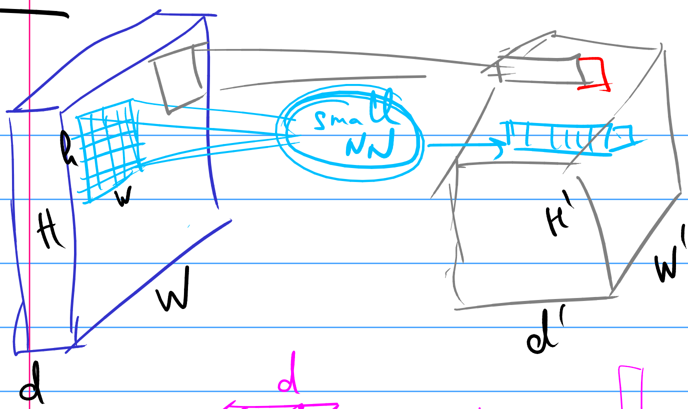
CNN

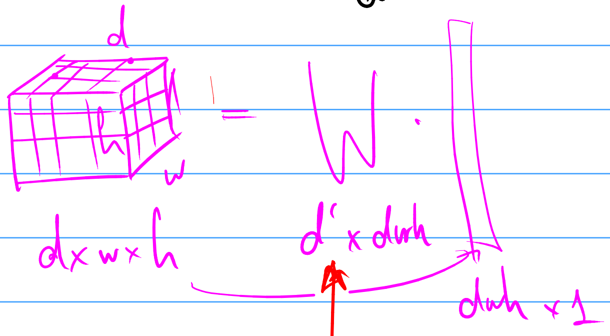


fully connected:

$(d \times H \times W) \cdot (d' \times H' \times W')$

convolutional:

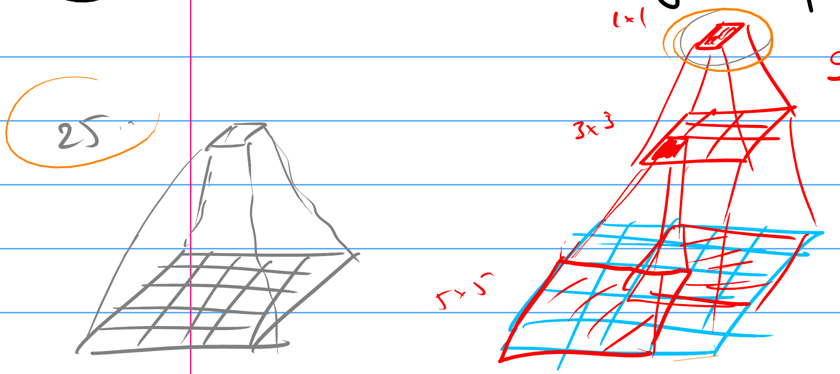
$d \times h \times w \times d'$



$\boxed{d \times w \times h \times d'}$

---

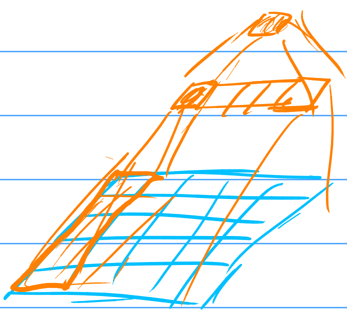
① VGG – Visual Geometry Group (Oxford)



9
+
9
18

$11 \times 11 = 121$

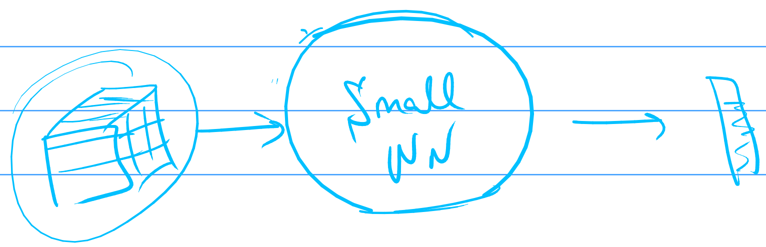
$5 \times (3 \times 3) = 45$



$5 \times 5 \quad (1 \times 5) + (5 \times 1) = 10$

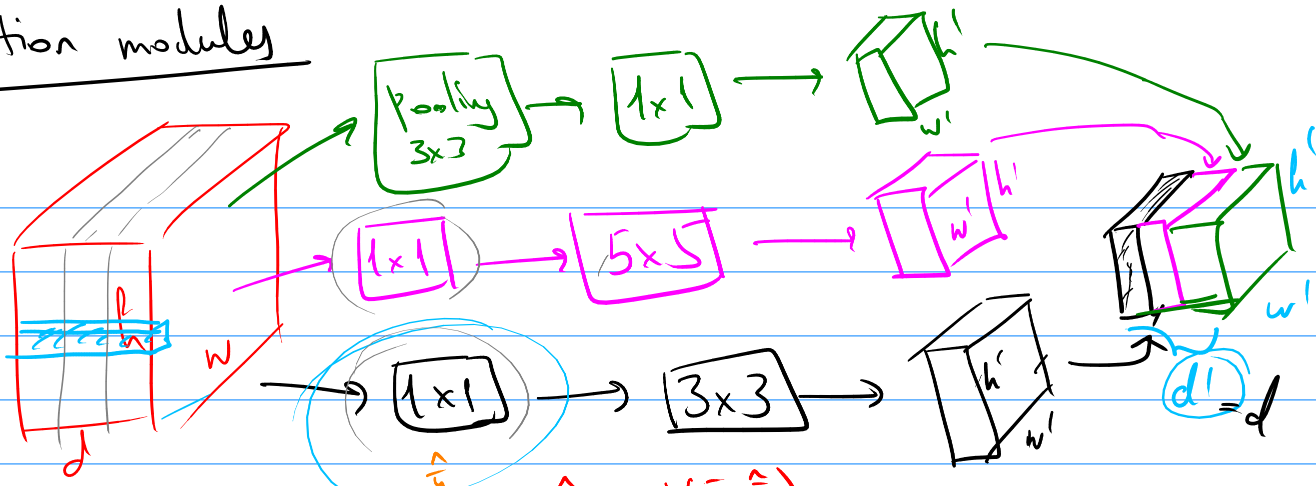
---

② Network in network



Inception modules

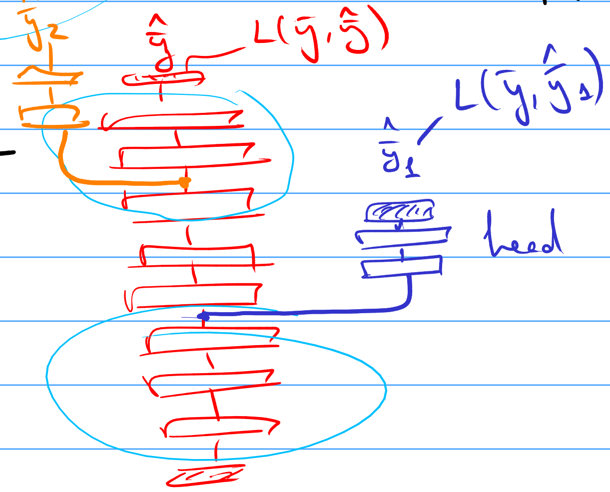
- 1x1
- filter concat



3 Auxiliary classifiers

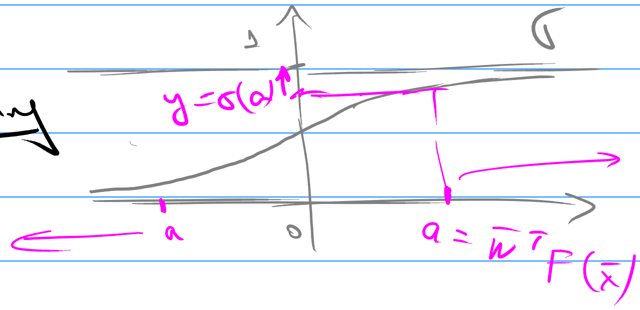
$$L' = L + \lambda_1 L_1 + \lambda_2 L_2$$

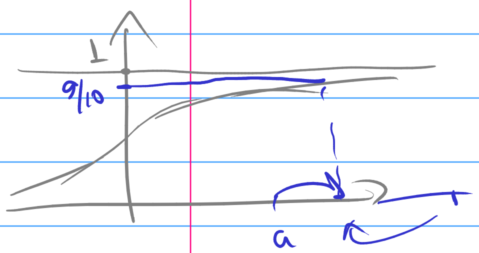
$\lambda_1 \to 0$, $\lambda_2 \to 0$



multi-head architecture

4 Label smoothing





one-hot – $[0 \ldots 1 \ldots 0]$

label smoothing $\left[\frac{\varepsilon}{d-1} \ldots 1-\varepsilon \ldots \frac{\varepsilon}{d-1}\right]$

$$p_{model} \approx p'_{data} = (1-\varepsilon)\, p_{data} + \varepsilon \cdot \text{Uniform}$$

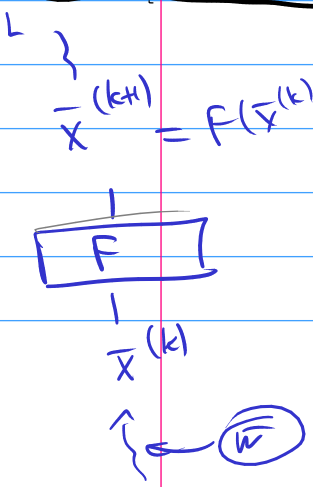
5 Residual connections Kaiming He, 2015

$$\bar{x}^{(k+1)} = F(\bar{x}^{(k)})$$

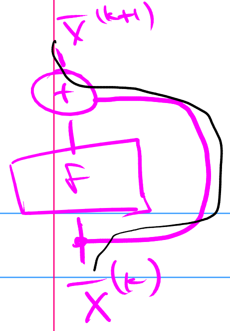
$$\frac{\partial L}{\partial \bar{w}} = \frac{\partial L}{\partial \bar{x}^{(k)}} \frac{\partial \bar{x}^{(k)}}{\partial \bar{w}} = \frac{\partial L}{\partial \bar{x}^{(k+1)}} \boxed{\frac{\partial \bar{x}^{(k+1)}}{\partial \bar{x}^{(k)}}} \frac{\partial \bar{x}^{(k)}}{\partial \bar{w}}$$

F'

$\to 0$



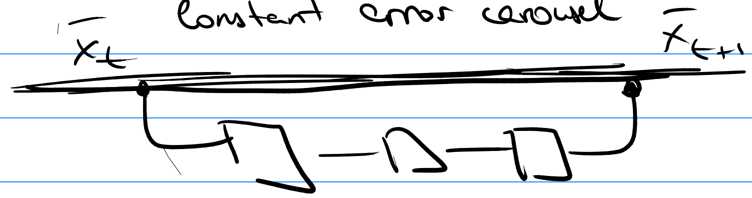
$F: \bar{x}^{(k)} \to \bar{x}^{(k+1)}$



$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + \underline{F}(\bar{x}^{(k)})$$

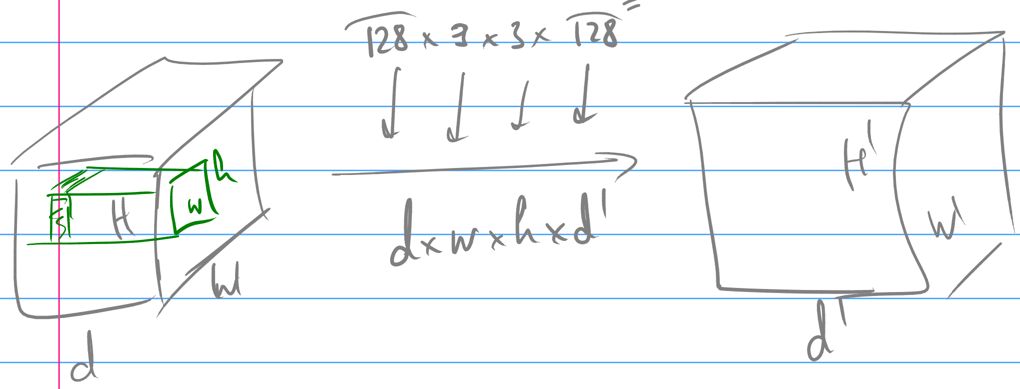
$$F: \bar{x}^{(k)} \mapsto \underbrace{(\bar{x}^{(k+1)} - \bar{x}^{(k)})}_{\text{residue}} \qquad \frac{\partial \bar{x}^{(k+1)}}{\partial \bar{x}^{(k)}} = \underline{F}' + I$$

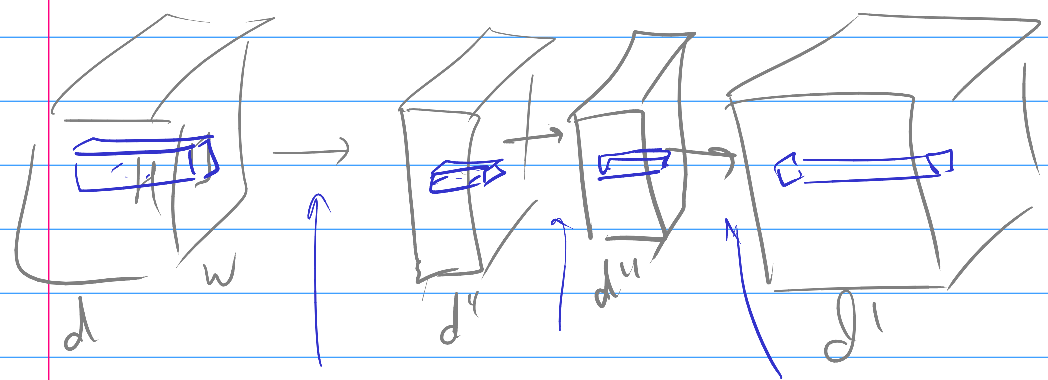
constant error carousel



6 Bottleneck

$$\overline{128} \times 3 \times 3 \times \overline{128} = 2^{14} \cdot 27 \approx 450k$$





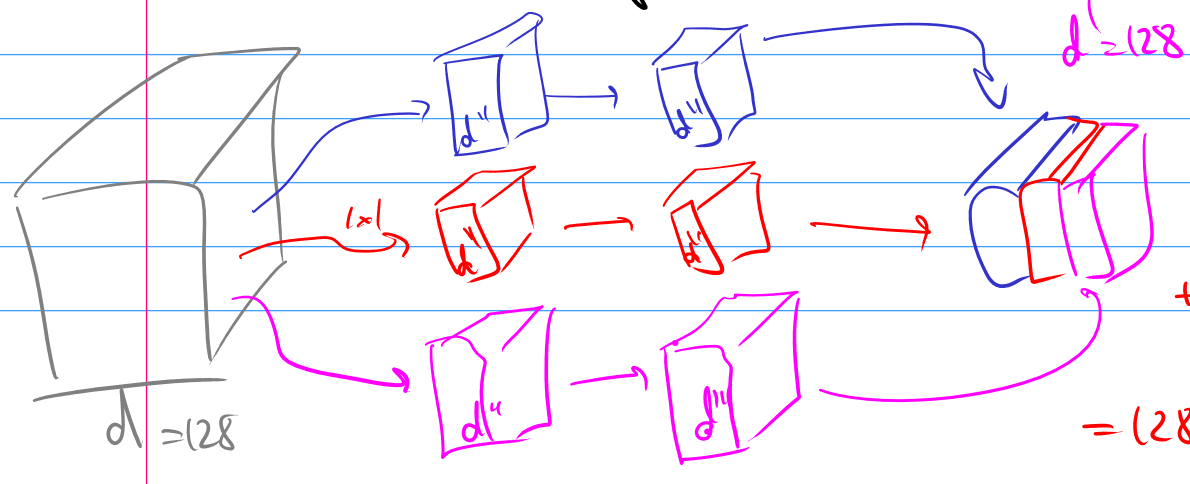
$$\underset{128}{d} \times 1 \times 1 \times \underset{16}{d''} + \underset{16}{d''} \times \underset{3}{w} \times \underset{3}{h} \times \underset{16}{d''} + \underset{16}{d''} \times 1 \times 1 \times \underset{128}{d'}$$

$$128 \cdot 32 + 128 \cdot 2 \cdot 27 = 128 \cdot 86 \approx 10k$$

7 Split – transform – merge

$128^2 \times 9$

$d' = 128$



$$8 \times (128 \times 1 \times 1 \times 16) + 8 \times (16 \times 3 \times 3 \times 16)$$

$$= 128^2 + 128 \times 27 \times 16 \approx 35k$$

(8) DenseNet