Limitations of Deep Learning

Oliver Richter

Artificial intelligence / Machine learning

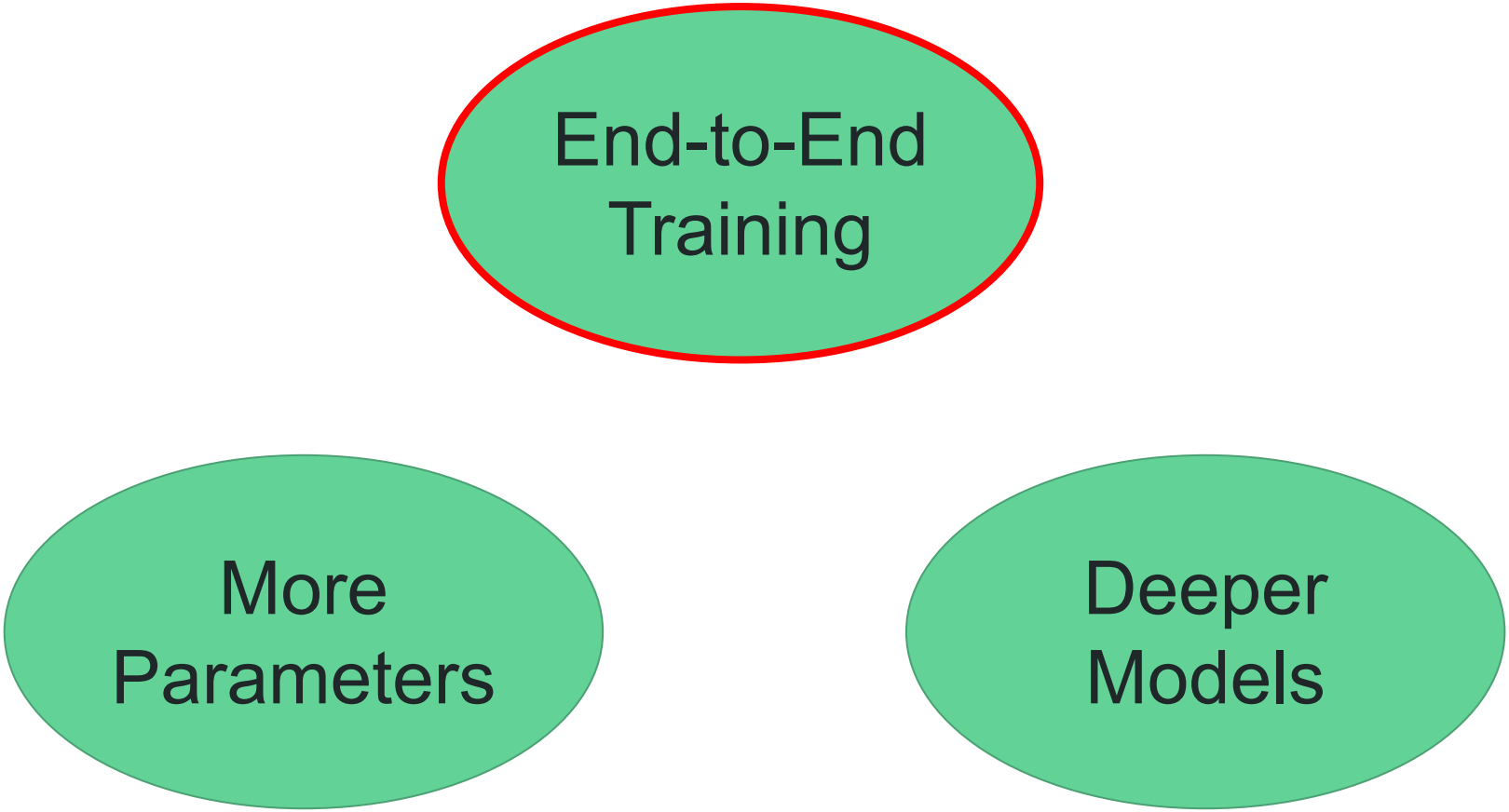
AI pioneer Geoff Hinton: "Deep learning is going to be able to do everything"

Thirty years ago, Hinton's belief in neural networks was contrarian. Now it's hard to find anyone who disagrees, he says.

by **Karen Hao**

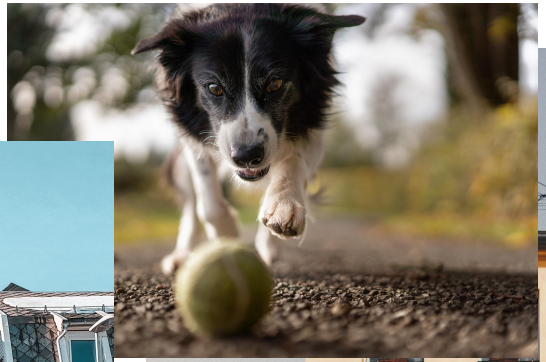
November 3, 2020

Source: https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything/


End-to-End Training
More Parameters
Deeper Models

Why should you do it?

End-to-End Training



Why should you do it?

End-to-End Training

“Cat”

“Dog”

“House”

“Animal”

“No animal”

What happened?

End-to-End Training

“Cat”

“Dog”

“House”

“Animal”

“No animal”

What happened?

End-to-End Training

Impossible?

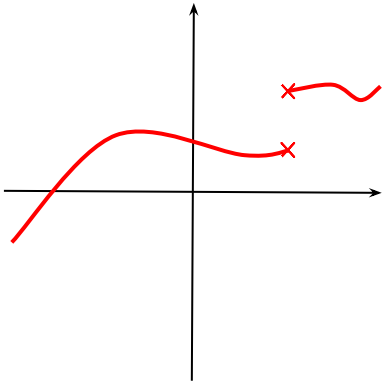
“Animal”

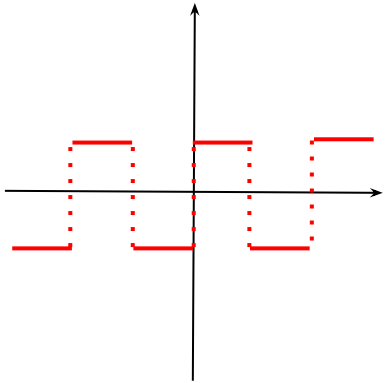
“No animal”

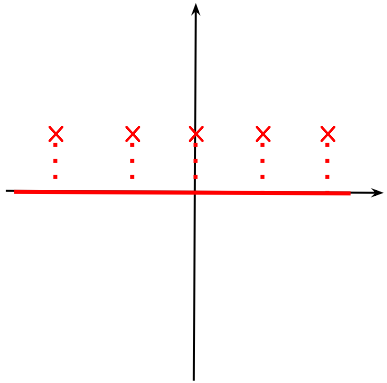
What's the catch?

End-to-End Training

Neural networks cannot learn to represent discontinuous functions.

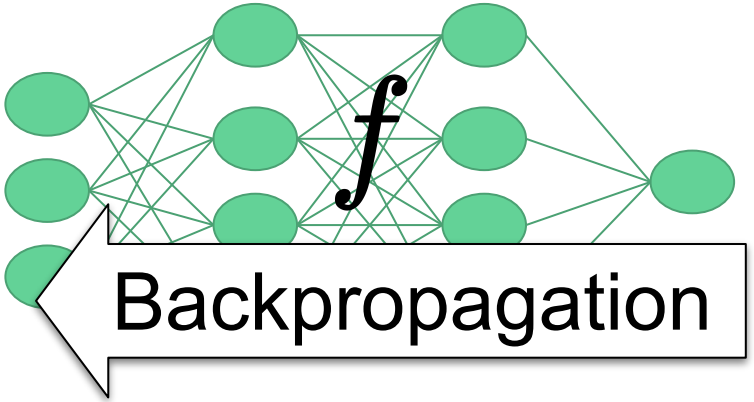






Why?

End-to-End Training

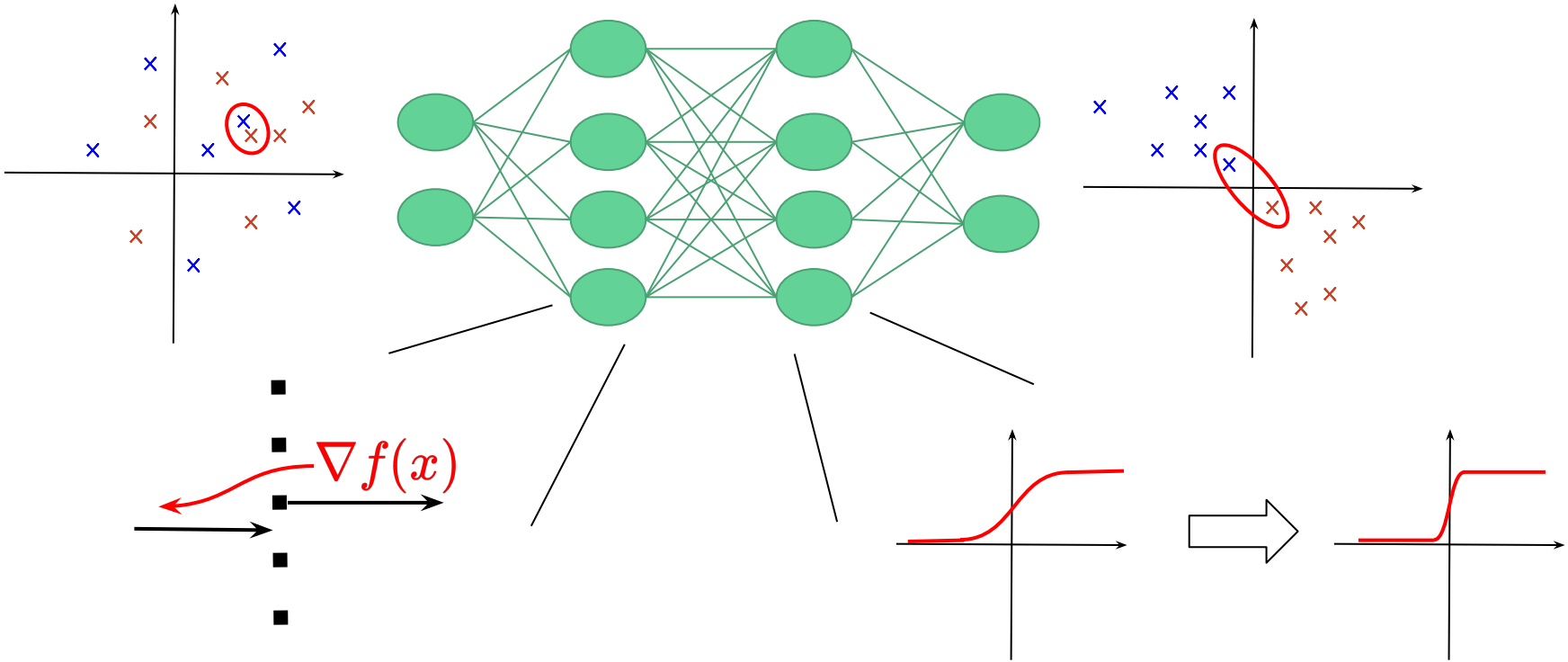


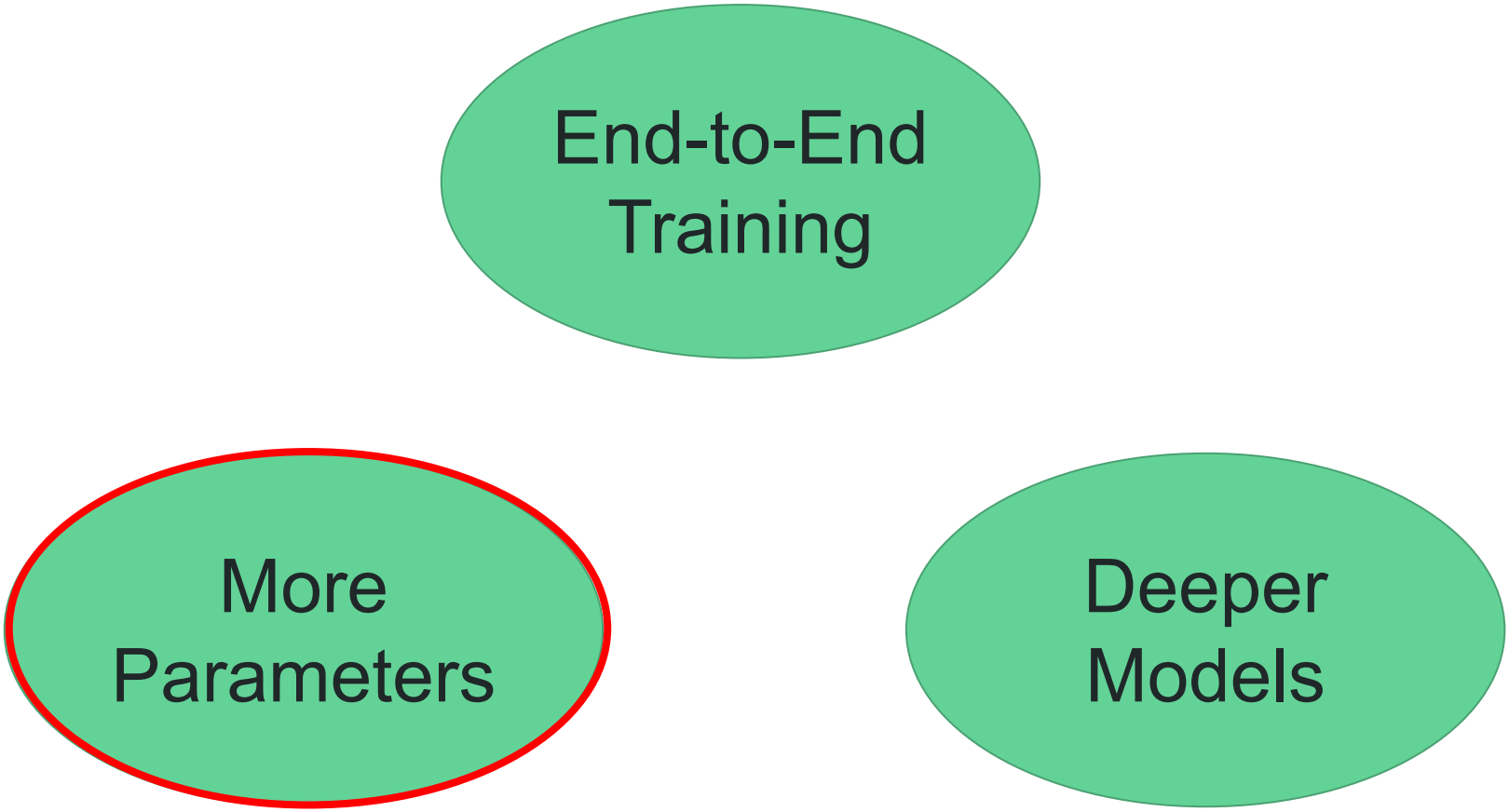
f continuous

$$\exists \nabla_x f(x) \ \forall x$$

Is it that bad?

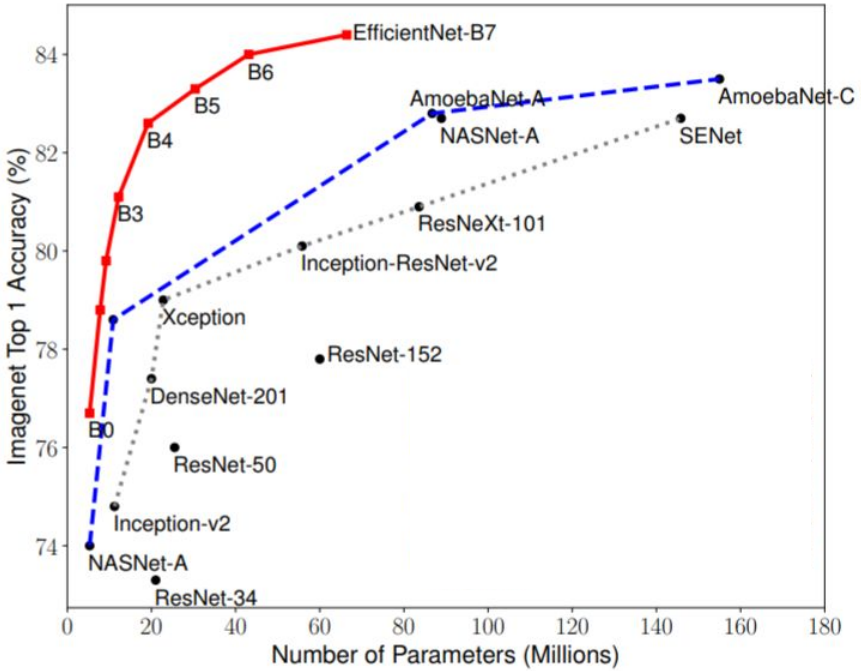
End-to-End Training

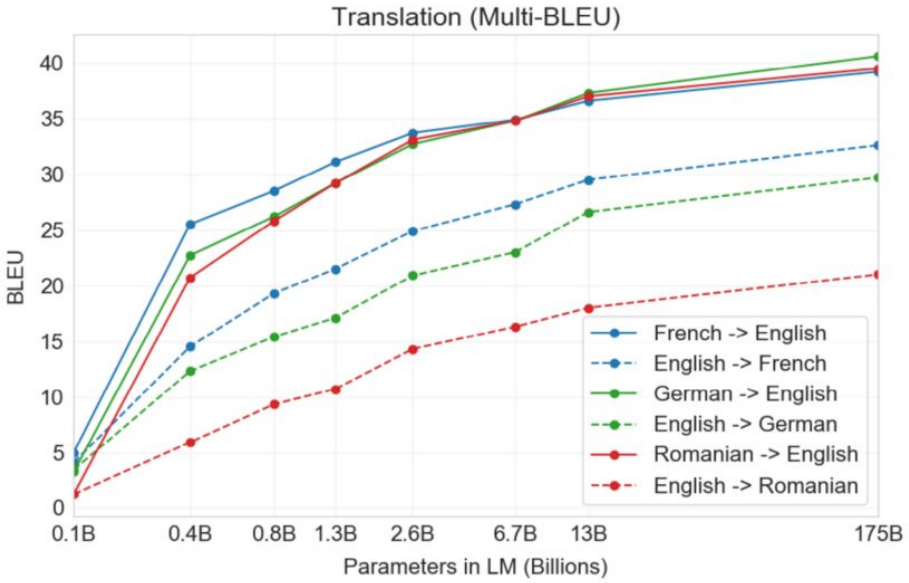



End-to-End Training
More Parameters
Deeper Models

Why should you do it?

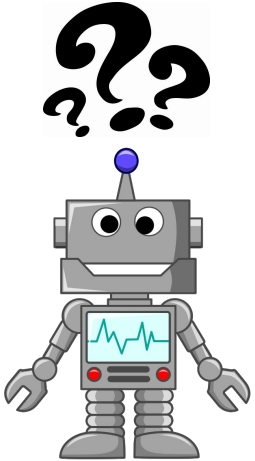
More Parameters

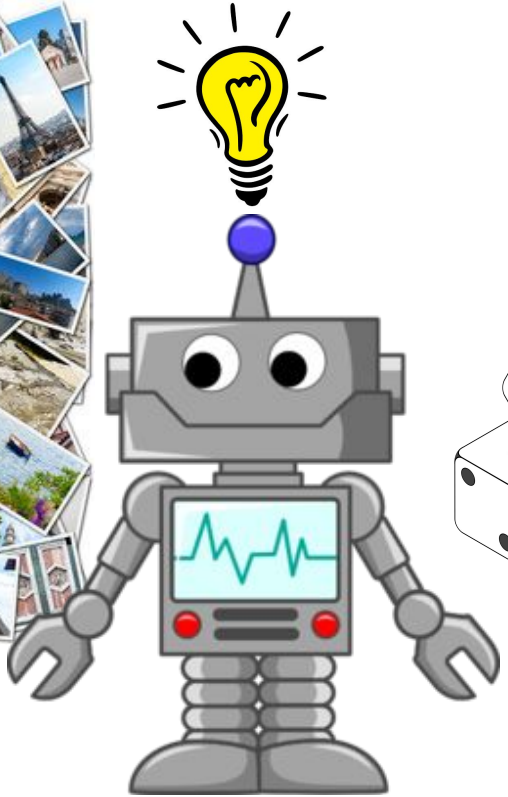


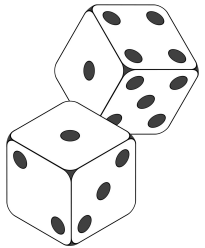


Why should you do it?

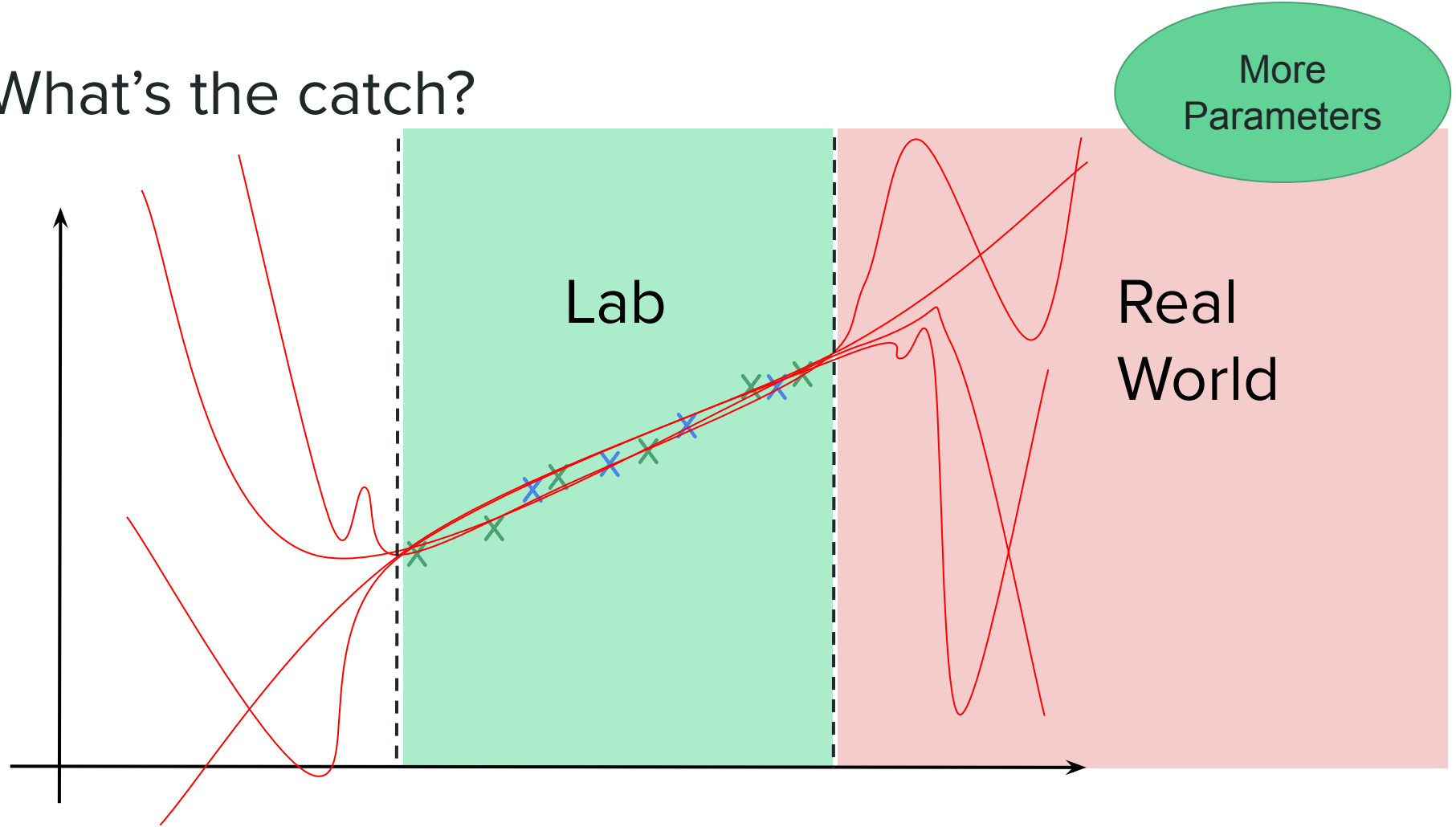
More Parameters

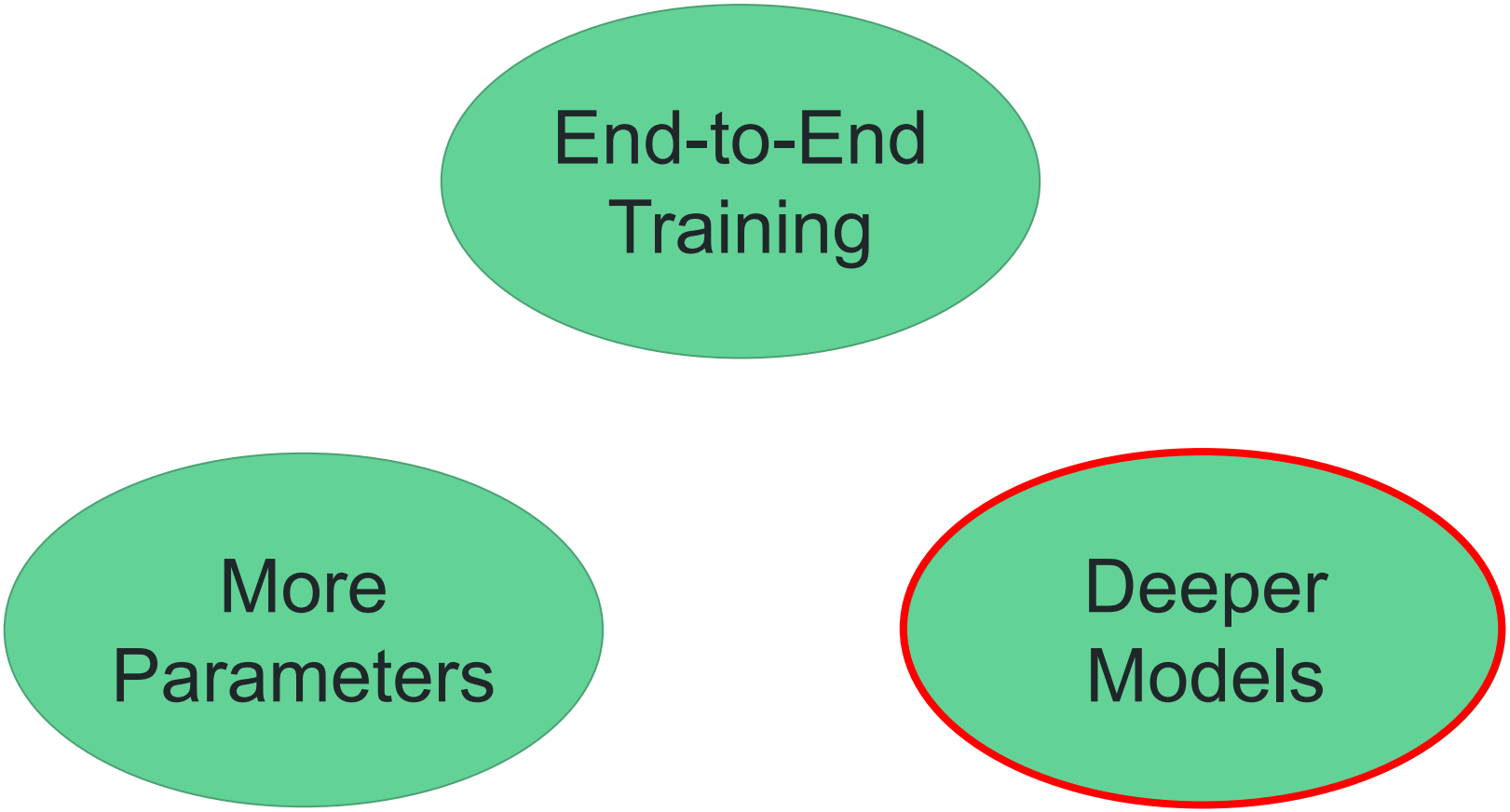






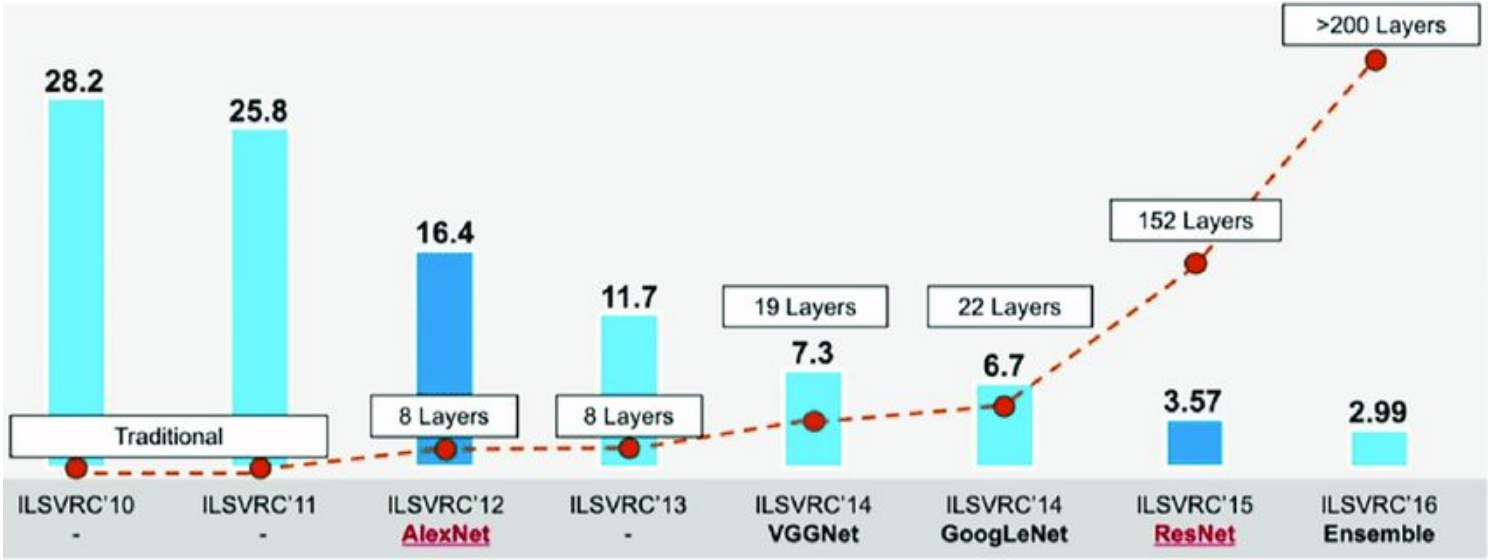
What's the catch?




End-to-End Training
More Parameters
Deeper Models

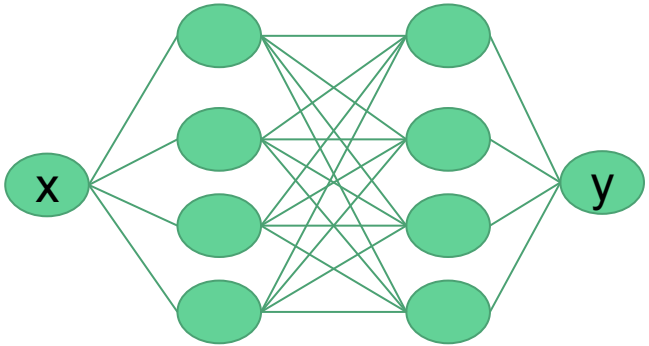
Why should you do it?

Deeper Models

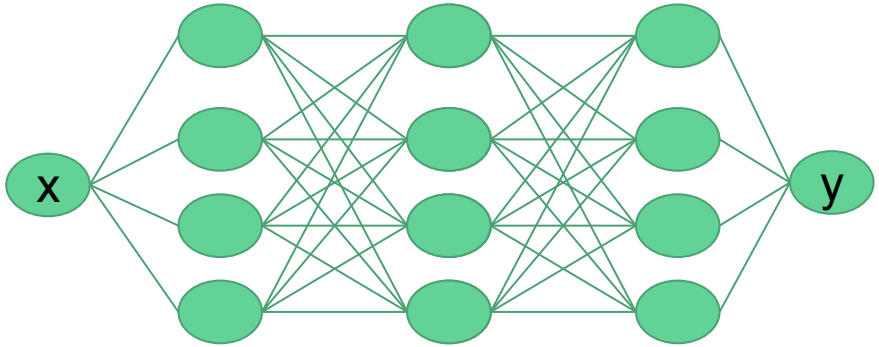


Why should you do it?

Deeper Models



n^2 paths



n^3 paths

$$\#params \in \mathcal{O}(L \cdot n^2)$$

$$\#paths \in \mathcal{O}(n^L)$$

Deeper Models

What's the catch?

$$||A||_{op} = c \rightarrow ||Av|| \leq c||v||$$

$$c_l = ||W_l^T||_{op} \qquad c_\sigma \leq 1$$

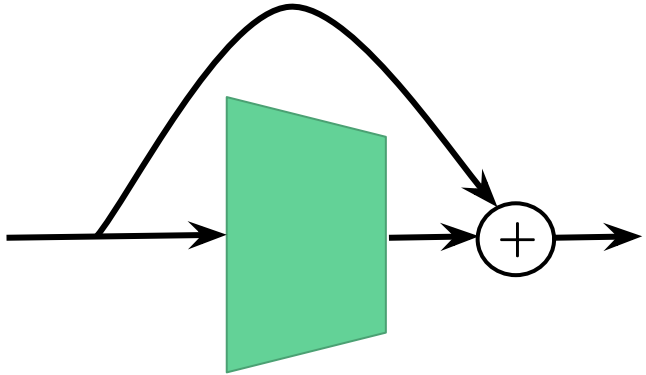
$$||\nabla_x f(x)|| \leq ||\nabla_y f(x)|| \cdot \prod_{l=1}^{L} c_l \cdot c_\sigma$$

$$c_l < 1 \rightarrow \text{vanishing gradients}$$

Solutions?

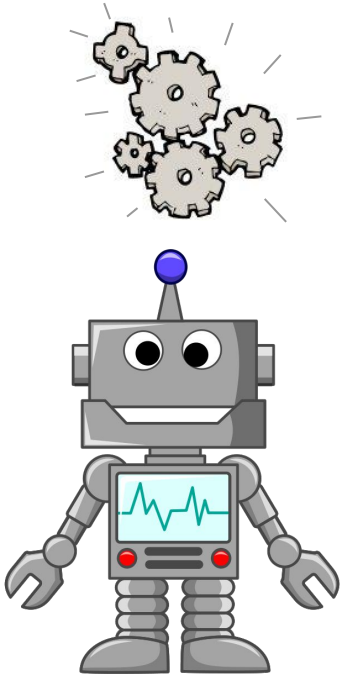
Deeper Models

- Initialization
- Constraint Optimization
- Residual Connections



Other Problems with Depth?

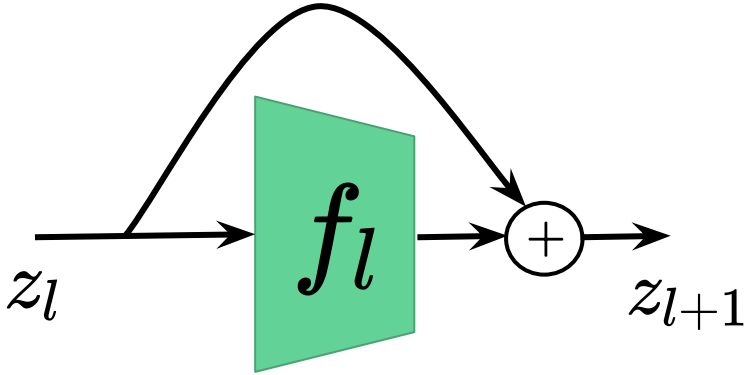
Deeper Models



$$\mathcal{O}(L)$$

Implicit Layers

Deeper Models



$$z_{l+1} = z_l + f_l(z_l)$$

$$z_L = z_0 + \sum_{l=0}^{L} f_l(z_l)$$

Implicit Layers

Deeper Models

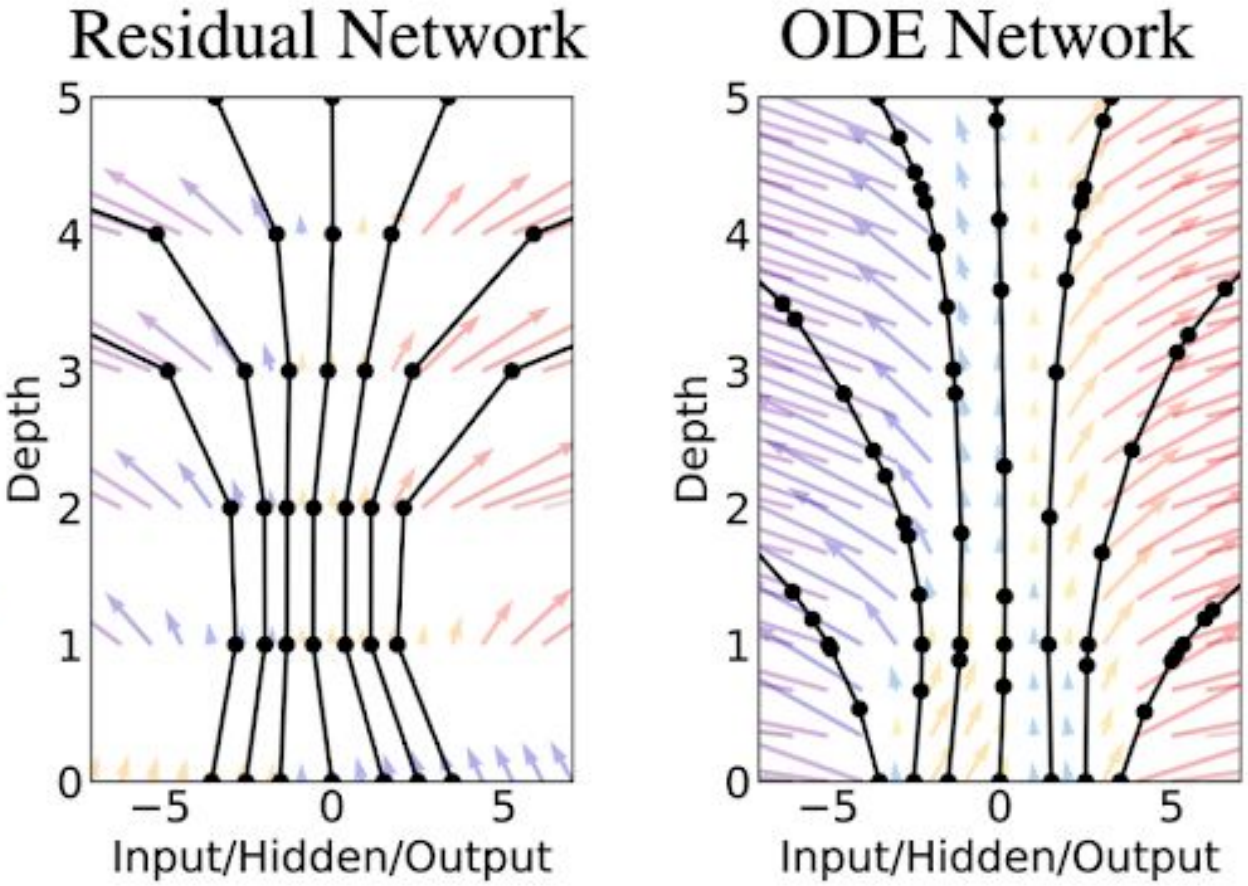
$$z(L) = z(0) + \sum_{l=0}^{L} f(z(l), l)$$

$$z(L) = z(0) + \int_{0}^{L} f(z(l), l) dl$$

$$\frac{dz}{dl} = f(z(l), l)$$

Given f, z_0 use ODE solver to find z_L

Deeper Models



How to train?

Deeper Models

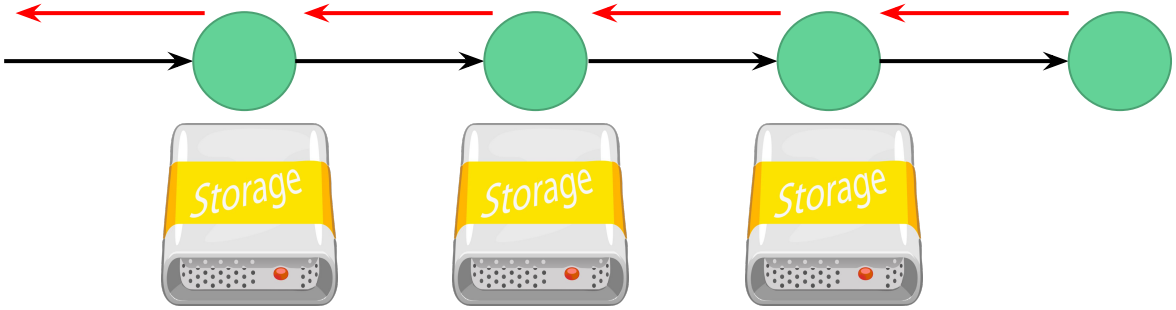
$$\frac{\delta \mathcal{L}}{\delta \theta} = \sum_l \frac{\delta \mathcal{L}}{\delta z_{l+1}} \frac{\delta f(z_l, \theta)}{\delta \theta}$$

$$\frac{\delta \mathcal{L}}{\delta \theta} = \int_L^0 \frac{\delta \mathcal{L}}{\delta z(l)} \frac{\delta f(z(l), \theta)}{\delta \theta} dl$$

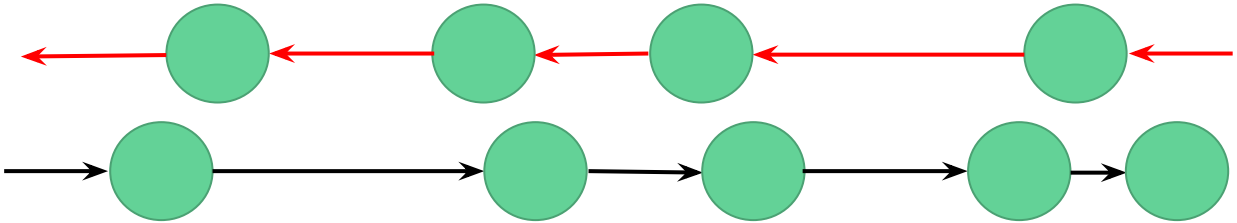
➔ Use ODE solver

$\mathcal{O}(1)$

Deeper Models



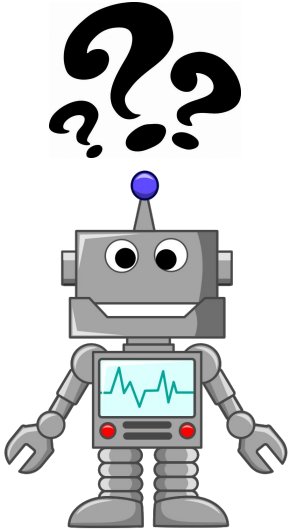
$\mathcal{O}(L)$



$\mathcal{O}(1)$

End-to-End Training

More Parameters



Deeper Models