# 6 Convolutional Networks

Data comes in many different shapes. Reshaping this data into a table, as is done with images for training fully-connected networks, destroys the locality of the data. For example, when an image of a person's face is reshaped into a single row, the person's face will be sliced up and interlaced with non-face pixels destroying the locality of the pixels of the person's face. Convolutional networks do not require that data be reshaped into a single row, thus preserving the locality of the data.

92 Example (CIFAR-10 Dataset)
Load the CIFAR-10 Dataset `CIFAR-10_train.npz`.

(a) Each image is a tensor. What is the dimension (or rank) of each image tensor?
(b) The entire dataset is also a tensor. What is the dimension of the dataset tensor?
(c) How many images are contained in the dataset?
(d) What is the resolution in pixels of each image?
(e) How many color channels (or bands) does each image contain?
(f) Determine what each index of the dataset tensor represents.
(g) Display a representative image and corresponding label for each class in the dataset.

93 Example (Landsat-8 Satellite Images)
Landsat-8 Operational Land Imager and Thermal Infrared Sensor[9]

| Band Number | Description | Wavelength | Resolution |
|---|---|---|---|
| Band 1 | Coastal / Aerosol | 0.433 to 0.453 m | 30 meter |
| Band 2 | Visible blue | 0.450 to 0.515 m | 30 meter |
| Band 3 | Visible green | 0.525 to 0.600 m | 30 meter |
| Band 4 | Visible red | 0.630 to 0.680 m | 30 meter |
| Band 5 | Near-infrared | 0.845 to 0.885 m | 30 meter |
| Band 6 | Short wavelength infrared | 1.56 to 1.66 m | 30 meter |
| Band 7 | Short wavelength infrared | 2.10 to 2.30 m | 60 meter |
| Band 8 | Panchromatic | 0.50 to 0.68 m | 15 meter |
| Band 9 | Cirrus | 1.36 to 1.39 m | 30 meter |
| Band 10 | Long wavelength infrared | 10.3 to 11.3 m | 100 meter |
| Band 11 | Long wavelength infrared | 11.5 to 12.5 m | 100 meter |

[9] https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands/

Data can come as tensors of various shapes and dimensions. Tensors generalize the notion of scalars, vectors and matrices to higher dimensions. Tensors in `TensorFlow` are treated simply as n-dimensional arrays.

**Tensor Characteristics:**

**dimension (or rank)** number of indices needed to access the elements of a tensor.

**shape** dimensions of the tensor.

**data type** `float32`, `int32`, etc. All the elements of a tensor must have the same data type.

94 Example (Tensors)

| tensor | dimension | shape | data type |
|---|---|---|---|
| 5 | 0 | 1 | int32 |
| $\begin{pmatrix} 2.1 \\ 0.5 \\ 1.1 \end{pmatrix}$ | 1 | 3 | float32 |
| $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ | 2 | $3 \times 2$ | int32 |
| $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ $\begin{pmatrix} 3 & 5 \\ 0 & 1 \end{pmatrix}$ $\begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}$ | 3 | $2 \times 2 \times 3$ | int32 |

95 Example (Data With Special Structure)
Which of the following data sets have special structure?

(a) Home Prices Data Set: Features include number of rooms, home age, ranking of school district.

(b) Stock Prices Data Set: Daily stock prices for S&P 500 stocks from 2000-2010.

(c) RGP Color Images of Faces Data Set

---

Text is one-dimensional. Images are two dimensional (three with color). Video is three dimensional (four with color). Convolutions preserve the locality of information.

Dot products can be extended to tensors.

96 Example (Dot Product of Tensors)

(a) 1-D tensors of shape 3.

$$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} = 3$$

(b) 2-D tensors of shape $3 \times 2$.

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 5 & 1 \end{pmatrix} = 9$$

(c) 3-D tensors of shape $3 \times 2 \times 2$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \circ \begin{pmatrix} 0 & 0 \\ 3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} = 7$$

---

In comparison to fully-connected, feedforward networks, convolutional networks (ConvNets) have far fewer parameters that must be learned and involve much less computation. ConvNets preserve the locality of information inherent in text, images and video by performing only local dot products.

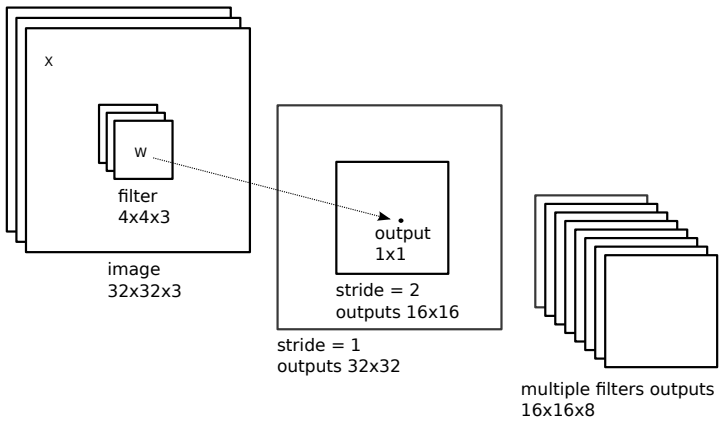


Figure 1: Observe the locality of the dot-products of the 3-D tensors shown in the diagram above.

**Key Hyper-Parameters of Convolutional Networks:**[10]

**patch size** dimensions of filter.

**stride** number of pixels to slide filter.

**padding** should the image be padded with zeros when filter is off of the edge of the image.

**channels** number of input layers.

**depth** number of output layers (filters).

97 Example (LeNet Convolutional Network)
LeNet[11] is on of the earliest convolutional networks. It was designed for OCR (optical character recognition) tasks. A block diagram of LeNet is

[10] Convolution Demo
http://cs231n.github.io/convolutional-networks/

[11] LeCun et al., Gradient-Based Learning Applied to Document Recognition, Proc. IEEE 86(11): 2278–2324, 1998.