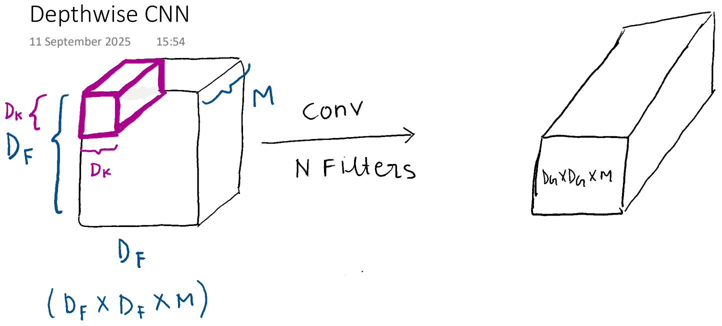
Depthwise CNN

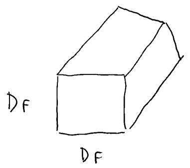
11 September 2025 15:54



Computation of normal convolution

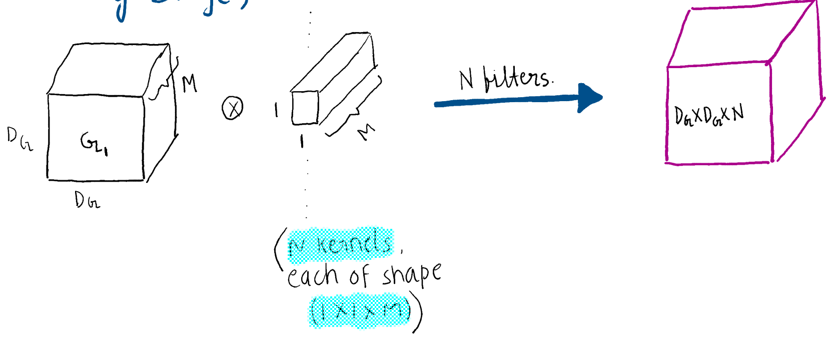
- Mults once $= D_K^2 \times M$
- Mults per Kernel $= D_G^2 \times D_K^2 \times M$
- Mults N Kernels $= N \times D_G^2 \times D_K^2 \times M$

Depthwise Convolution



- We only apply one kernel to a single input channel.
- Hence we require M such kernels over the entire F.
- Each kernel is $D_K \times D_K \times 1$
- Each kernel's output is $D_G \times D_G \times 1$
- Stacking all the kernels (M) we get $D_G \times D_G \times M$

Pointwise Convolution (Filtering stage)



Computations

Depthwise Separable Convolution:

- Mults once $= D_K^2$
- Mults one channel $= D_G^2 \times D_K^2$
- DC Mults $= M \times D_G^2 \times D_K^2$

Pointwise Convolution:

- Mults once $= M$
- Mults 1 kernel $= D_G^2 \times M$
- PC Mults $= N \times D_G^2 \times M$

Total = DC Mults + PC Mults

$$= M \times D_G^2 \times D_K^2 + N \times M \times D_G^2$$

$$= M \times D_G^2 \left(D_K^2 + N\right)$$

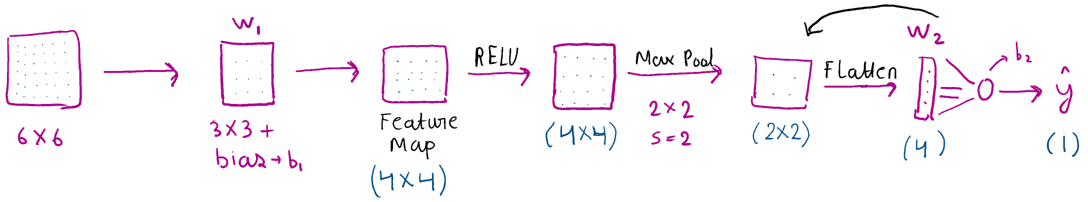
Comparsion Standard Vs. Depthwise

$$\frac{\text{No. of mults in Depthwise}}{\text{No. of mults in Standar}} = \frac{\cancel{M} \times \cancel{D_G^2} \left(D_K^2 + N\right)}{\cancel{M} \times N \times \cancel{D_G^2} \times D_K^2}$$

$$= \frac{D_K^2 + N}{(D_K^2 \times N)} = \boxed{\frac{1}{N} + \frac{1}{D_K^2}}$$

1) Group Convolution

- A normal conv would look accross all 64 input channels. That's PRICEY
- GroupConv splits 64 channels into groups (paper uses the half the channels as groups → 32 groups each size 2).



Trainable Parameters

$W_1 = (3,3)$ $W_2 = (1,4)$

$b_1 = (1,1)$ $b_2 = (1,1)$

= 15 trainable parameters.

$$L = -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$$

Logical Flow

$$X \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{ReLU} A_1 \xrightarrow[Pool]{Max} P_1 \xrightarrow{Flat} F \xrightarrow{[W_2, b_2]} Z_2 \xrightarrow{sigmoid} A_2 \longrightarrow L$$

Forward Prop

$Z_1 = \text{conv}(X, W_1) + b_1$	$F = \text{flatten}(P_1)$
$A_1 = \text{relu}(Z_1)$	$Z_2 = W_2 F + b_2$
$P_1 = \text{MaxPool}(A_1)$	$A_2 = \sigma(Z_2)$

$$\frac{\delta L}{\delta W_2}$$

Gradient Decent on ANN

$$\frac{\delta L}{\delta W_2} = \frac{\delta L}{\delta A_2} \times \frac{\delta A_2}{\delta Z_2} \times \frac{\delta Z_2}{\delta W_2}$$

Ch-1
Ch-2
Ch-4

The Gaussian Distribution

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

likelyhood function

$$p(D \mid \mu, \sigma^2) = \prod_{n=1}^{N} p(x_n \mid \mu, \sigma^2) \quad [D = [x_1, x_2, \ldots, x_n]]$$

N

$$p(\mathcal{D}|\mu,\sigma^2) = \prod_{n=1} p(x_n|\mu,\sigma^2). \quad [\mathcal{D} = \{x_1, x_2, \ldots, x_N\}]$$

$$\log p(\mathcal{D}|\mu,\sigma^2) = \sum_{n=1}^{N} \log p(x_n|\mu,\sigma^2)$$

$$\log p(\mathcal{D}|\mu,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$$

Maximising w.r.t to μ,

$$\frac{\partial}{\partial\mu}\log p(\mathcal{D}|\mu,\sigma^2) = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu) = 0$$

$$\Rightarrow \quad \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Similarly,

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$

Ex-2.1

Given from book:

- True +ve: $\alpha = p(T=1 \mid C=1) = 0.9$
- False +ve: $\beta = p(T=1 \mid C=0) = 0.03$
- NEW prior (prevelance): $p(C=1) = \pi = 0.001$

$$\Rightarrow p(C=0) = 1-\pi = 0.999$$

We want to calculate $p(C=1 \mid T=1)$

Bayes' Theorem:

$$p(C=1|T=1) = \frac{p(T=1 \mid C=1)\, p(C=1)}{p(T=1 \mid C=1)\, p(C=1) + p(T=1 \mid C=0)\, p(C=0)}$$

$$= \frac{\alpha\pi}{\alpha\pi + \beta(1-\pi)} = \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.003 \times 0.999}$$

$$= \frac{0.0009}{0.03087} \approx 0.0292$$

Ex-2.4

$p(x) = 1/(d-c), \quad x \in (c,d)$

Integrating over x we obtain

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{c}^{d}\frac{1}{d-c}\,dx = \frac{d-c}{1} = 1$$

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{c}^{d} \frac{1}{d-c}\,dx = \frac{d-c}{d-c} = 1$$

$$\mathbb{E}[x] = \int_a^b \frac{1}{b-a} x\,dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathbb{E}[x^2] = \int_a^b \frac{1}{b-a} x^2\,dx = \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

Ex- 2.10

For a continuous random variable X

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x p(x)\,dx$$

$$\mathbb{E}[x+z] = \iint (x+z)\, p(x,z)\,dx\,dz$$

$$= \iint (x+z)\, p(x)\, p(z)\,dx\,dz$$

$$= \int x\,p(x)\,dx \underbrace{\int p(z)\,dz}_{=1} + \underbrace{\int p(x)\,dx}_{=1} \int z\,p(z)\,dz$$

$$= \mathbb{E}[x] + \mathbb{E}[z]$$