We propose Deep Supervision. In order to tackle diminishing gradients for early layers, deep supervision uses auxiliary pathways with weights unused at test time. In fractal networks, deep supervision emerges naturally with no through-ways. We use an even mixture between Local and Global.

We compare test error (%) with other leading methods, trained with either no data augmentation, or with some form of augmentation (+) or more substantial augmentation (++).

FractalNet is a neural network macro-architecture based on truncated fractals. A common technique is to have modules of parallel paths with different receptive field sizes (Szegedy et al. 2015b). Fractal networks exhibit a similar property since the deeper paths have increasingly large receptive fields. A block expanded once can be seen as a parallel module with a 3 x 3 convolution and a non-linear 5 x 5 convolution. A block expanded twice is a parallel module with a 5 x 5 convolution and a non-linear 5 x 5 convolution. A block expanded three times is a parallel module with a 7 x 7 convolution and a non-linear 5 x 5 convolution.