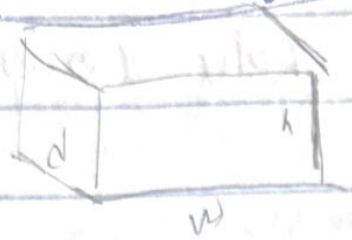
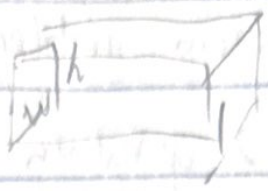
Resnet

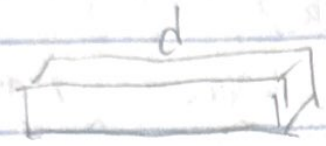
more parameters (w), should more likely to fit, better performance

address ☆vanishing gradience and (degradation)



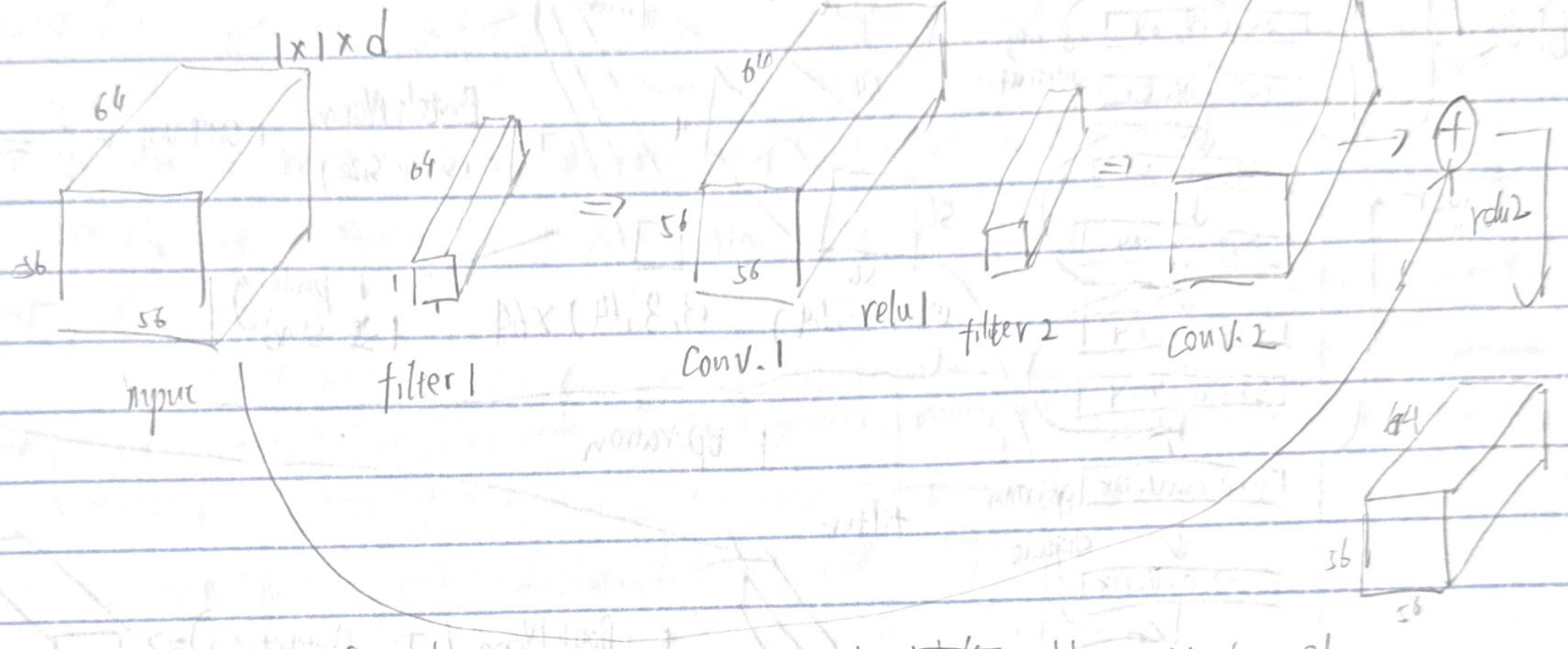


vanishing gradience:
for deep network, the gradience from where the loss function calculated easily shrink to 0 after several application of chain rule



through skip connection, gradience can flow directly backward from later layers to initial filters

pointwise addition



☆ we just skip 1, since second ~~don't~~ add to X for relu

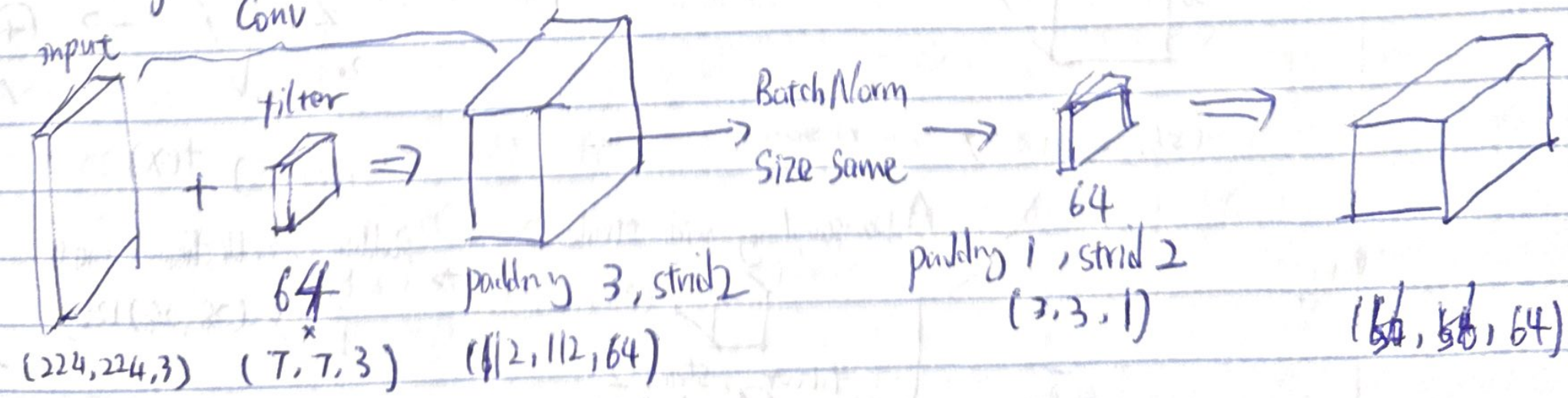
☆ what we try to learn: f(x)

Resnet makes Leaning Identity easier

Leaning f(x) + x => 0 + x is easier than teaming H(x) => X

Gradients can flow directly through the skip connections backward from ~~initial~~ later layers to initial filters

initial process



Only 1 max pooling (at the beginning first ConV1 layer) and 1 average pooling layer at the end of Resnet

⚠ Downsampling of the volume is acieved by increasing stride x pooling

Resnet Layer

1 Operation = a convolution + a Batch Norm + a Relu (Activation)

the last operation in block doesn't have Relu (exception)

- Basic Block = 2 operation in block
- Bottleneck Block = 3 operation in block