



Contribution ID: 27

Type: **not specified**

Adeline Fermanian- Scaling ResNets in the Large-depth Regime

Monday, October 3, 2022 3:00 PM (1 hour)

Deep ResNets are recognized for achieving state-of-the-art results in complex machine learning tasks. However, the remarkable performance of these architectures relies on a training procedure that needs to be carefully crafted to avoid vanishing or exploding gradients, particularly as the depth L increases. No consensus has been reached on how to mitigate this issue, although a widely discussed strategy consists in scaling the output of each layer by a factor α_L . We show in a probabilistic setting that with standard i.i.d. initializations, the only non-trivial dynamics is for $\alpha_L = 1/\sqrt{L}$ (other choices lead either to explosion or to identity mapping). This scaling factor corresponds in the continuous-time limit to a neural stochastic differential equation, contrarily to a widespread interpretation that deep ResNets are discretizations of neural ordinary differential equations. By contrast, in the latter regime, stability is obtained with specific correlated initializations and $\alpha_L=1/L$. Our analysis suggests a strong interplay between scaling and regularity of the weights as a function of the layer index. Finally, in a series of experiments, we exhibit a continuous range of regimes driven by these two parameters, which jointly impact performance before and after training.