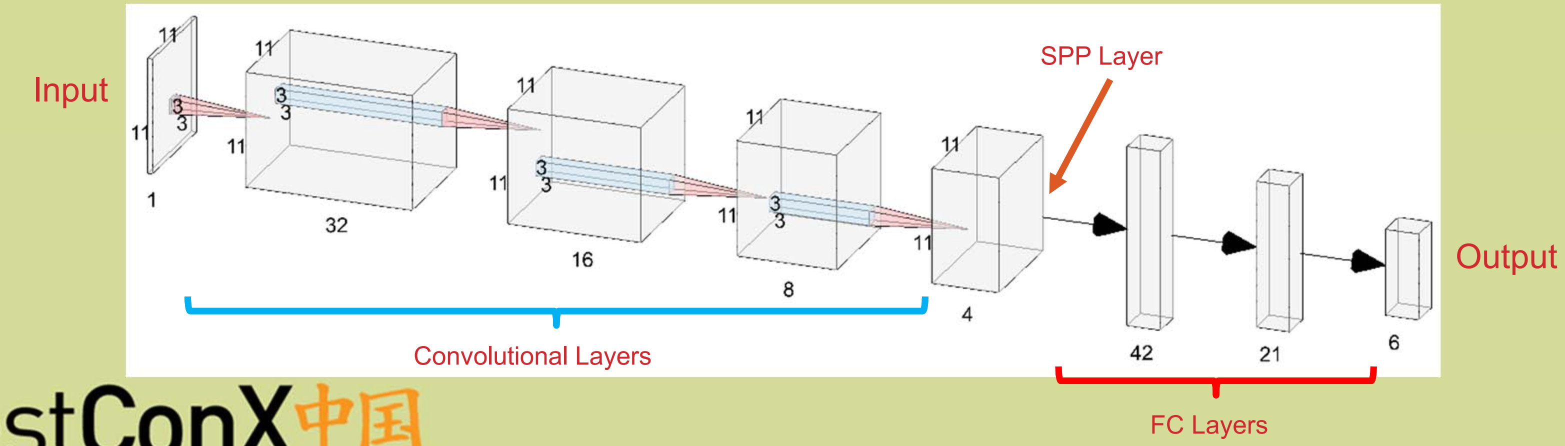
CNN Structure of Shmoo Detect Tool

- Spatial Pyramid Pooling(SPP) layer is a pooling strategy to result in an output of fixed size
- It will turn a 2D input of arbitrary size into an output of fixed dimension
- Convolutional layer of a CNN can be connected to a FC layer with a fixed number of nodes even if the dimensions of the input image are unknown
- In my tool, a 3-level SPP (4, 2, 1) is applied, the output of SPP is a fixed 21*channel element vector (16*ch, 4*ch and 1*ch)



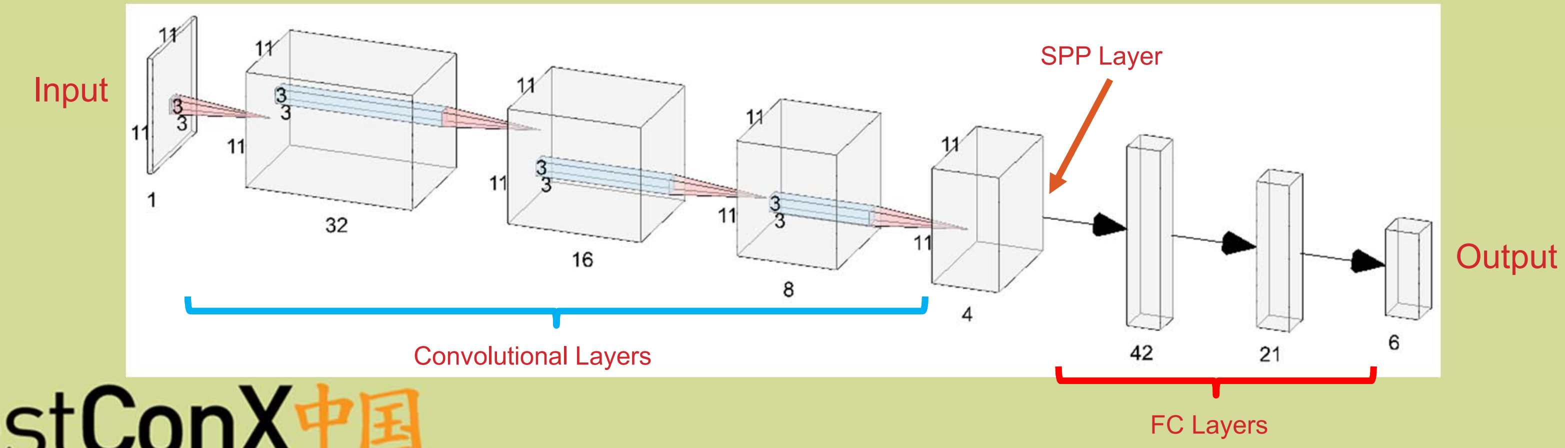
CNN Structure of Shmoo Detect Tool

- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Then, there are 4 convolutional layers to extract the features of Shmoo
  - 1 Spatial Pyramid Pooling(SPP) layer is applied after convolutional layers to process variable size of input
  - There are 3 fully-connected(FC) layers to detect the result of Shmoo



CNN Structure of Shmoo Detect Tool

- The net contains 8 layers with weights, which based on AlexNet
  - The first in the very beginning converts Shmoo plot to tensor as input
  - Then, there are 4 convolutional layers to extract the features of Shmoo
  - 1 Spatial Pyramid Pooling(SPP) layer is applied after convolutional layers to process variable size of input
  - There are 3 fully-connected(FC) layers to detect the result of Shmoo
- The output is a 1x6 label vector, each element stands for Pass, Fail, Vol-Wall, Freq-Wall, Marginal, Hole



CNN Structure of Shmoo Detect Tool

- Activation function
  - It defines the output of the neuron for a given input
  - It is like a gate that checks that an incoming value is greater than a critical number, and decides which information of the model should pass to sub-sequence layer
  - In this application, Rectified Linear Unit (ReLU) was chosen over Sigmoid because ReLU can help to prevent the situation where the gradient returns to zero during back propagation, hence accelerating the training

Note: Image retrieved from Google Image

Optimize Test Accuracy

- Loss function
  - MultiLabelSoftMarginLoss() is used in this application as loss function, the formula is shown as below
  - $l_n = -w_n[y_n \cdot \log z_n + (1 - y_n) \cdot \log(1 - z_n)]$, and $z_n = 1 + \exp(-x[i]))^{-1}$, the total loss is the mean of all l_n
- Batch Normalization
  - Using batch normalization makes the network more stable during training, and accelerates training
  - Can make training deep networks less sensitive to the choice of weight initialization method
  - Batch normalization offers some regularization effect, reducing generalization error
  - Do not use batch normalization and dropout in the same network
- Pooling
  - Pooling layers provide an approach to down sampling feature maps by summarizing the presence of features in patches of the feature map, which means it leads to information loss
  - There are 2 major types: average pooling and max pooling
  - In this application, each pixel is sensitive to the result detection, so this layer is removed from the tool's network

Fighting with Overfitting

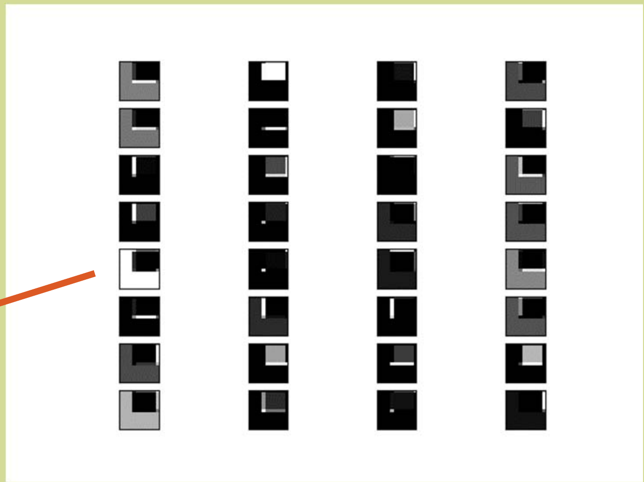
- There are more than 1000 parameters in this neural network architecture. The tool imposes 6 bits of constraint to mapping Shmoo to labels, so this leads to insufficient to learn so many parameters without overfitting. There are 4 methods available to eliminate overfitting.

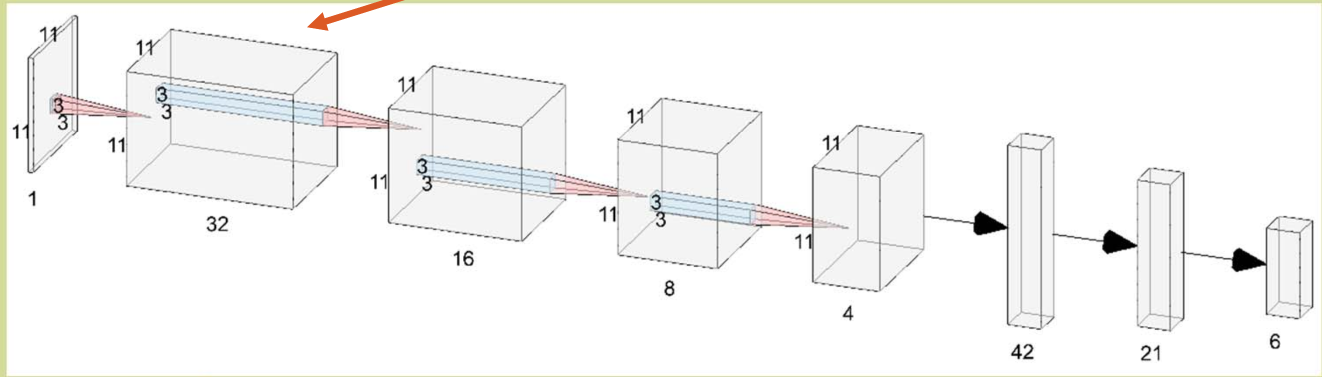
- Enlarge the training data
  - The most effective and common way to reduce overfitting on image data is to enlarge the dataset using label-preserving transformations
- Replace FC layer with convolutional layer
  - Convolutional layer can be helpful to reduce the number of parameters, hence reduce the overfitting
- Regularization
  - Dropout method: some number of layer outputs are randomly ignored or “dropped out.”, which force other nodes within a layer to probabilistically take on more responsibility for the inputs
  - Batch normalization method is deployed instead of Dropout in this application for it can compatibles with convolutional layer and better performance
- Weight decay
  - Weight decay is a regularization technique by adding a small penalty, usually the L2 norm of the weights (all the weights of the model), to the loss function
  - $loss = loss + weight\ decay\ parameter * L2\ norm\ of\ the\ weights$

TestConX China 2021

Exam the network

- Feature maps of middle layer
  - The first Convolutional layer extract the special position and edge

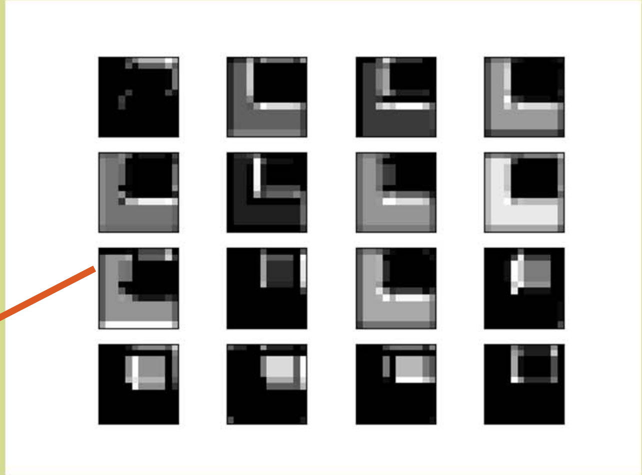


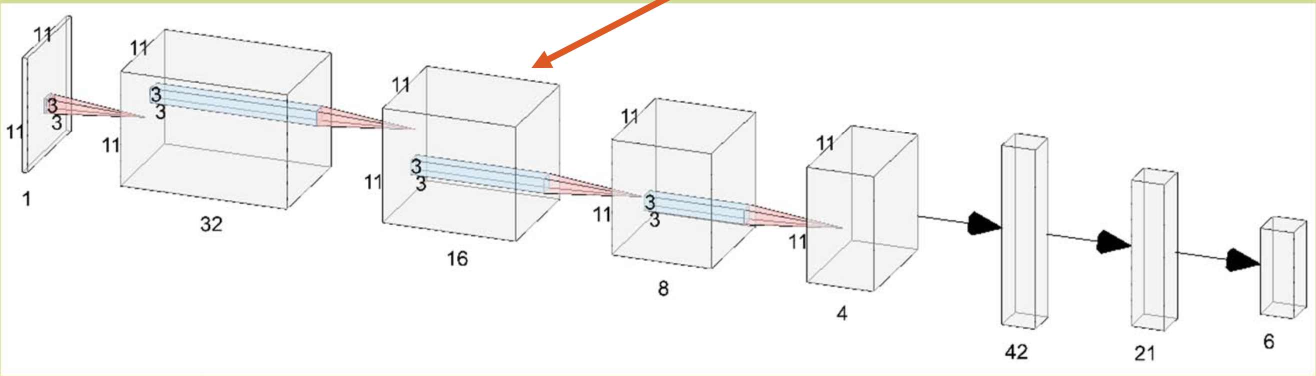


TestConX China 2021

Exam the network

- Feature maps of middle layer
  - In this layer net combine the output of first layer to some simple patterns





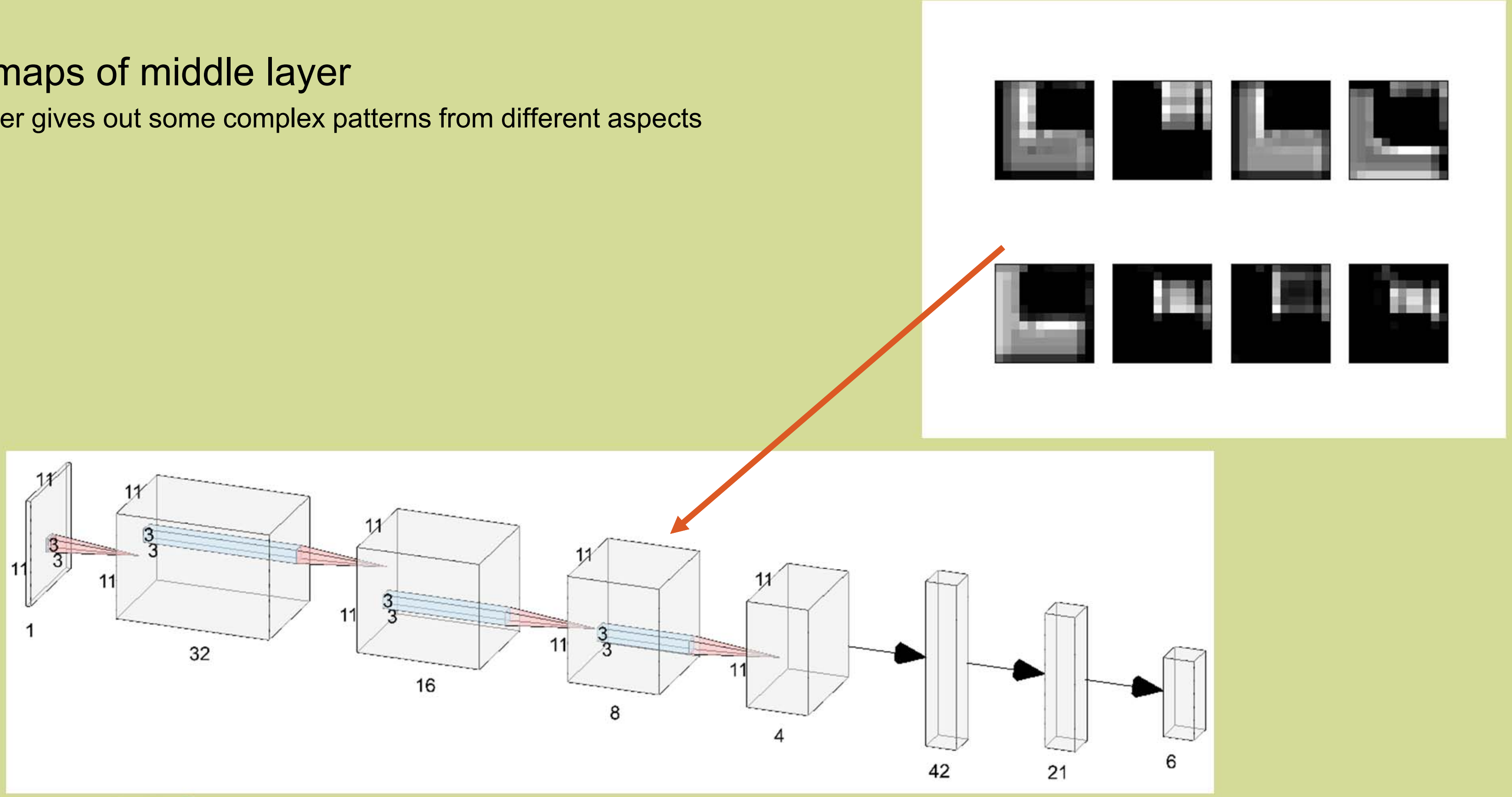
TestConX 中国 China

Application of Deep Learning on Shmoo Data Analysis

2021

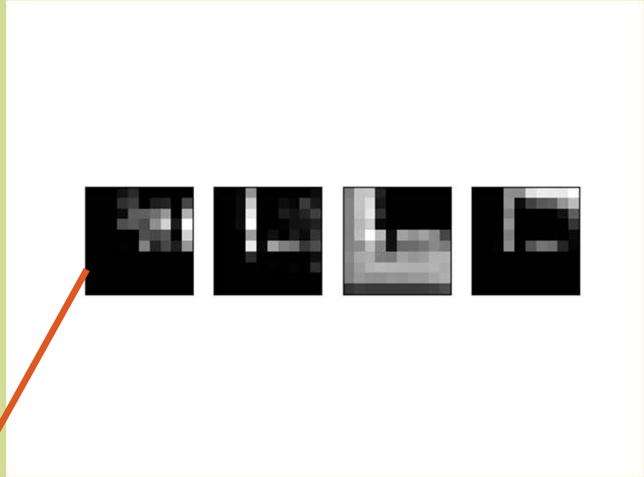
Exam the network

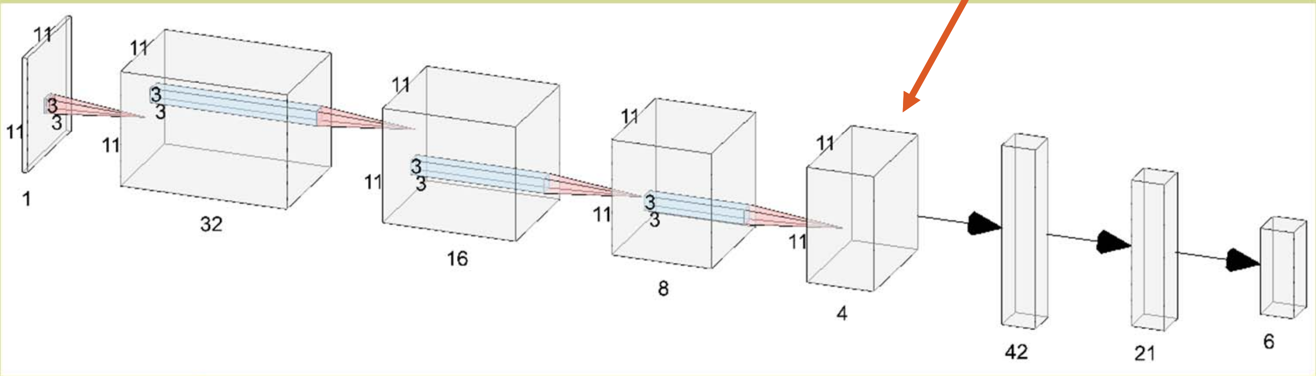
- Feature maps of middle layer
  - This layer gives out some complex patterns from different aspects



Exam the network

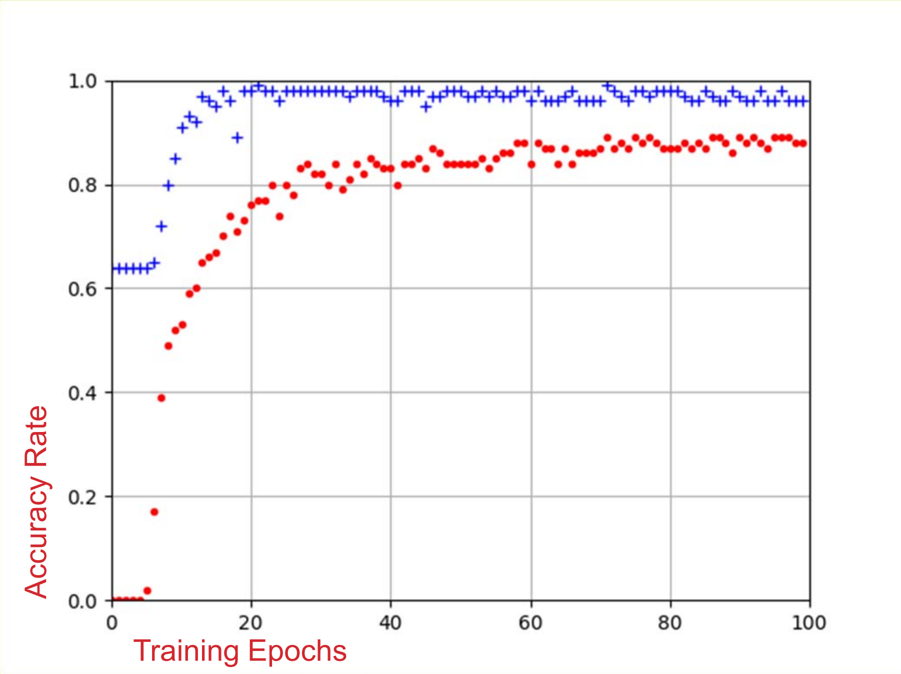
- Feature maps of middle layer
  - As the layers become deeper, the last layer can output more complicate patterns





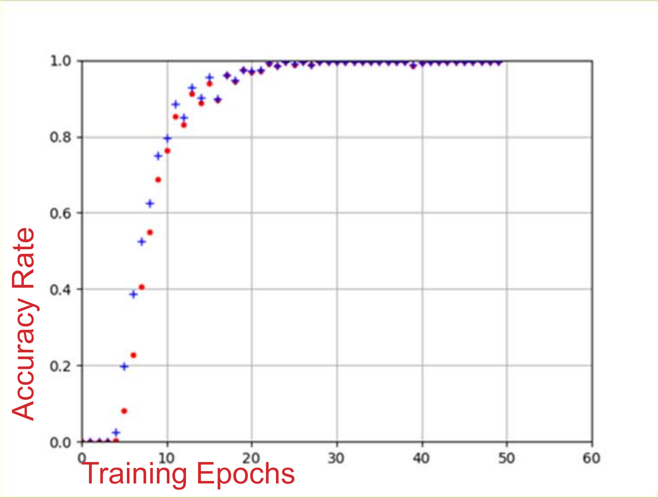
Result & Future Work

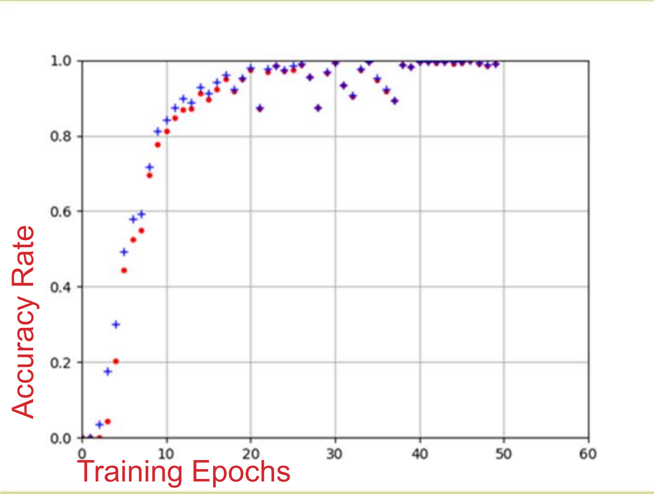
- Take 100 samples as batch size, learning rate as 0.0014, and weight decay of 0.0004
- After 100 epochs:
  - the total loss is 0.0002
  - train accuracy is 0.993
  - test accuracy with multiple labels is around 0.89
  - test accuracy with P/F flag is about 0.97



Result & Future Work

- The stability of test accuracy affected by the number of convolutional layer
- Removing one convolutional layer leads to an accuracy drop at epoch 28 under the same training conditions by about 3%, and leads to accuracy fluctuations during training
- It can be seen that "depth" is the key to this neural network model to meet expectations





Result & Future Work

- To support the Shmoo center point (test base point) recognition function
- To support more type of X and Y axis in Shmoo
- Adapt other effective deep learning model to reduce overfit, like ResNet, YOLO
- Add the sample in training data set

Conclusion

- By apply deep learning technique in Shmoo data analysis can help to increase the review efficiency dramatically
- This tool still during beta period, there are more work to do to optimize its performance

Reference

1. A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural net-works. In *NIPS*, 2012.
2. Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, et al. Hand-written digit recognition with a back-propagation network. In *Advances in neural information processing systems*, 1990.
3. Jason Brownlee.Loss and Loss Functions for Training Deep Learning Neural Networks[EB/OL]. https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/ ,2019-01-28.
4. Sijan Bhandari.Intuition behind Gradient Descent for Machine Learning Algorithms[EB/OL]. http://sijanb.com.np/posts/intuition-behind-gradient-descent-for-machine-learning-algorithms/ ,2019-08-05.
5. 强波.梯度下降算法[EB/OL]. https://paul.pub/gradient-descent/ ,2019-06-19.
6. Ashwin Singh.Optimization Algorithms for Deep Learning[EB/OL]. https://medium.com/@ashwin8april/optimization-algorithms-in-deep-learning-4f2c3b53f9f ,2019-09-01.
7. 7 Types of Neural Network Activation Functions: How to Choose?, https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/
8. Jason Brownlee.A Gentle Introduction to Batch Normalization for Deep Neural Networks[EB/OL]. https://machinelearningmastery.com/batch-normalization-for-training-of-deep-neural-networks/ ,2019-01-16.
9. K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *ECCV*. 2014.

COPYRIGHT NOTICE