Partial derivatives in ConvNets

a) $$\frac{\partial L}{\partial w^{(1)}_{1111}} = -\sum_o (t_o - \hat{y}_o) \frac{\partial \hat{y}_o}{\partial w_{1111}}$$

$$= -\sum_o (t_o - \hat{y}_o) \sum_{ijk} w^{(2)}_{ijko}\, \sigma'(a_{ijk}) \frac{\partial a_{ijk}}{\partial w_{1111}}$$

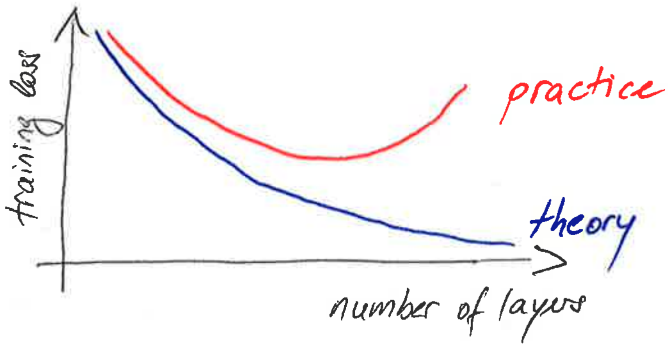
$$= -\sum_o (t_o - \hat{y}_o) \sum_{ijk} w^{(2)}_{ijko}\, \sigma'(a_{ijk})\, I_{ij1} \quad (\star)$$

b) $$\frac{\partial L}{\partial w^{(1)}_{1111}} = -\sum_o (t_o - \hat{y}_o) \sum_{k} w^{(2)}_{ok}\, \sigma'(a_k)\, I_{111}$$

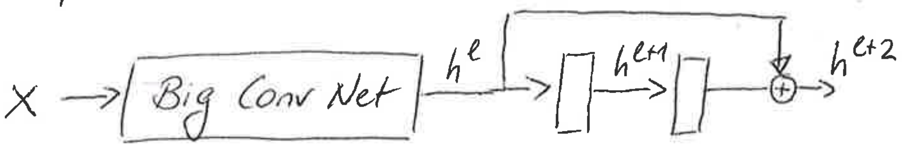
c) Sum in $(\star)$ runs only over i and j that where the arg max for some max operation.

ResNet

without skip connections



with skip connections

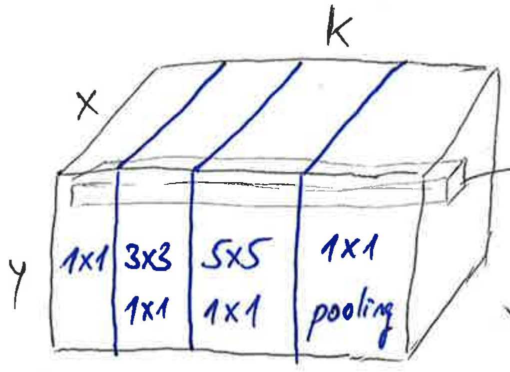


$$h^{l+2} = \sigma(a^{l+2} + h^l)$$

$$= \sigma(\underbrace{W^{l+2} \cdot h^{l+1} + b^{l+2}} + h^l)$$

$= 0$	$\neq 0$
$= \sigma(h^l) = h^l$ ($\sigma = \text{relu}$)	$h^{l+2} \neq h^l$
network with l-layers	network with $l+2$-layers

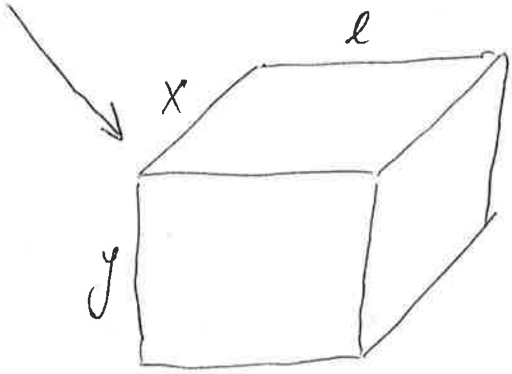
Inception module



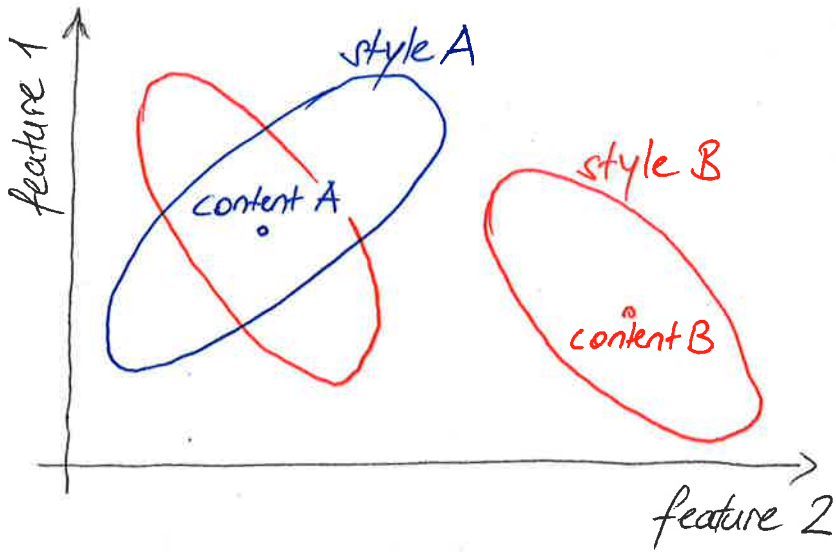
Filter concatenation

1x1 conv = 1x1xk filter

if we take ℓ such filters



Neural style



Automatic reverse mode differentiation

Example:

$$f(x) = g(h(k(x), x), m(x), x)$$

$$\frac{df}{dx} = \frac{\partial g}{\partial h}\frac{\partial h}{\partial x} + \frac{\partial g}{\partial m}\frac{\partial m}{\partial x} + \frac{\partial g}{\partial x}$$

$$= \frac{\partial g}{\partial h}\left(\frac{\partial h}{\partial k}\frac{\partial k}{\partial x} + \frac{\partial h}{\partial x}\right) + \frac{\partial g}{\partial m}\frac{\partial m}{\partial x} + \frac{\partial g}{\partial x}$$