Sprint4

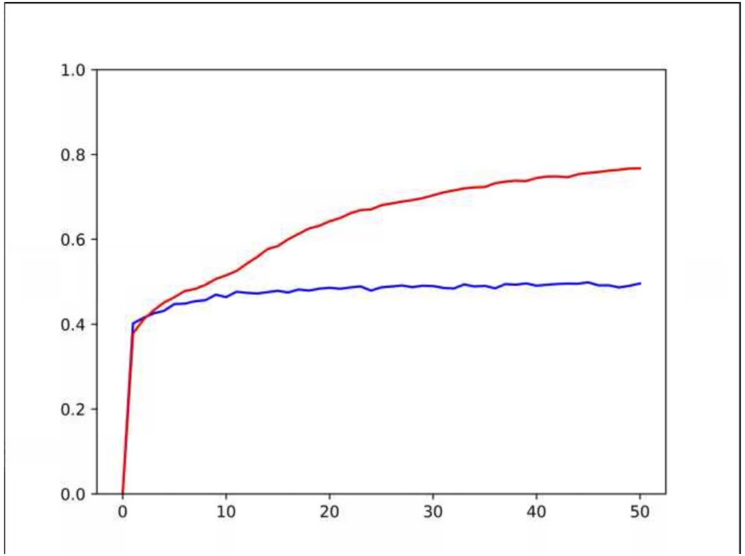
MASS-VQA

What our goals for Sprint 4

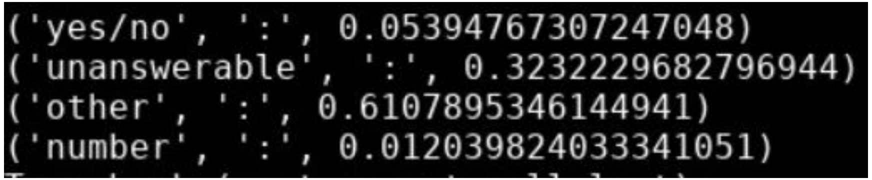
- Preprocessed the dataset for the module
- Analyzed the structure of different pretrained models for image featurization (mainly Resnet 152)
- Trained VizWiz training and validation dataset on chosen model.
- Tested our model for the Test Images from the dataset.

I:Training result

Training result



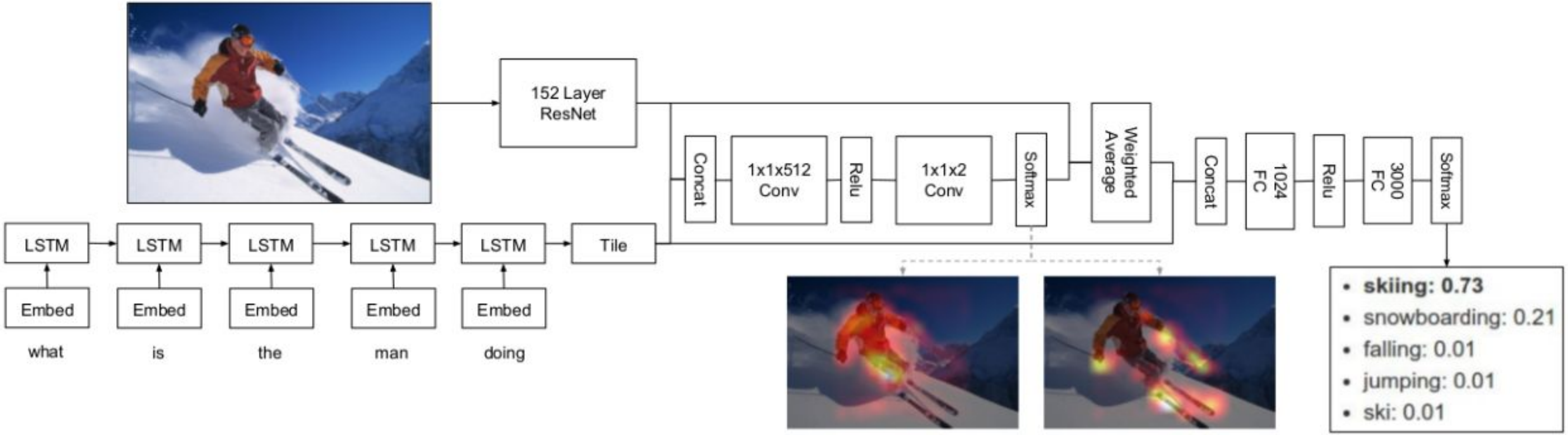
Training accuracy : 0.54
Validation Accuracy: 0.46



```
('yes/no', ':', 0.05394767307247048)
('unanswerable', ':', 0.3232229682796944)
('other', ':', 0.6107895346144941)
('number', ':', 0.012039824033341051)
```

%split of answers

II: Model analyse



$$\hat{a} = \arg\max P(a|I, q)$$

$$\phi = \text{CNN}(I)$$

$$E_q = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_P\} \text{ where } \mathbf{e}_i \in \mathcal{R}^D,$$

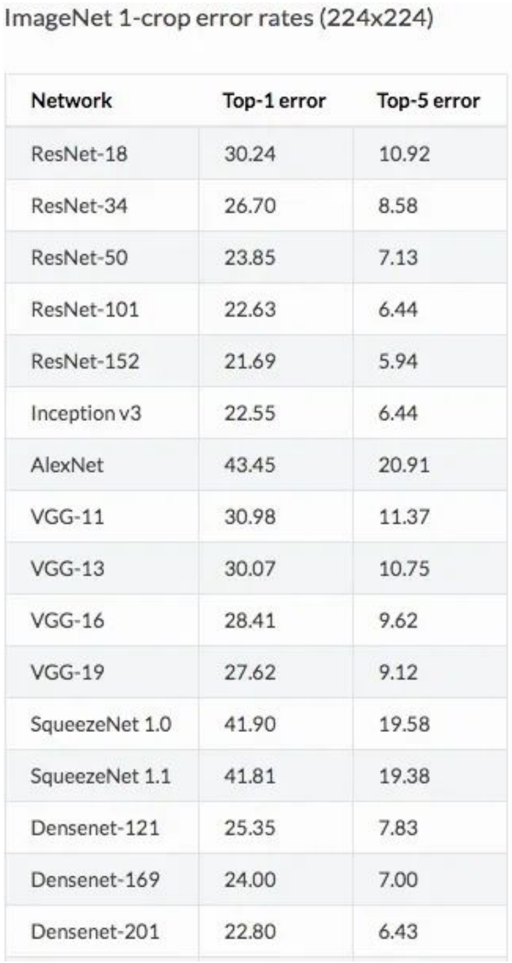
$$\mathbf{s} = \text{LSTM}(E_q)$$

$$\mathcal{L} = \frac{1}{K}\sum_{k=1}^{K} -\log P(a_k|I, q)$$

$$Acc(a) = \frac{1}{K}\sum_{k=1}^{K}\min(\frac{\sum_{1\leq j\leq K, j\neq k}\mathbb{1}(a = a_j)}{3}, 1)$$

Module ResNet-152

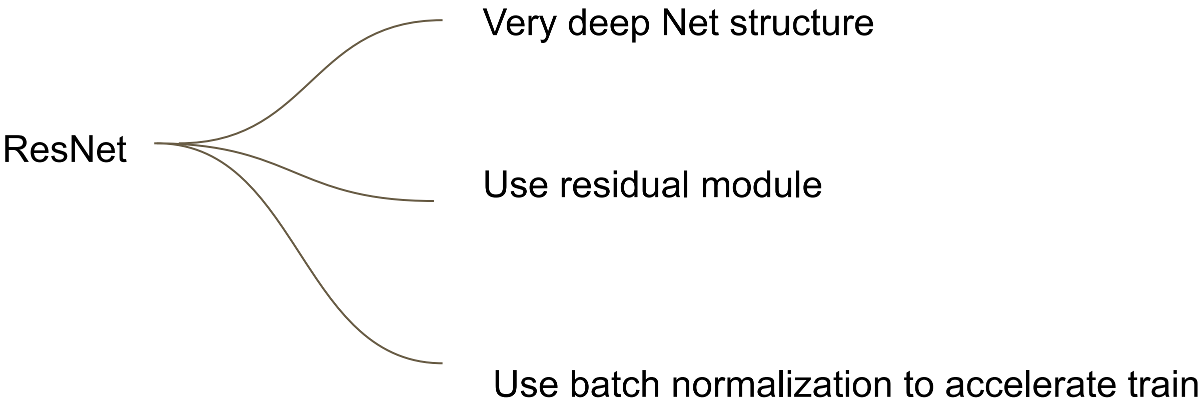
ResNet152 is one of the best-performing neural networks in the current image classification task. The specific performance of each network is shown in the official table given by pytorch.



ImageNet 1-crop error rates (224x224)

Network	Top-1 error	Top-5 error
ResNet-18	30.24	10.92
ResNet-34	26.70	8.58
ResNet-50	23.85	7.13
ResNet-101	22.63	6.44
ResNet-152	21.69	5.94
Inception v3	22.55	6.44
AlexNet	43.45	20.91
VGG-11	30.98	11.37
VGG-13	30.07	10.75
VGG-16	28.41	9.62
VGG-19	27.62	9.12
SqueezeNet 1.0	41.90	19.58
SqueezeNet 1.1	41.81	19.38
Densenet-121	25.35	7.83
Densenet-169	24.00	7.00
Densenet-201	22.80	6.43

Module ResNet-152

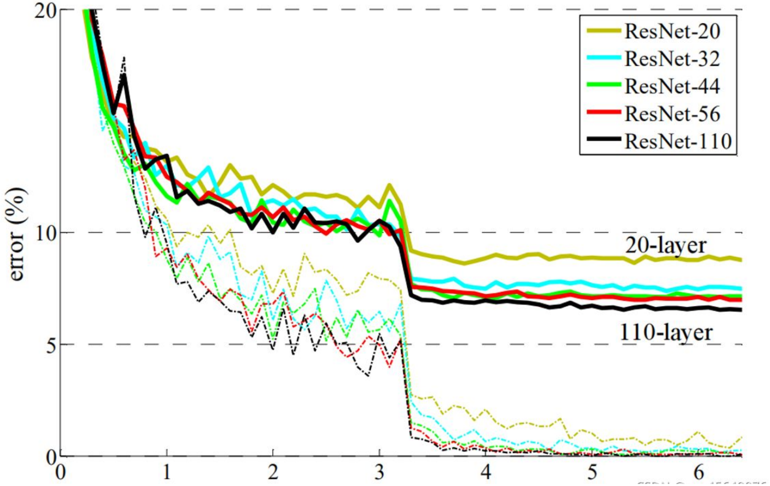


ResNet

Very deep Net structure

Use residual module

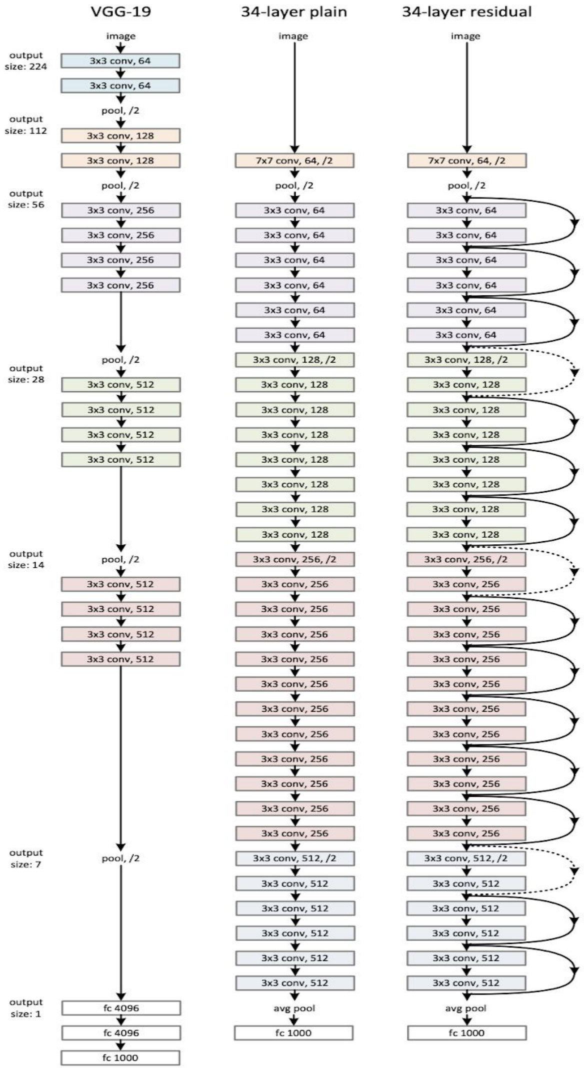
Use batch normalization to accelerate train

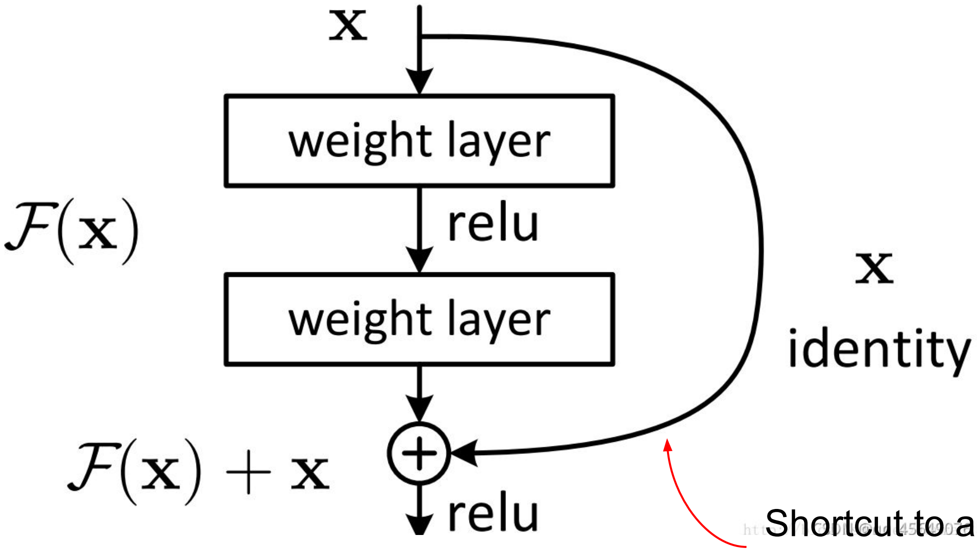


In order to solve the degradation problem in the deep network, some layers of the neural network are made to skip the next layer to weaken the strong connection between each layer. Such neural networks are called Residual Networks (ResNets).

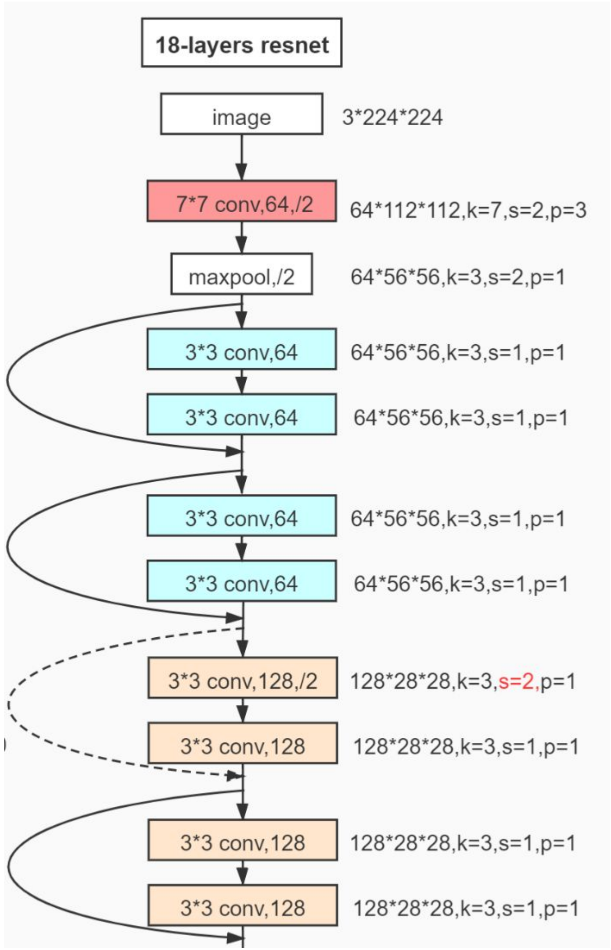
The figure is a convolutional network using a residual structure. It can be seen that as the network continues to deepen, the effect becomes better.
(Dotted line is train error, solid line is test error)

ResNet (Residual Network) is a convolutional neural network that democratized the concepts of residual learning and skip connections. This enables to train much deeper models.

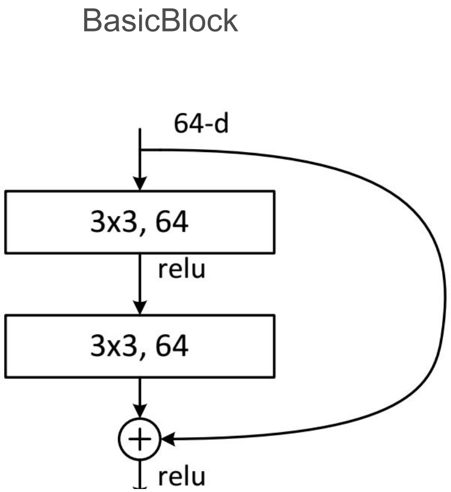


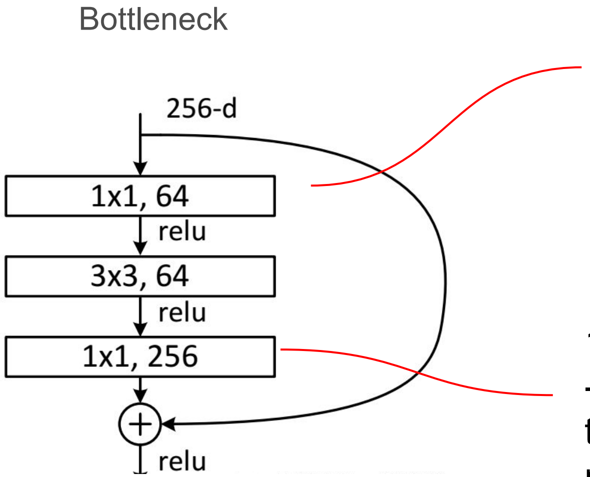


Shortcut to add
Input with output



Two different residuals





1×1 convolution kernel
-dimension reduction operation
-reducing the depth of the feature matrix from 256 to 64;

1×1 convolution kernel
-increase the dimension the depth of the feature matrix is increased from 64 to 256.

Future works

- Improve on unanswerable questions
- Preprocess a different dataset (VQA 2.0) so that it aligns with our model.
- Do the training with VQA 2.0 dataset and compare.
- Use custom test images instead of using the test dataset.