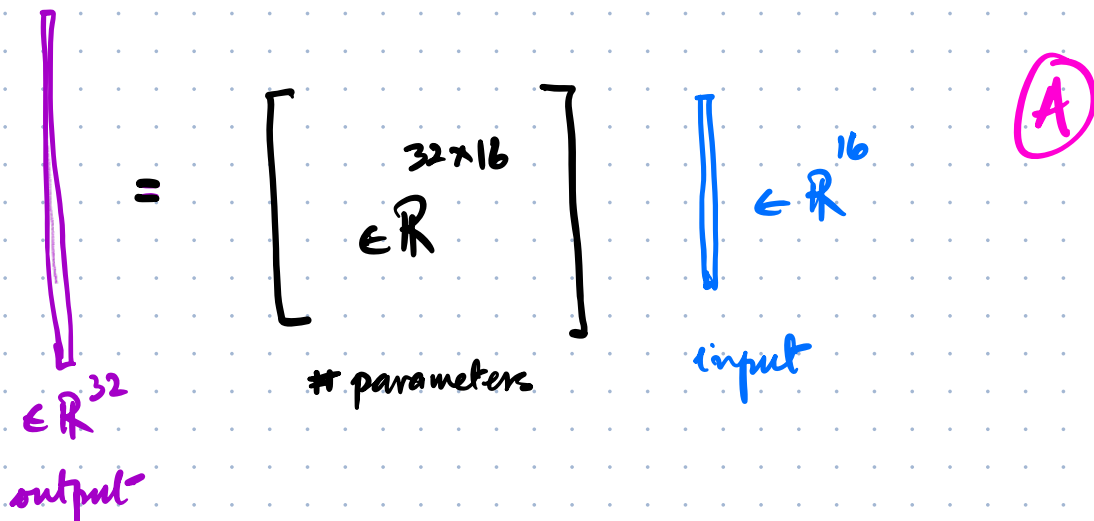Topics in CS 1 (Fall 2024)

Oct 30, 2024

Quiz. $64 \times 64 \times 16 \longrightarrow 64 \times 64 \times 32$

↑
channels

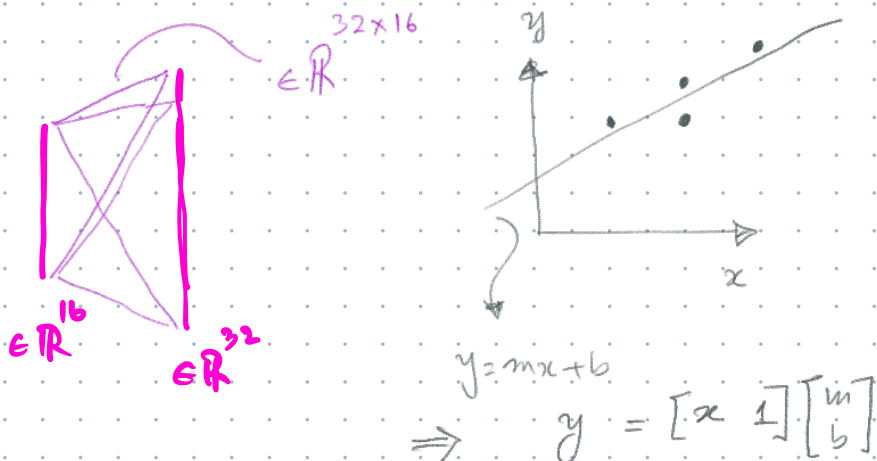We are expanding the number of channels

Point-wise convolution (1x1 kernel)
  Depthwise convolution.

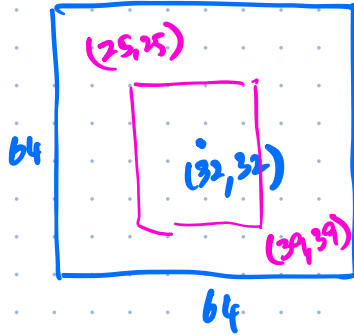$$\Big| = \begin{bmatrix} \in \mathbb{R}^{32 \times 16} \end{bmatrix} \Big| \in \mathbb{R}^{16}$$

$\in \mathbb{R}^{32}$ # parameters   input

output

(A)

Total number of multiplications = $\dfrac{(64)(64)}{} \underbrace{(32)(16)}_{\text{at each location}}$

↓
total # of locations

(A) Ensure that you understand that you have a <u>linear</u> layer.
Or an MLP that applies to individual feature
independantly.

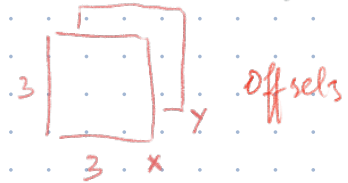$\in \mathbb{R}^{32 \times 16}$

$\in \mathbb{R}^{16}$ $\in \mathbb{R}^{32}$

$y = mx + b$

$\Rightarrow \quad y = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$

Quiz

input   64 x 64

(25,25)

(32,32)

(39,39)
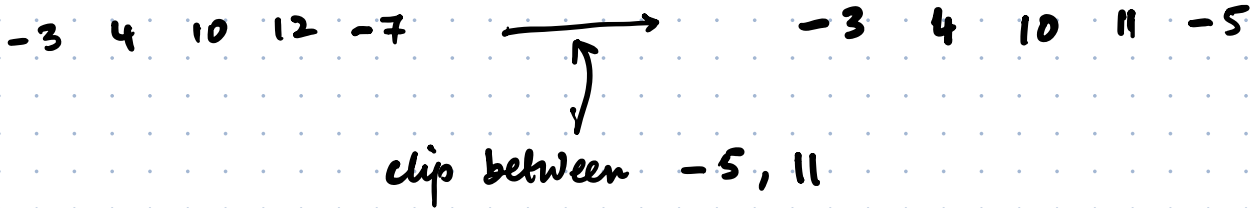
64

64

Kernel : 3x3 ⟶ # parameters = 9

parameters : 34

how many parameters available : 34-9=25

Use 5x5 kernel to compute offsets.

3

3   x

−y

Offsets

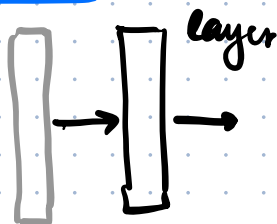Ensure that the minimum and maximum values are <u>clipped</u> at |7|.

<u>Clipping values</u>

−3   4   10   12   −7   ⟶   −3   4   10   11   −5
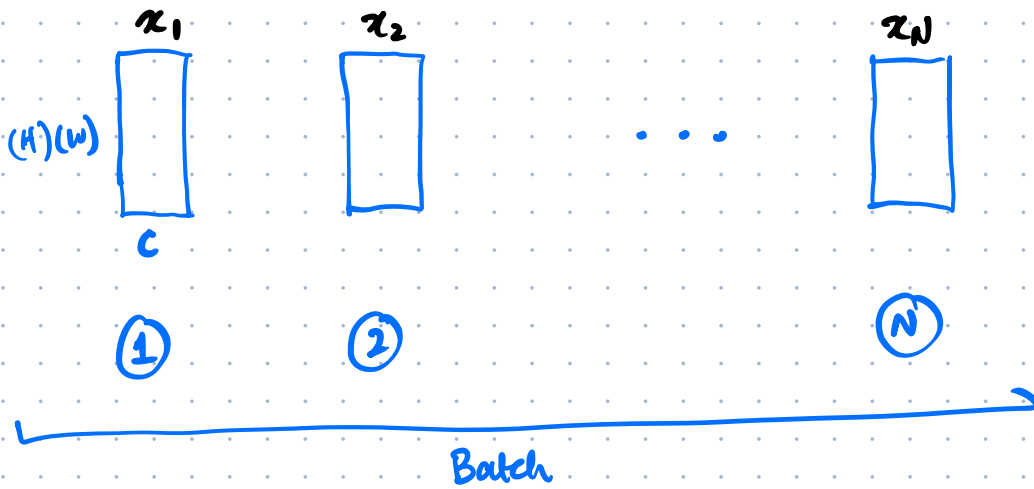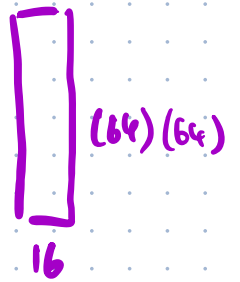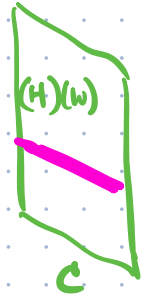
clip between −5, 11

<u>Building Blocks</u>

1. Linear layers (MLP)

2. Convolutional layers

3. <u>Normalization</u>

BATCH NORMALIZATION

layer

(H)(W)

c channel

N Batch

Feature at batch k $\in \mathbb{R}^C$
k location j

(H)(W)

c

64   16
64

(64)(64)

16

x_1 x_2 $\dots$ x_N

(H)(W)

c

① ② Ⓝ
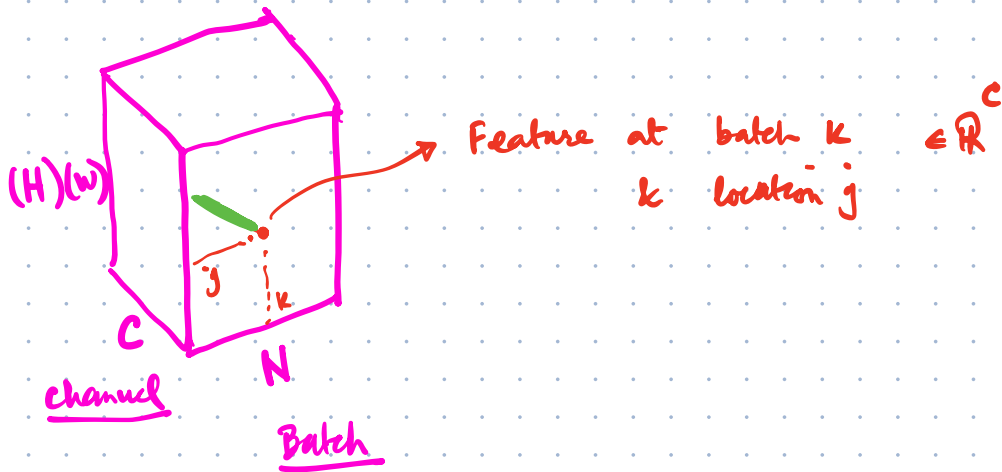
Batch

$$\mu_B = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

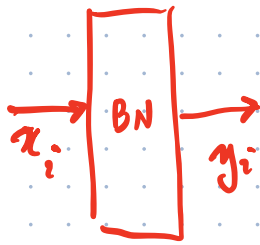$$\sigma_B^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}$$

$\uparrow$ to avoid the divide by zero.

scale   shift

$$y_i = \gamma \hat{x}_i + \beta$$

$\uparrow$ $\uparrow$

the parameters of the Batch Normalization layer.

$x_i \rightarrow$ BN $\rightarrow y_i$

how many parameters BN has.

Batch 1

current
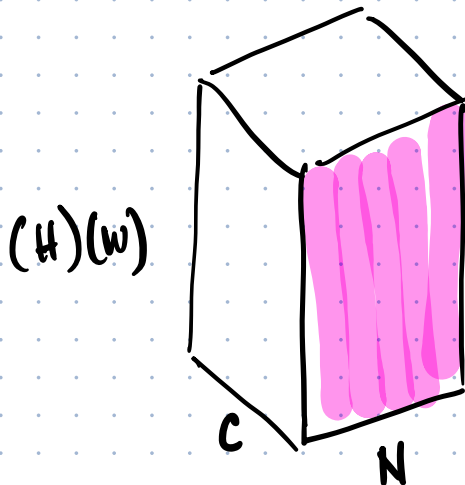estimate
of γ and β
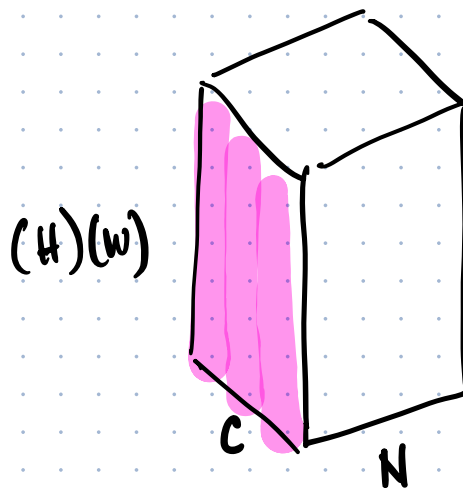
y_i, Batch 1

Batch 2

y_i, batch 2

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 7 \\ 6 \end{pmatrix} \rightarrow \begin{pmatrix} 13/3 \\ 14/3 \end{pmatrix}$$
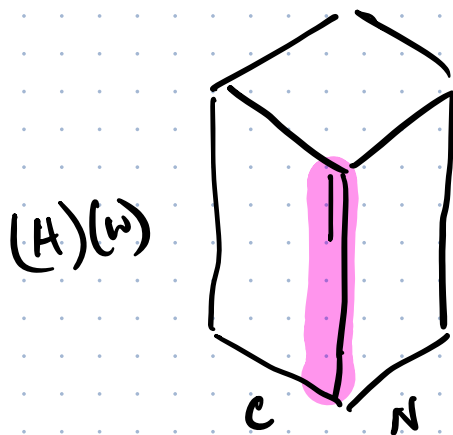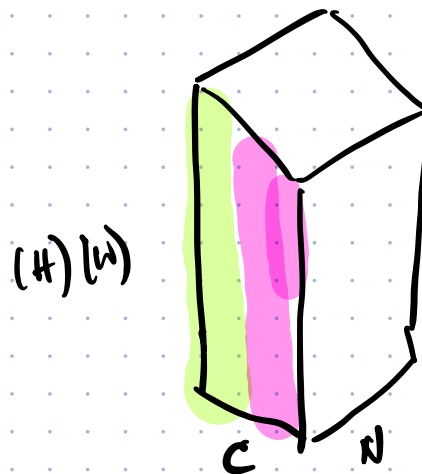
mean

Batch Normalization

$(H)(W)$

C   N

Layer Normalization

$(H)(W)$

C   N

Instance normalization

$(H)(w)$

C   N

Group Normalization

$(H)(w)$

C   N

You should be able to express the difference between training & <u>testing</u> when it comes to <u>batch normalization.</u>
validation