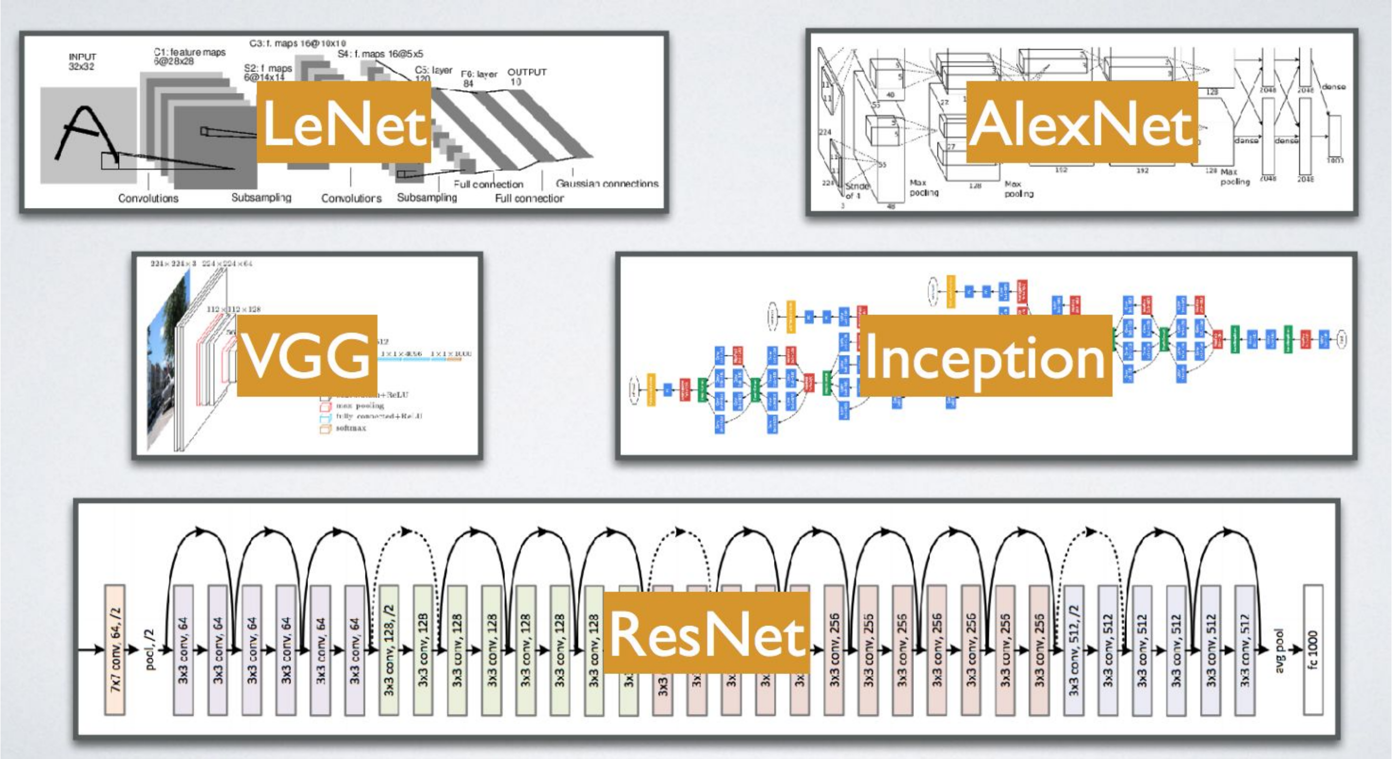
Densely Connected Convolutional Networks

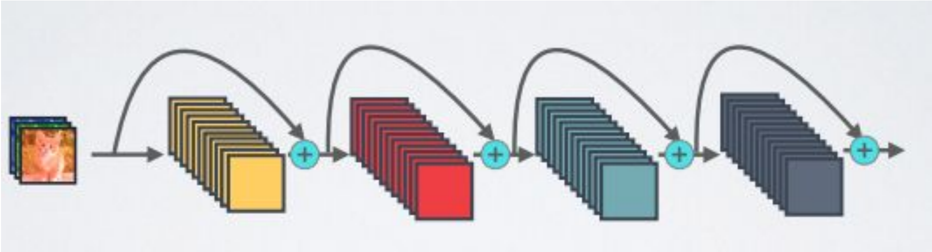
Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger
Presentation: Xingyi Zhou

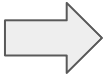
Task: CNN architecture design

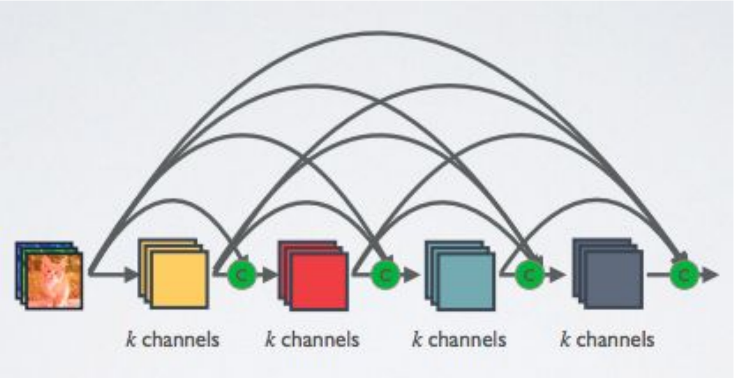


slides credit: Gao Huang

Idea: Skip connection is helpful, do it throughly







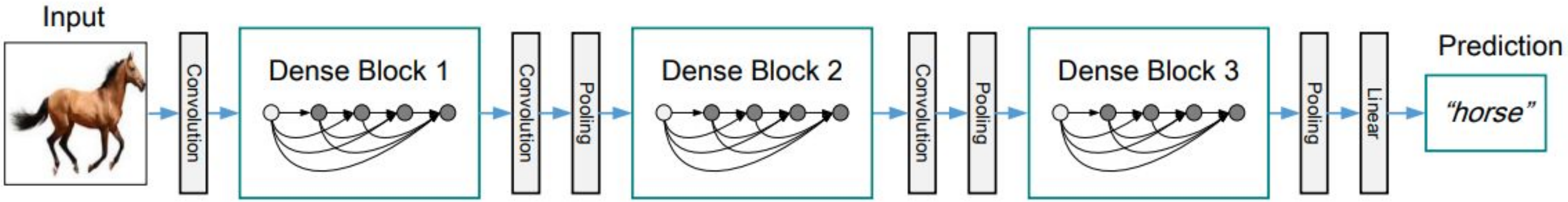
He et al. 2017, ResNet

DenseNet

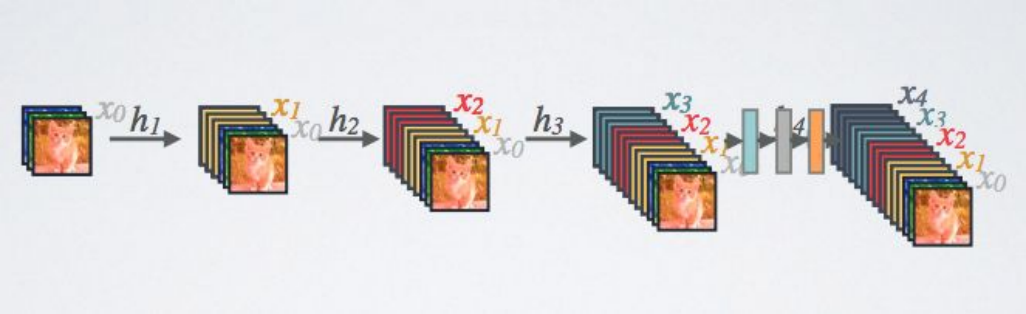
Concatenate all features from previous layers. Each layer outputs fewer Feature channels.

images credit: Gao Huang

Overall architecture

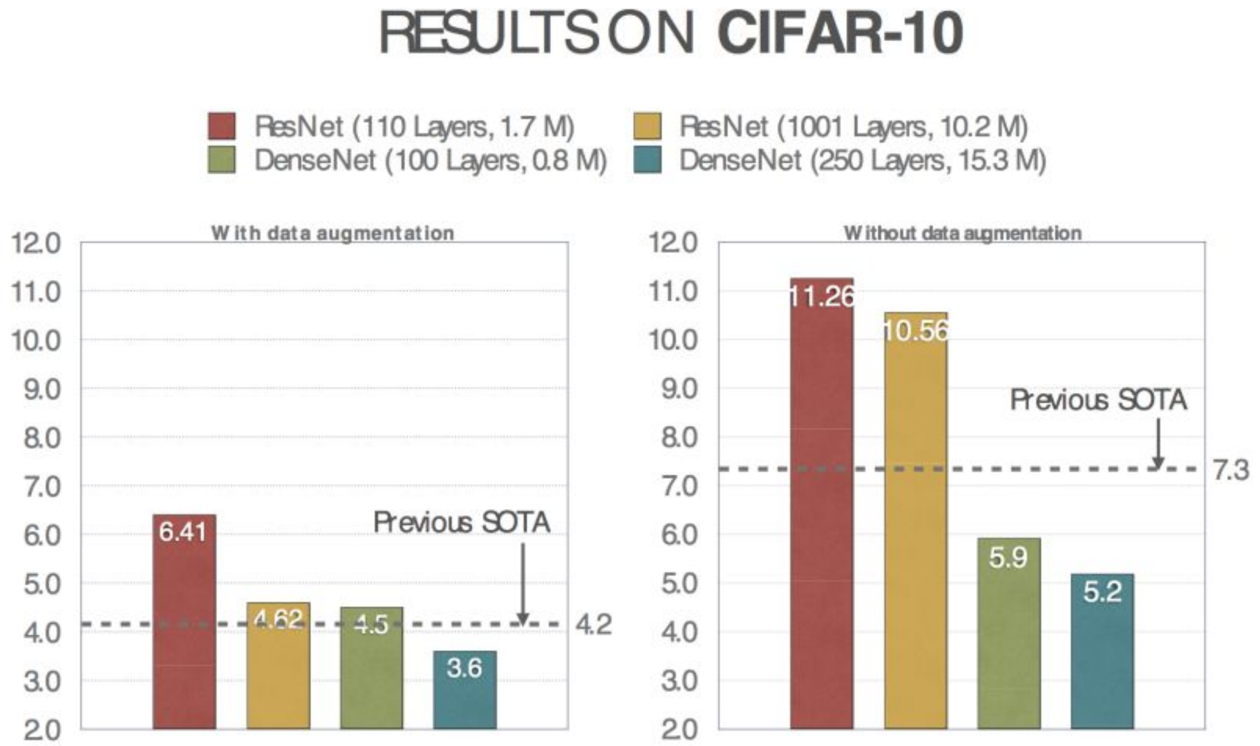


Feed forward in each dense block



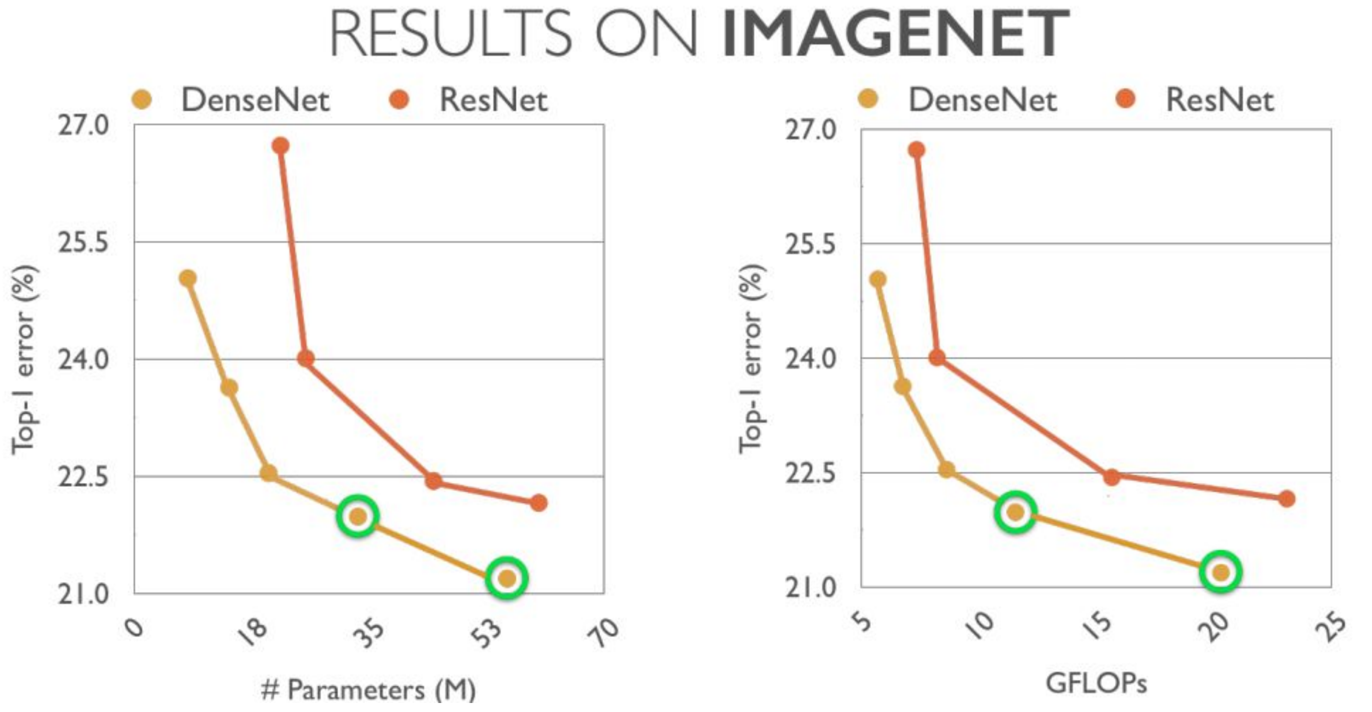
images credit: Gao Huang

Results: Cifar 10



images credit: Gao Huang

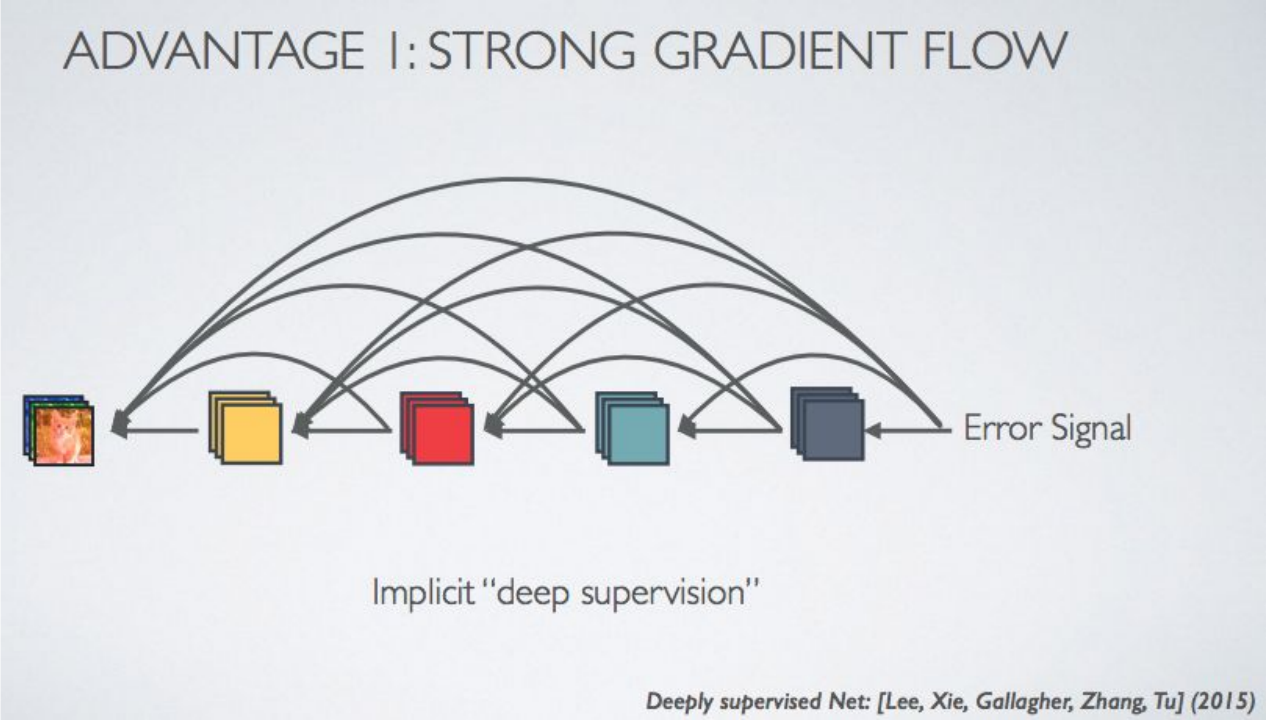
Results: Image Net



images credit: Gao Huang

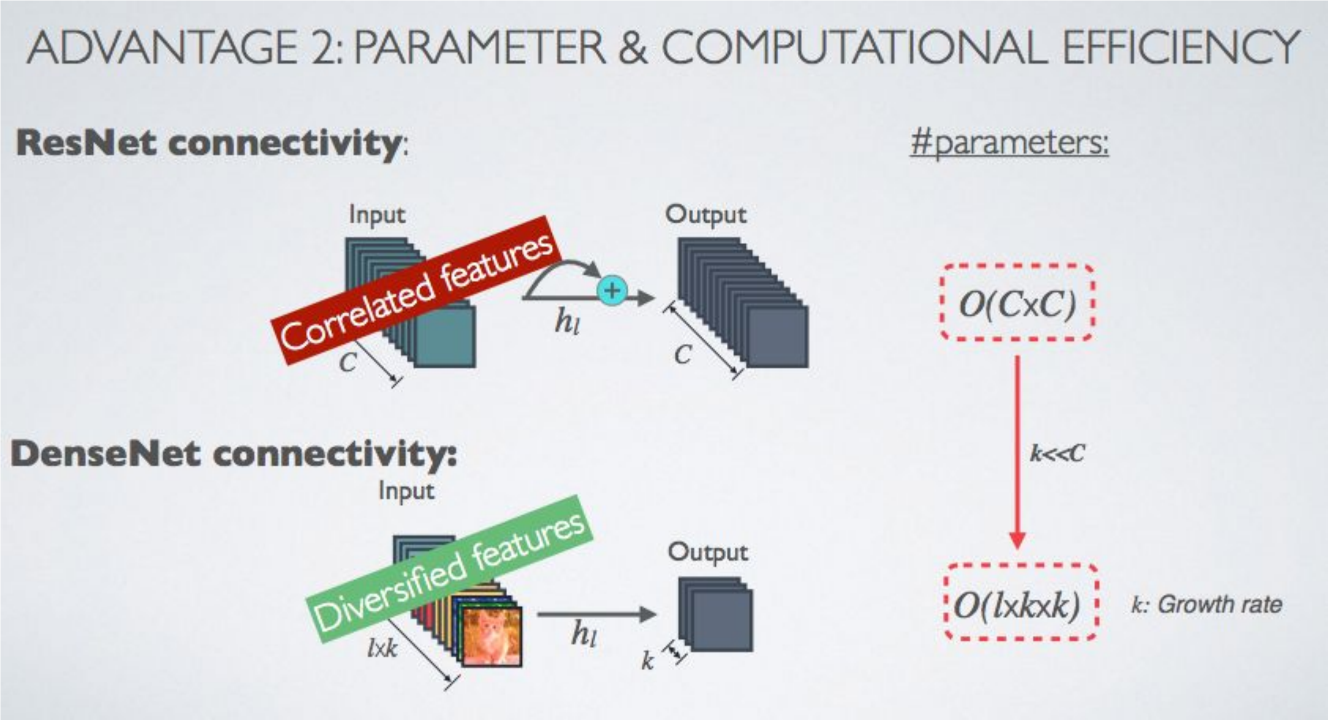
Pros

Pros from the paper author:



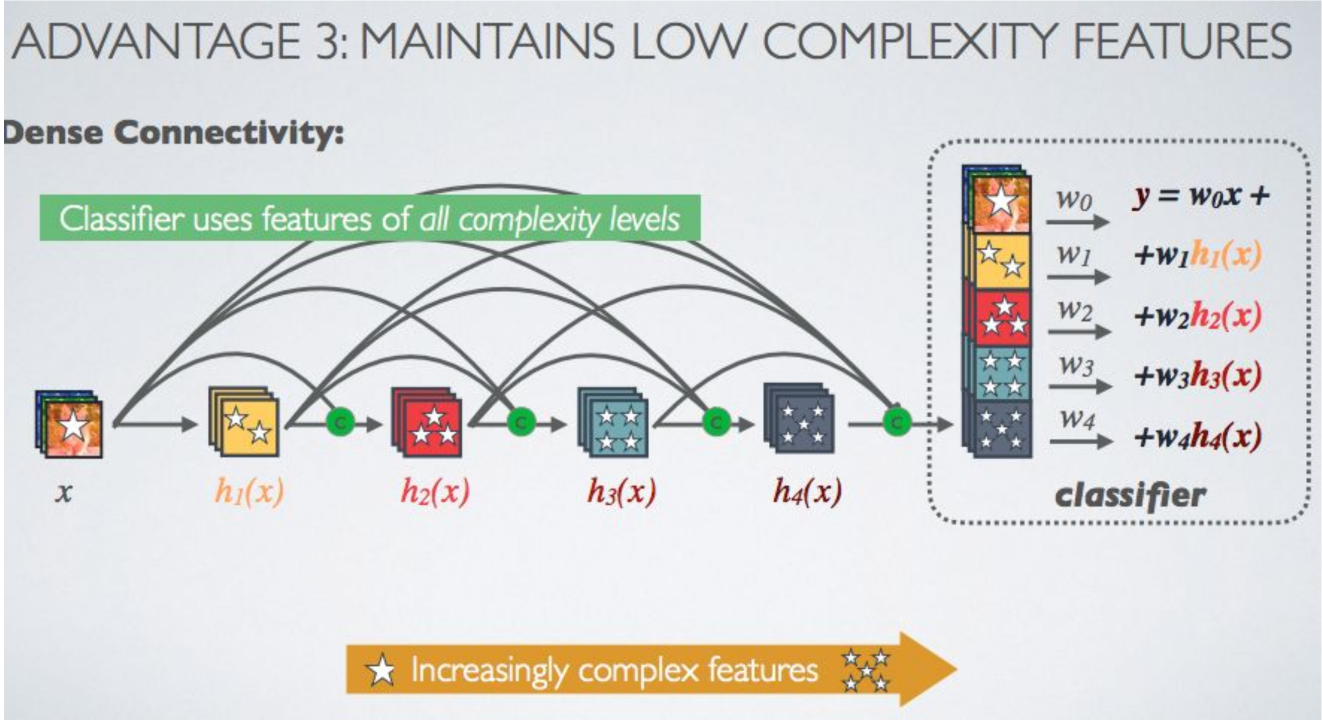
images credit: Gao Huang

Pros from the paper author:



images credit: Gao Huang

Pros from the paper author:

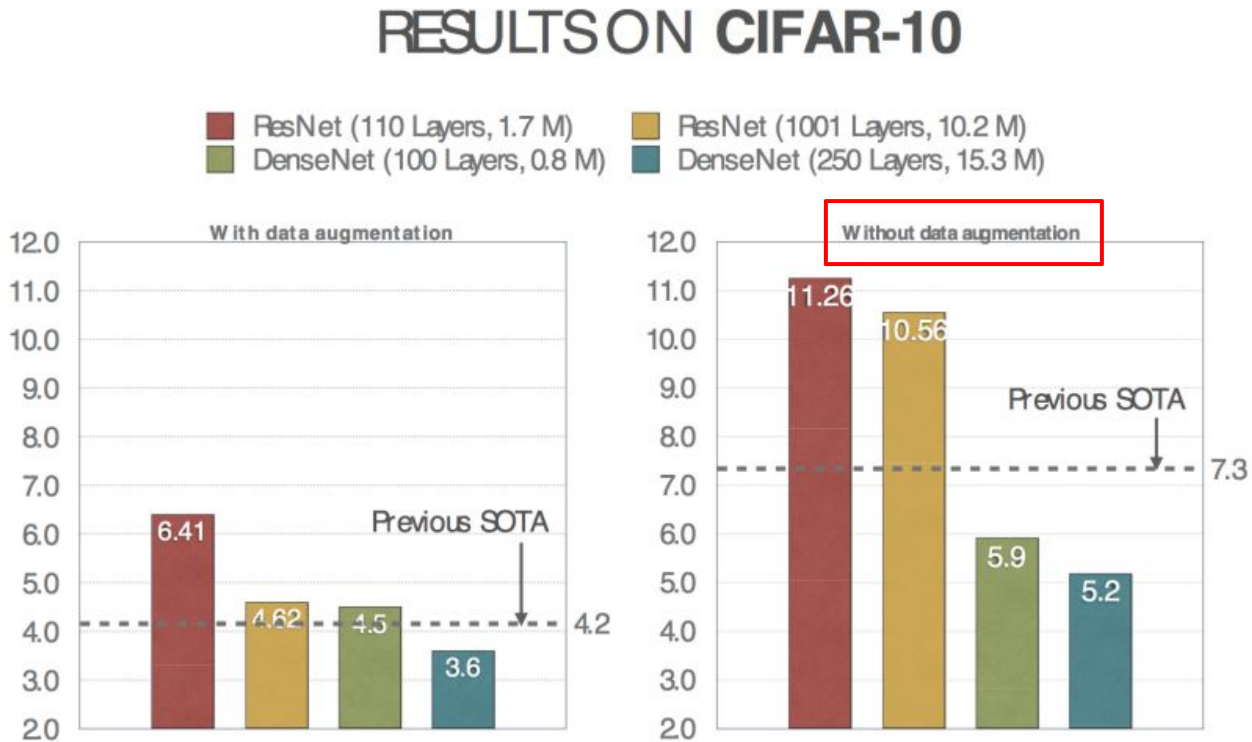


images credit: Gao Huang

Pros #4: Enable traning from scratch for small datasets

- ResNet > 50 will not converge if training from scratch in a small dataset.
- DenseNet 200 can be trained from scratch for image classification, with the same accuracy comparing to ImageNet-pretrained.
- Can also be used to train object detection from scratch. (Shen et al. ICCV17, DSOD: Learning Deeply Supervised Object Detectors from Scratch)
- Rethinking the effectiveness of pretraining: is it from more ‘information input’ or just provide a better initialization for optimization?
- Is long range skip connection a key factor for training from scratch?

Pro #5: Generilization ability



images credit: Gao Huang

Pro #5: Generilization ability (Continued)

Issues on learning deep models

- **Representation** ability
  - Ability of model to fit training data, if optimum could be found
  - If model A's solution space is a superset of B's, A should be better.
- **Optimization** ability
  - Feasibility of finding an optimum
  - Not all models are equally easy to optimize
- **Generalization** ability
  - Once training data is fit, how good is the test performance

images credit: Kaiming He

Pro #5: Generilization ability (Continued)

How do ResNets address these issues?

- **Representation** ability
  - No explicit advantage on representation (only re-parameterization), but
  - Allow models to go deeper
- **Optimization** ability
  - Enable very smooth forward/backward prop
  - Greatly ease optimizing deeper models
- **Generalization** ability
  - Not explicitly address generalization, but
  - Deeper+thinner is good generalization

DenseNet

- It is more likely that densenet decreases the representation ability.

- Pro #1 #4, deep supervision via skip supervision.

- Low representation ability actually increase the generalization ability. (compare to weight decay)

images credit: Kaiming He

Pro #6: Discussion of the connection to previous works.

- Model compactness:
- Implicit Deep Supervision:
- Stochastic vs. deterministic connection
- Feature Reuse ⇒ network slim