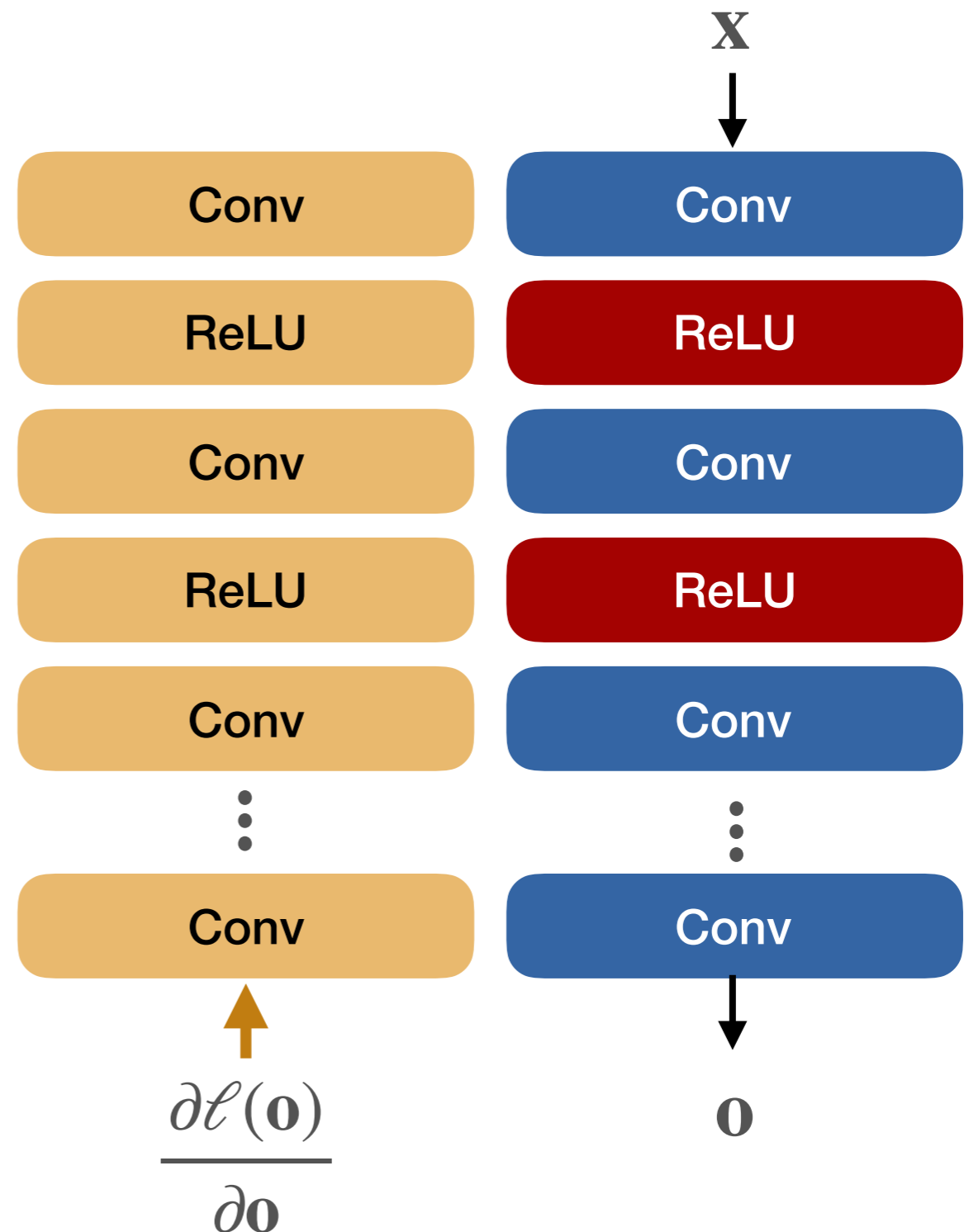


Vanishing and exploding gradients

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

Exploding gradients

- Weight of one layer grows
- Gradient of all other layers grow
- Weights of other layers grow
- Bad feedback loop



Exploding gradients

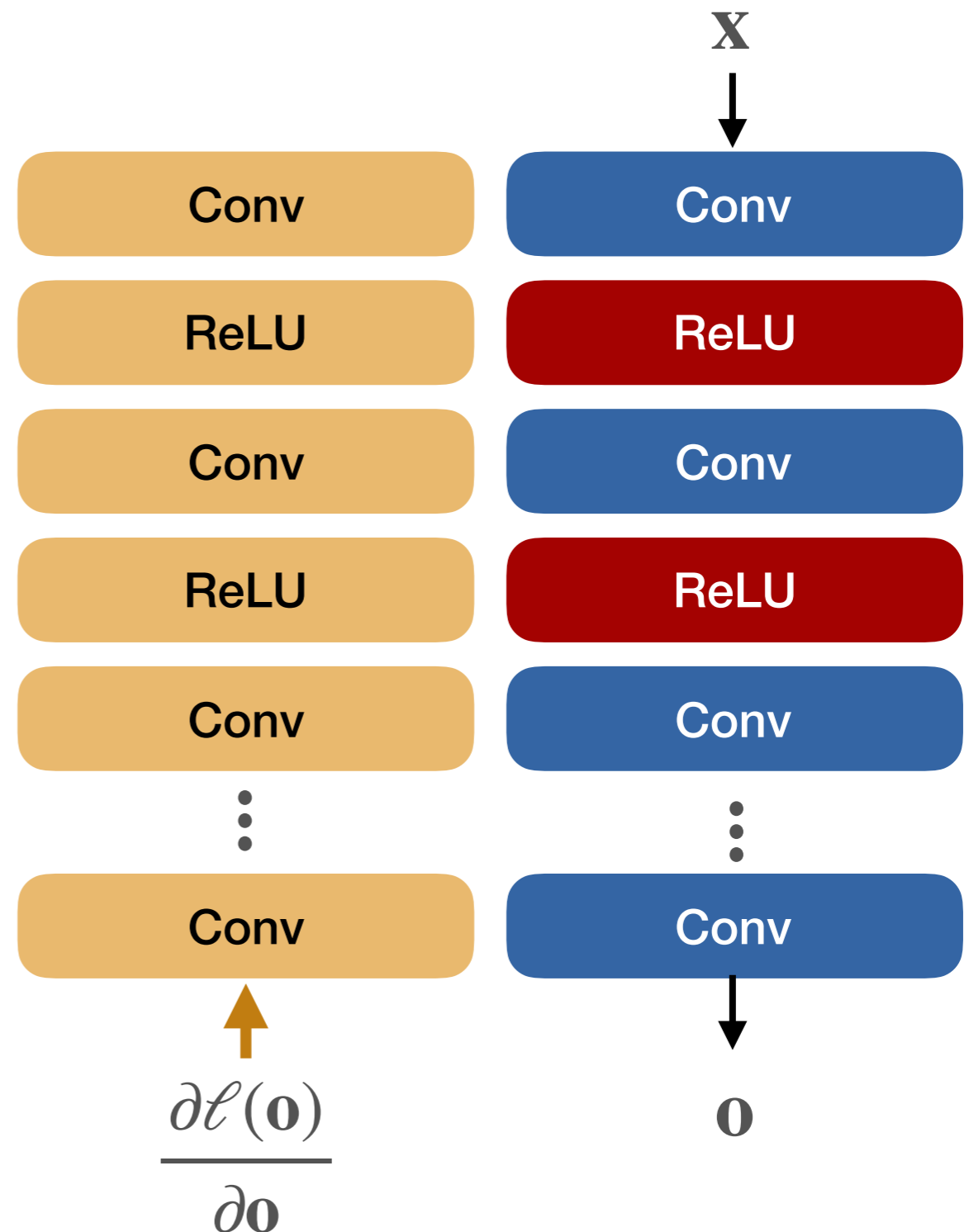
- Not a bit issue for most networks
- An issue for recurrent networks, and networks that share parameters across layers

Detecting exploding gradients

- Plot loss
- Plot weight and gradient magnitude per layer

Vanishing gradients

- Weight of one layer shrink
- Gradient of all other layers shrink
- Weights of other layers stay constant
- No progress



Vanishing gradient

- Big issue for larger networks
- Issue for recurrent networks and weights tied across layers

Detecting vanishing gradients

- Plot loss
- Plot weight and gradient magnitude per layer