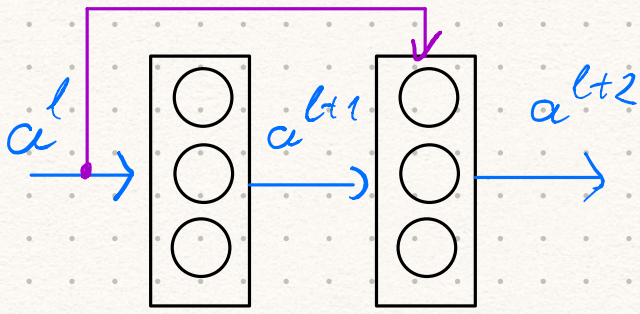
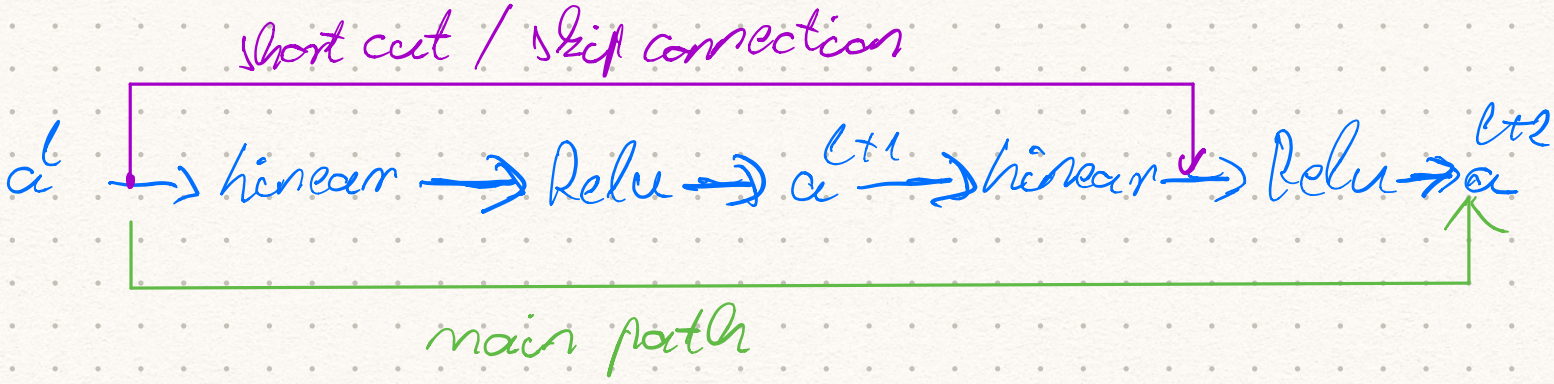
Residual block

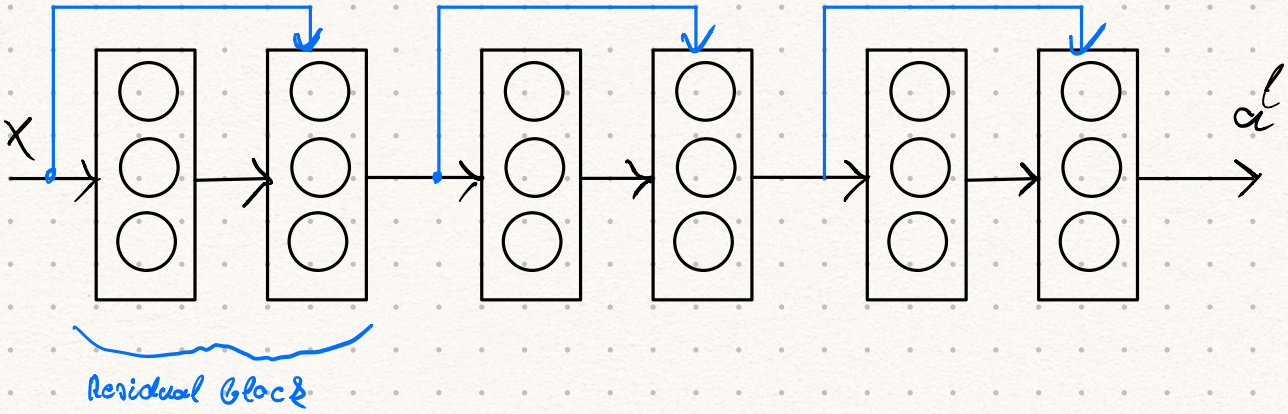


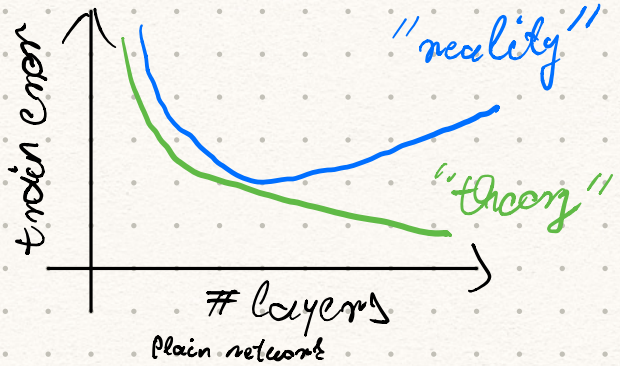


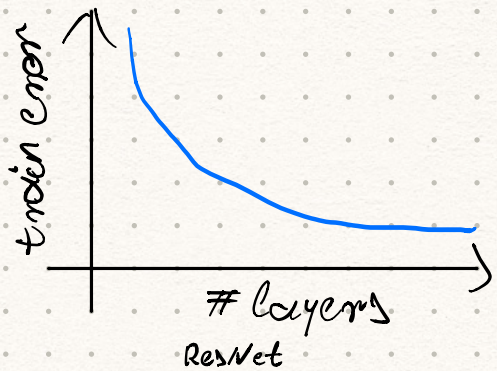
$$z^{l+1} = w^{l+1} a^l + b^{l+1}, \quad a^{l+1} = g(z^{l+1})$$

$$z^{l+2} = w^{l+2} a^{l+1} + b^{l+2}, \quad a^{l+2} = g(z^{l+2}) \Rightarrow a^{l+2} = g(z^{l+2} + a^l)$$

Residual Network (ResNet)

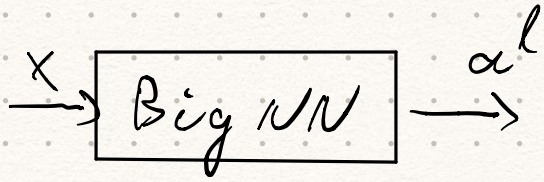


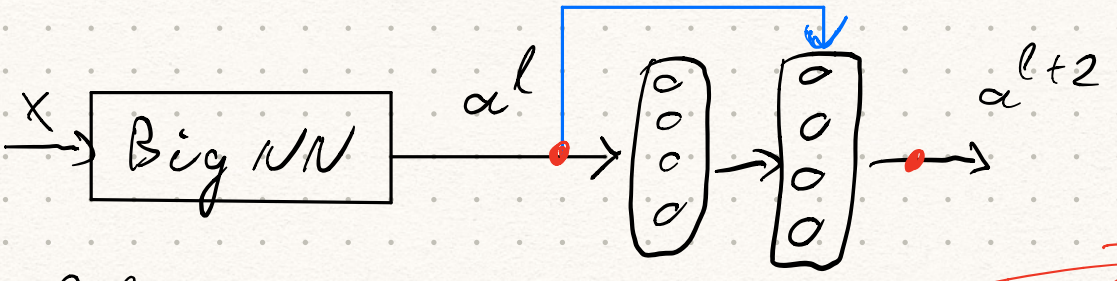




Skip connections → allow to train much deeper networks

Why ResNets work





Relu $\Rightarrow a \geq 0$

$$a^{l+2} = g\left(z^{l+2} + a^l\right) = g\left(w^{l+2}a^{l+1} + b^{l+2} + a^l\right)$$

use "same" convolutions to get the same size

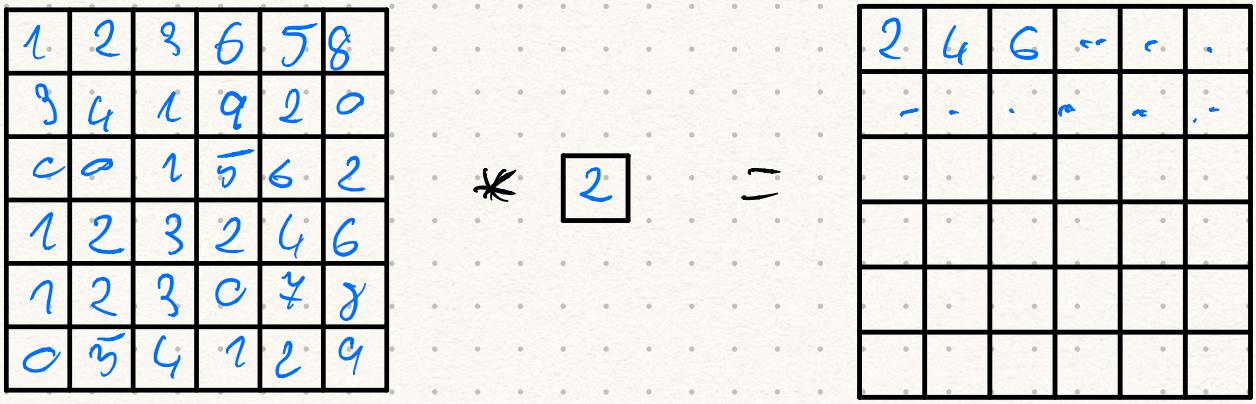
If we use L2 regularization or weight decay
$\Rightarrow$ shrink the value of w^{l+2} and b^{l+2}

If $w^{l+2} = 0$ and $b^{l+2} = 0$

$\hookrightarrow a^{l+2} = g(a^l) = a^l$ ← because of Relu

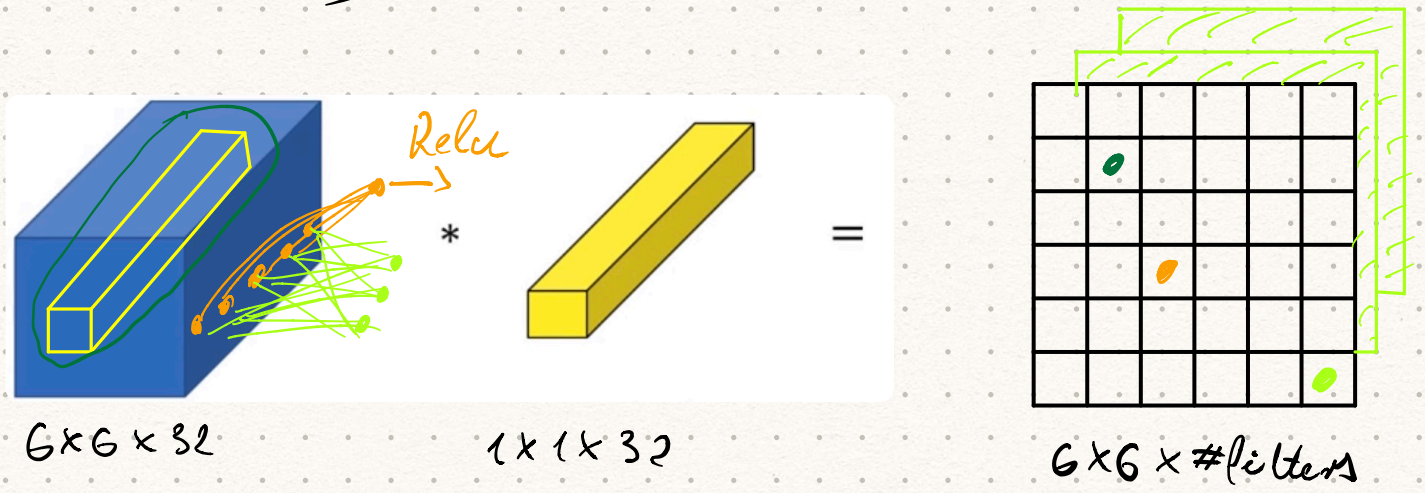
Because of skip connection $\Rightarrow$ identity function is easy to learn $\Rightarrow$ adding more layers do not harm the original structures

Network in network — 1×1 convolutions



with one filter => just a multiplication

but =>



6×6×32 1×1×32 6×6×#filters

take element-wise product and apply Relu => single real number

- if # filters = 1
- if # filters > 1 => fully connected network

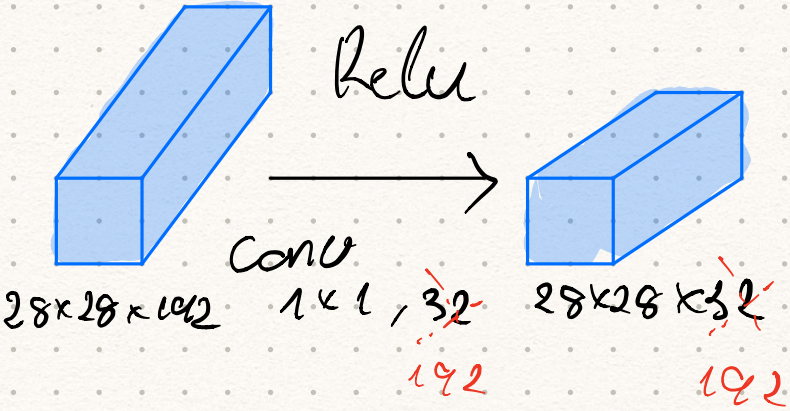
32 input filters -> # filters output value

↳ network in network

Using 1×1 convolution

Pooling -> shrinking width and height

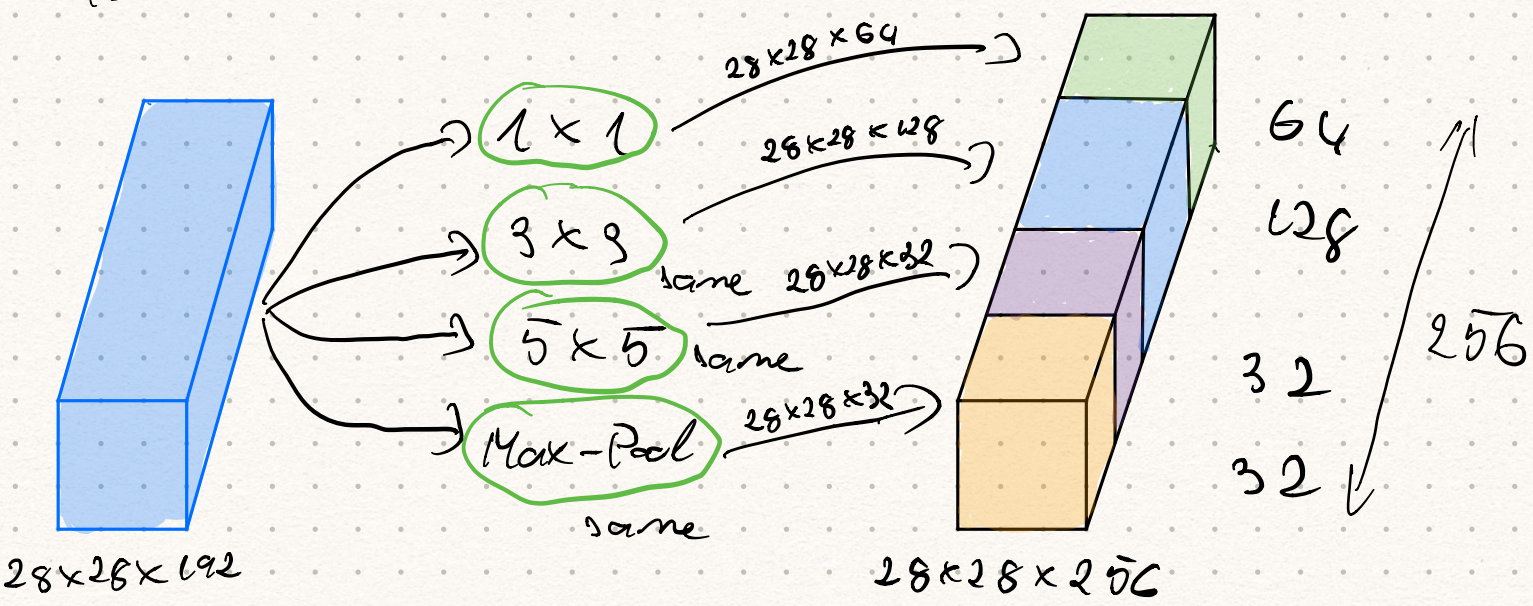
To shrink # filters -> use 1×1 convolutions



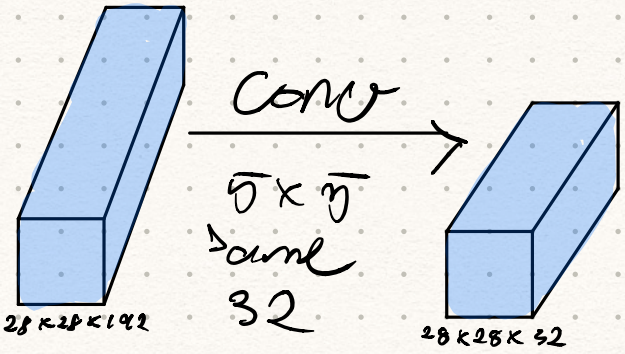
If # filters is the same -> to learn more complex functions

Inception network

Choosing the conv, and pooling filter size is difficult -> don't choose use more



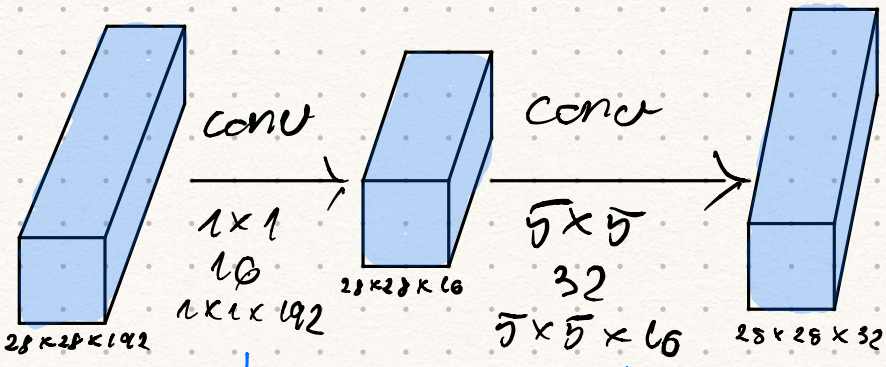
The problem of computational cost



$\Rightarrow 28 \times 28 \times 32 * 5 \times 5 \times 192$

$\approx 120M$

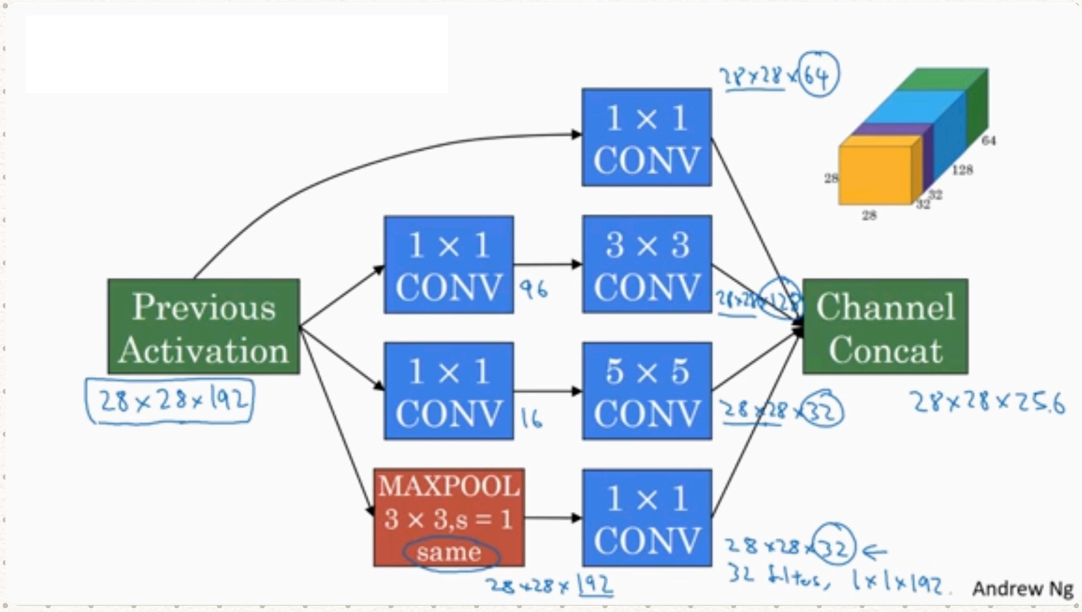
=> reduce computation cost



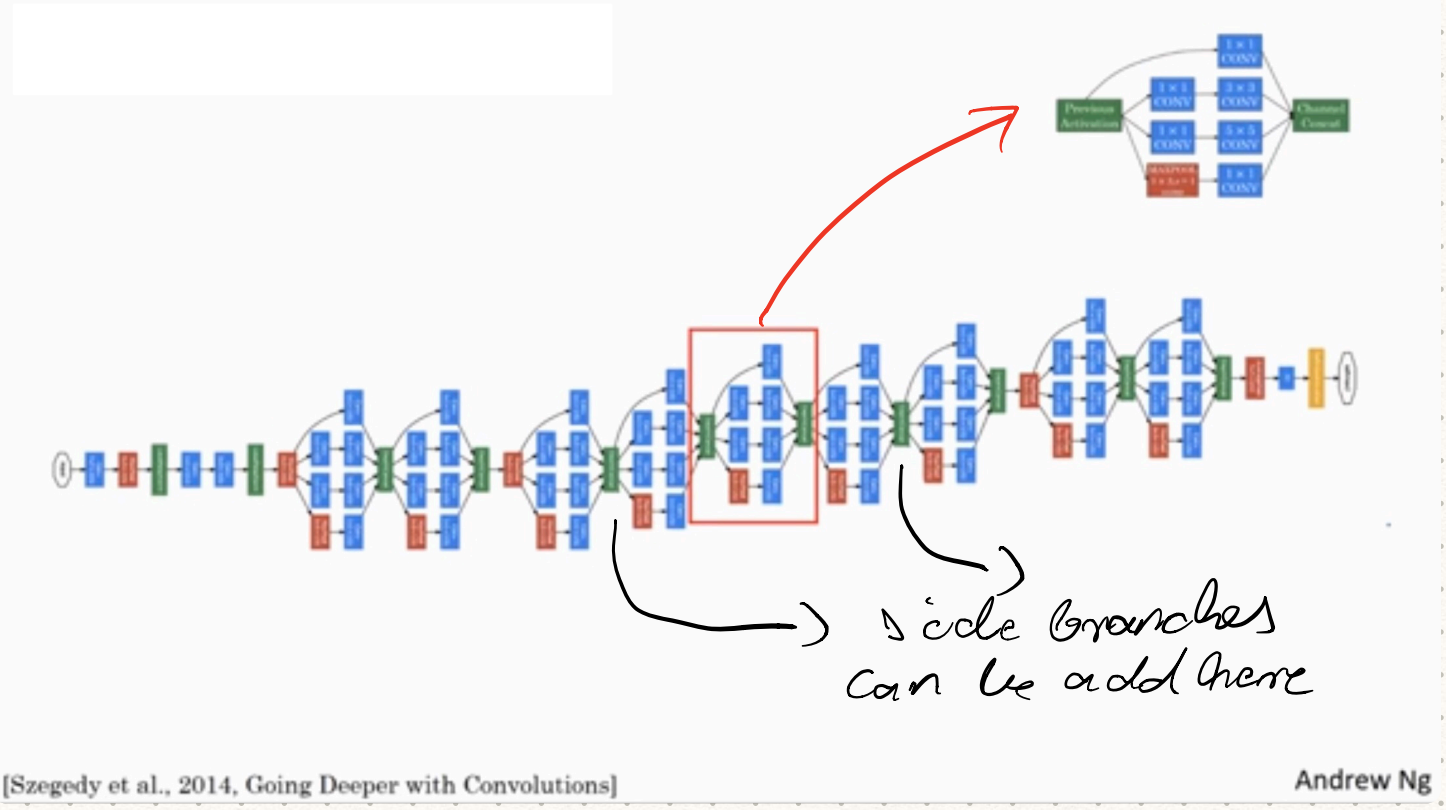
$28 \times 28 \times 16 \times 1 \times 1 \times 192 \cong 2.4M$ | $28 \times 28 \times 32 \times 5 \times 5 \times 16 \cong 10M$

$\cong 12.4M$ => 10 times smaller

Inception module



Inception network ~ google net



[Szegedy et al., 2014, Going Deeper with Convolutions]

Andrew Ng