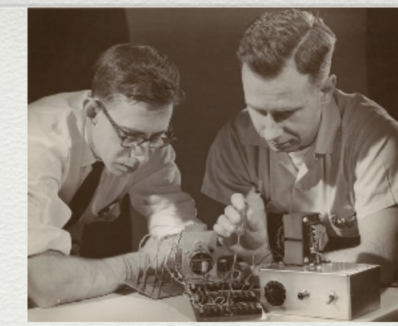
RECENT NN MODELS

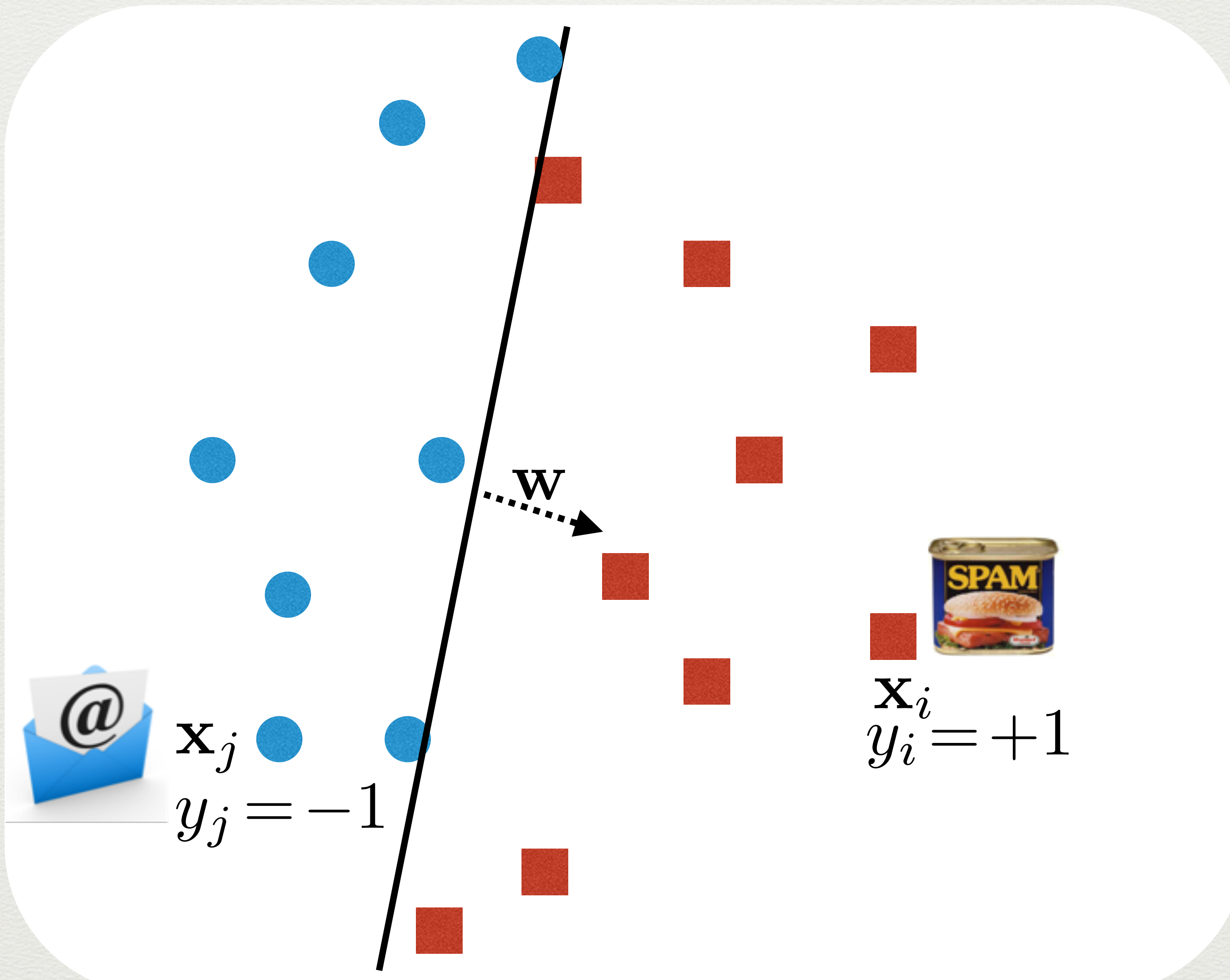
A 3 Minutes Introduction to

"Deep Learning"

Perceptron

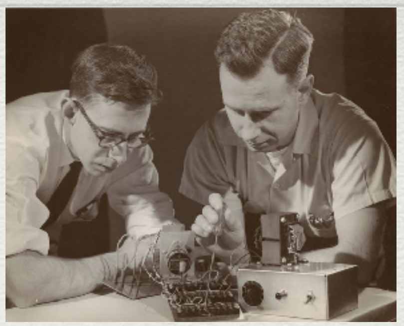


[Rosenblatt 1957]

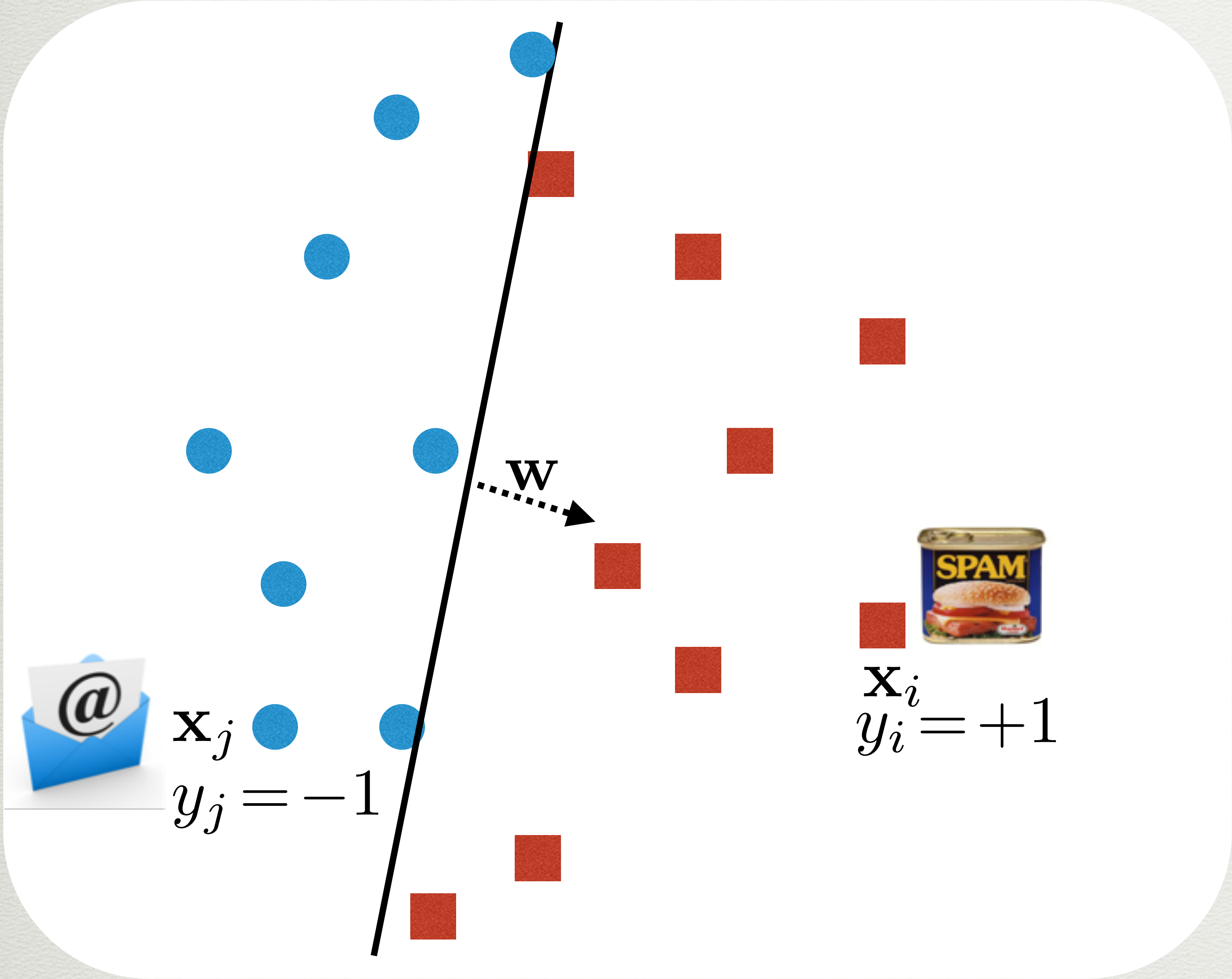


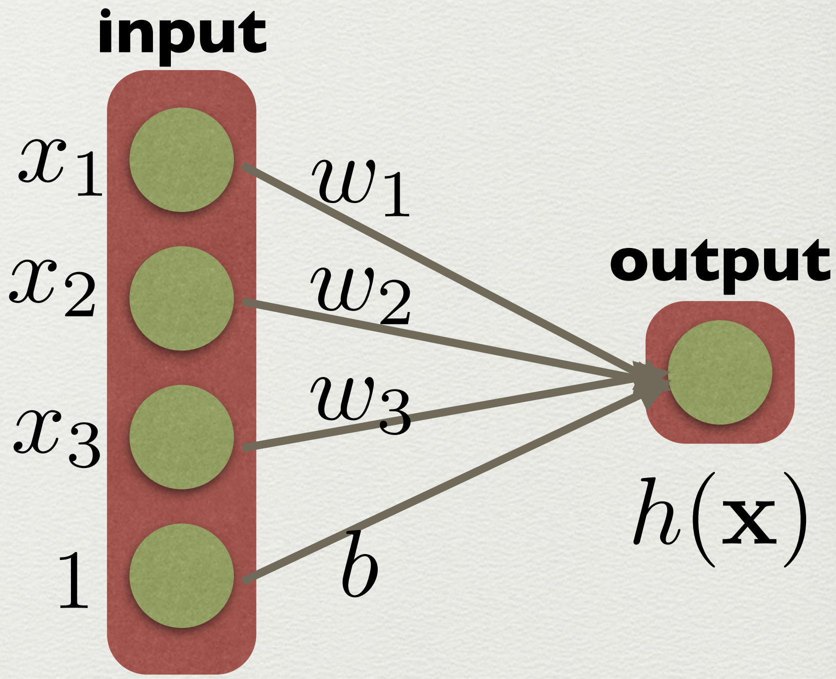
$$h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

Perceptron



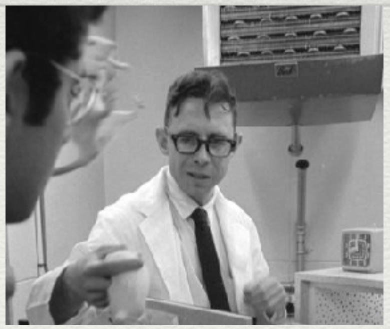
[Rosenblatt 1957]





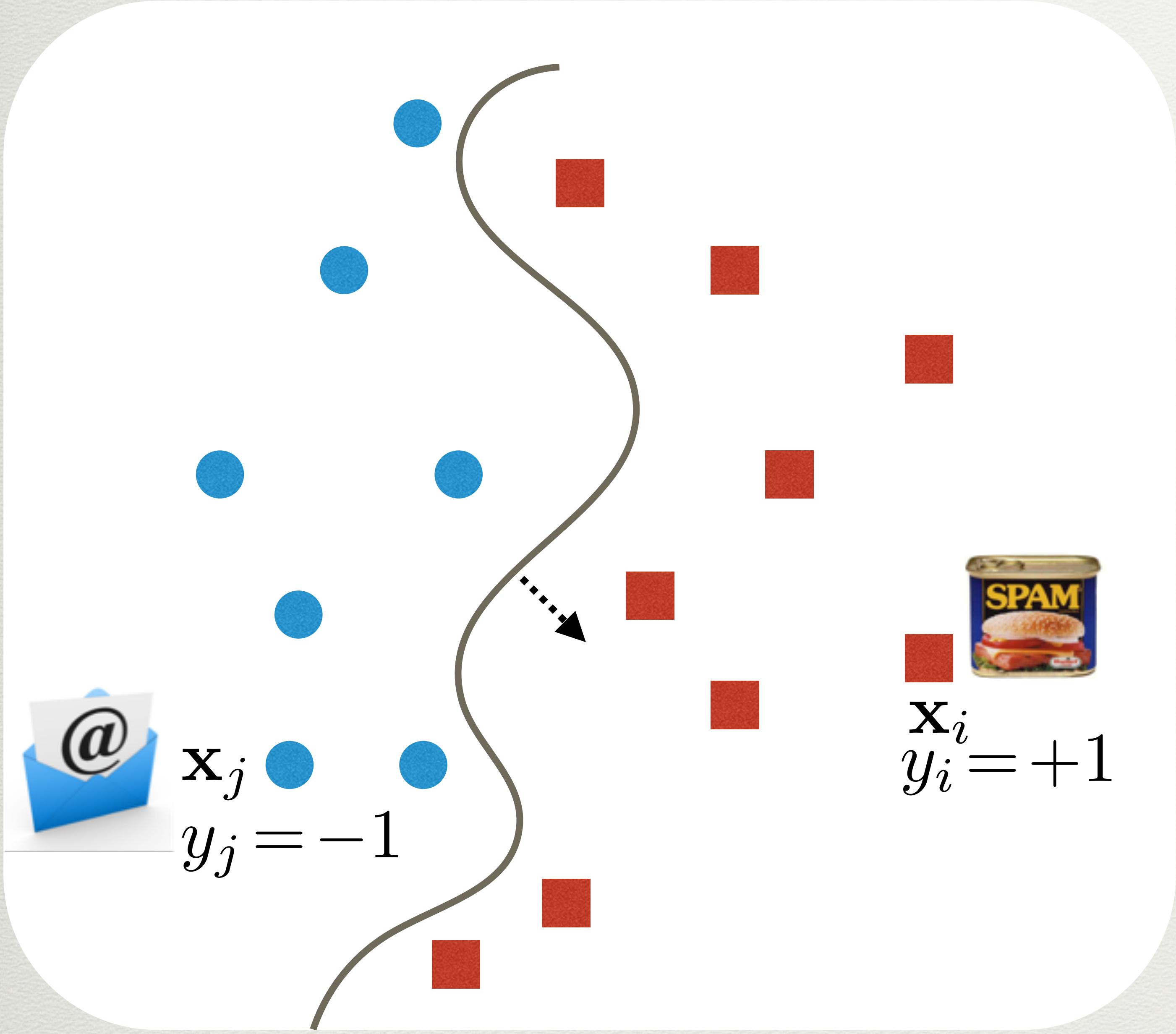
$$h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

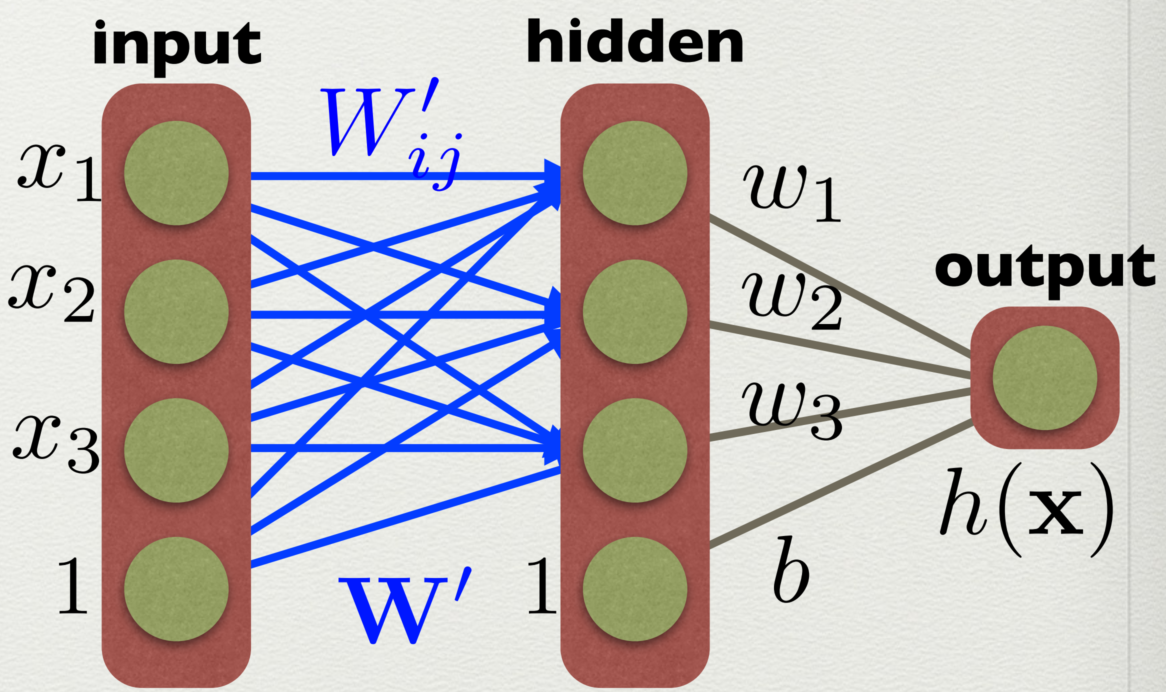
Multi-Layer Perceptron



(a.k.a. Neural Networks)

[Rosenblatt 1961]





$$\sigma(a) = \max(a, 0)$$

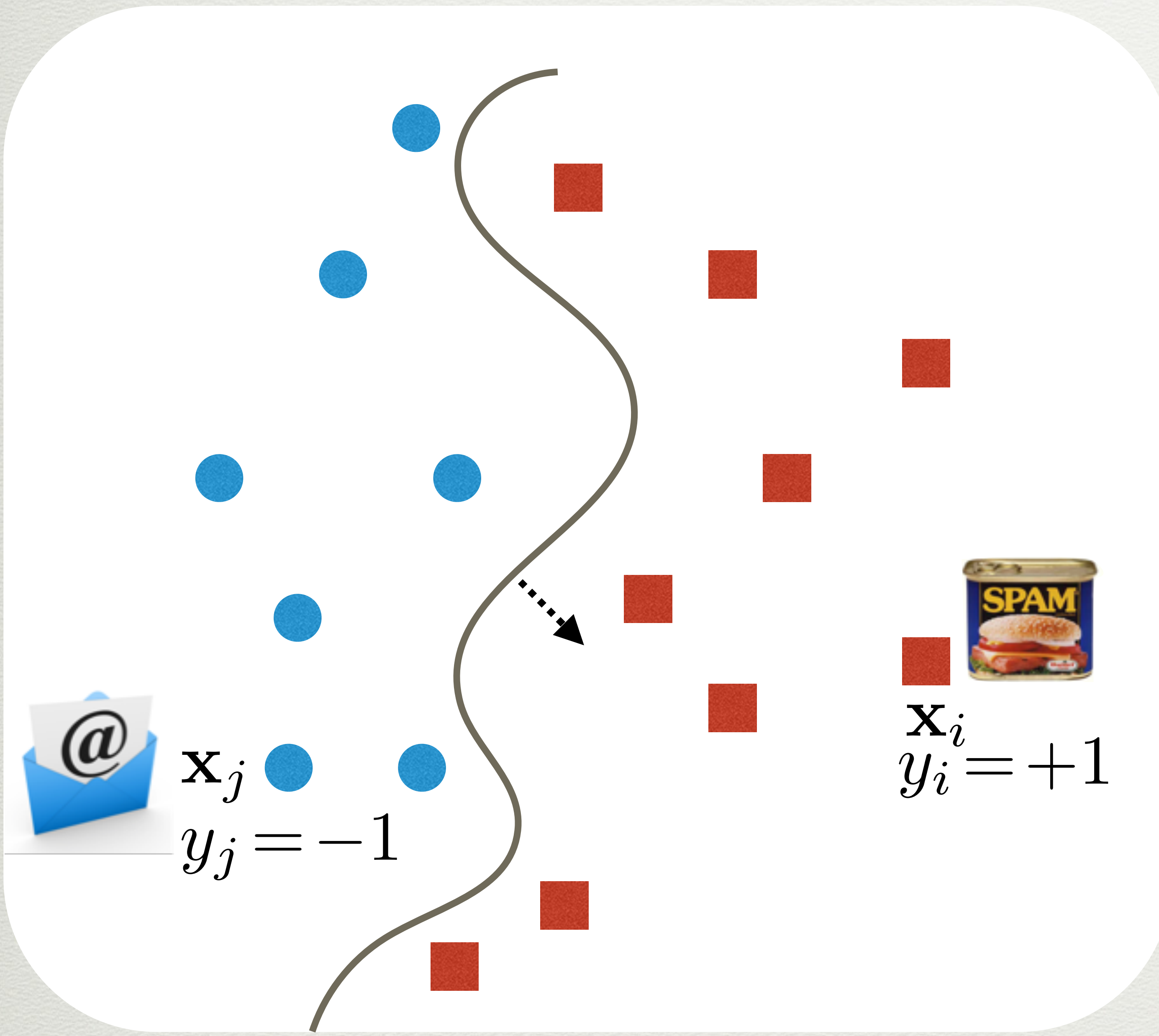
Rectified Linear Units

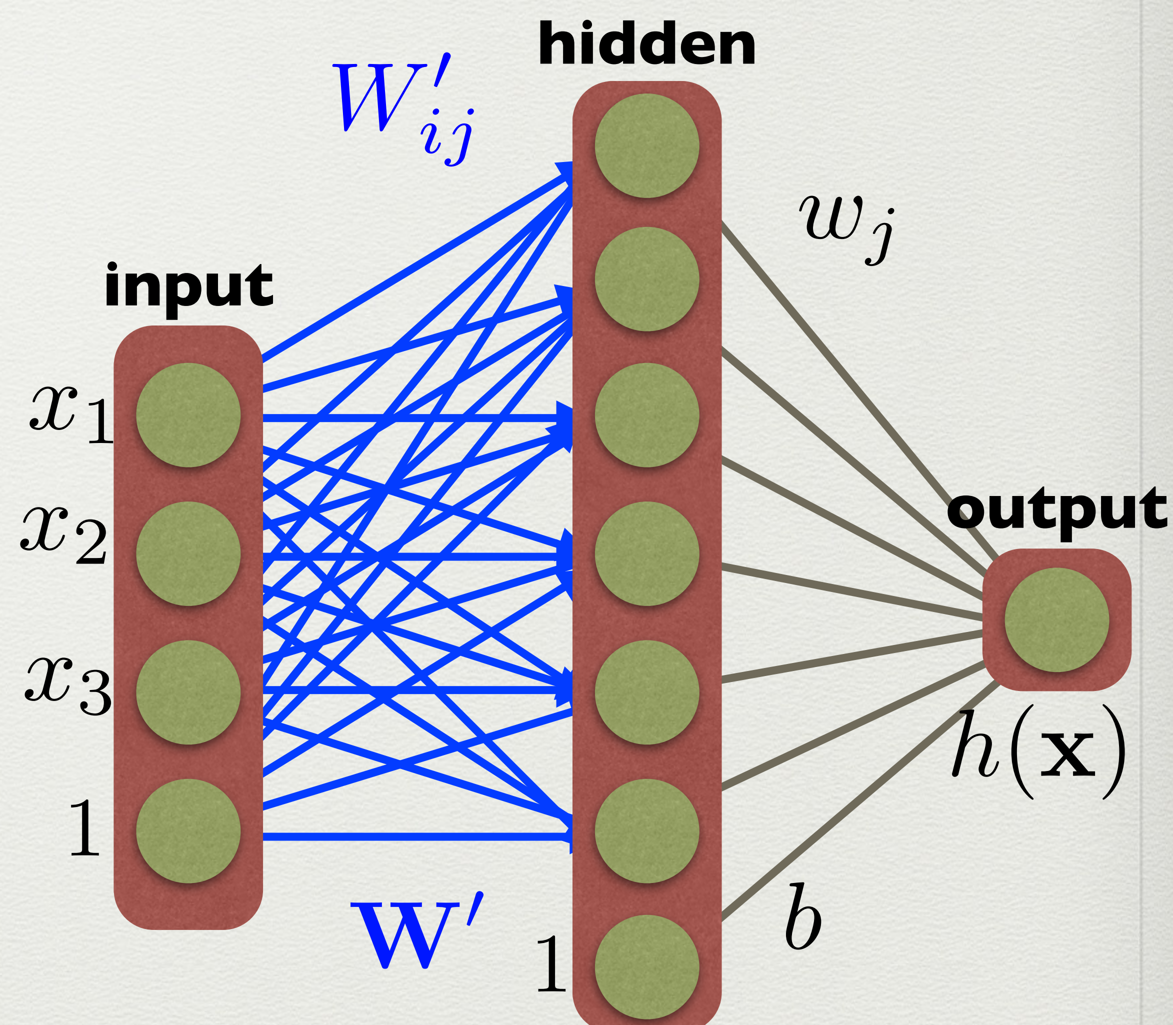
$$h(\mathbf{x}) = \mathbf{w}^\top \sigma\left(\mathbf{W}'\mathbf{x} + \mathbf{c}\right) + b$$

Multi-Layer Perceptron

(a.k.a. Neural Networks)

[Rosenblatt 1961]





$$\sigma(a) = \max(a, 0)$$

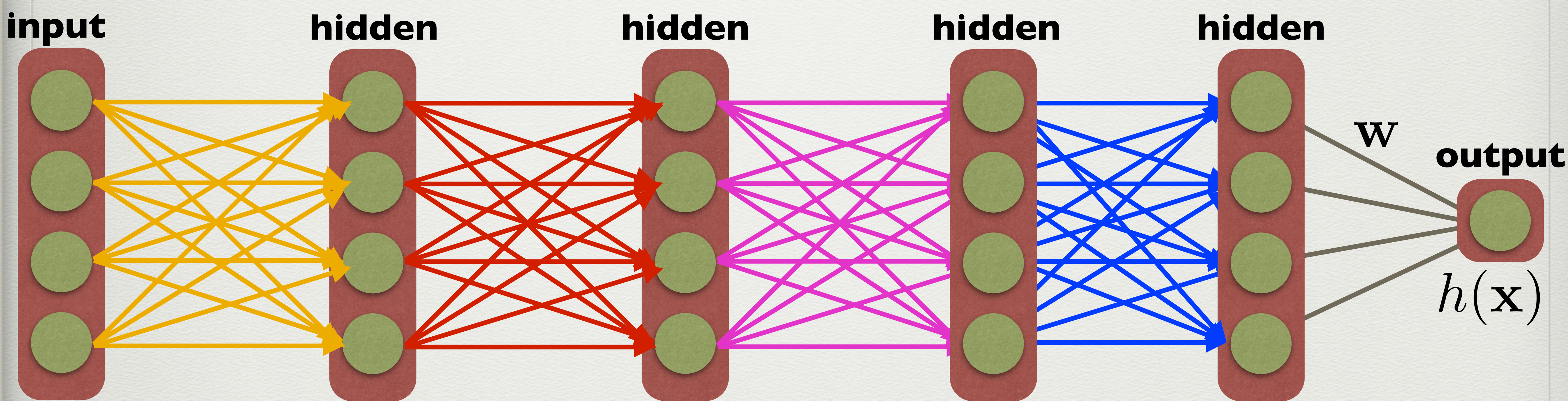
Rectified Linear Units

$$h(\mathbf{x}) = \mathbf{w}^\top \sigma\left(\mathbf{W}'\mathbf{x} + \mathbf{c}\right) + b$$

Multi-Layer Perceptron

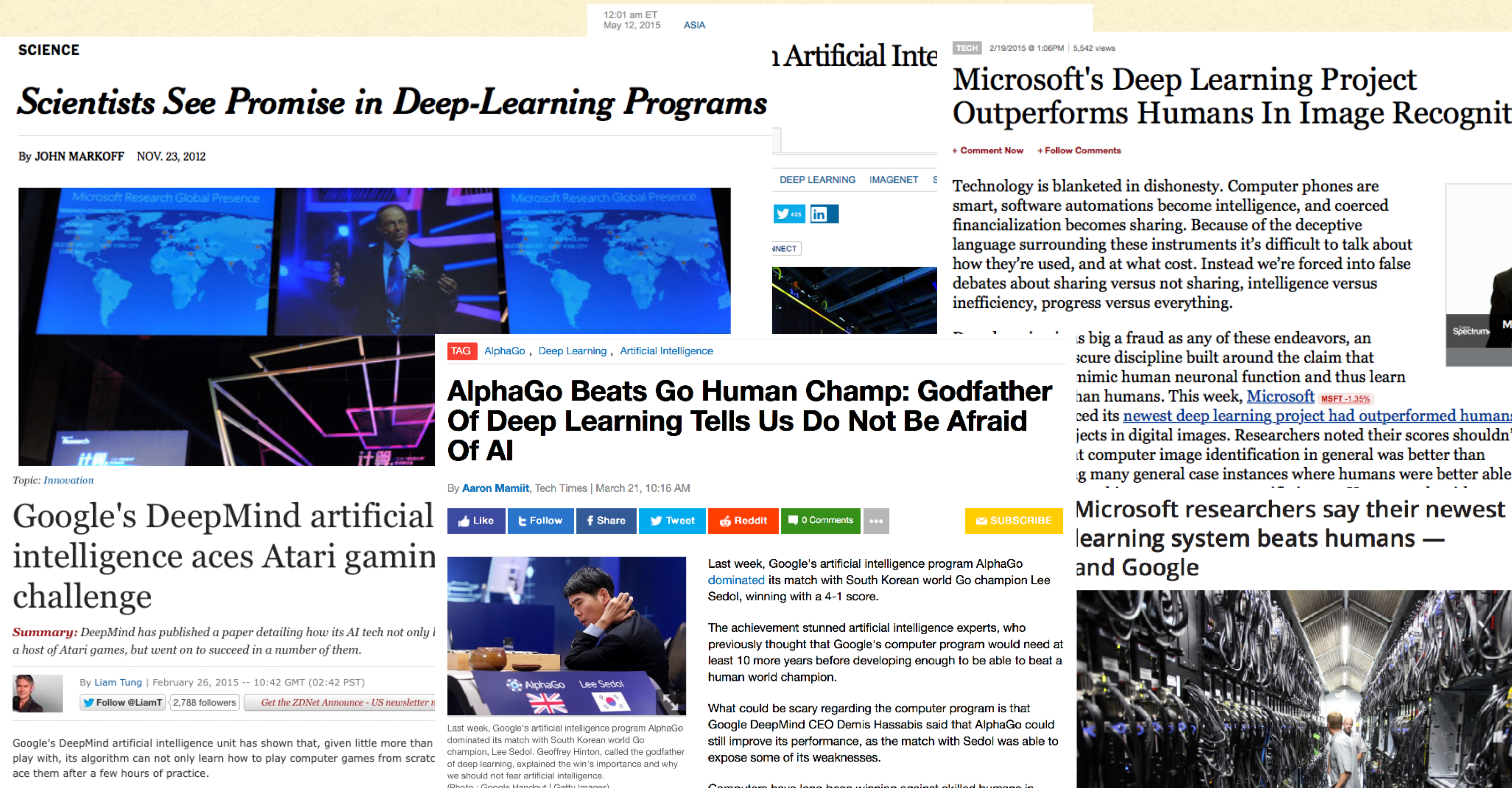
(a.k.a. Neural Networks, Deep Learning)

[Rosenblatt 1961]



$$h(\mathbf{x}) = \mathbf{w}^\top \sigma(\mathbf{W}' \sigma(\mathbf{W}^2 \sigma(\mathbf{W}^3 \sigma(\mathbf{W}^4 \mathbf{x}))))$$

DEEP LEARNING WORKS



12:01 am ET
May 12, 2015 ASIA

SCIENCE

Scientists See Promise in Deep-Learning Programs

By JOHN MARKOFF NOV. 23, 2012

Topic: Innovation

Google's DeepMind artificial intelligence aces Atari gamin challenge

Summary: DeepMind has published a paper detailing how its AI tech not only l a host of Atari games, but went on to succeed in a number of them.

By Liam Tung | February 26, 2015 -- 10:42 GMT (02:42 PST)

Follow @LiamT 2,788 followers Get the ZDNet Announce - US newsletter n

Google's DeepMind artificial intelligence unit has shown that, given little more than play with, its algorithm can not only learn how to play computer games from scratc ace them after a few hours of practice.

ı Artificial Inte

DEEP LEARNING IMAGENET

TAG AlphaGo, Deep Learning, Artificial Intelligence

AlphaGo Beats Go Human Champ: Godfather Of Deep Learning Tells Us Do Not Be Afraid Of AI

By Aaron Mamiit, Tech Times | March 21, 10:16 AM

Like Follow Share Tweet Reddit 0 Comments SUBSCRIBE

Last week, Google's artificial intelligence program AlphaGo dominated its match with South Korean world Go champion Lee Sedol, winning with a 4-1 score.

The achievement stunned artificial intelligence experts, who previously thought that Google's computer program would need at least 10 more years before developing enough to be able to beat a human world champion.

What could be scary regarding the computer program is that Google DeepMind CEO Demis Hassabis said that AlphaGo could still improve its performance, as the match with Sedol was able to expose some of its weaknesses.

Last week, Google's artificial intelligence program AlphaGo dominated its match with South Korean world Go champion, Lee Sedol. Geoffrey Hinton, called the godfather of deep learning, explained the win's importance and why we should not fear artificial intelligence.

TECH 2/19/2015 @ 1:06PM | 5,542 views

Microsoft's Deep Learning Project Outperforms Humans In Image Recognit

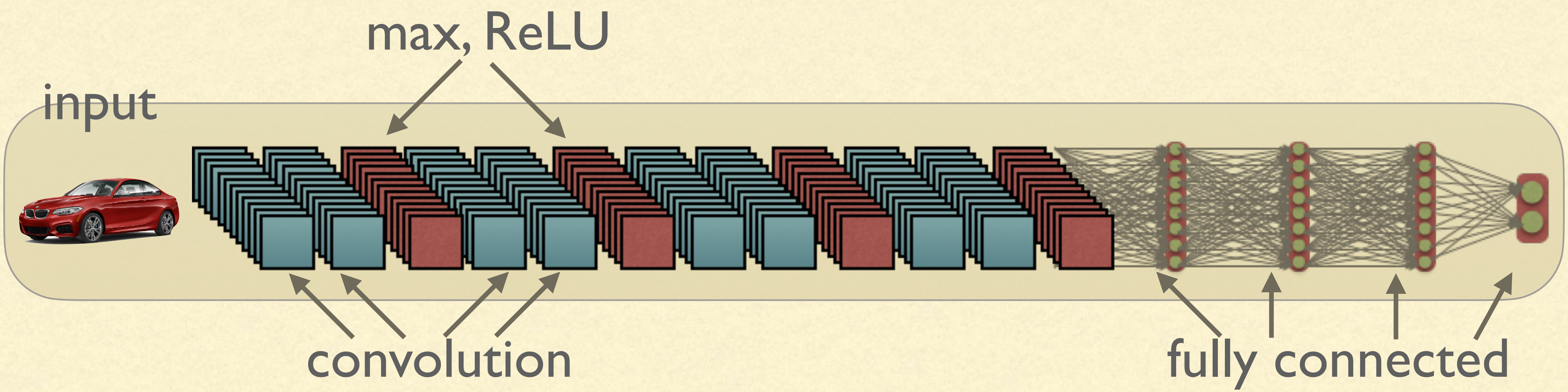
+ Comment Now + Follow Comments

Technology is blanketed in dishonesty. Computer phones are smart, software automations become intelligence, and coerced financialization becomes sharing. Because of the deceptive language surrounding these instruments it's difficult to talk about how they're used, and at what cost. Instead we're forced into false debates about sharing versus not sharing, intelligence versus inefficiency, progress versus everything.

s big a fraud as any of these endeavors, an scure discipline built around the claim that nimic human neuronal function and thus learn han humans. This week, Microsoft MSFT -1.35% ced its newest deep learning project had outperformed human jects in digital images. Researchers noted their scores shouldn' t computer image identification in general was better than g many general case instances where humans were better able

Microsoft researchers say their newest learning system beats humans — and Google

VERY DEEP NETWORKS



TELEPHONE

(input)

training

It’s a car!!
(output)

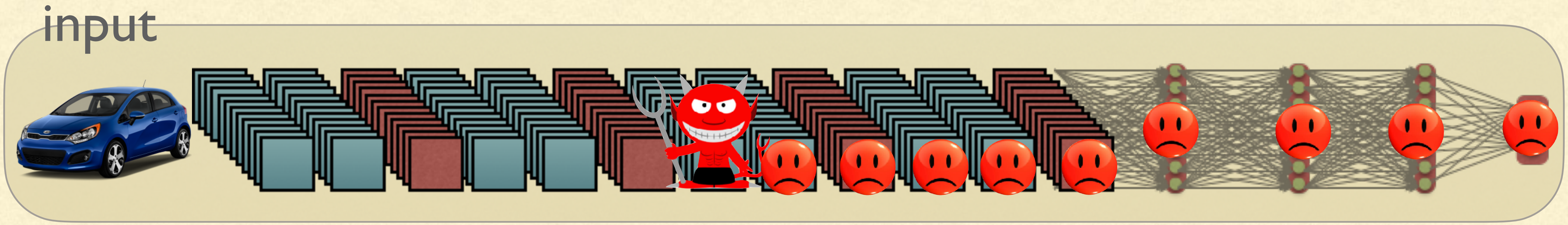
TELEPHONE

(input)

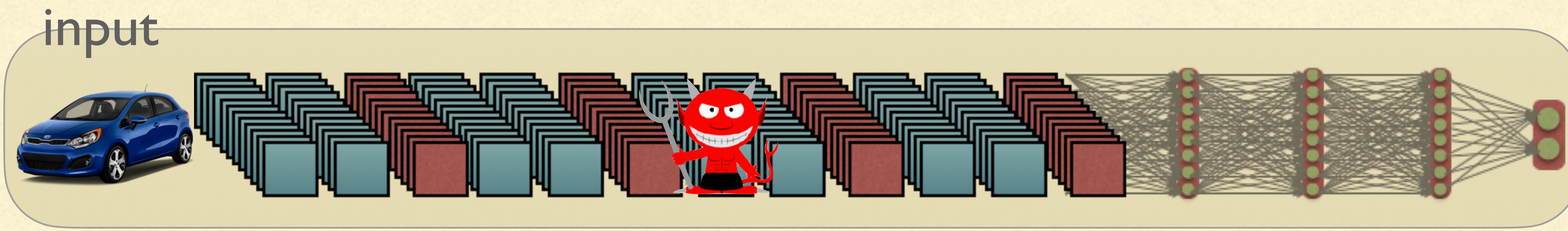
testing

It’s a bar!!
(output)

VERY DEEP NETWORKS



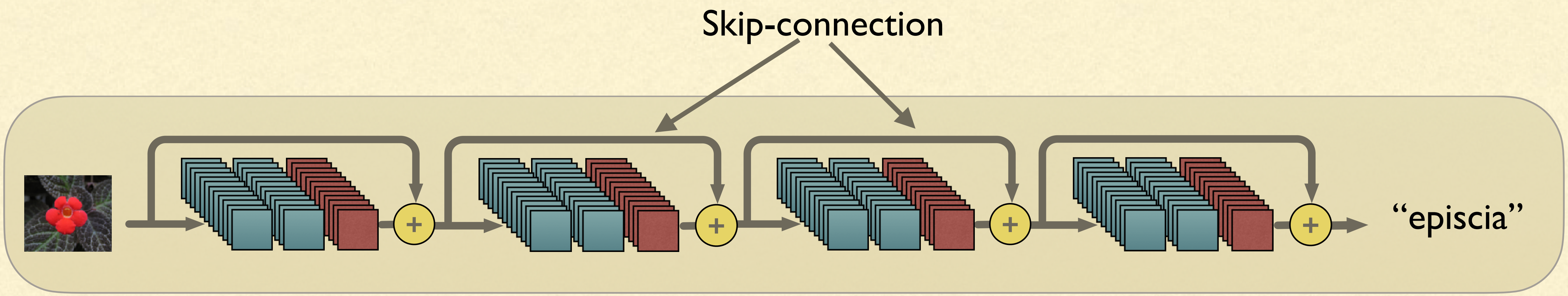
VERY DEEP NETWORKS



- More likely with more layers.
- Undetectable during training.

But you can **reduce the impact of bad layers**!

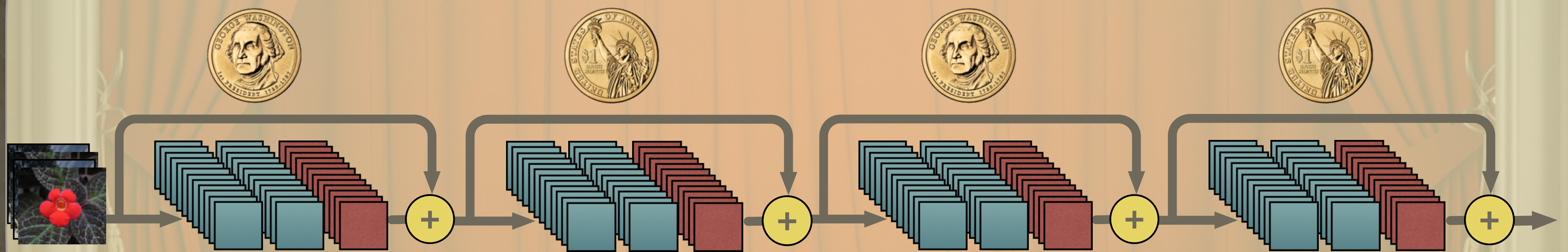
RESIDUAL NETWORKS



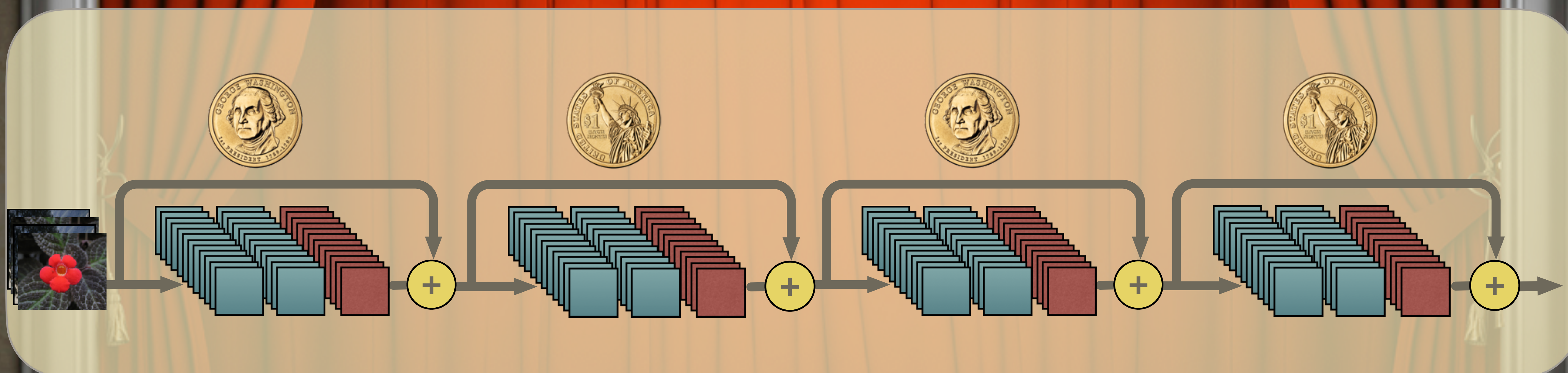
Layers still work too perfectly during training.

ResNet Architecture: [He, Zhang, Ren, Sun, CVPR'16]

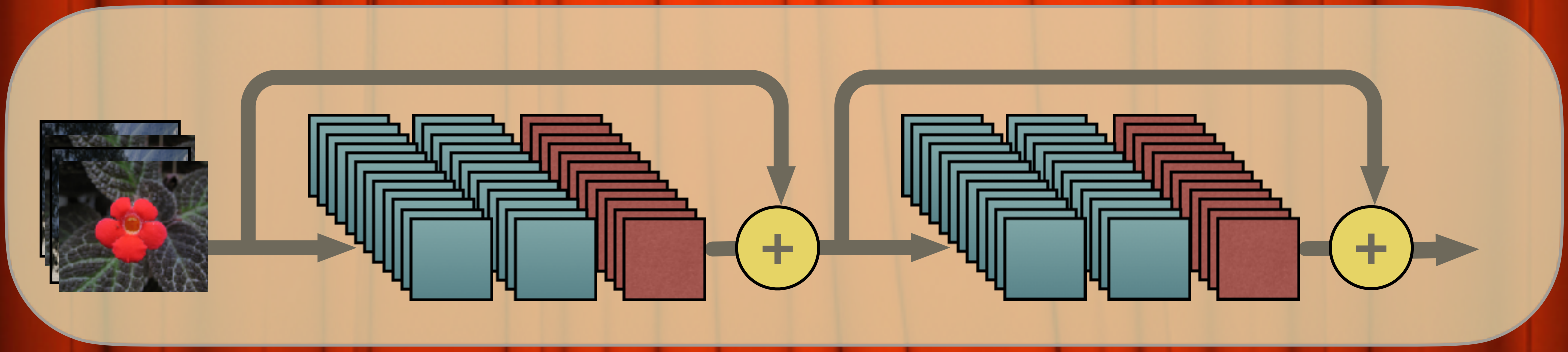
STOCHASTIC DEPTH



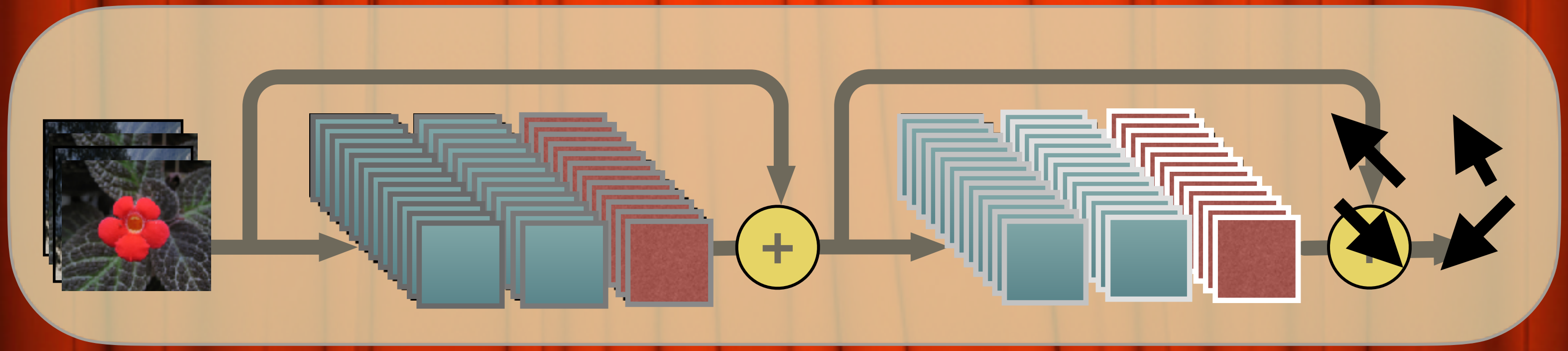
STOCHASTIC DEPTH



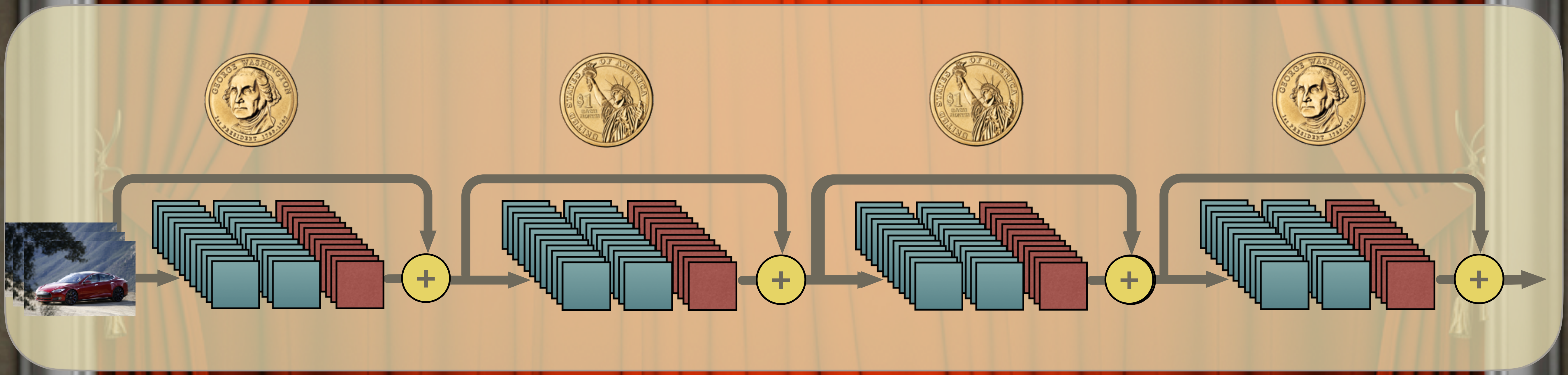
STOCHASTIC DEPTH



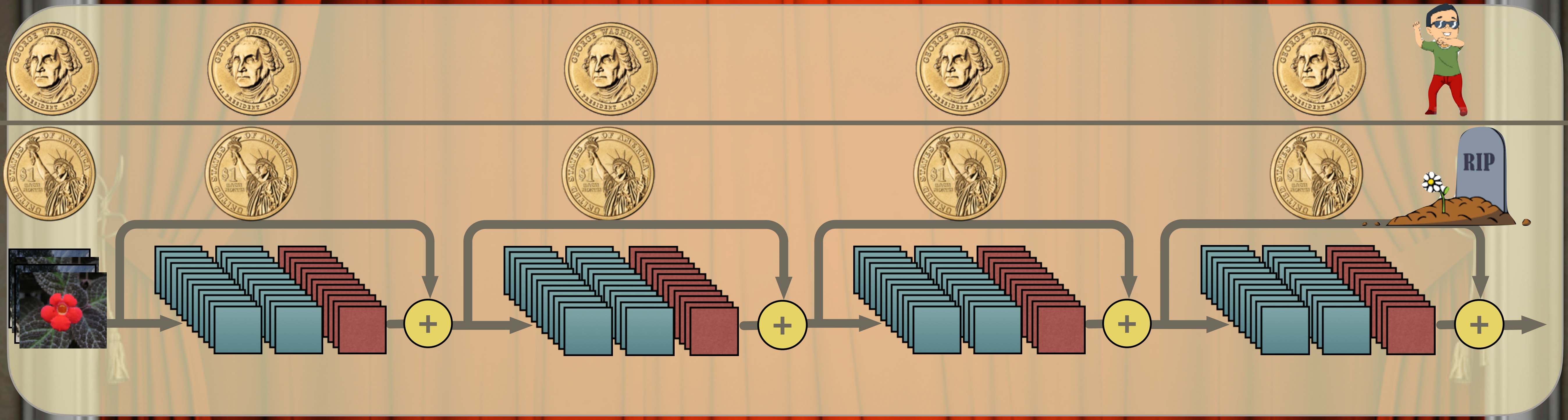
STOCHASTIC DEPTH



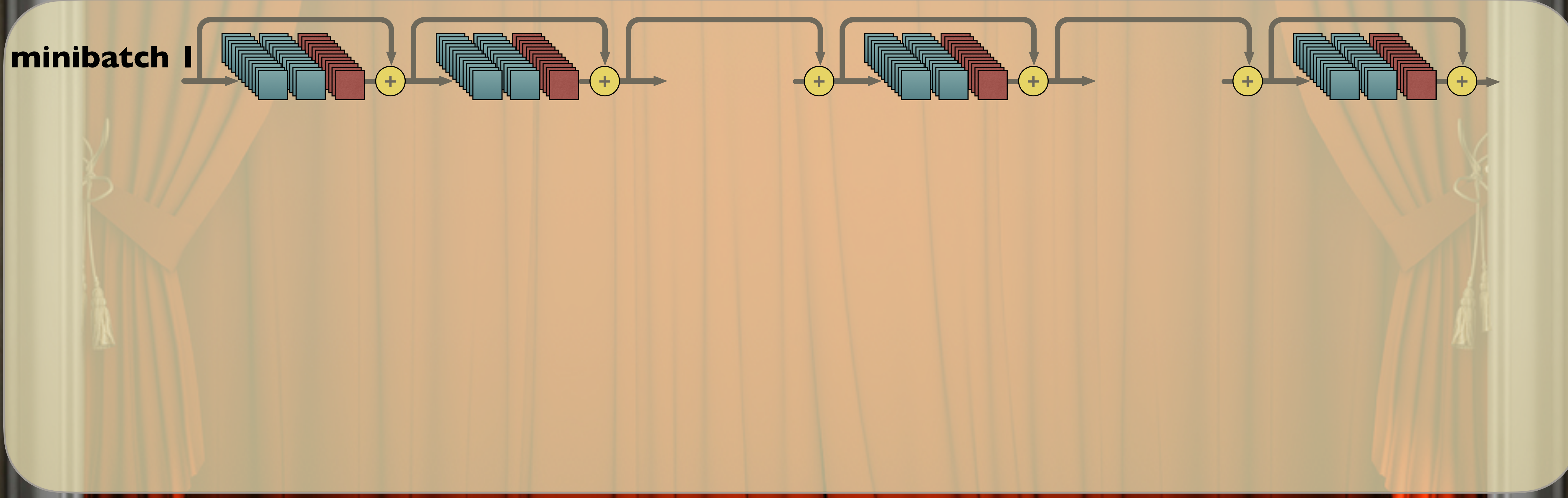
STOCHASTIC DEPTH



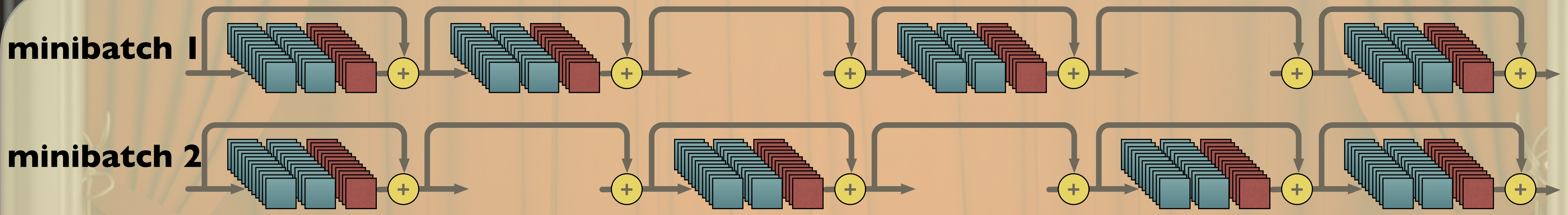
STOCHASTIC DEPTH



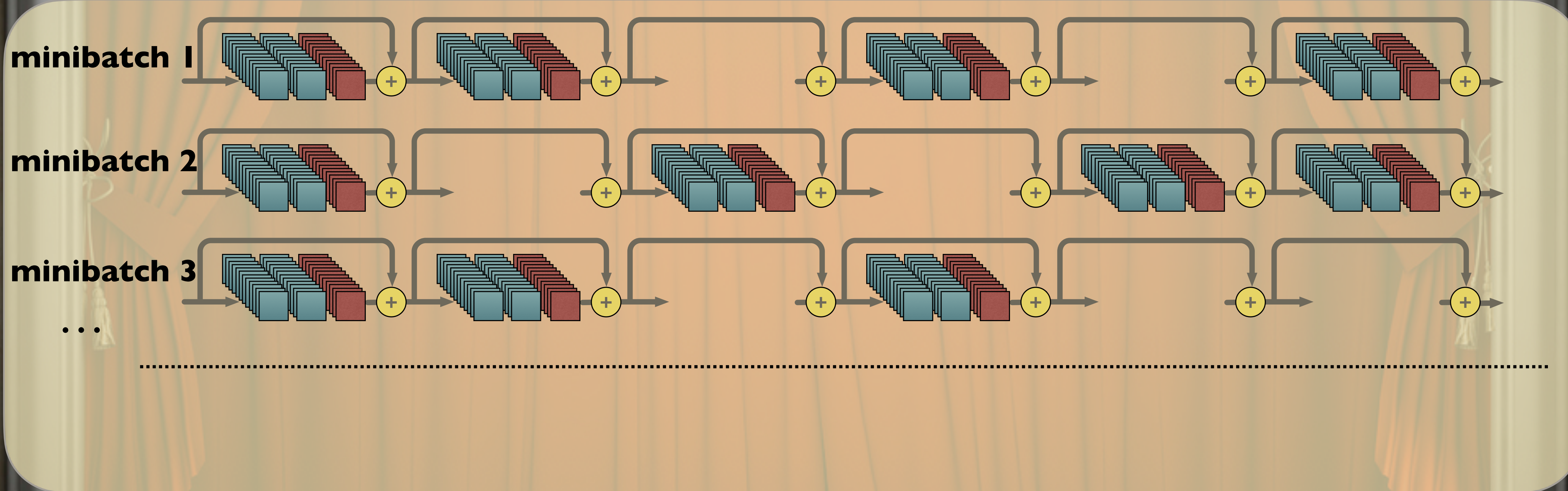
STOCHASTIC DEPTH



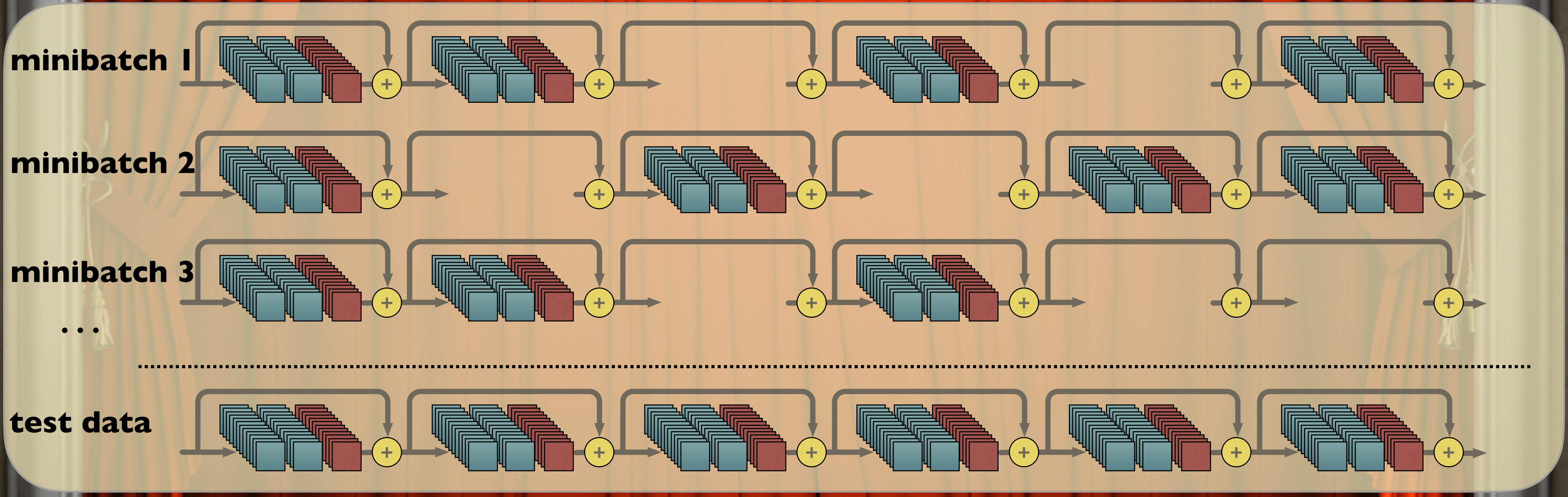
STOCHASTIC DEPTH



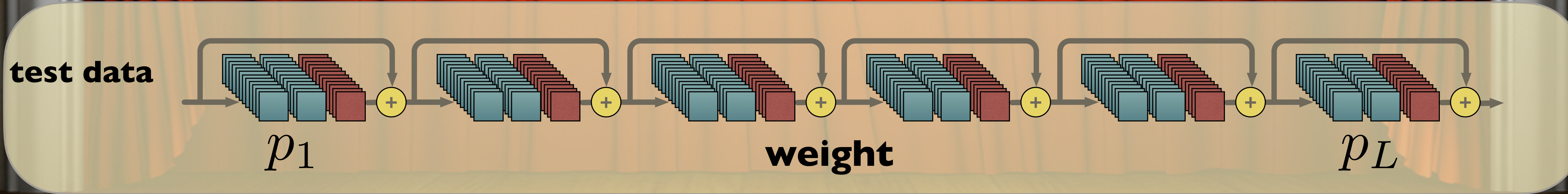
STOCHASTIC DEPTH



STOCHASTIC DEPTH

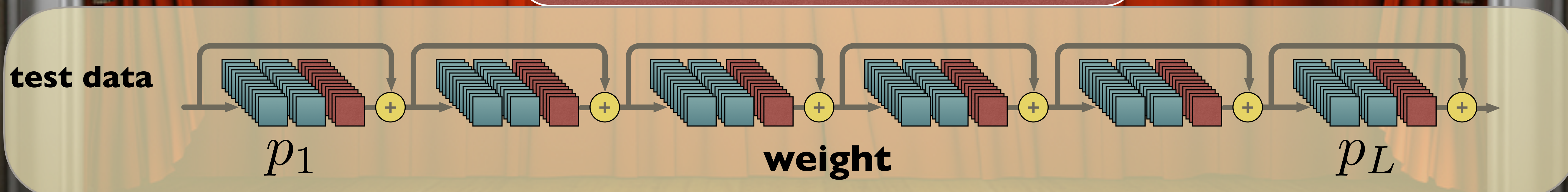


STOCHASTIC DEPTH

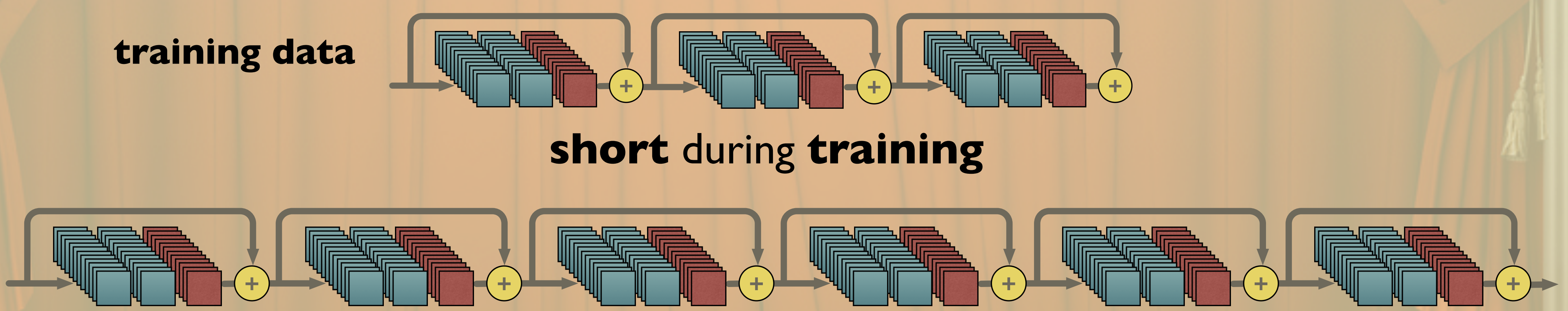


STOCHASTIC DEPTH

Implicit ensemble of 2^L models



STOCHASTIC DEPTH



training data

short during **training**

test data

deep during **testing**

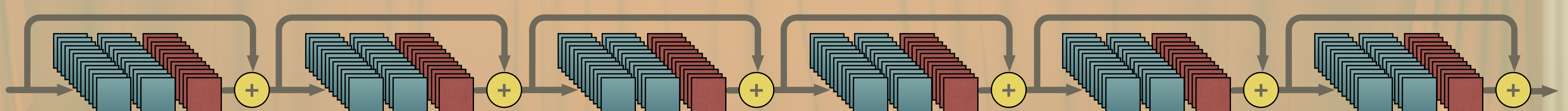
STOCHASTIC DEPTH

improved information flow

training data

short during **training**

test data



deep during **testing**

STOCHASTIC DEPTH

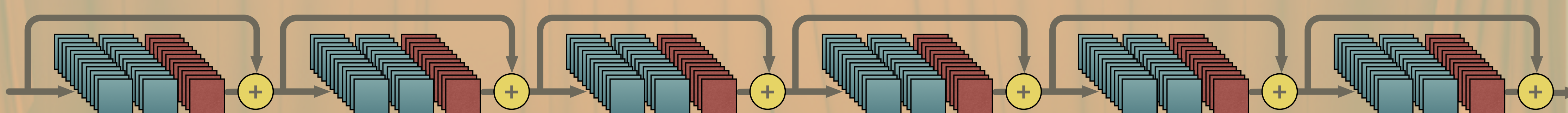
25% speedup during training

improved information flow

training data

short during **training**

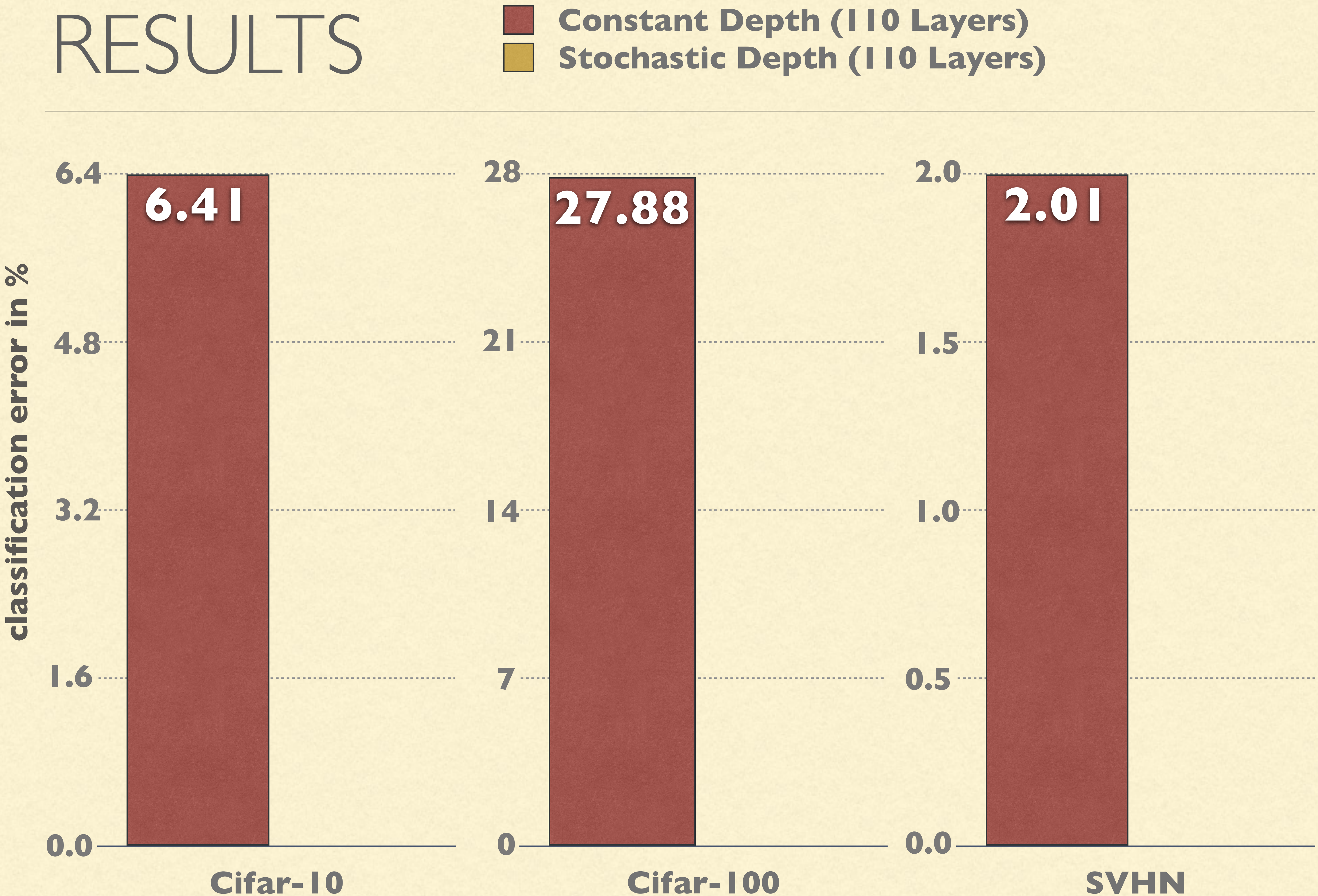
test data



deep during **testing**

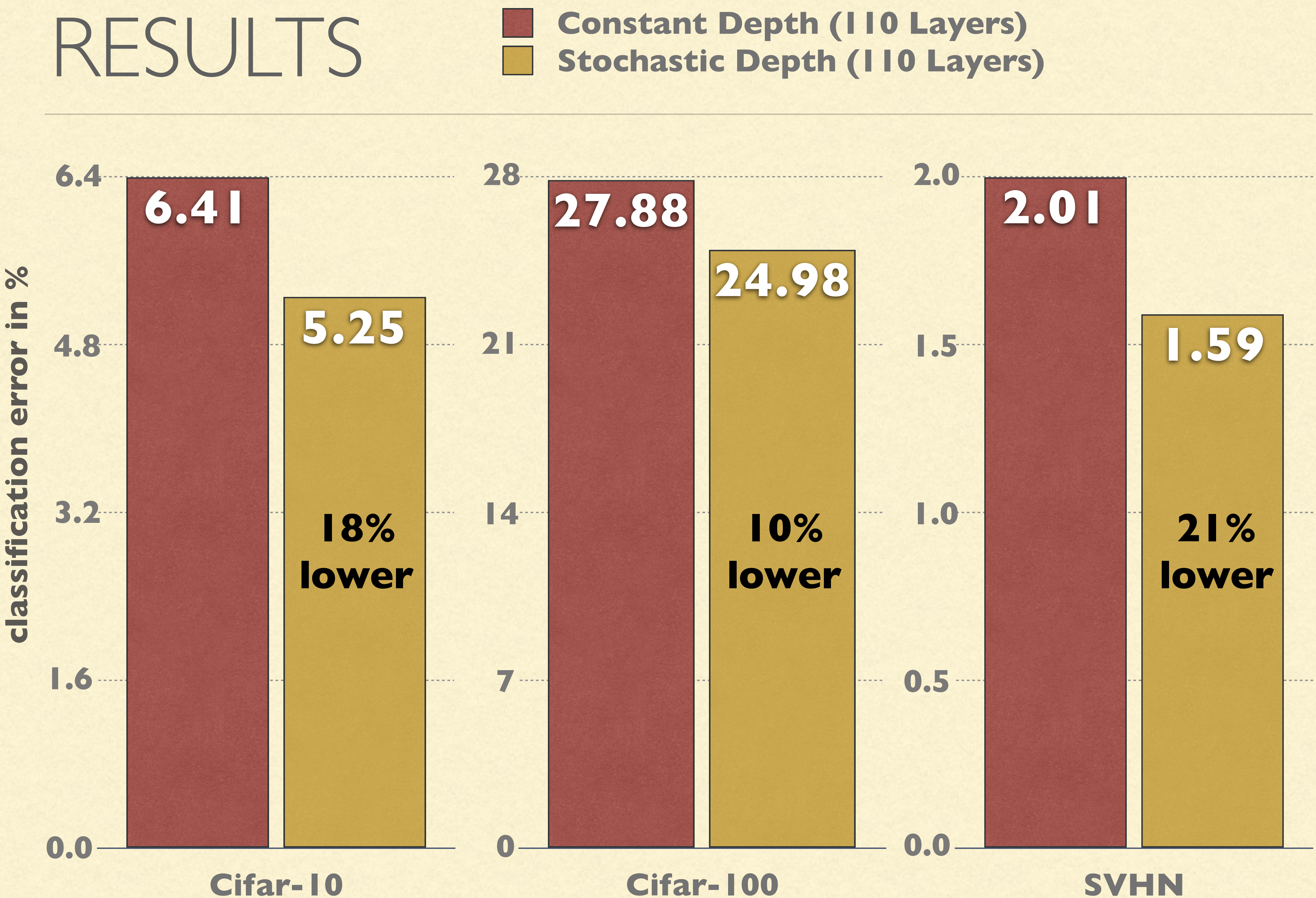
Results

RESULTS



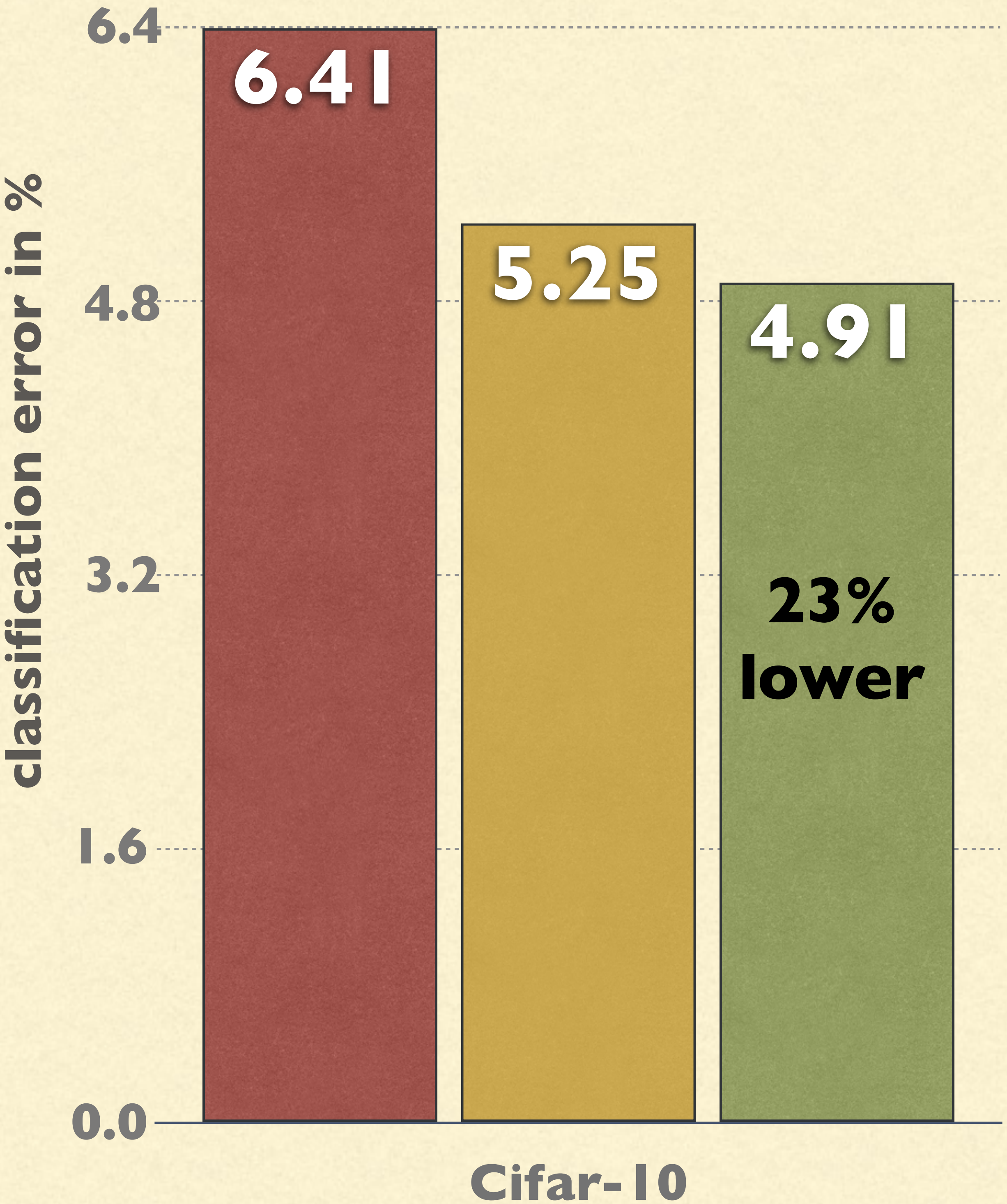
Constant Depth (110 Layers)
Stochastic Depth (110 Layers)
classification error in %
6.4
4.8
3.2
1.6
0.0
6.41
Cifar-10
28
21
14
7
0
27.88
Cifar-100
2.0
1.5
1.0
0.5
0.0
2.01
SVHN

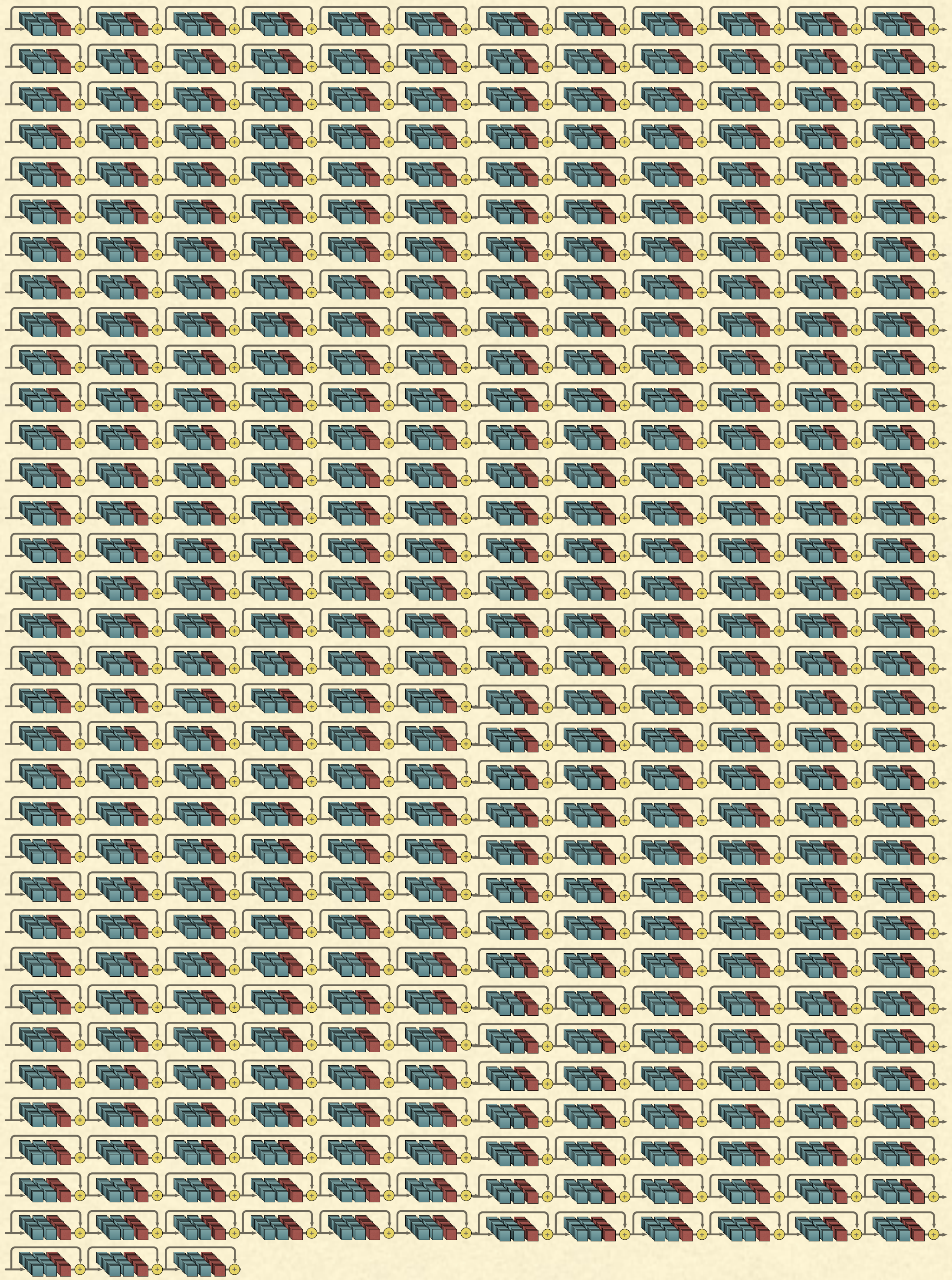
RESULTS



Constant Depth (110 Layers)
Stochastic Depth (110 Layers)
classification error in %
6.4
4.8
3.2
1.6
0.0
6.41
5.25
18%
lower
Cifar-10
28
21
14
7
0
27.88
24.98
10%
lower
Cifar-100
2.0
1.5
1.0
0.5
0.0
2.01
1.59
21%
lower
SVHN

LUDICROUS DEPTH





Swap Out [Singh et al., NIPS 2016]

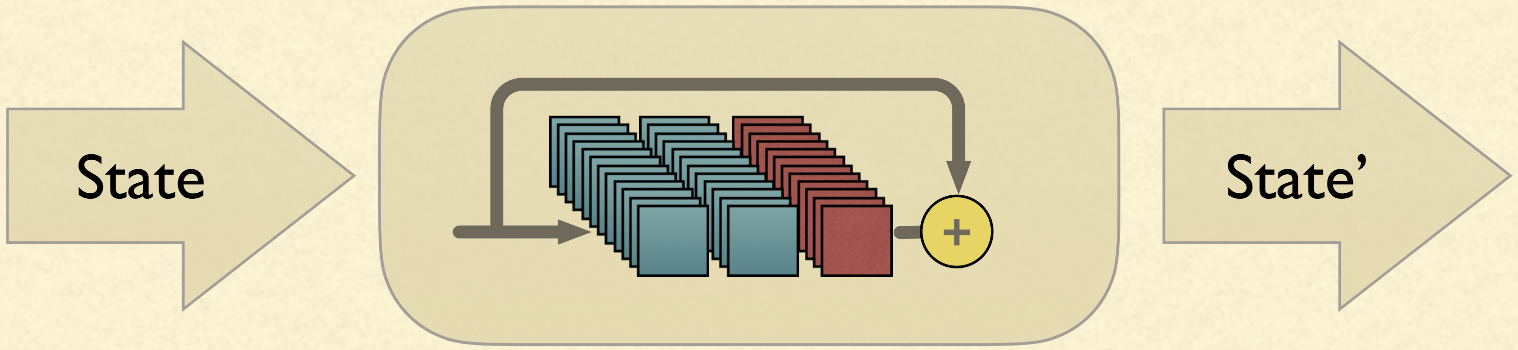
You can now

- train **faster**
- train **deeper** networks
- obtain **lower** test error.

But, can we achieve all of this without so many layers?

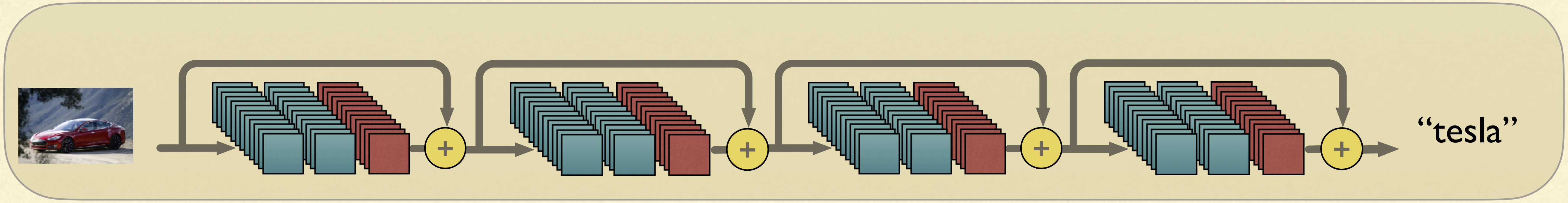
IMPOTENT LAYERS

Robustness because **no layer is too important**!



IMPOTENT LAYERS

Robustness because **no layer is too important**!

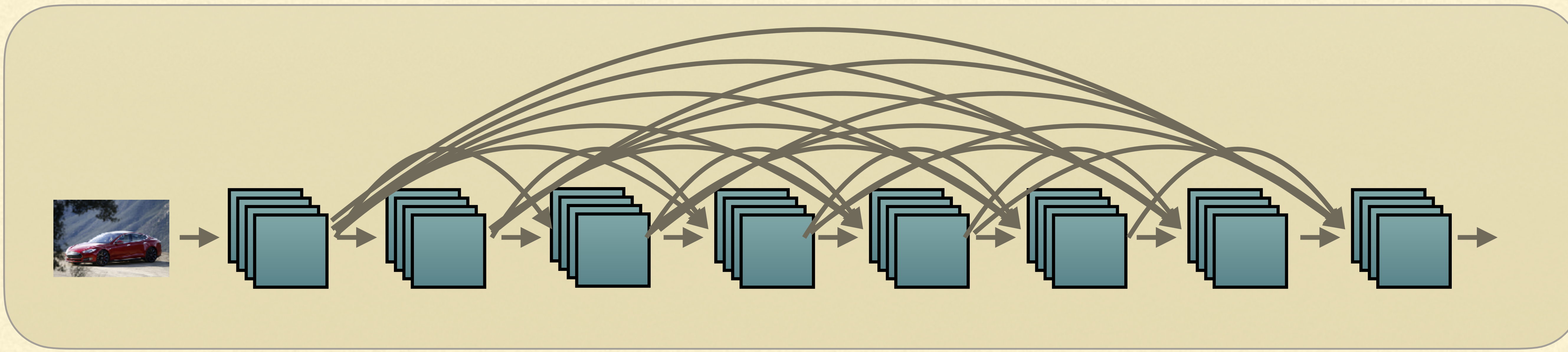


Enormous amount of **redundancy** across layers

DENSE CONNECTIVITY

[arXiv:1608.06993]

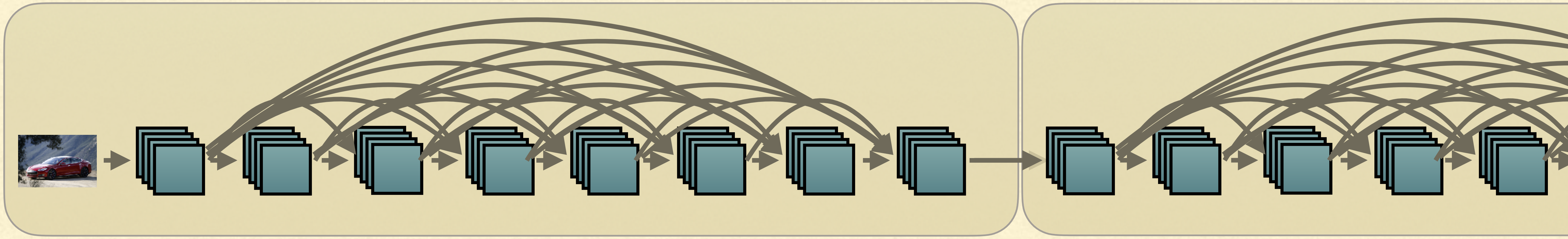
Connect every layer to every other layer of the same filter size.



DENSE CONNECTIVITY

[arXiv:1608.06993]

Connect every layer to every other layer of the same filter size.

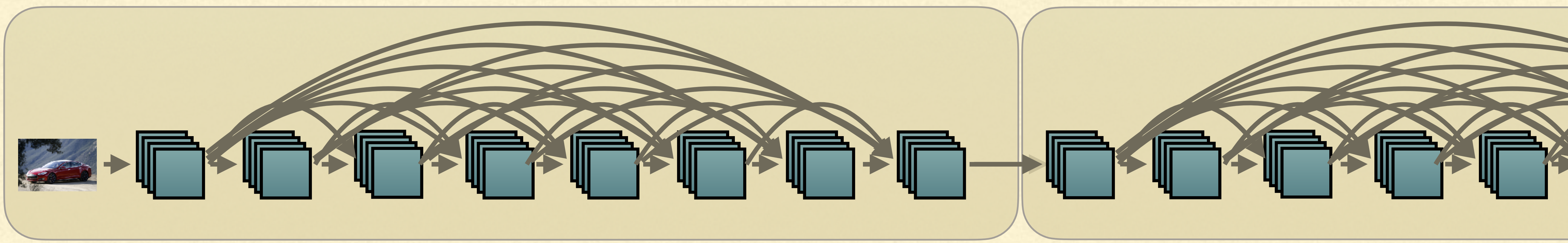


AdaNet: [arXiv:1607.01097]

DENSE CONNECTIVITY

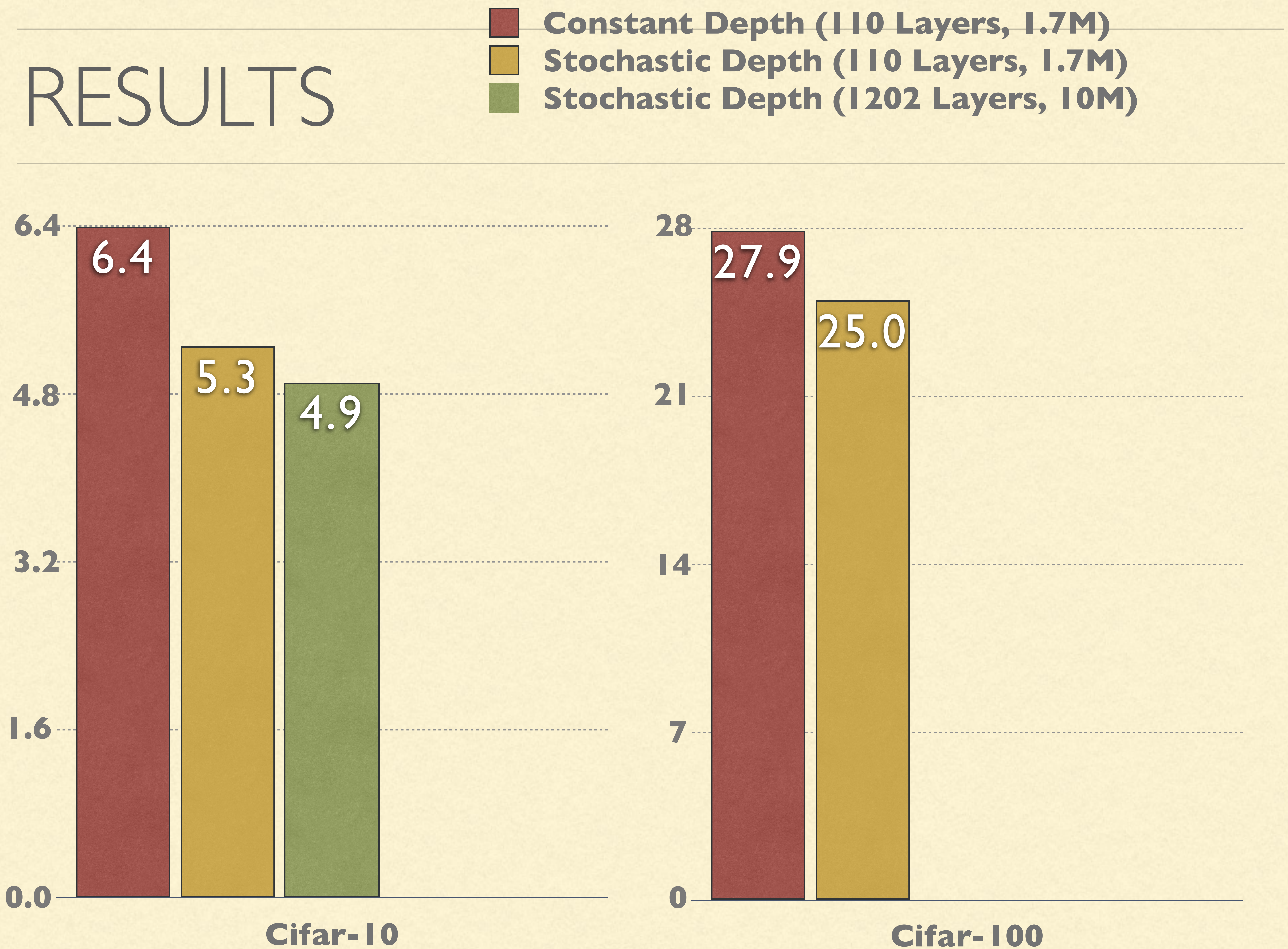
[arXiv:1608.06993]

Connect every layer to every other layer of the same filter size.



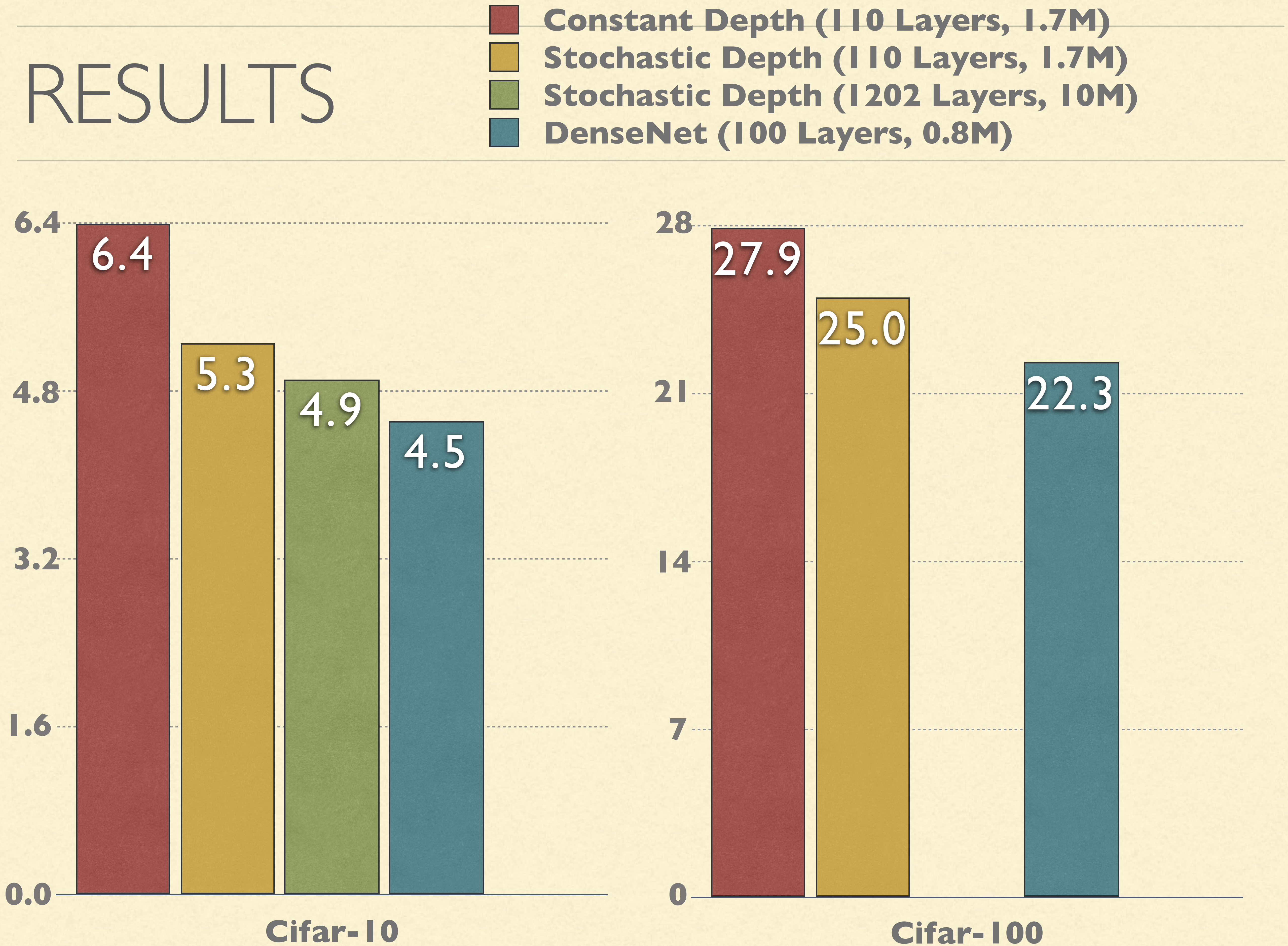
- Also removes reliance on any single layer
- Also solve information flow through direct connections
- **Do not introduce redundancy**

RESULTS



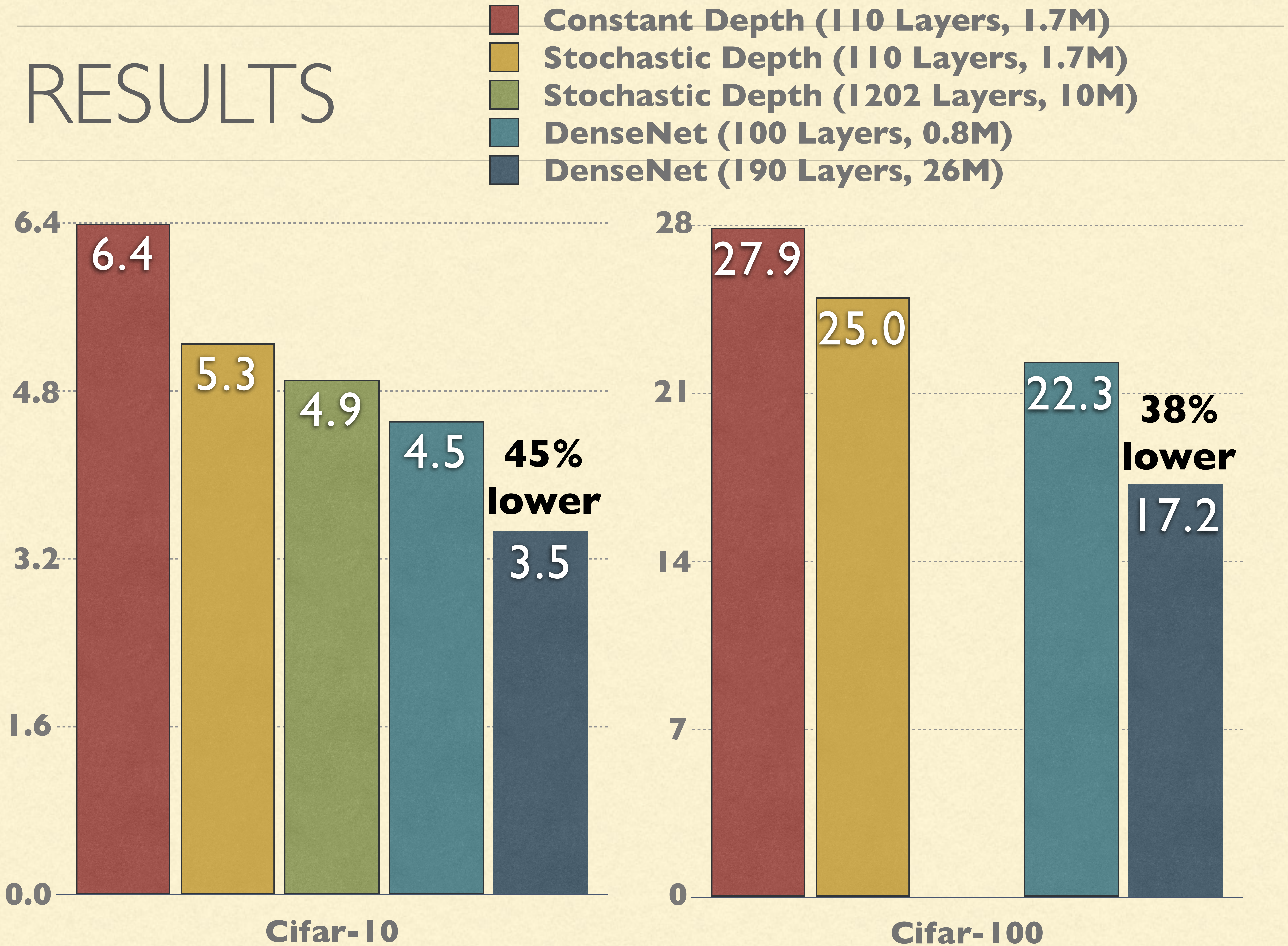
Constant Depth (110 Layers, 1.7M)
Stochastic Depth (110 Layers, 1.7M)
Stochastic Depth (1202 Layers, 10M)
6.4
4.8
3.2
1.6
0.0
6.4
5.3
4.9
Cifar-10
28
21
14
7
0
27.9
25.0
Cifar-100

RESULTS

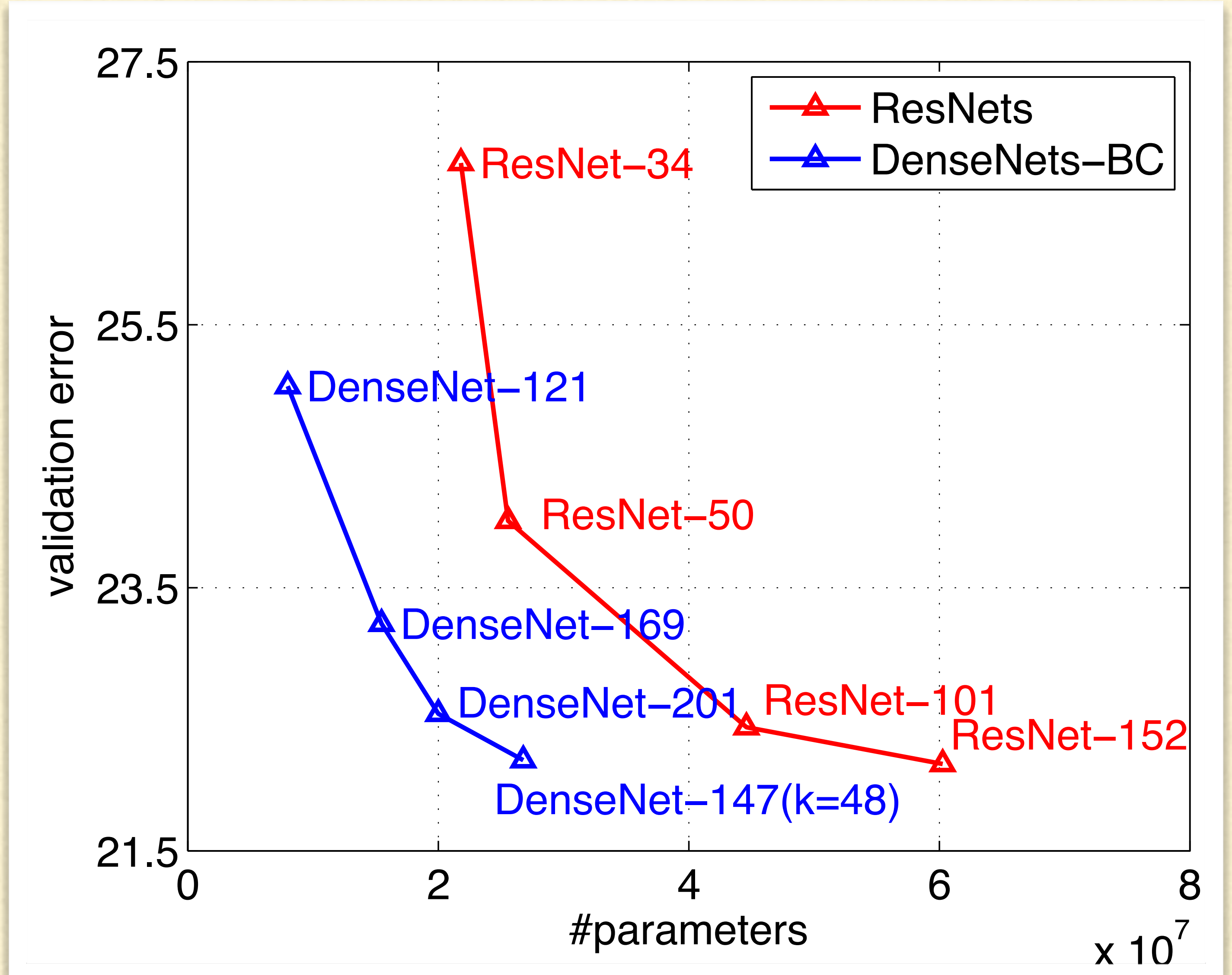


Constant Depth (110 Layers, 1.7M)
Stochastic Depth (110 Layers, 1.7M)
Stochastic Depth (1202 Layers, 10M)
DenseNet (100 Layers, 0.8M)
6.4
4.8
3.2
1.6
0.0
6.4
5.3
4.9
4.5
Cifar-10
28
21
14
7
0
27.9
25.0
22.3
Cifar-100

RESULTS



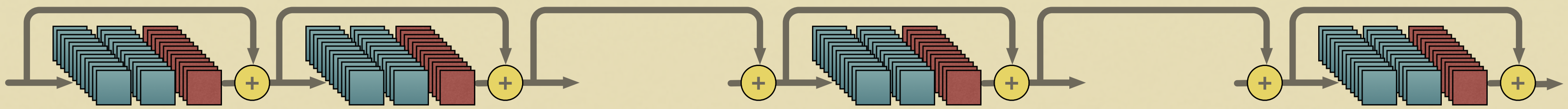
Constant Depth (110 Layers, 1.7M)
Stochastic Depth (110 Layers, 1.7M)
Stochastic Depth (1202 Layers, 10M)
DenseNet (100 Layers, 0.8M)
DenseNet (190 Layers, 26M)
6.4
4.8
3.2
1.6
0.0
6.4
5.3
4.9
4.5
45%
lower
3.5
Cifar-10
28
21
14
7
0
27.9
25.0
22.3
38%
lower
17.2
Cifar-100



DENSENET VS. RESNET (IMAGENET)

CONCLUSION

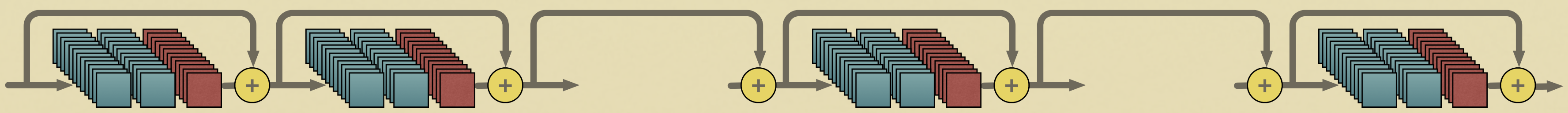
Use ResNets with Stochastic Depth!



- **lower training time**
- **lower testing error**

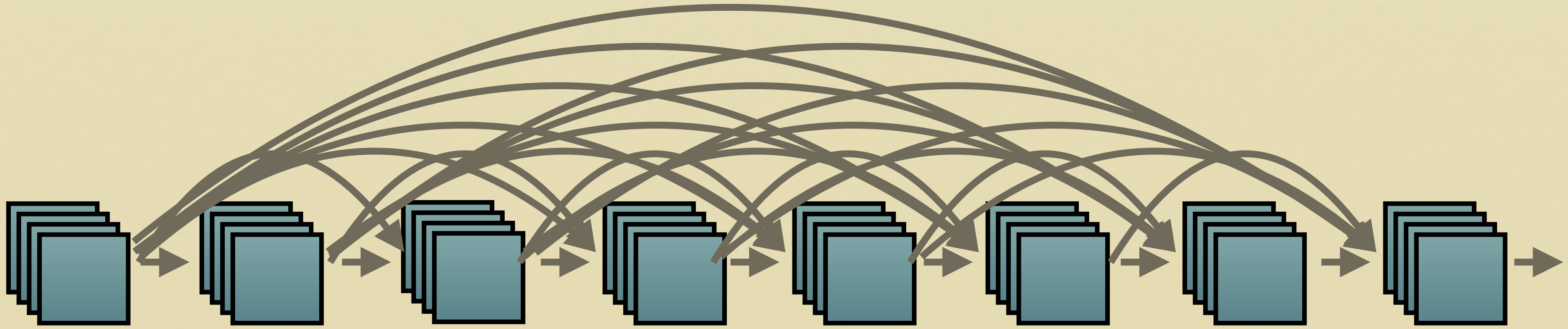
CONCLUSION

Use ResNets with Stochastic Depth!



- **lower training time**
- **lower testing error**

Improved via Dense Connectivity!



- **Explicit long term connections**
- **Best generalization performance**