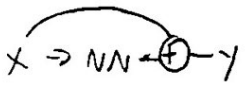
Resnet | December 2015

So after Alex net nf's Just Keep Just adding depth With Idea that first set of layers higher res with Less feature the as the Down scale the imy we add more features. But they run into issue where at some amount of depth the traing Stops Working. it's intursting to not tha this is training error going up So it's not a function of over fitting & we can So that a Bigger Network can do what the Smaller Network Can Since the Smallers one param Space is a Sub Space of the param Space of the Larger NN. it could learn the Same as the Smaller NN for Some Set of layers & for the rest it Would learn the Identity. Note this hard Because The identiy is 1 & we init weight's with gausian near zero & if we do Weight decay then it harder to learn Identity or as hard as to learn any other function. So Why is it so hard for these very Deep N N to learn?

We as the input goes forward the multiplys & Nonlinearitys transform the Input alot. this results inthe Deep layers getting an input thats not representative of the original input. This makes it hard for them to learn. So We add the skip connection. Where the network insted of needing to learn the complicated $H(x)$ where $H(x) = f(x) + x$ We can learn the simple change $f(x)$ & just add in the x. If you this $x \approx y$ Like Low res to high res image the 

but if $x \neq y$ Shape/dim not Similar then we have residual connections within the network