Deep Learning

4.1 Convolutional Neural Networks (CNN)

Dr. Konda Reddy Mopuri
kmopuri@iittp.ac.in
Dept. of CSE, IIT Tirupati

CNNs

1. Neurons are similar to that of MLP

CNNs

1. Neurons are similar to that of MLP
    - Perform a linear (dot product) operation and have a nonlinearity

CNNs

1. Neurons are similar to that of MLP
    - Perform a linear (dot product) operation and have a nonlinearity
2. Architecture will have a differentiable loss function, backpropagation is used

CNNs

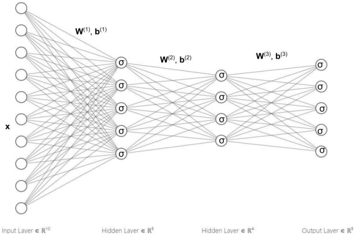
1. Neurons are similar to that of MLP
    - Perform a linear (dot product) operation and have a nonlinearity
2. Architecture will have a differentiable loss function, backpropagation is used
3. Same tips and tricks apply

CNNs

1. Neurons are similar to that of MLP
   - Perform a linear (dot product) operation and have a nonlinearity
2. Architecture will have a differentiable loss function, backpropagation is used
3. Same tips and tricks apply
4. So, what changes?

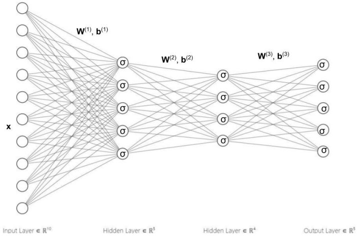
An MLP

1. Input is a vector



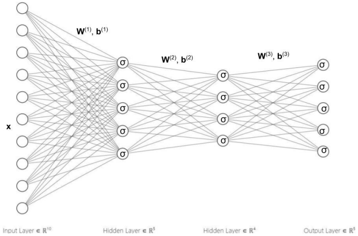
An MLP

1. Input is a vector
2. Series of densely connected hidden layers



An MLP

1. Input is a vector
2. Series of densely connected hidden layers
3. Neurons in each layer are independent



An MLP for processing an image

1. Say, we want to process a 200×200 RGB image

An MLP for processing an image

1. Say, we want to process a 200×200 RGB image
2. Vectorizing leads to $200 \times 200 \times 3 \rightarrow 120K$ neurons in the input layer

An MLP for processing an image

1. Say, we want to process a 200×200 RGB image
2. Vectorizing leads to $200 \times 200 \times 3 \rightarrow 120K$ neurons in the input layer
3. A hidden layer of same size leads to $\approx 1.44e^{10}$ weights $\rightarrow \approx 58GB$

An MLP for processing an image

1. Say, we want to process a 200×200 RGB image
2. Vectorizing leads to $200 \times 200 \times 3 \rightarrow 120K$ neurons in the input layer
3. A hidden layer of same size leads to $\approx 1.44e^{10}$ weights $\rightarrow \approx 58GB$
4. Full connectivity blows the number of weights $\rightarrow$ hardware limits, overfitting, etc.

An MLP for processing an image

1. Say, we want to process a 200×200 RGB image
2. Vectorizing leads to $200 \times 200 \times 3 \rightarrow 120K$ neurons in the input layer
3. A hidden layer of same size leads to $\approx 1.44e^{10}$ weights $\rightarrow \approx 58GB$
4. Full connectivity blows the number of weights $\rightarrow$ hardware limits, overfitting, etc.
5. Flattening removes the structure

Large Signals

1. Have invariance in translation

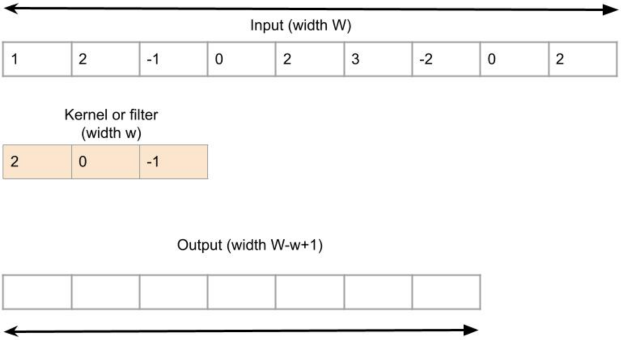
Large Signals

1. Have invariance in translation
2. Features may occur at different locations in the signal

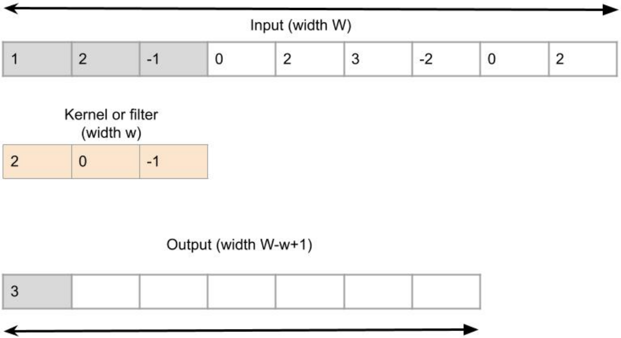
Large Signals

1. Have invariance in translation
2. Features may occur at different locations in the signal
3. Convolution incorporates this idea: Applies same linear operation at all the locations and preserves the structure

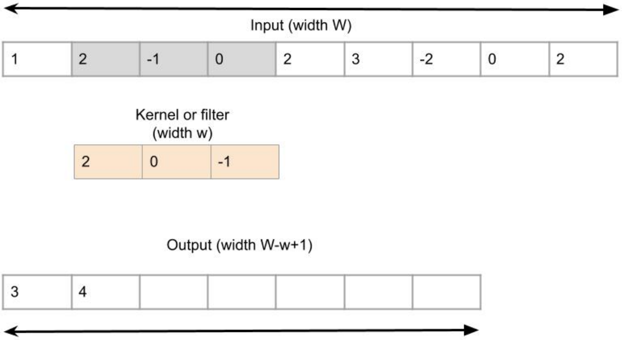
Convolution



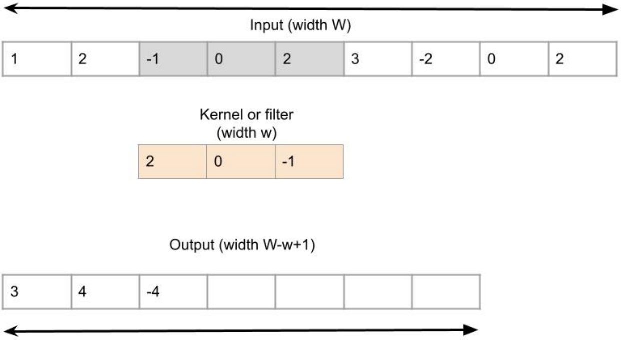
Convolution



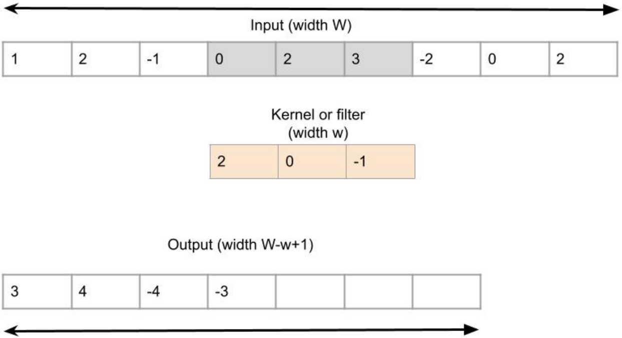
Convolution



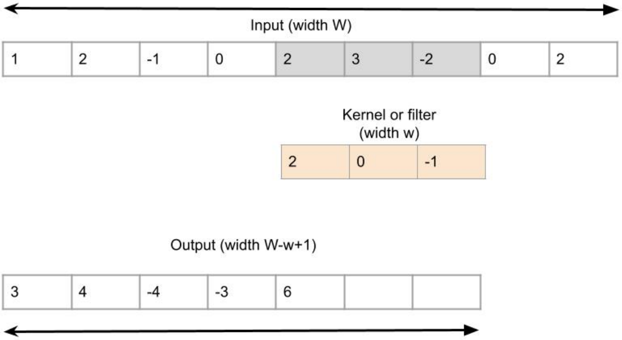
Convolution



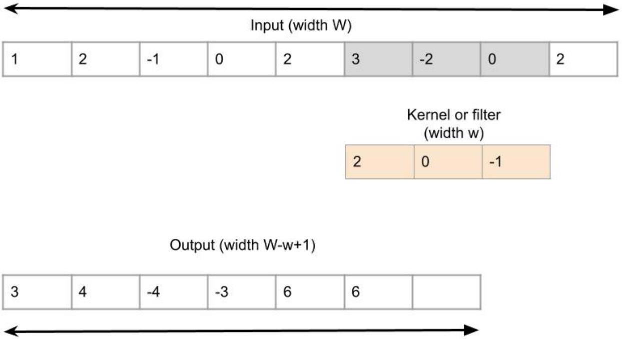
Convolution



Convolution



Convolution



Convolution

Input (width W)

1	2	-1	0	2	3	-2	0	2

Kernel or filter
(width w)

2	0	-1

Output (width W-w+1)

3	4	-4	-3	6	6	-6

Convolution

1. Preserves the structure

Convolution

1. Preserves the structure
   - if the i/p is a 2D tensor $\rightarrow$ o/p is also a 2D tensor

Convolution

1. Preserves the structure
   - if the i/p is a 2D tensor $\rightarrow$ o/p is also a 2D tensor
   - There exist a relation between the locations of i/p and o/p values

Convolution

1. Let $\mathbf{x} = (x_1, x_2, \ldots x_W)$ is the input, $\mathbf{k} = (k_1, k_2, \ldots k_w)$ is the kernel

Convolution

1. Let $\mathbf{x} = (x_1, x_2, \ldots x_W)$ is the input, $\mathbf{k} = (k_1, k_2, \ldots k_w)$ is the kernel
2. The result $(x \circledast k)$ of convolving $\mathbf{x}$ with $\mathbf{k}$ will be a 1D tensor of size $W - w + 1$

$$
\begin{aligned}
(x \circledast k)_i =& \sum_{j=1}^{w} x_{i-1+j} k_j \\
=& (x_i, \ldots x_{i+w-1}) \cdot \mathbf{k}
\end{aligned}
$$

Convolution

1. Powerful feature extractor

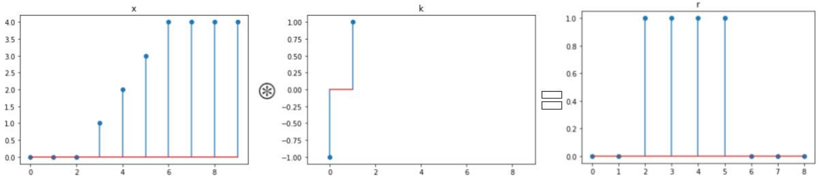
Convolution

1. Powerful feature extractor
2. For instance, it can perform differential operation and look for interesting patterns in the input

Convolution

1. Powerful feature extractor
2. For instance, it can perform differential operation and look for interesting patterns in the input
3.

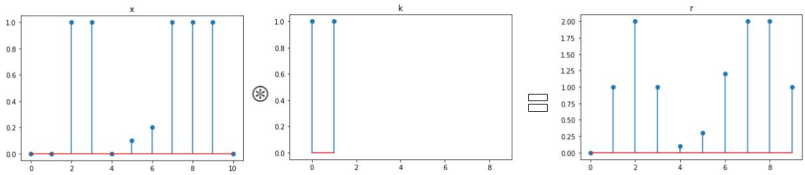
$$(0,0,0,1,2,3,4,4,4,4) \circledast (-1,1) = (0,0,1,1,1,1,0,0,0)$$



Convolution

1. Powerful feature extractor
2. For instance, it can perform differential operation and look for interesting patterns in the input
3.

$$(0,0,1,1,0,0.1,0.2,1,1,1,0) \circledast (1,1) = (0,1,2,1,0.1,0.3,1.2,2,2,1)$$



Convolution

1. Naturally generalizes to multiple dimensions

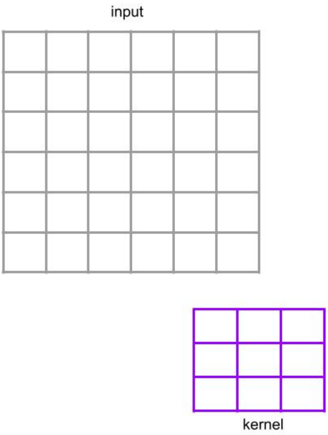
Convolution

1. Naturally generalizes to multiple dimensions
2. In their most usual form, CNNs process 3D tensors of size $C \times H \times W$ with kernels of size $C \times h \times w$ and result in 2D tensors of size $H - h + 1 \times W - w + 1$

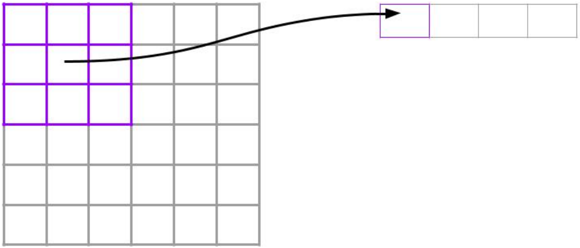
Convolution

1. Naturally generalizes to multiple dimensions
2. In their most usual form, CNNs process 3D tensors of size $C \times H \times W$ with kernels of size $C \times h \times w$ and result in 2D tensors of size $H - h + 1 \times W - w + 1$
3. Note that we generally refer to these inputs as 2D signal (despite having C channels), because, they are referenced as vectors indexed by 2d locations without structure in the channel dimension

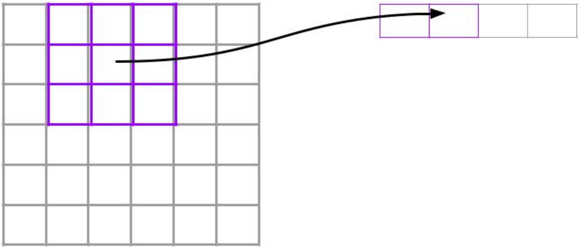
2D Convolution



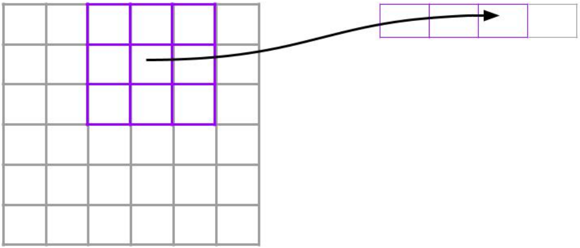
2D Convolution



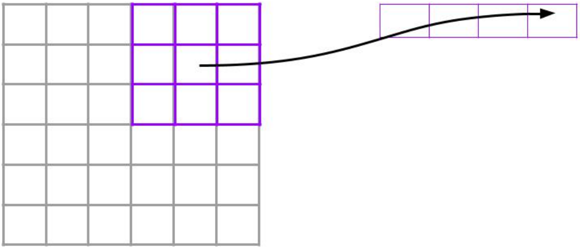
2D Convolution



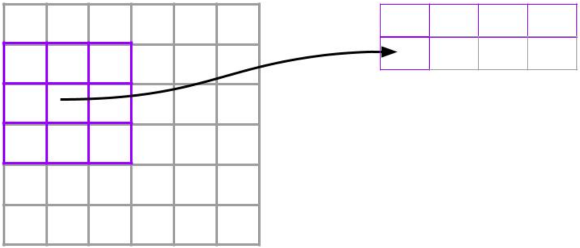
2D Convolution



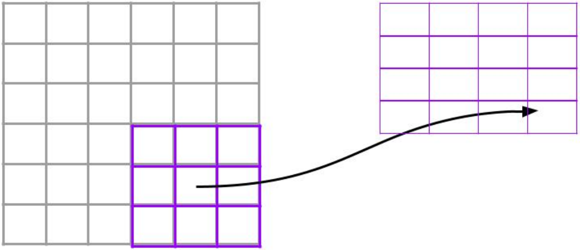
2D Convolution



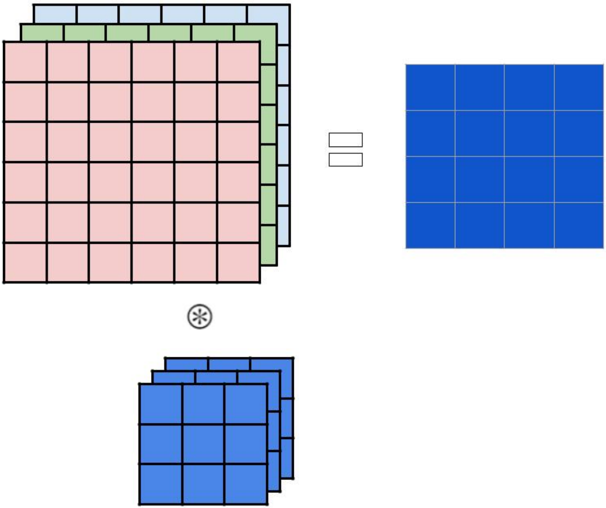
2D Convolution



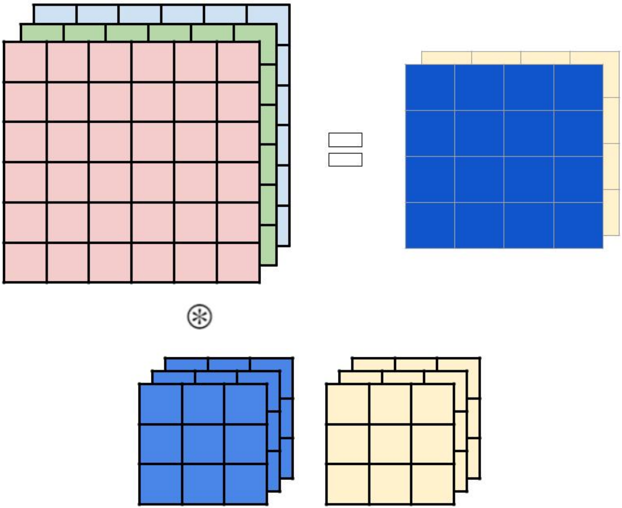
2D Convolution



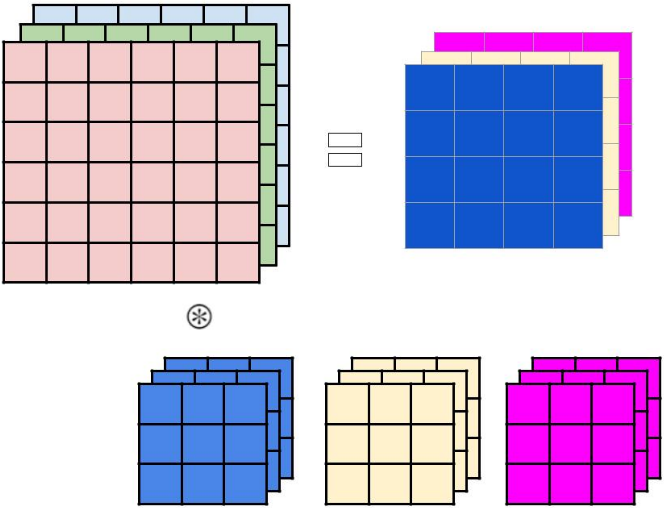
2D Convolution



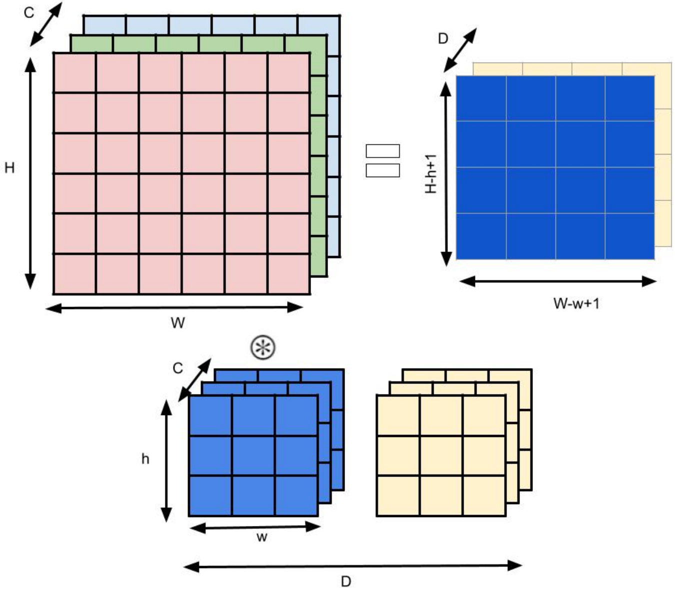
2D Convolution



2D Convolution



2D Convolution

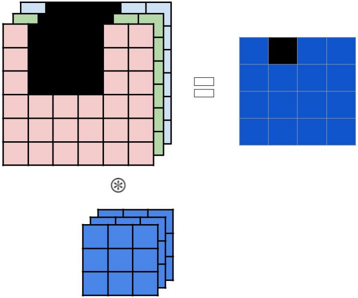


2D Convolution

1. Kernel is not convolved in the channel dimension

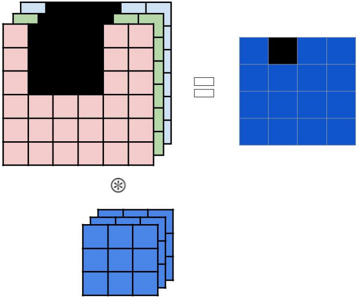
2D Convolution

1. Kernel is not convolved in the channel dimension
2. Another way to interpret convolution is that an affine function is applied on an input block of size $C \times h \times w$ and results in output of size $D \times 1 \times 1$



2D Convolution

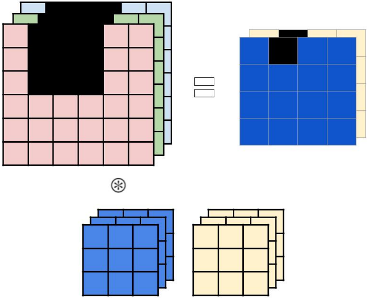
1. Kernel is not convolved in the channel dimension
2. Another way to interpret convolution is that an affine function is applied on an input block of size $C \times h \times w$ and results in output of size $D \times 1 \times 1$



3. Same affine function is applied on all such blocks in the input

2D Convolution

1. Kernel is not convolved in the channel dimension
2. Another way to interpret convolution is that an affine function is applied on an input block of size $C \times h \times w$ and results in output of size $D \times 1 \times 1$



3. Same affine function is applied on all such blocks in the input

Convolution

1. Preserves the input structure

Convolution

1. Preserves the input structure
   - 1D signal outputs 1D signal, 2D signal outputs 2D signal

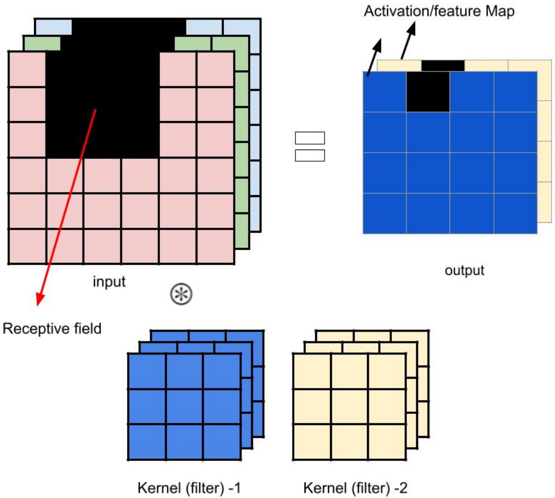
Convolution

1. Preserves the input structure
   - 1D signal outputs 1D signal, 2D signal outputs 2D signal
   - Adjacent components in o/p are influenced by adjacent parts in the i/p

Convolution

1. Preserves the input structure
   - 1D signal outputs 1D signal, 2D signal outputs 2D signal
   - Adjacent components in o/p are influenced by adjacent parts in the i/p
2. If the channel dimension has a metric meaning (e.g. time) 3D convolution can be employed (e.g. frames in a video)

Terminology in Convolution



Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`

Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`
2. `weight` is $D \times C \times h \times w$ dimensional kernels

Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`
2. `weight` is $D \times C \times h \times w$ dimensional kernels
3. `bias` D dimensional

Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`
2. `weight` is $D \times C \times h \times w$ dimensional kernels
3. `bias` D dimensional
4. `input` is $N \times C \times H \times W$ dimensional signal

Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`
2. `weight` is $D \times C \times h \times w$ dimensional kernels
3. `bias` D dimensional
4. `input` is $N \times C \times H \times W$ dimensional signal
5. Output is $N \times D \times (H - h + 1) \times (W - w + 1)$ tensor

Convolution function in PyTorch

1. `F.conv2d(input, weight, bias=None, stride=1, padding=0, dilation=1, groups=1)`
2. `weight` is $D \times C \times h \times w$ dimensional kernels
3. `bias` D dimensional
4. `input` is $N \times C \times H \times W$ dimensional signal
5. Output is $N \times D \times (H - h + 1) \times (W - w + 1)$ tensor
6. Autograd compliant

Convolution function in PyTorch

```
input = torch.empty(128, 3, 20, 20).normal_()
weight = torch.empty(5, 3, 5, 5).normal_()
bias = torch.empty(5).normal_()
output = F.conv2d(input, weight, bias)
output.size()
torch.Size([128, 5, 16, 16])
```

Look/Access the filters

```
weight[0,0]
tensor([[-0.6974, 0.1342, -0.2632, -0.4672, 0.1827],
[-0.1184, -0.2164, 0.2772, -0.1099, 0.0103],
[-0.8272, 0.3580, 0.2398, -0.5795,-0.9472],
[-1.1734, -0.1019, 0.7394, 0.3342, 0.1699],
[ 1.9271, 0.1250, 0.4222, 0.2014, 1.1100]])
```

Conv layer in PyTorch

1. `Class torch.nn.Conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True)`

Conv layer in PyTorch

1. Class `torch.nn.Conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True)`
2. `kernel_size` cane be either a pair `(h, w)` or a single value k interpreted as `(k, k)`.

Conv layer in PyTorch

1. Class `torch.nn.Conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True)`
2. `kernel_size` cane be either a pair `(h, w)` or a single value k interpreted as `(k, k)`.
3. Encloses the convolution as a module

Conv layer in PyTorch

1. `Class torch.nn.Conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True)`
2. `kernel_size` cane be either a pair `(h, w)` or a single value k interpreted as `(k, k)`.
3. Encloses the convolution as a module
4. Initializes the kernel parameters and biases as random

Conv layer in PyTorch

```
f = nn.Conv2d(in_channels = 3, out_channels = 5,
kernel_size = (2, 3))
for n, p in f.named_parameters():
...print(n, p.size())

weight torch.Size([5, 3, 2, 3])
bias torch.Size([5])
```

Padding in Convolution

1. Adds number of zeros around the input

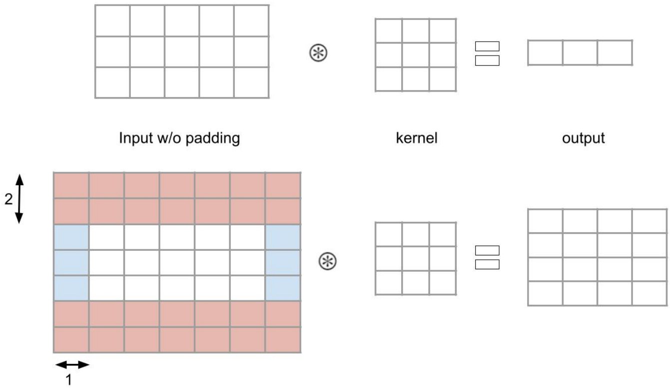
Padding in Convolution

1. Adds number of zeros around the input
2. Takes cares of size reduction after convolution

Padding in Convolution

1. Adds number of zeros around the input
2. Takes cares of size reduction after convolution
3. Instead of zeros, one may pad with signal values at the edges

Stride in Convolution



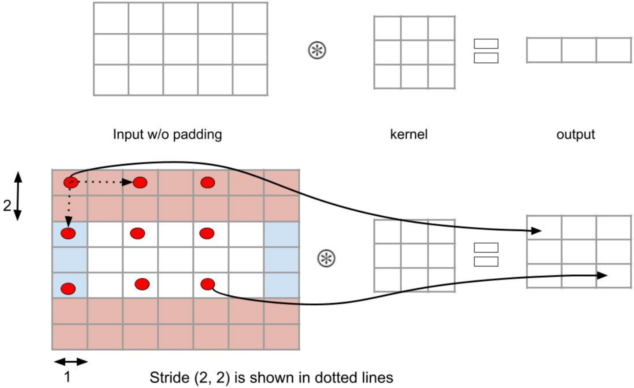
Stride in Convolution

1. Specifies the step size taken while performing convolution

Stride in Convolution

1. Specifies the step size taken while performing convolution
2. Default value is 1, i.e., move the kernel across the signal densely (without skipping)

Stride in Convolution



Stride (2, 2) is shown in dotted lines

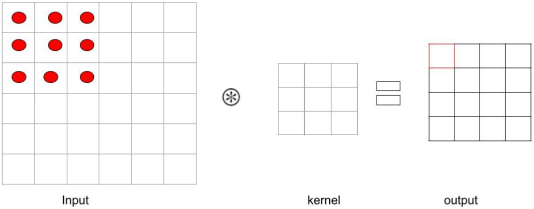
Dilation in Convolution

1. Manipulates the size of the kernel via expanding its size without adding weights.

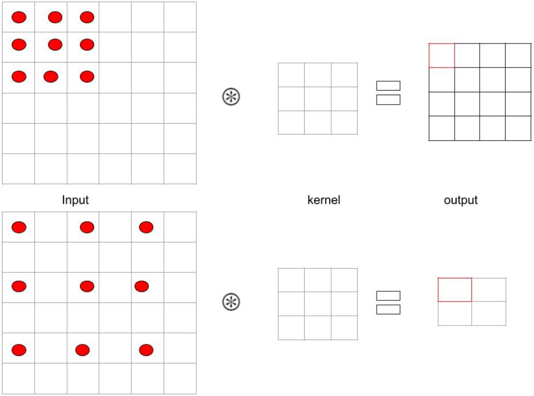
Dilation in Convolution

1. Manipulates the size of the kernel via expanding its size without adding weights.
2. In other words, it inserts 0s in between the kernel values

Without Dilation



Dilation (2, 2)



Dilation

1. Expands the kernel by adding rows and columns of zeros

Dilation

1. Expands the kernel by adding rows and columns of zeros
2. Default value for dilation is 1, i.e., no zeros placed

Dilation

1. Expands the kernel by adding rows and columns of zeros
2. Default value for dilation is 1, i.e., no zeros placed
3. Any higher value of dilation makes the kernel sparse

Dilation

1. Expands the kernel by adding rows and columns of zeros
2. Default value for dilation is 1, i.e., no zeros placed
3. Any higher value of dilation makes the kernel sparse
4. Dilation increases the receptive field

Dilation

1. Expands the kernel by adding rows and columns of zeros
2. Default value for dilation is 1, i.e., no zeros placed
3. Any higher value of dilation makes the kernel sparse
4. Dilation increases the receptive field
5. It is referred to as 'atrous' convolution

Title

1

Slide Title

1. Slide content