Model fusion at score level for image classification

Aakash Jignesh Modi

CB.EN.P2CEN20001

Introduction

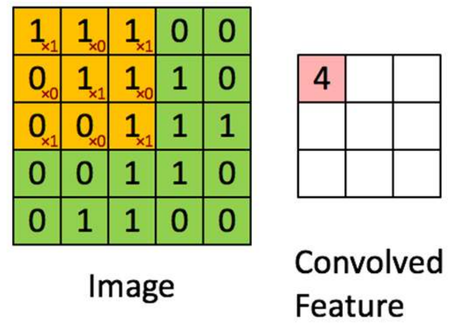
- *Image classification* is a fundamental task in *computer vision*.
- The goal is to classify & assign the image to a specific label or class.
- Convolutional Neural Networks (CNN) is a deep learning method used extensively in processing & classifying images for multiple applications like digital pathology, traffic image recognition, face recognition to name a few.

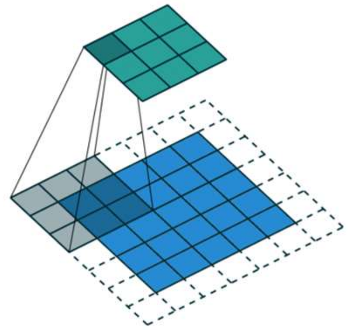
Traffic image recognition using computer vision

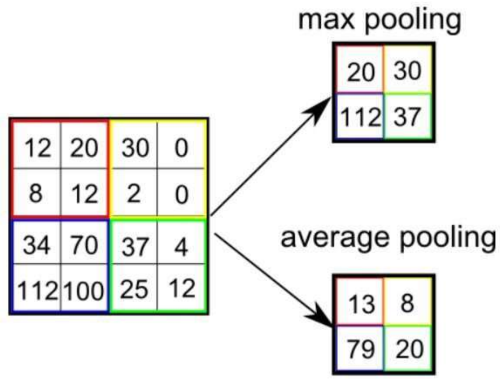
- Convolutional Neural Network (CNN) is the most classical multilayer neural network & the most common deep learning framework inspired by the visual perception mechanism of human beings.
- In 1990, LeCun et al. published a paper about a multi-layer neural network called LeNet-5 which could classify handwritten digits with little or no preprocessing using a backpropagation algorithm.
- Krizhevsky et al. developed a neural architecture, called AlexNet which was similar to LeNet-5 but had a deeper structure. It showed a significant improvement in the image classification task. With the success of AlexNet, many works have been done to increase performance like VGGNet, GoogleNet, ResNet, etc.
- With the advancement in neural networks a new machine learning method in which utilizing knowledge acquired from one task to solve related ones emerged known as Transfer Learning.
- Despite several advancements in neural networks & learning methods for image classification, robust & accurate classification of the target object in images remains unsolved because of difficulties posed by interclass & intraclass similarities, noisy images, etc. To overcome this difficulty a multimodal classification using fusion is used.
- This project aims to investigate the fusion of different models at the score level for the Fashion-MNIST dataset.

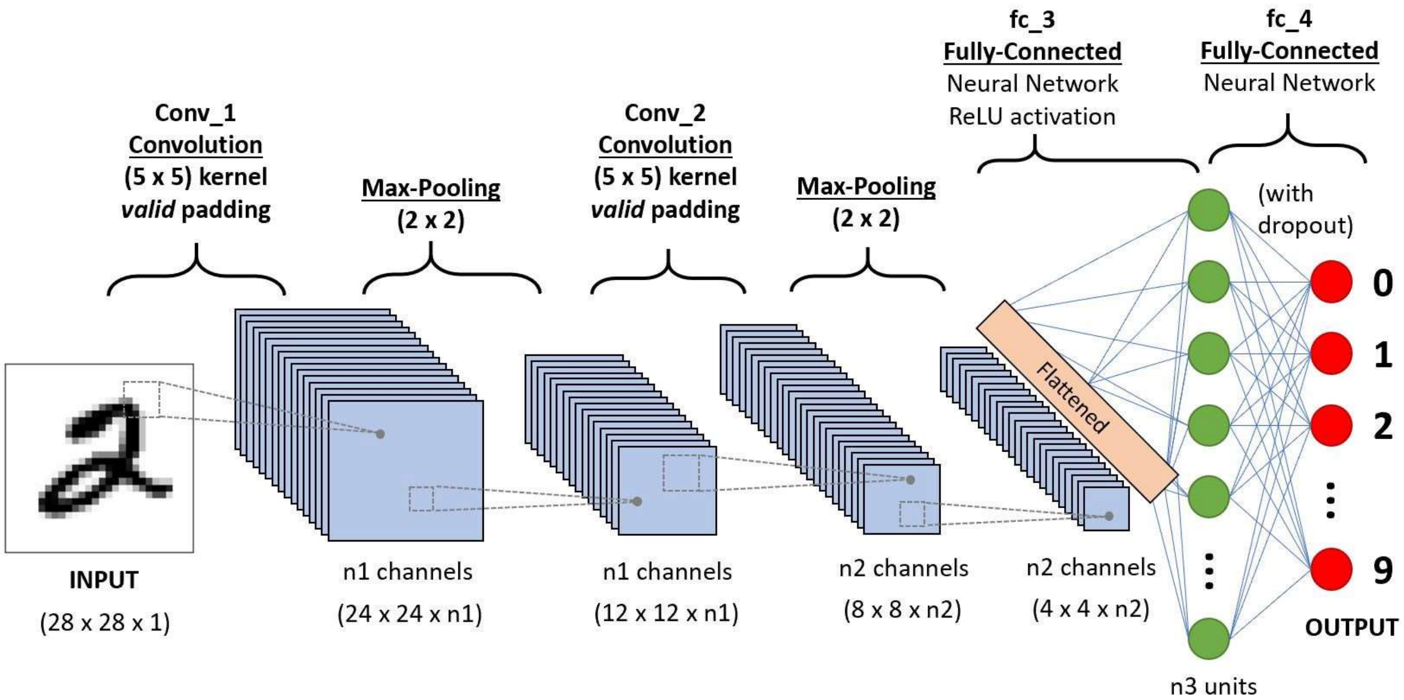
Theoretical Background

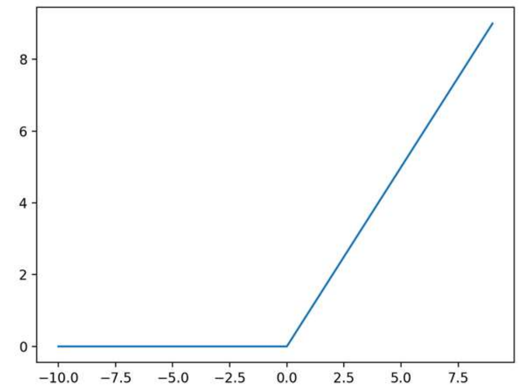
- Convolutional Neural Network (CNN)



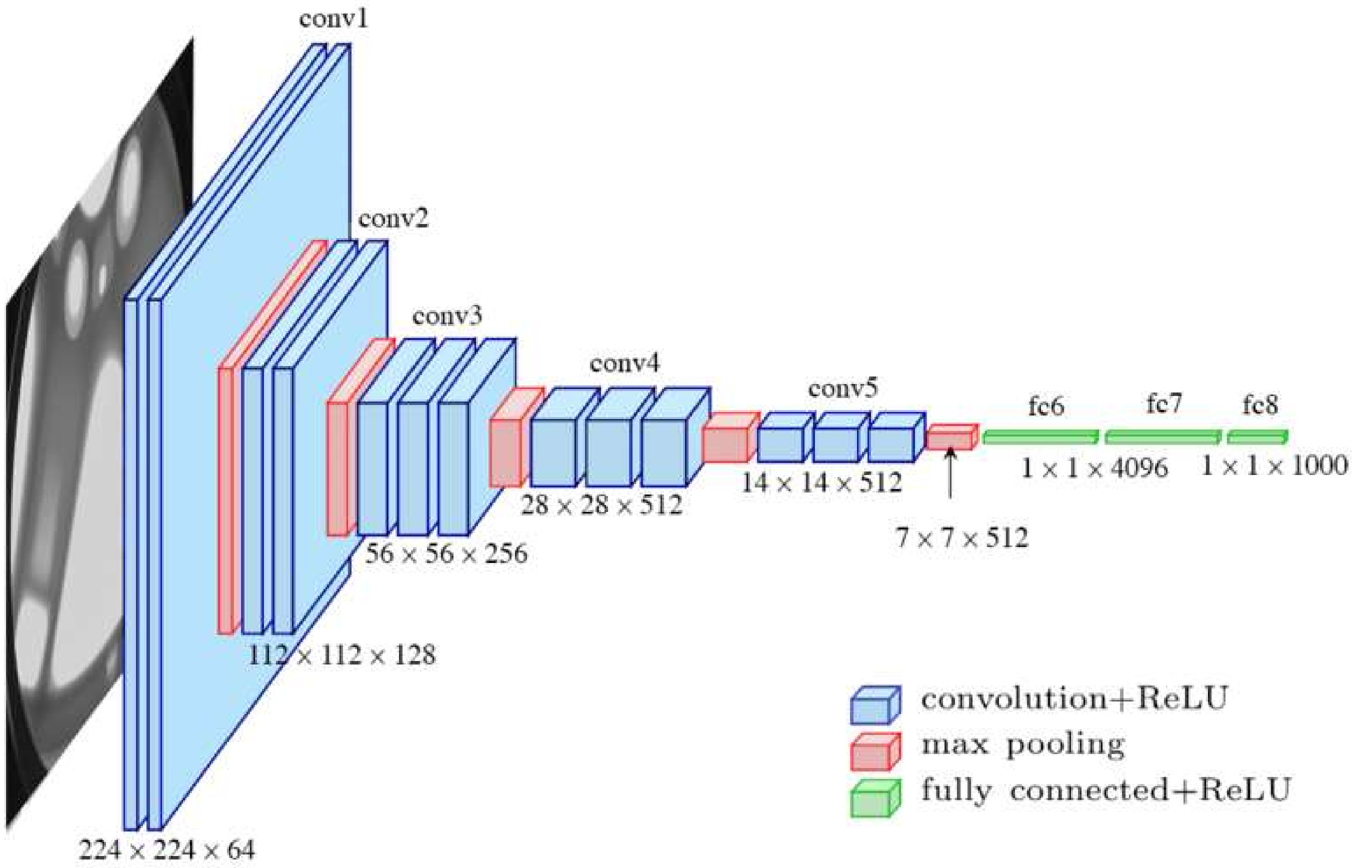




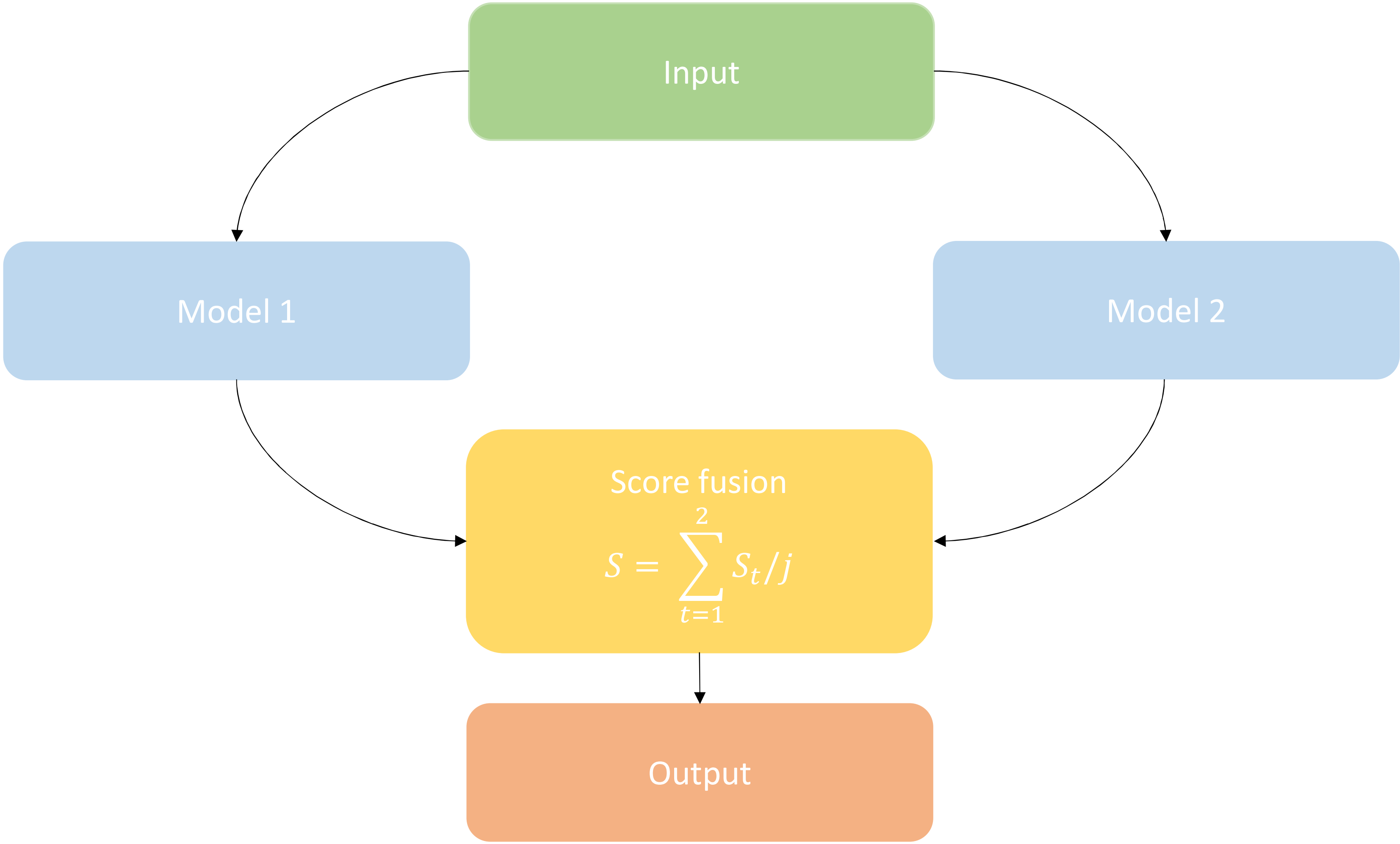




• VGG16


conv1
conv2
conv3
conv4
conv5
fc6
fc7
fc8
224 × 224 × 64
112 × 112 × 128
56 × 56 × 256
28 × 28 × 512
14 × 14 × 512
7 × 7 × 512
1 × 1 × 4096
1 × 1 × 1000
convolution+ReLU
max pooling
fully connected+ReLU

- Model fusion at score level

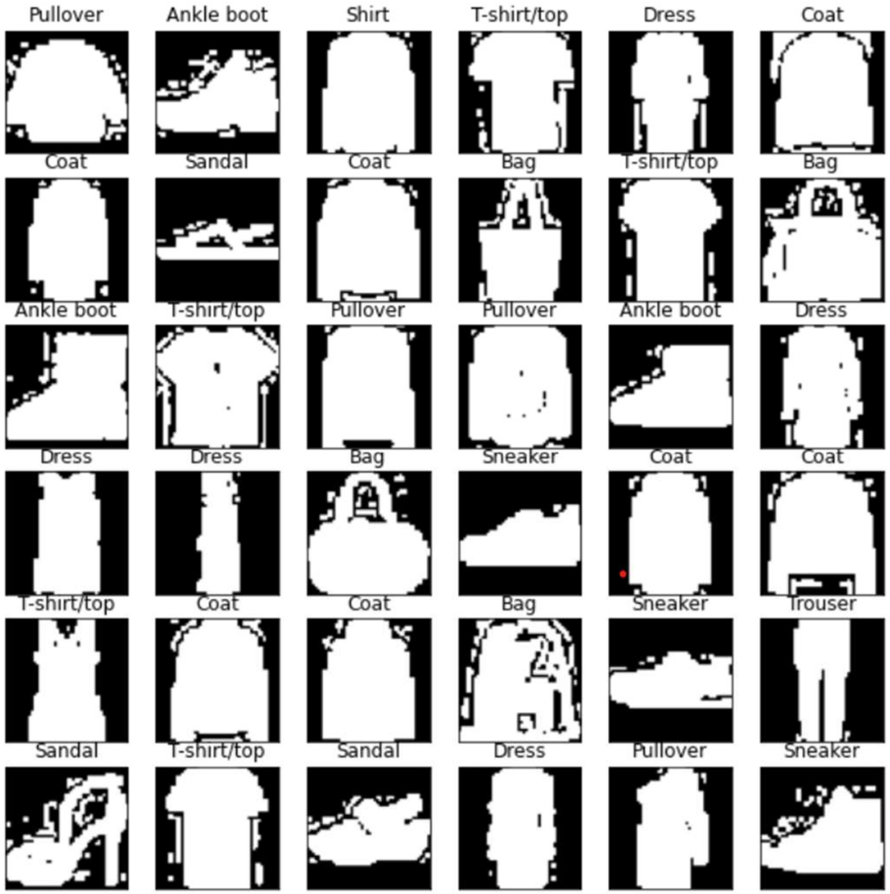


Methodology

- Understanding the dataset (Fashion-MNIST)
  - Fashion-MNIST is a dataset of Zalando's article images.
  - Training set: 60,000 examples; Test set: 10,000 examples.
  - Each example is a 28x28 grayscale image, associated with a label from 10 classes.

- Data preparation & preprocessing

The images were resized from 28*28 to 48*48 & 3 channels as required for transfer learning



- Data normalization
- Converting labels to one-hot encoder

```
[[0. 0. 1. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 1.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 1. 0.]
 [0. 0. 0. ... 0. 1. 0.]
 [0. 0. 0. ... 1. 0. 0.]]
[[1. 0. 0. ... 0. 0. 0.]
 [0. 1. 0. ... 0. 0. 0.]
 [0. 0. 1. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 1. 0.]
 [0. 0. 0. ... 0. 1. 0.]
 [0. 1. 0. ... 0. 0. 0.]]
```

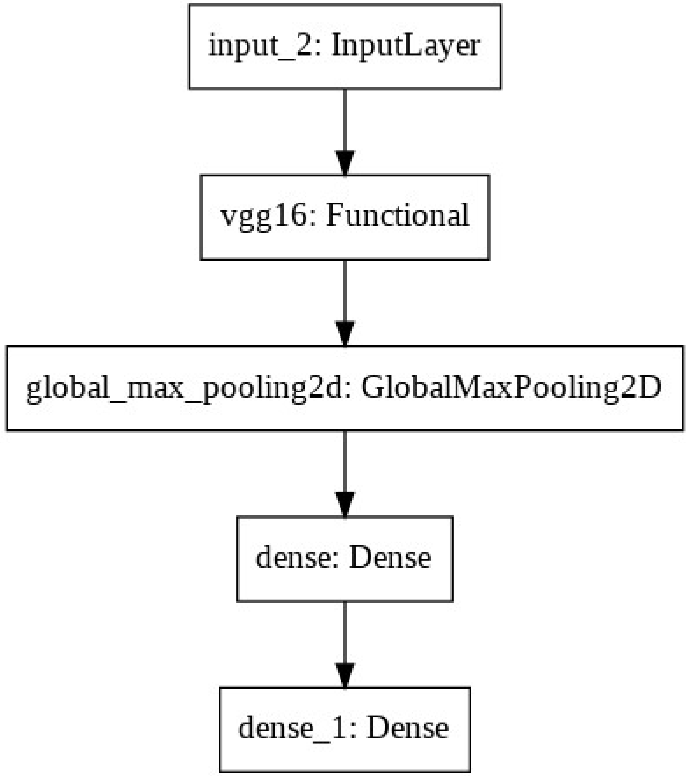
- Splitting train data into training and validation data

Train Validation

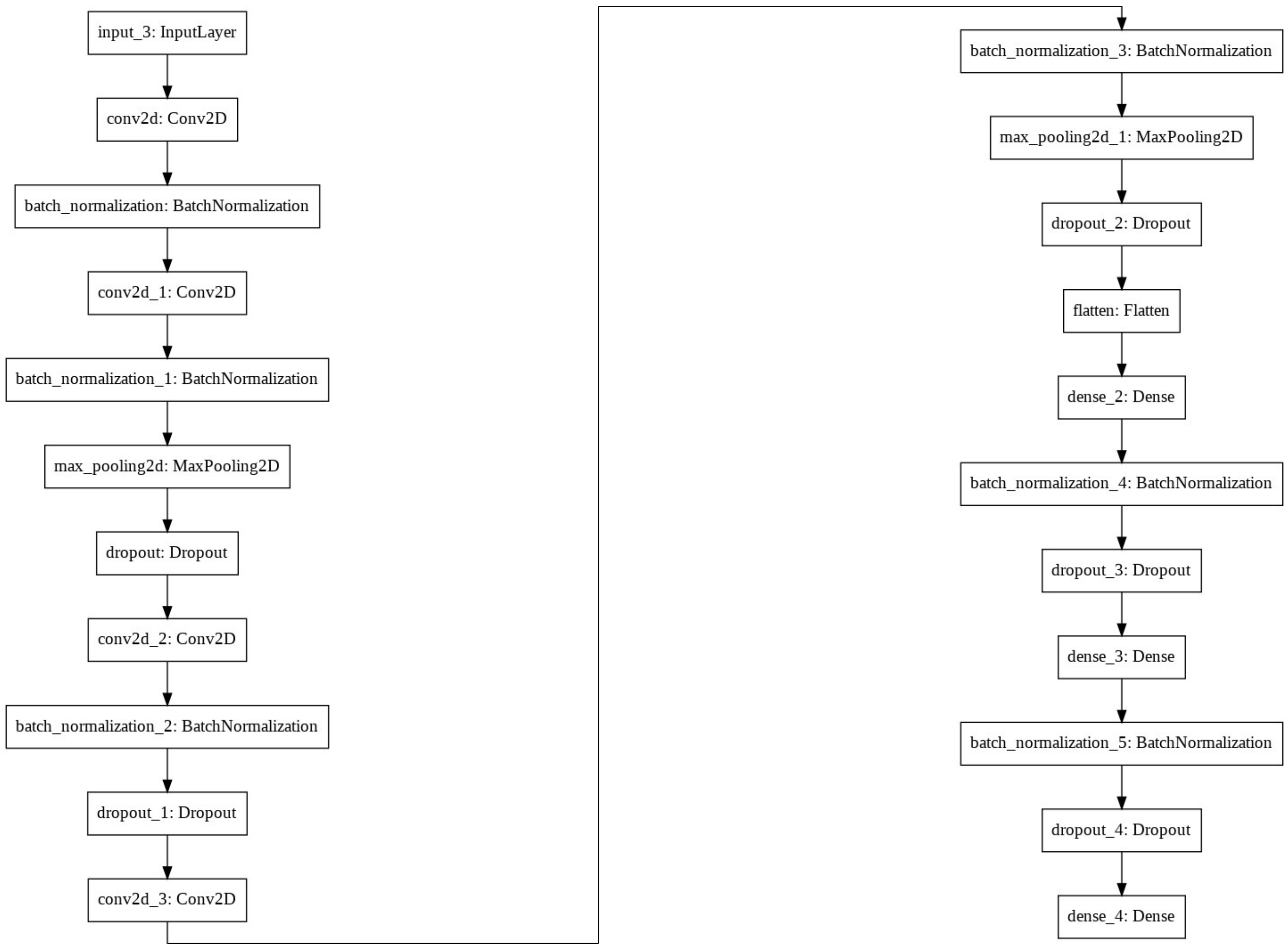
```
((57000, 48, 48, 3), (3000, 48, 48, 3))
```

- Model Definition

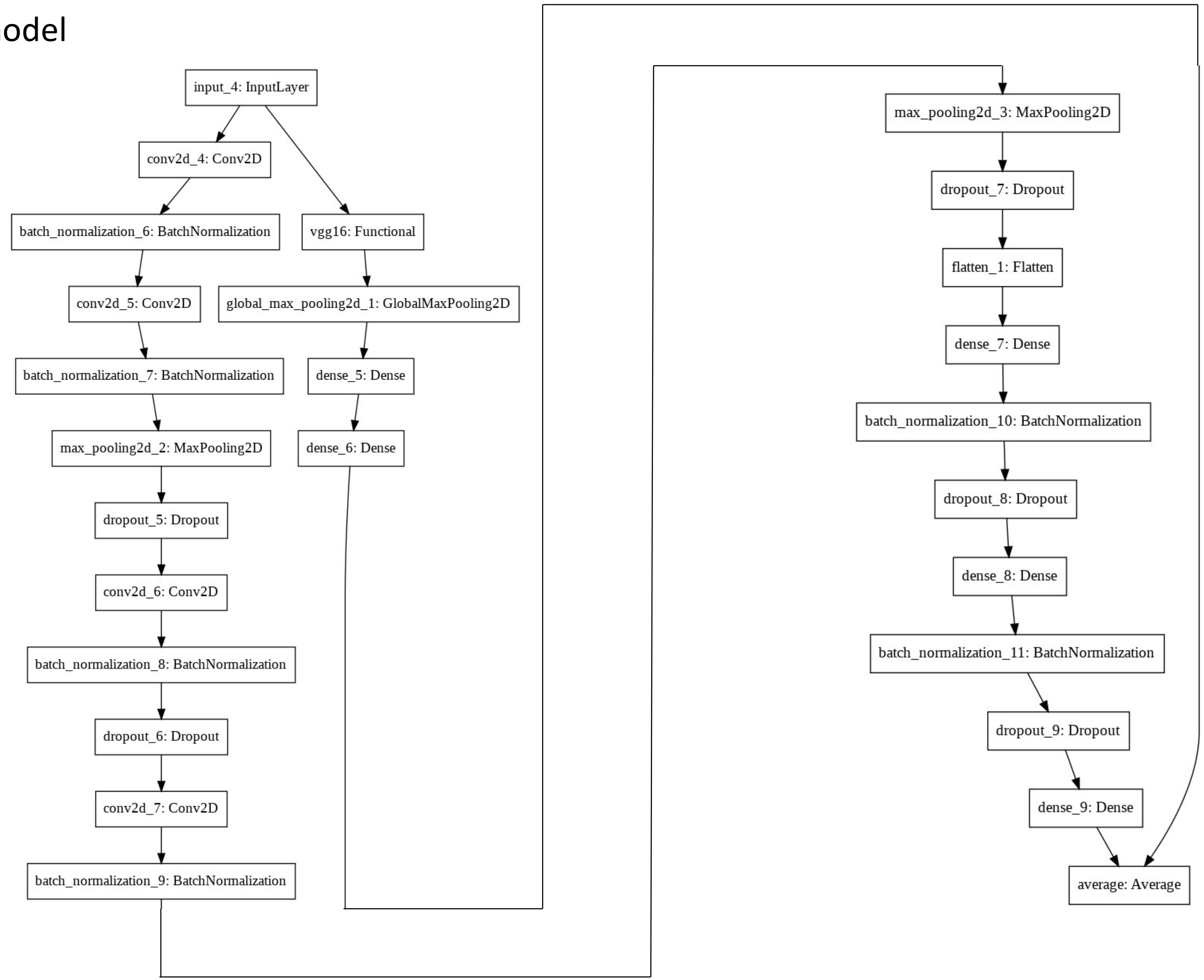
  - Uni-model (Transfer learning model VGG16)



- Uni-model (CNN with 4 layers)

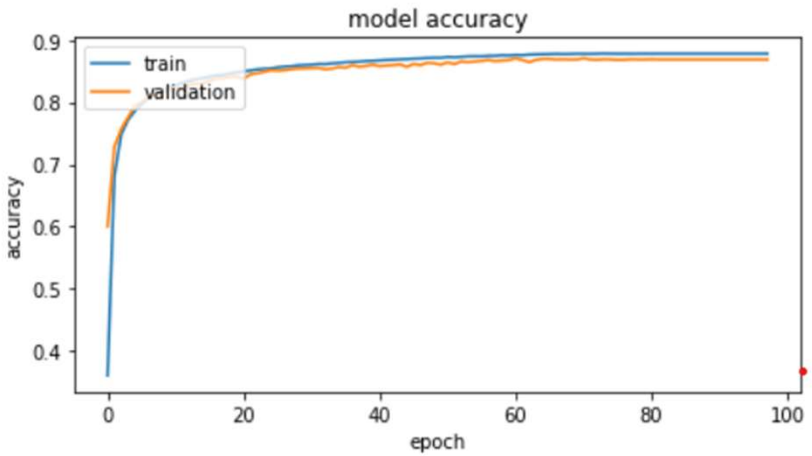


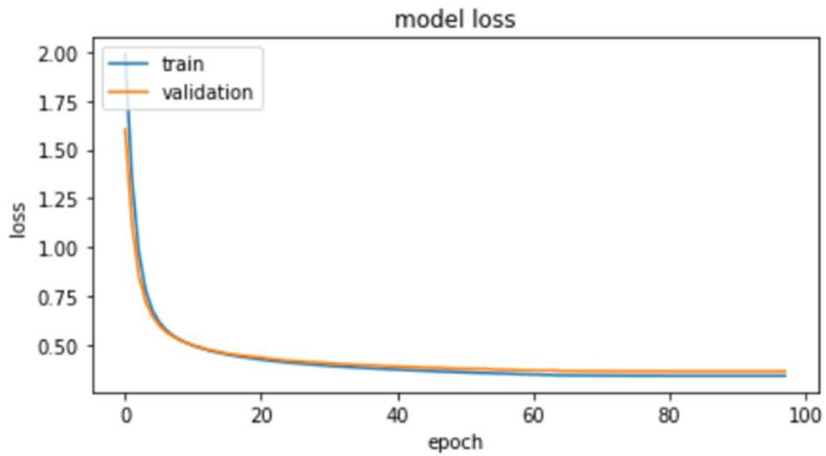
○ Fusion model



Results

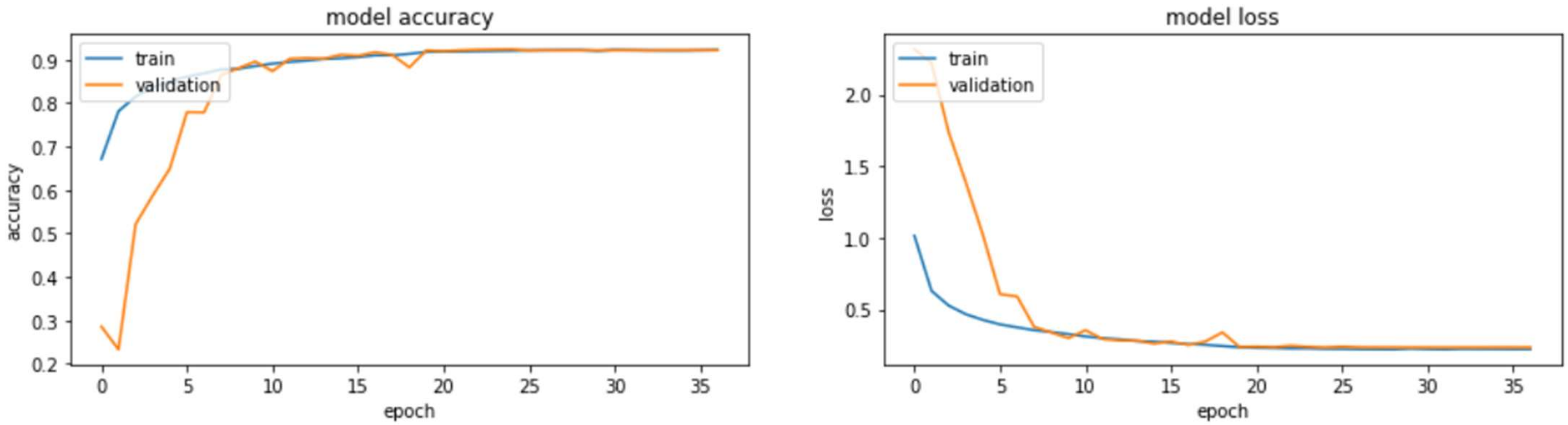
- Visualizing accuracy & loss
  - Uni-model (Transfer learning model VGG16)





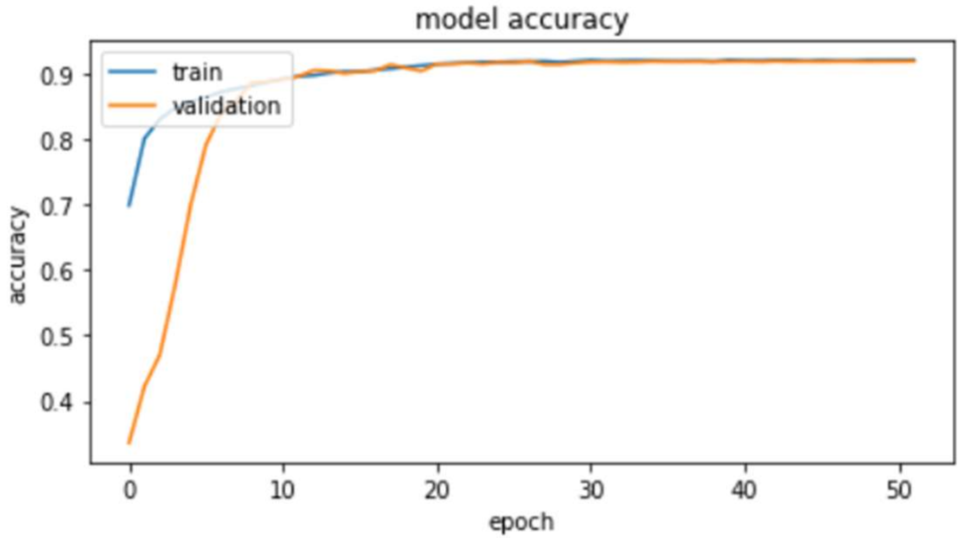

```
Evaluate VGG16 Model
313/313 [==============================] - 4s 12ms/step - loss: 0.3499 - accuracy: 0.8746
[0.3499150276184082, 0.8745999932289124]
```

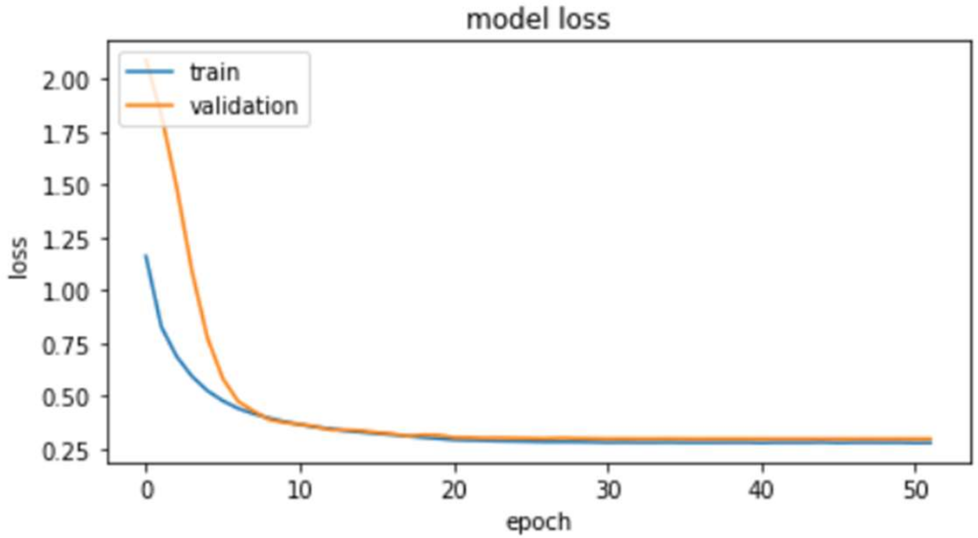
○ Uni-model (CNN with 4 layers)




```
Evaluate CNN Model
313/313 [==============================] - 1s 4ms/step - loss: 0.2246 - accuracy: 0.9232
[0.22458283603191376, 0.9232000112533569]
```

○ Fusion model

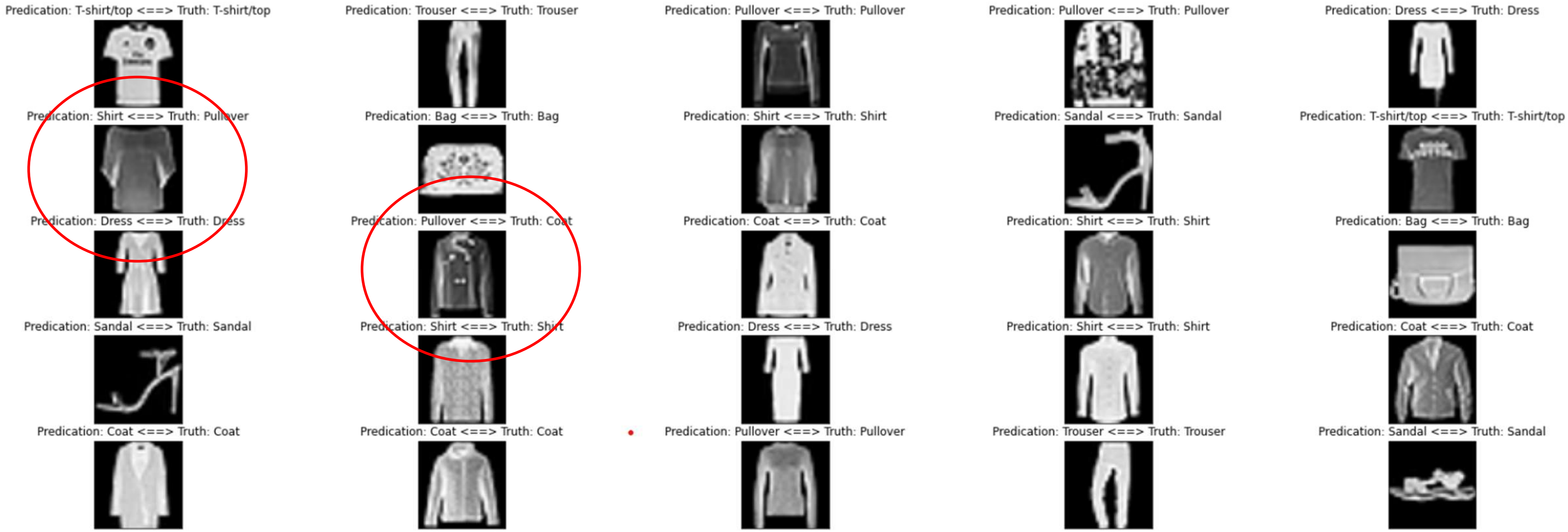




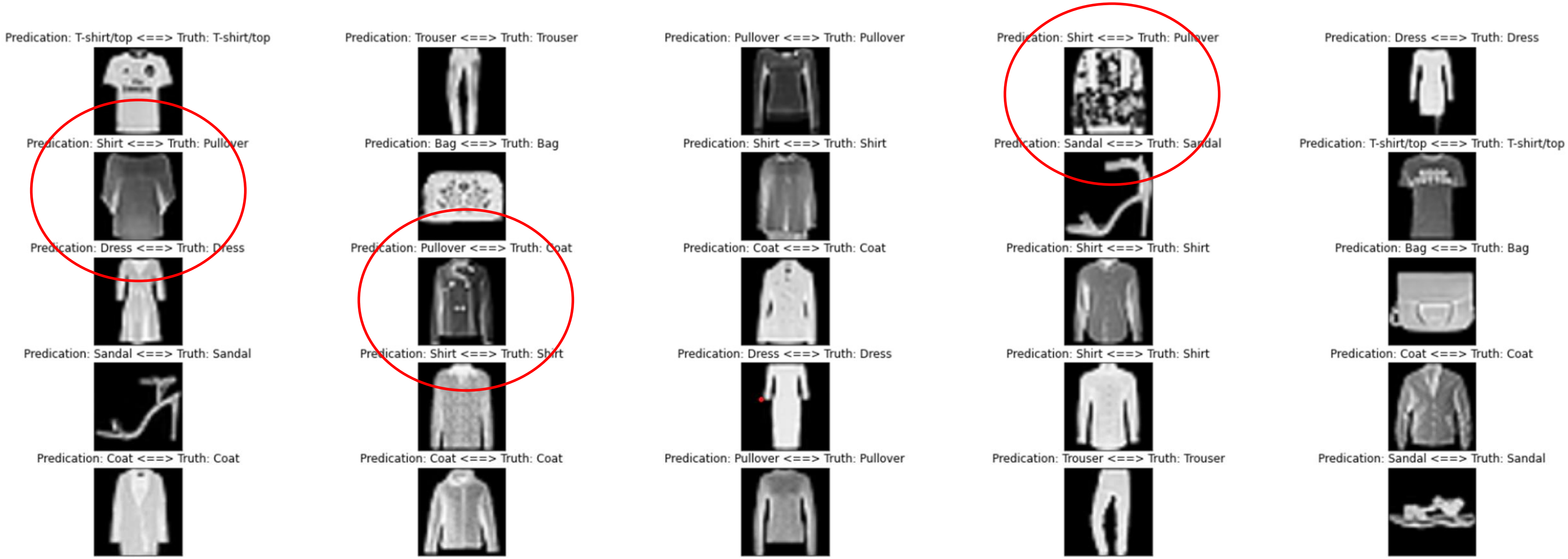

```
Evaluate Fusion Model
313/313 [==============================] - 5s 14ms/step - loss: 0.2793 - accuracy: 0.9200
[0.2793245017528534, 0.9200000166893005]
```

- Model prediction
  - Uni-model (Transfer learning model VGG16)

- Uni-model (CNN with 4 layers)

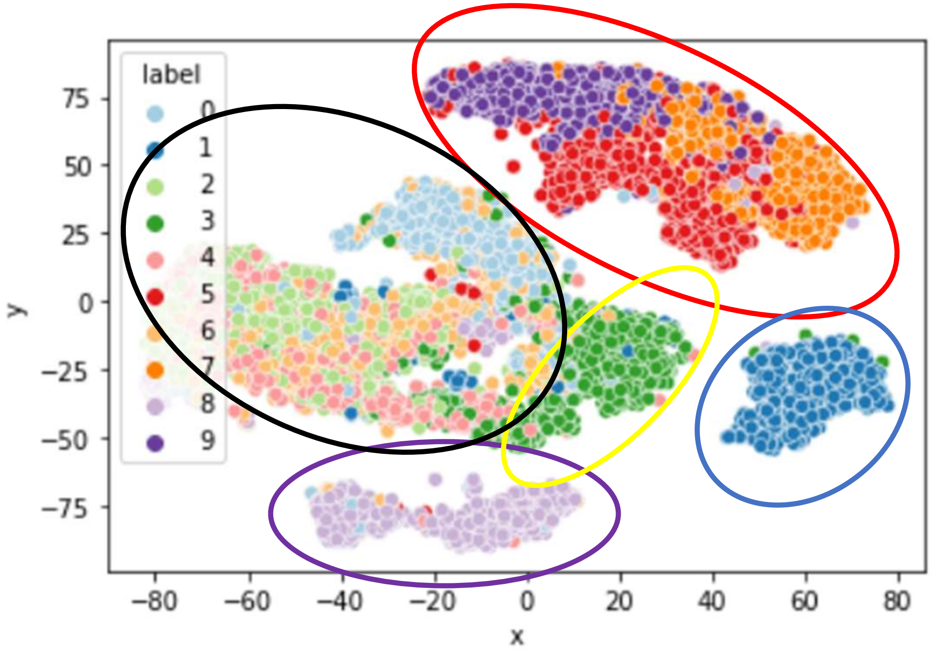


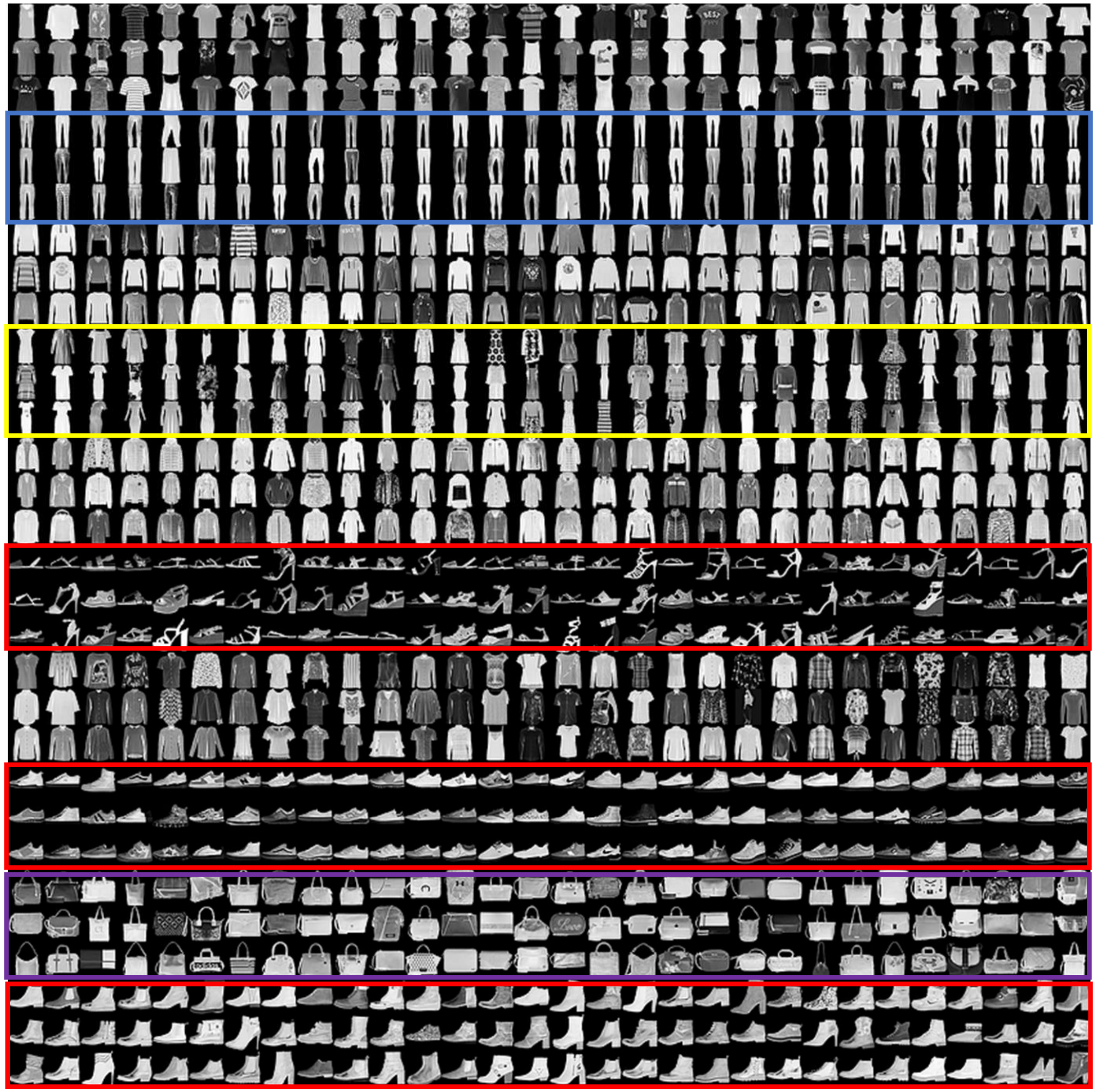
○ Fusion model



- t-SNE plot

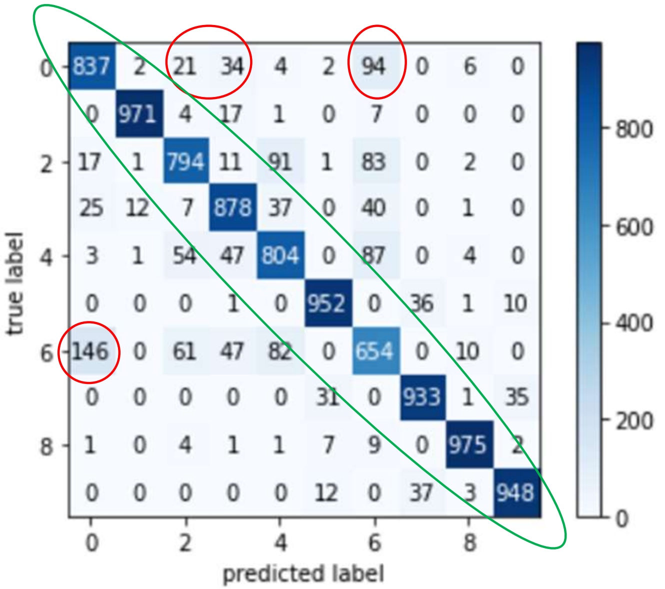
index	0	1	2	3	4	5	6	7	8	9
Label	T-shirt/Top	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Ankle Boot



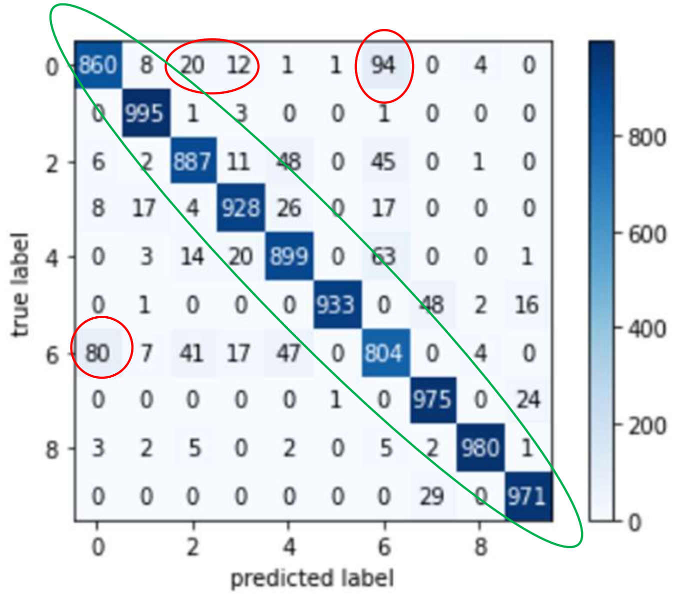


Classes which can be classified distinctly:
Sandal, Sneaker & Ankle Boot
Trouser
Bag
Dress
Classes which cannot be classified distinctly:
T-shirt/Top, Pullover, Coat, Shirt

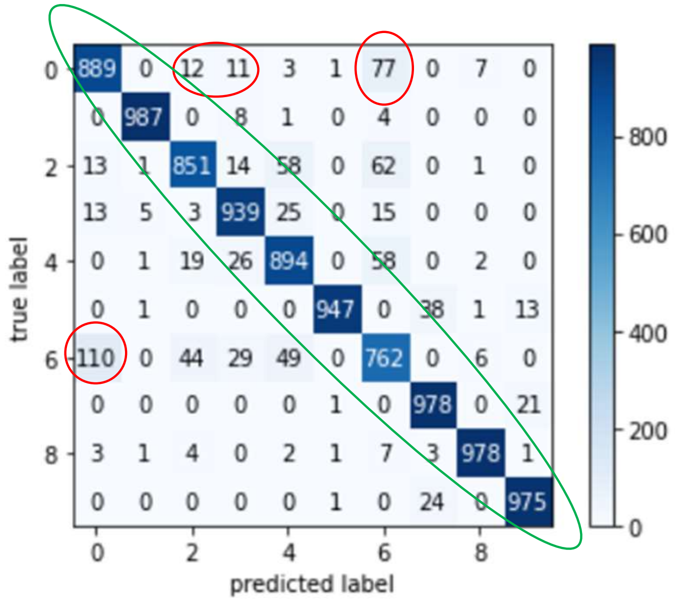
- Confusion Matrix



VGG16



CNN



Fusion

- Classification Report

	precision	recall	f1-score	support
0	0.81	0.84	0.83	1000
1	0.98	0.97	0.98	1000
2	0.84	0.79	0.82	1000
3	0.85	0.88	0.86	1000
4	0.79	0.80	0.80	1000
5	0.95	0.95	0.95	1000
6	0.67	0.65	0.66	1000
7	0.93	0.93	0.93	1000
8	0.97	0.97	0.97	1000
9	0.95	0.95	0.95	1000
accuracy			0.87	10000
macro avg	0.87	0.87	0.87	10000
weighted avg	0.87	0.87	0.87	10000

VGG16

	precision	recall	f1-score	support
0	0.90	0.86	0.88	1000
1	0.96	0.99	0.98	1000
2	0.91	0.89	0.90	1000
3	0.94	0.93	0.93	1000
4	0.88	0.90	0.89	1000
5	1.00	0.93	0.96	1000
6	0.78	0.80	0.79	1000
7	0.93	0.97	0.95	1000
8	0.99	0.98	0.98	1000
9	0.96	0.97	0.96	1000
accuracy			0.92	10000
macro avg	0.92	0.92	0.92	10000
weighted avg	0.92	0.92	0.92	10000

CNN

	precision	recall	f1-score	support
0	0.86	0.89	0.88	1000
1	0.99	0.99	0.99	1000
2	0.91	0.85	0.88	1000
3	0.91	0.94	0.93	1000
4	0.87	0.89	0.88	1000
5	1.00	0.95	0.97	1000
6	0.77	0.76	0.77	1000
7	0.94	0.98	0.96	1000
8	0.98	0.98	0.98	1000
9	0.97	0.97	0.97	1000
accuracy			0.92	10000
macro avg	0.92	0.92	0.92	10000
weighted avg	0.92	0.92	0.92	10000

Fusion

Future Work

- This fusion method can be extended further with
  - Product rule

$$S = \prod_{t=1}^{j} S_t$$

  - Weight rule

$$W = \sum_{t=1}^{j} acc_t$$

$$w_j = \frac{acc_j}{W}, where\ j\ is\ the\ no\ of\ models$$

$$S = \sum_{t=1}^{j} w_t S_t$$

Conclusion

The proposed method of fusing two different models trained on same dataset seems promising. It takes into account both the positive and negative of both the models. From the experiment it is evident that the accuracy of fusion model is close to CNN model which compared to the VGG16 model is high. On further fine tuning of models & proper model selection with respect to data the score level fusion method can result in high accuracy.

Q/A

Aakash Jignesh Modi
CB.EN.P2CEN20001