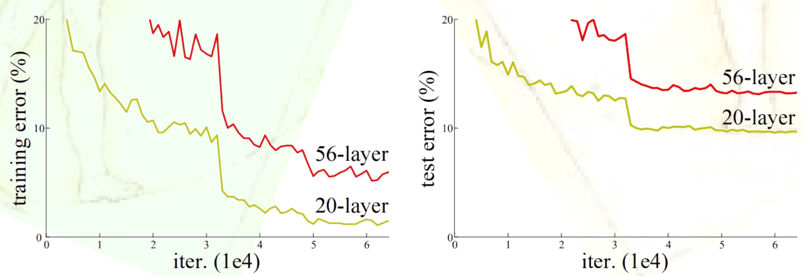
ResNet
DenseNet

Pattern Recognition & Machine Learning Laboratory
Geonjun Yang, Aug. 3rd, 2021

Deep Residual Learning for Image Recognition [K. He et al., 2016] (1/4)

- **Goal**
  - Provide evidence that residual network (ResNet) is easier to optimize
  - Gain high level of accuracy from considerably increased depth
- **Motivation**
  - The more layers are stacked, the higher accuracy you get
    - Problem of vanishing/exploding gradients
  - Offer a substantially deeper and simpler model
- **Contribution**
  - Won 1st place on the ILSVRC 2015
  - Achieved 3.57% error on the ImageNet test set using ensemble
  - Obtain 28% relative improvement on the COCO object detection dataset
  - ResNet is still used as backbone framework of lots of tasks



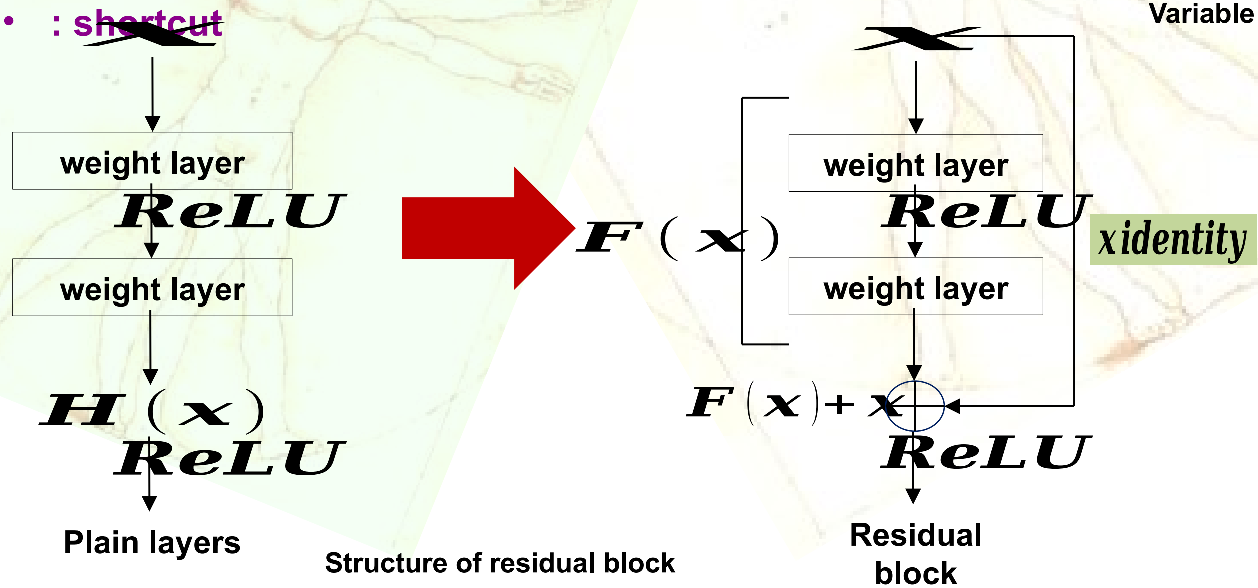
Training error and test error on CIFAR-10

Deep Residual Learning for Image Recognition [K. He et al., 2016] (1/4)

- **Residual block**
  - Reduce the optimization difficulty of network by using residual block
    - The deeper network is, the more difficult to train as you wanted
    - Simply stacking a lot of layers doesn't guarantee high performance
    - Instead of training , which is an inherent mapping and difficult to train immediately, train
  - Structure
    - : multiple convolutional layers
    - : shortcut

- input value

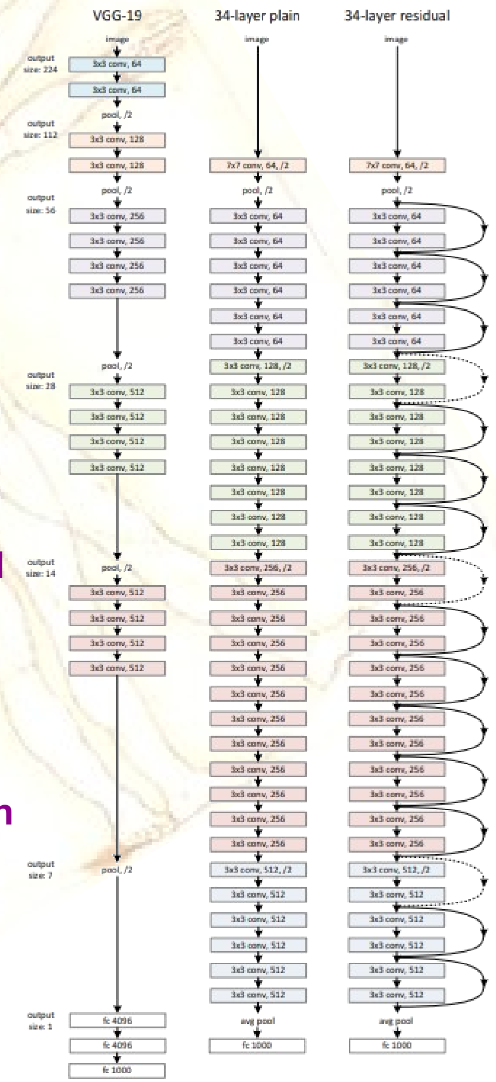
Variable explanation



Structure of residual block

Deep Residual Learning for Image Recognition [K. He et al., 2016] (1/4)

- **Network architectures**
  - **Plain Network**
    - Mainly inspired by the philosophy of VGG nets
      - 3x3 filters
      - If the feature map size is halved, the number of filters is doubled
    - Down sampling by conv layers that have a stride of 2
    - 1000 fully connected layer with softmax
  - **Residual network**
    - Insert shortcut connections based on plain network
    - Identity shortcuts can be directly used when input and output are of the same dimensions
      - Zero padding
      - Projection shortcut (done by 1x1 convolutions)
    - Adopt batch normalization
    - Learning rate starts from 0.1 and is divided by 10 when the error plateaus



Network architectures for ImageNet

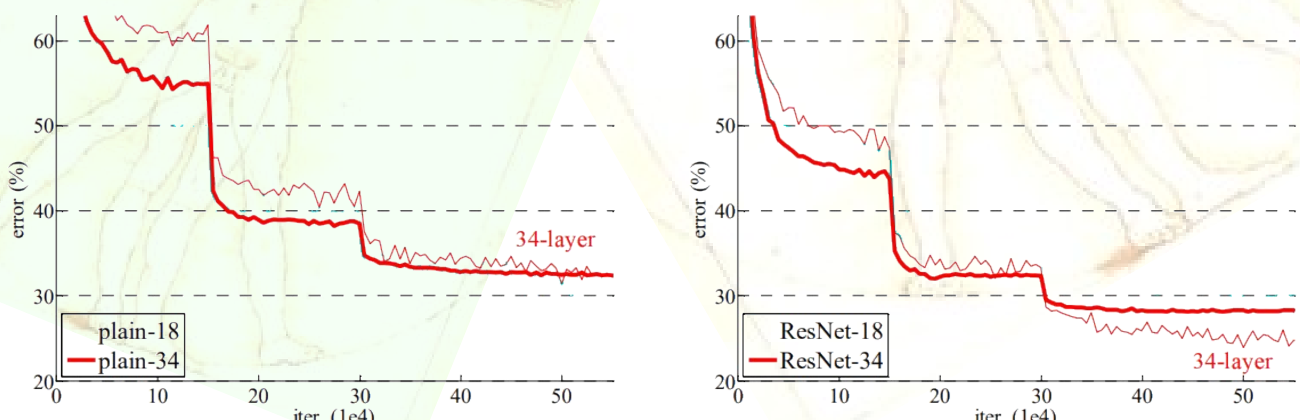
K. He, X. Zhang, S. Ren and J. Sun , " Deep Residual Learning for Image Recognition," *CVPR*, 2016.

Deep Residual Learning for Image Recognition [K. He et al., 2016] (1/4)

- **Experiments**
  - ImageNet Classification
    - Plain networks
      - 34-layer plain net has higher validation error than 18-layer net
      - Degradation problem observed
      - Unlikely caused by vanishing gradients
      - May have exponentially low convergence rates
    - Residual Networks
      - 34-layer ResNet is better than 18-layer ResNet
      - Considerably lower training error and generalizable to the validation data
      - Faster convergence at the early stage

Top-1 error on ImageNet validation

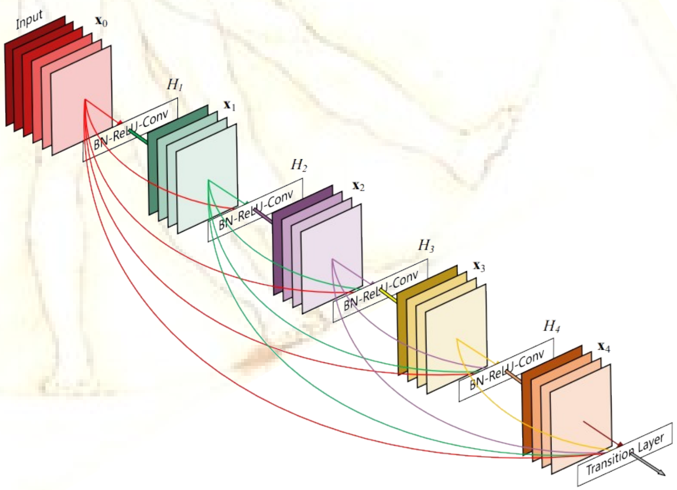
	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	**25.03**



Training on ImageNet

Densely Connected Convolutional Networks [G. Huang et al., 2017] (1/4)

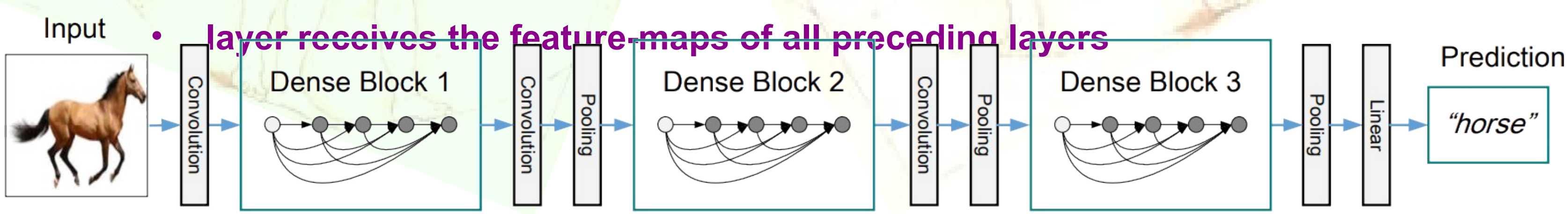
- **Goal**
  - Offer Dense Convolutional Network (DenseNet) which has less parameters and shows better performance than ResNet
  - Lower computation cost
- **Motivation**
  - All recent network topologies create short paths
  - DenseNet connects each layer to every layer
- **Contribution**
  - Alleviates the vanishing gradient problem, strengthen feature propagation
  - Encourages feature reuse, reduce the number of parameters
  - Obtain significant improvements over the state-of-the-art at that time



Dense block

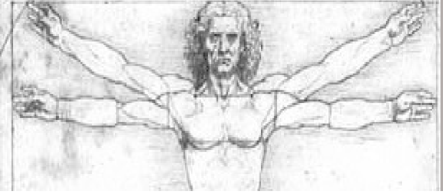
Densely Connected Convolutional Networks [G. Huang et al., 2017] (1/4)

- **Dense block**
  - Connect all layers directly with each other
    - preserve the feed-forward nature
  - Combine features by concatenating
    - To concatenate features, the sizes of all feature maps should be the same
    - Contains transition layer consisting of batch normalization, 1x1 conv, 2x2 average pooling
- **Connectivity**
  - ResNet
    - contains convolution layer or pooling, batch normalization, ReLU
    - May impede the information flow in the network
  - DenseNet
    - layer receives the feature-maps of all preceding layers



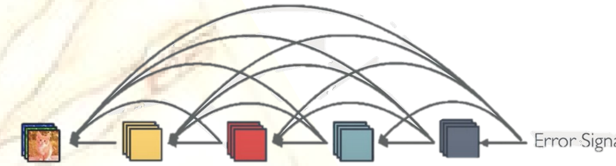
DenseNet with three dense blocks

G. Huang, Z. Liu and L. Maaten, " Densely Connected Convolutional Networks," *CVPR,* 2017.

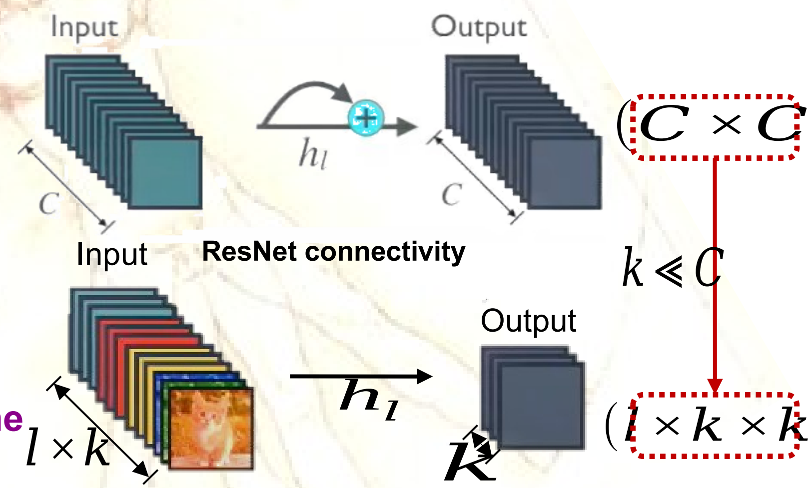


Densely Connected Convolutional Networks [G. Huang et al., 2017] (1/4)

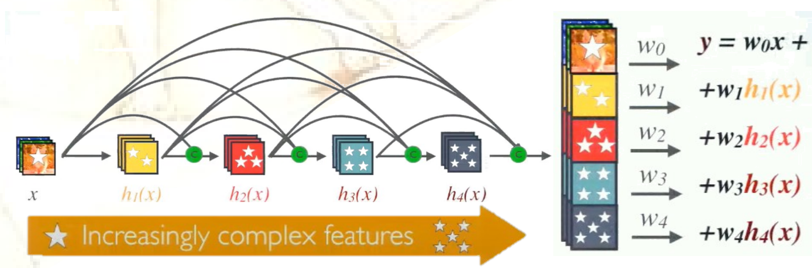
- **Advantages of DenseNet connectivity**
  - **Strong gradient flow**
    - **Arrow signal can be easily propagated**
    - **Implicit and direct deep supervision**
  - **Parameter & computational efficiency**
    - **In normal conv net,the number of parameters is proportional to (: layer width)**
    - **In DenseNet the number of parameters is proportional to (growth rate)**
  - **Maintains low complexity features**
    - **In standard conv net, classification is done based on the last layer**
    - **However, in DenseNet classifier uses features of all complexity levels**
    - **Uses both complex features and simple features**
    - **Gives smooth decision boundaries and high generalization performance**



Strong gradient flow



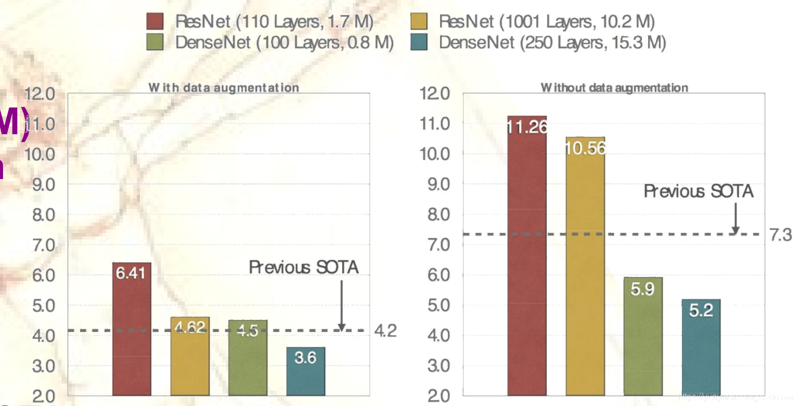
DenseNet connectivity **Number of parameters**



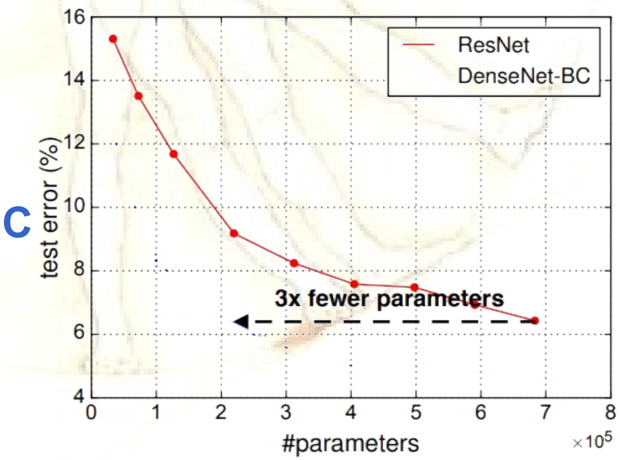
Use features of all complexity levels

Densely Connected Convolutional Networks [G. Huang et al., 2017] (1/4)

- **Results on CIFAR-10**
  - With data augmentation
    - DenseNet with much less parameters (0.8M) shows similar performance to ResNet with 10.2M parameters
  - Without data augmentation
    - DenseNet shows significantly lower test error than ResNet
    - Much better performance than previous SOTA
- **Results on ImageNet**
  - DenseNet-BC refers to DenseNet with
    - : # channels in transition layer
    - # output channel
  - At the same level of test error, DenseNet-BC shows three times less parameters than those of ResNet



Results on CIFAR-10



Results on ImageNet