Computer Vision

ACTL3143 & ACTL5111 Deep Learning for Actuaries

Patrick Laub

Lecture Outline

- **Images**
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

Shapes of data

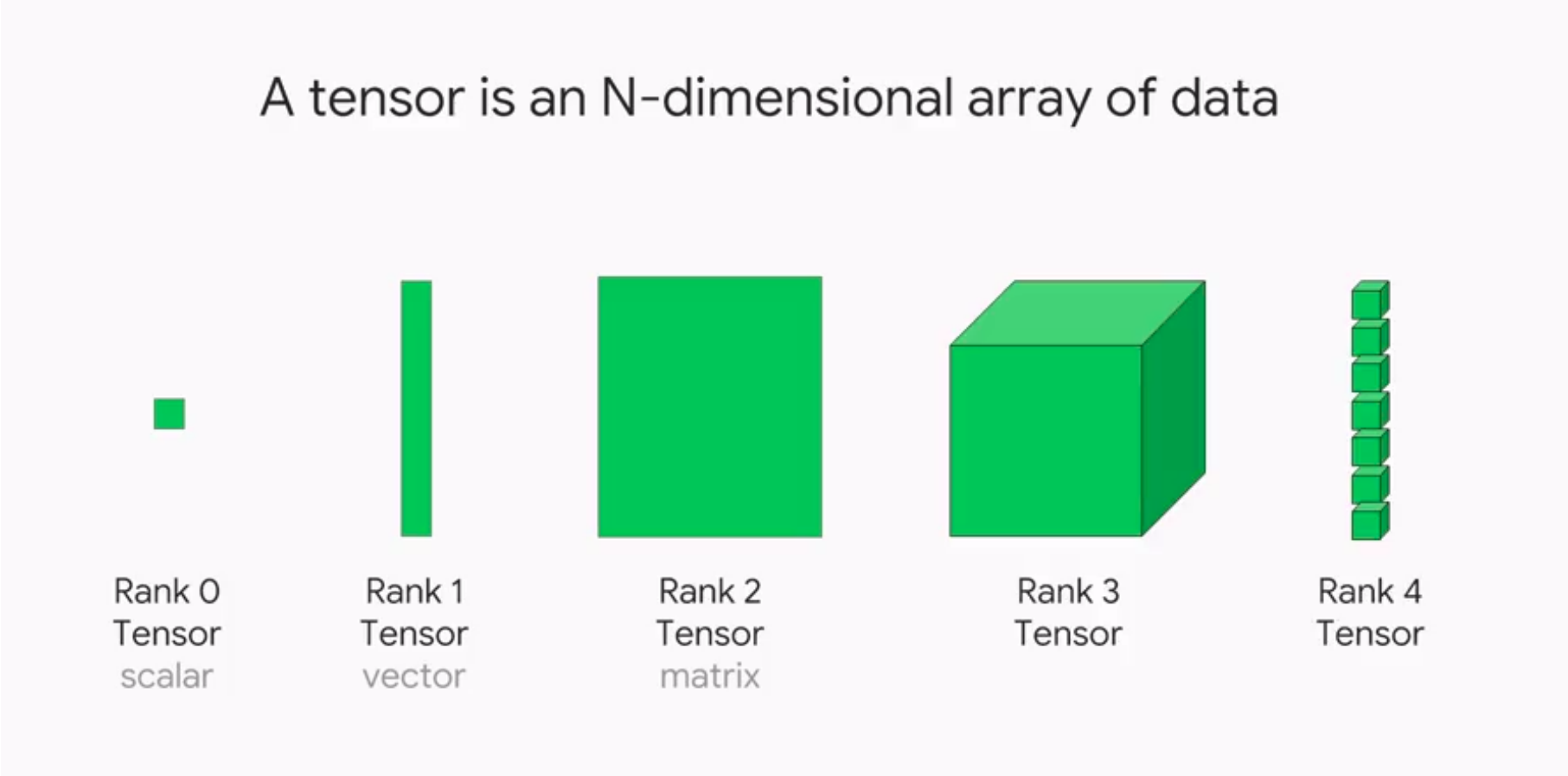
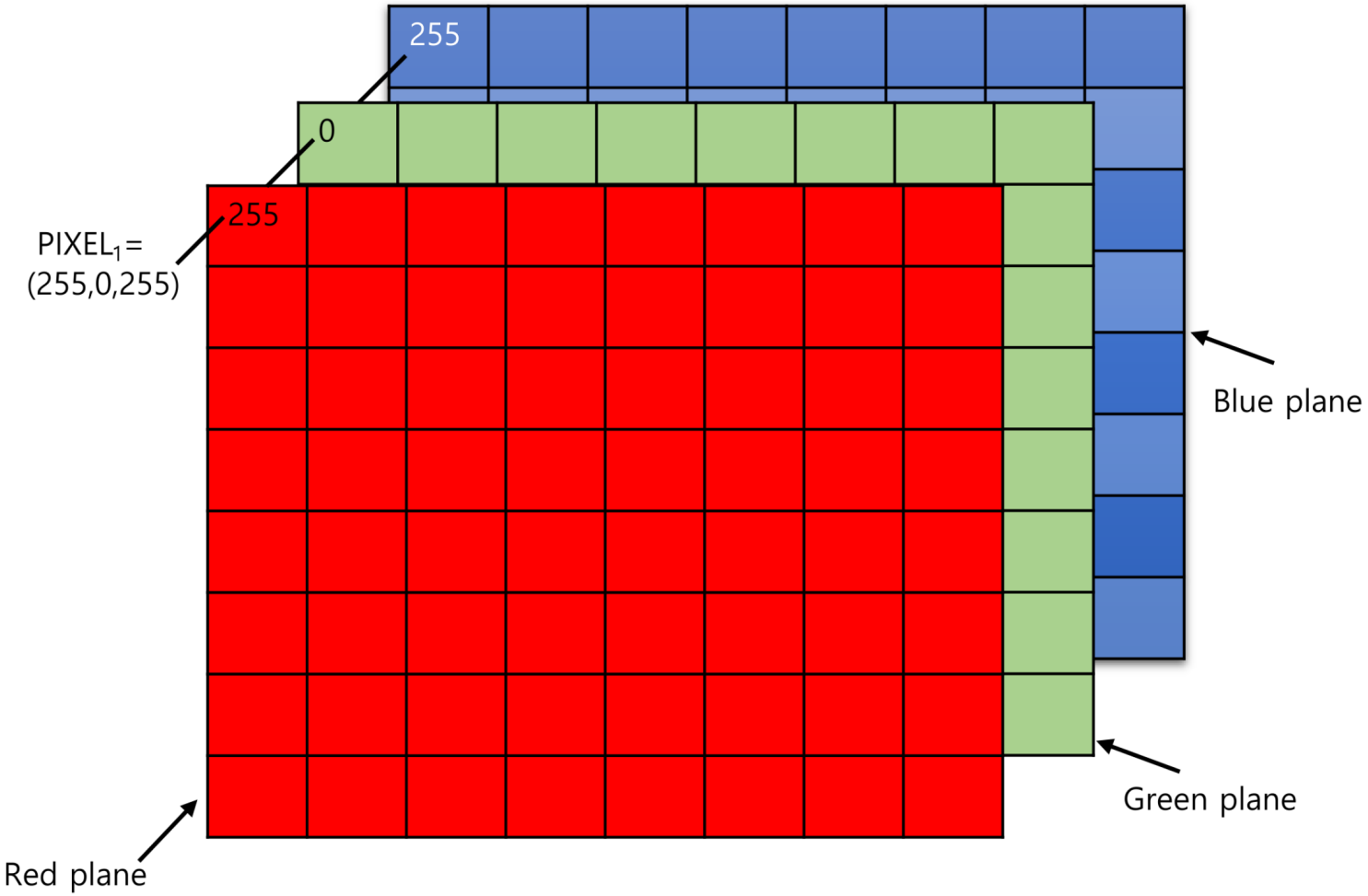


Illustration of tensors of different rank.

Source: Paras Patidar (2019), Tensors — Representation of Data In Neural Networks, Medium article.

Shapes of photos



A photo is a rank 3 tensor.

Source: Kim et al (2021), Data Hiding Method for Color AMBTC Compressed Images Using Color Difference, Applied Sciences.

How the computer sees them

```
1 from matplotlib.image import imread
2 img1 = imread('pu.gif'); img2 = imread('pl.gif')
3 img3 = imread('pr.gif'); img4 = imread('pg.bmp')
4 f"Shapes are: {img1.shape}, {img2.shape}, {img3.shape}, {img4.shape}."
```

```
'Shapes are: (16, 16, 3), (16, 16, 3), (16, 16, 3), (16, 16, 3).'
```

```
1 img1
```

```
array([[[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0]],

       [[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
```

```
1 img2
```

```
array([[[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0]],

       [[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
```

```
1 img3
```

```
array([[[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0]],

       [[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
        [255, 255,   0],
```

```
1 img4
```

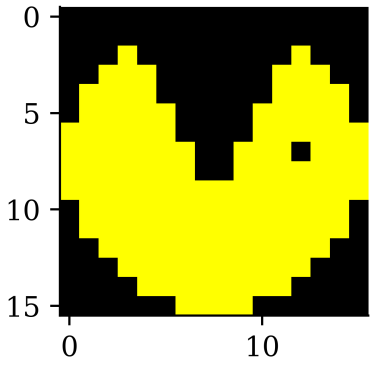
```
array([[[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0]],

       [[  0,   0,   0],
        [  0,   0,   0],
        [  0,   0,   0],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
        [255, 163, 177],
```

UNSW SYDNEY

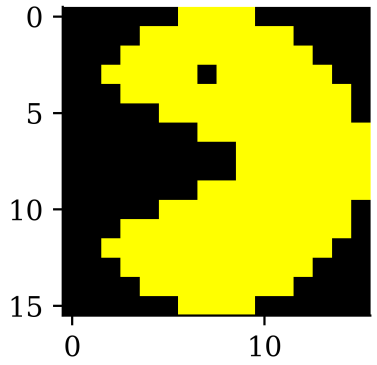
How we see them

```
1 from matplotlib.pyplot import imshow
```

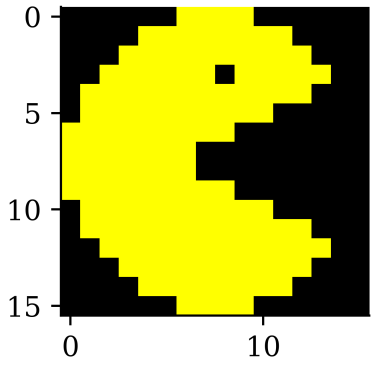
```
1 imshow(img1);
```



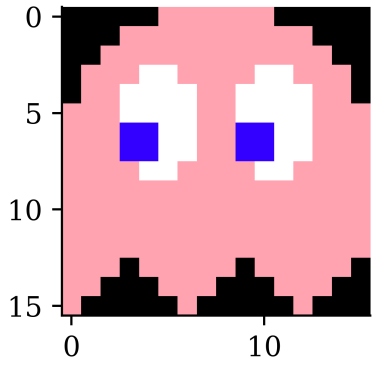

```
1 imshow(img2);
```




```
1 imshow(img3);
```




```
1 imshow(img4);
```



UNSW SYDNEY

Why is 255 special?

Each pixel's colour intensity is stored in one byte.

One byte is 8 bits, so in binary that is 00000000 to 11111111.

The largest *unsigned* number this can be is $2^8 - 1 = 255$.

```
1 np.array([0, 1, 255, 256]).astype(np.uint8)
```

```
array([  0,   1, 255,   0], dtype=uint8)
```

If you had *signed* numbers, this would go from -128 to 127.

```
1 np.array([-128, 1, 127, 128]).astype(np.int8)
```

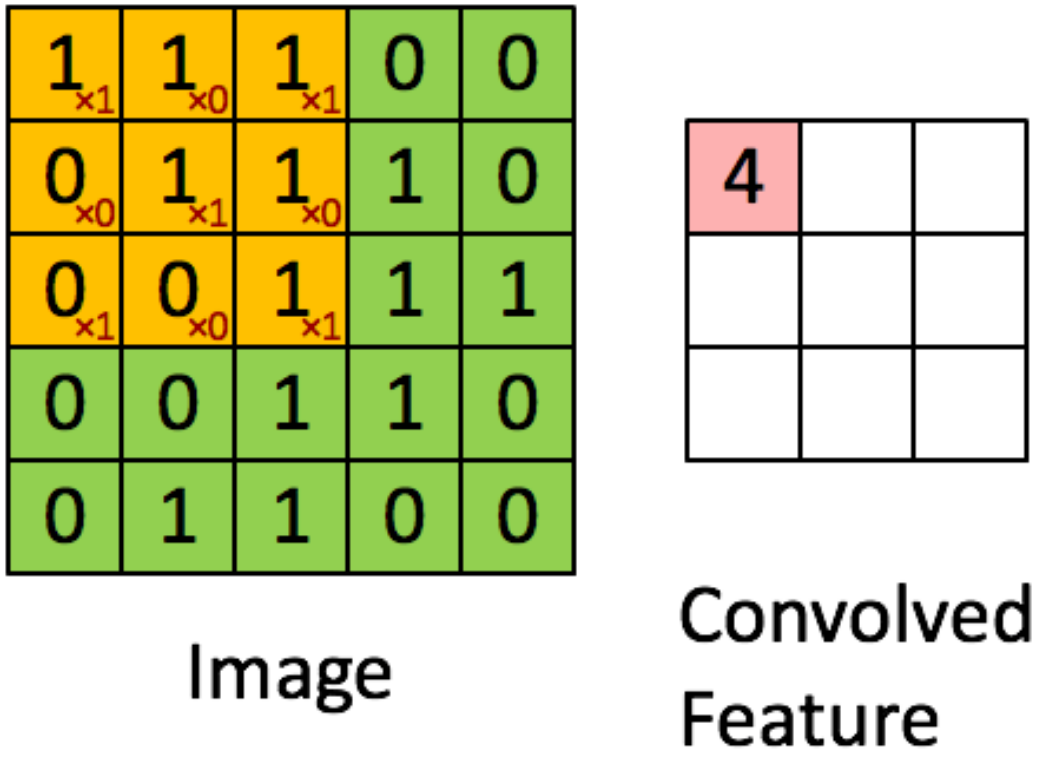
```
array([-128,    1,  127, -128], dtype=int8)
```

Alternatively, *hexidecimal* numbers are used. E.g. 10100001 is split into 1010 0001, and 1010=A, 0001=1, so combined it is 0xA1.

UNSW SYDNEY

Image editing with kernels

Take a look at https://setosa.io/ev/image-kernels/.



An example of an image kernel in action.

Source: Stanford's deep learning tutorial via Stack Exchange.

Lecture Outline

- Images
- **Convolutional Layers**
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

‘Convolution’ not ‘complicated’

Say $X_1, X_2 \sim f_X$ are i.i.d., and we look at $S = X_1 + X_2$.

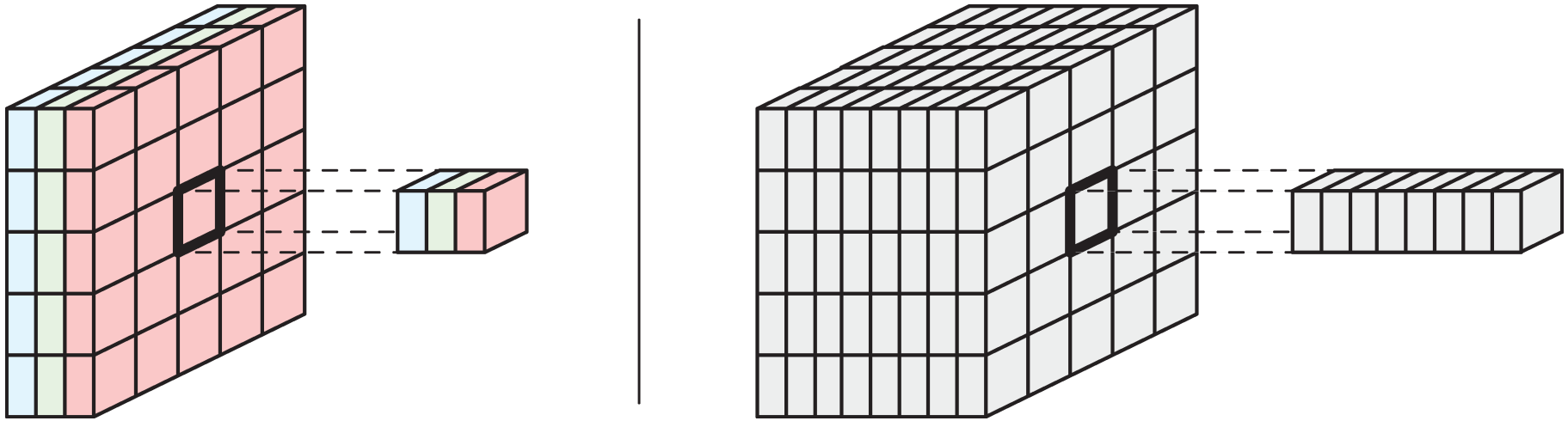
The density for S is then

$$f_S(s) = \int_{x_1=-\infty}^{\infty} f_X(x_1)\, f_X(s - x_1)\, \mathrm{d}s.$$

This is the *convolution* operation, $f_S = f_X \star f_X$.

Images are rank 3 tensors

Height, width, and number of channels.



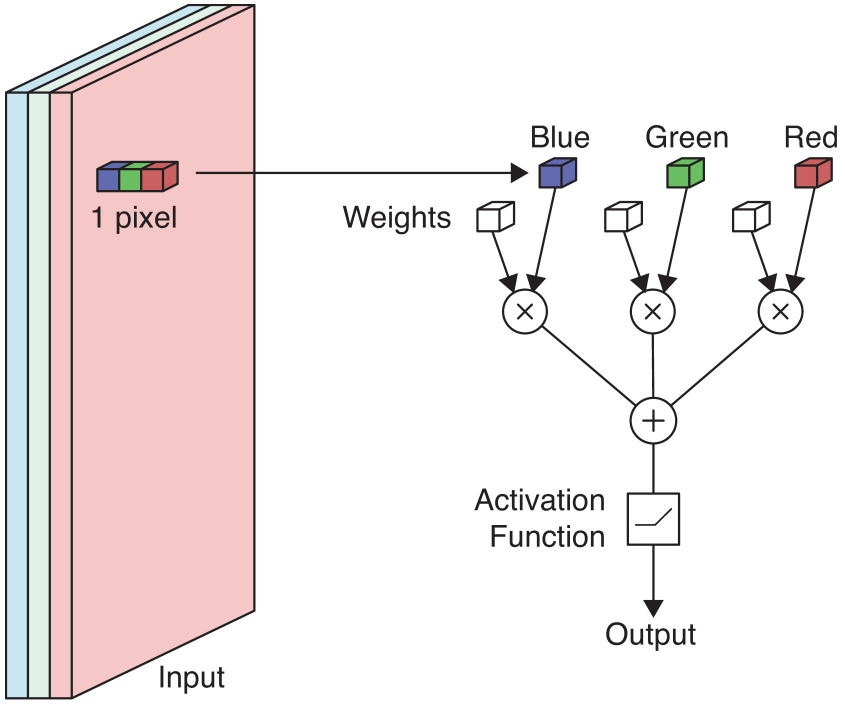
Examples of rank 3 tensors.

Grayscale image has 1 channel. RGB image has 3 channels.

Example: Yellow = Red + Green.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Example: Detecting yellow



Applying a neuron to an image pixel.

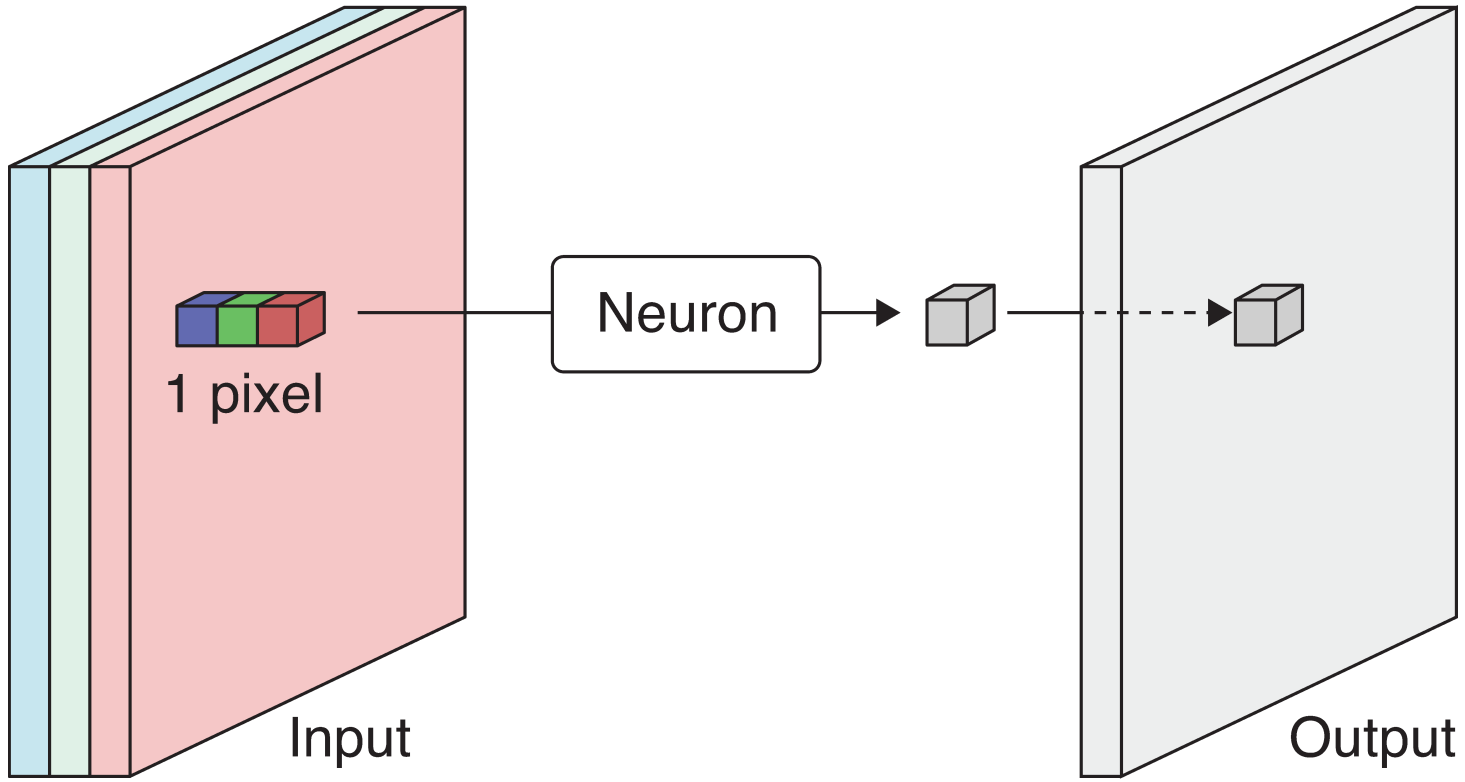
Apply a neuron to each pixel in the image.

If red/green ↗ or blue ↘ then yellowness ↗.

Set RGB weights to 1, 1, −1.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Example: Detecting yellow II



Scan the 3-channel input (colour image) with the neuron to produce a 1-channel output (grayscale image).

The output is produced by *sweeping* the neuron over the input. This is called **convolution**.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Example: Detecting yellow III

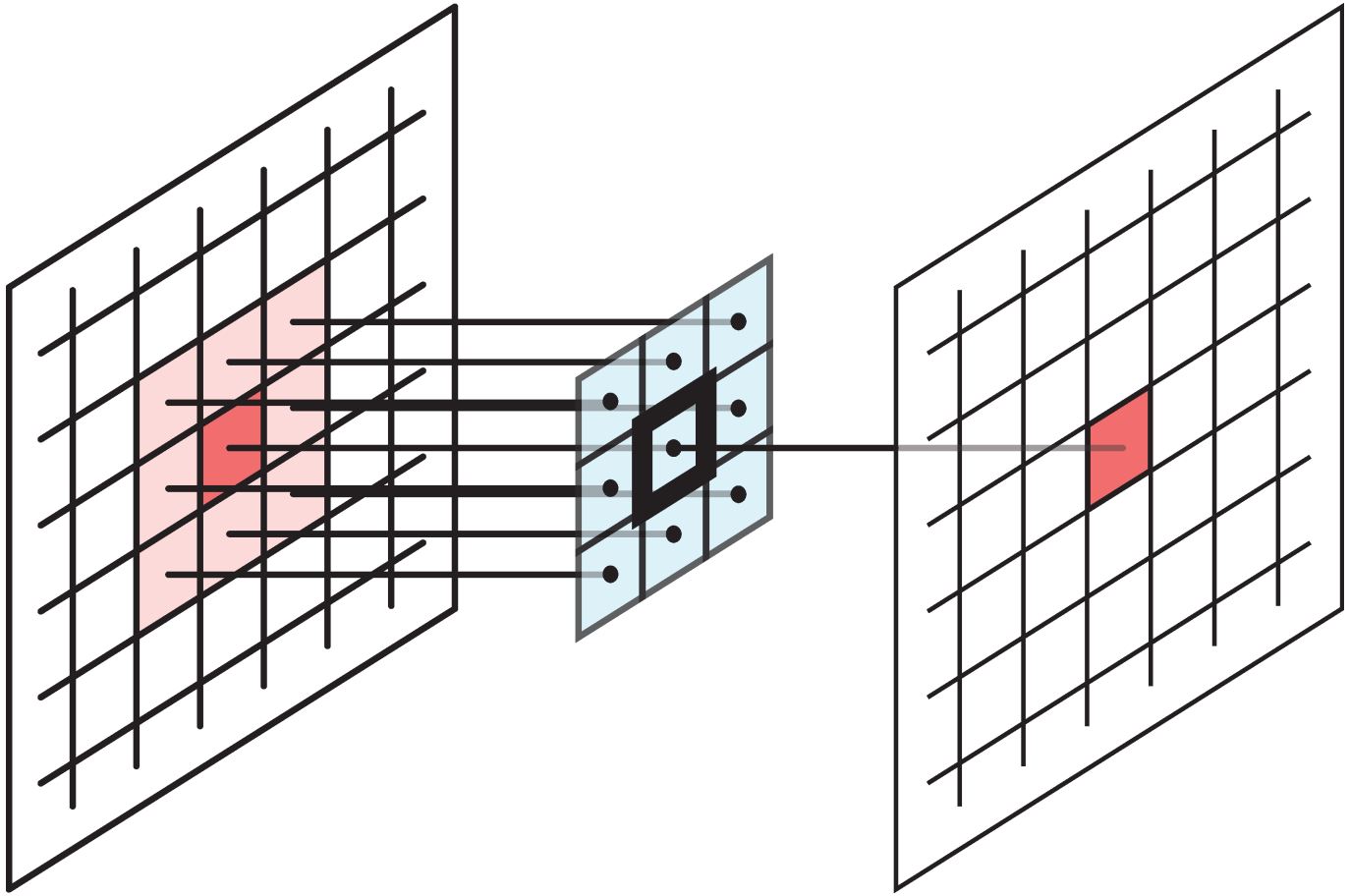
The more yellow the pixel in the colour image (left), the more white it is in the grayscale image.

The neuron or its weights is called a **filter**. We *convolve* the image with a filter, i.e. a **convolutional filter**.

Terminology

- The same neuron is used to sweep over the image, so we can store the weights in some shared memory and process the pixels in parallel. We say that the neurons are *weight sharing*.
- In the previous example, the neuron only takes one pixel as input. Usually a larger filter containing a *block of weights* is used to process not only a pixel but also its neighboring pixels all at once.
- The weights are called the filter **kernels**.
- The cluster of pixels that forms the input of a filter is called its *footprint*.

Spatial filter



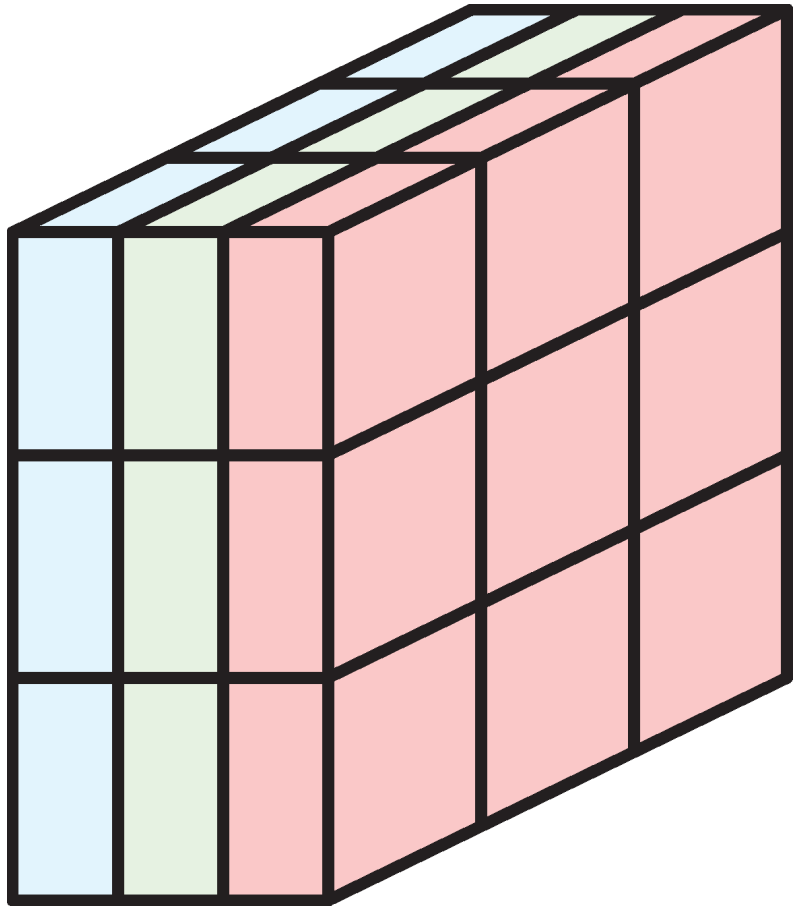
Example 3x3 filter

When a filter's footprint is > 1 pixel, it is a **spatial filter**.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Multidimensional convolution

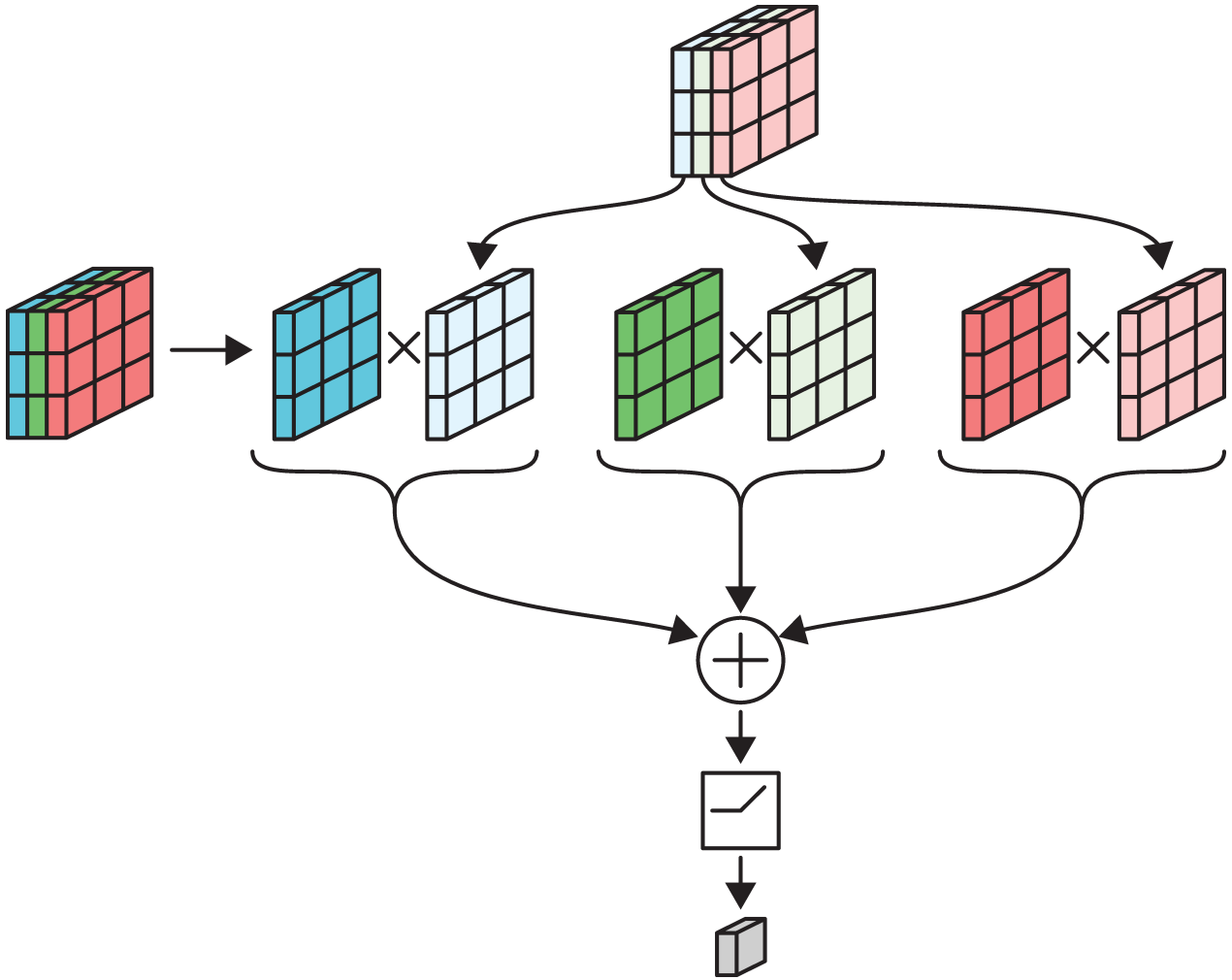
Need # Channels in Input = # Channels in Filter.



Example: a 3x3 filter with 3 channels, containing 27 weights.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

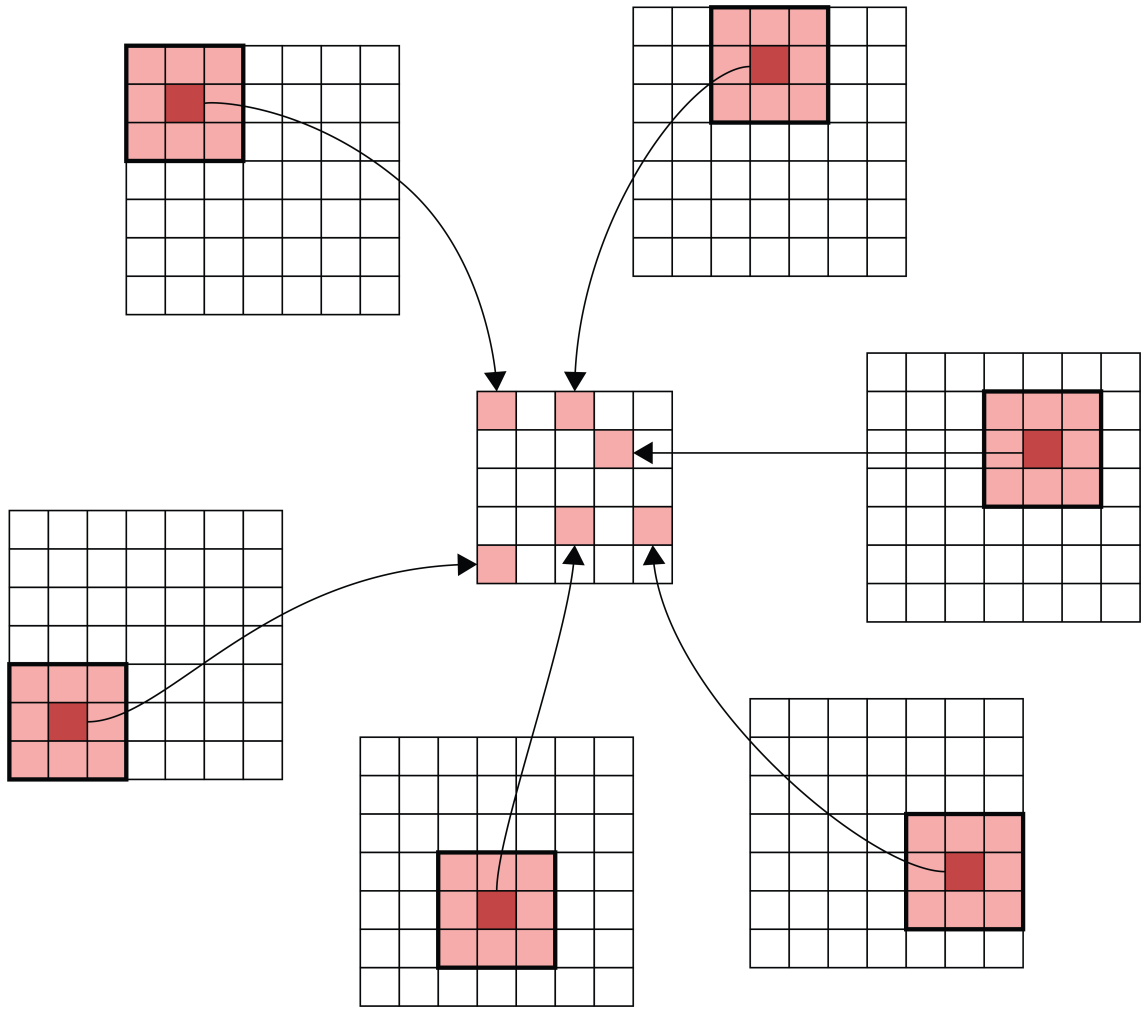
Example: 3x3 filter over RGB input



Each channel is multipled separately & then added together.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Input-output relationship



Matching the original image footprints against the output location.

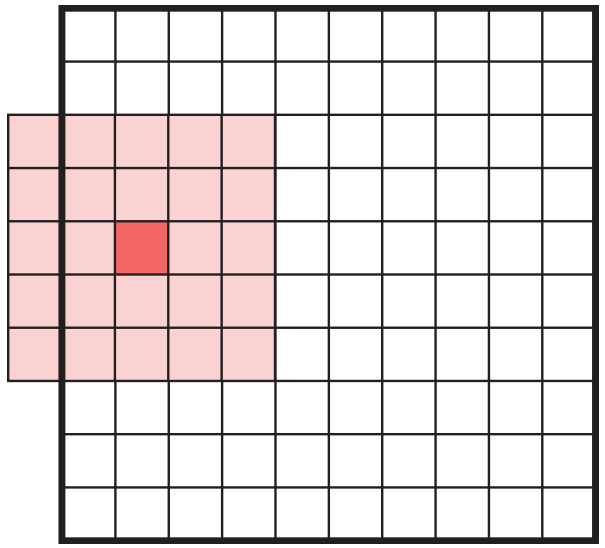
Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

UNSW SYDNEY

Lecture Outline

- Images
- Convolutional Layers
- **Convolutional Layer Options**
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

Padding



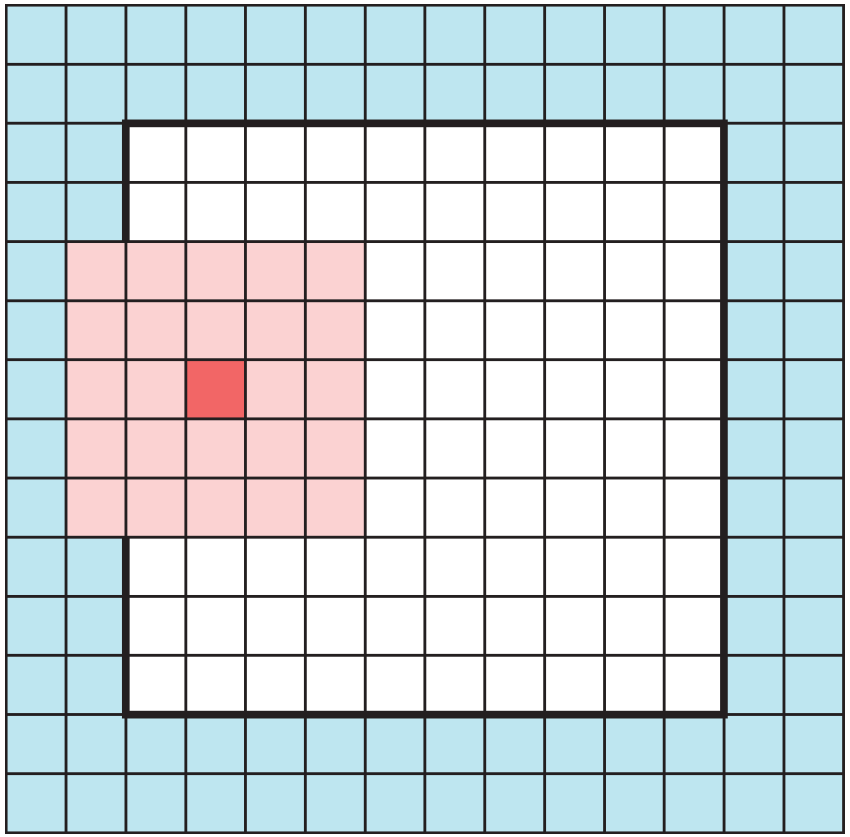
What happens when filters go off the edge of the input?

- How to avoid the filter's receptive field falling off the side of the input.
- If we only scan the filter over places of the input where the filter can fit perfectly, it will lead to loss of information, especially after many filters.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Padding

Add a border of extra elements around the input, called **padding**.
Normally we place zeros in all the new elements, called **zero padding**.



Padded values can be added to the outside of the input.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

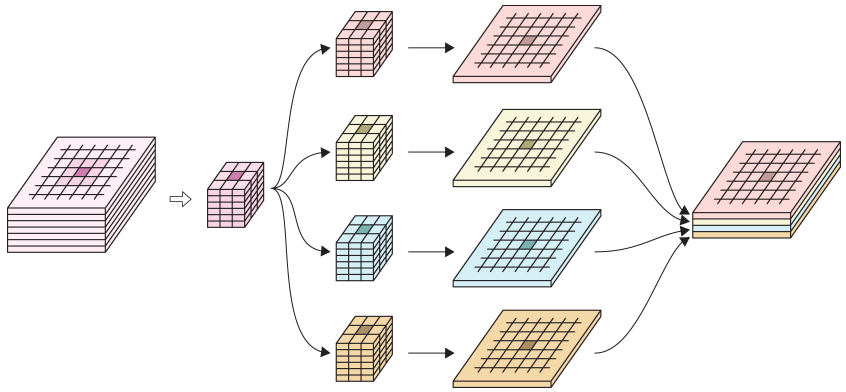
Convolution layer

- Multiple filters are bundled together in one layer.
- The filters are applied *simultaneously* and *independently* to the input.
- Filters can have different footprints, but in practice we almost always use the same footprint for every filter in a convolution layer.
- Number of channels in the output will be the same as the number of filters.

Example

In the image:

- 6-channel input tensor
- input pixels
- four 3x3 filters
- four output tensors
- final output tensor.



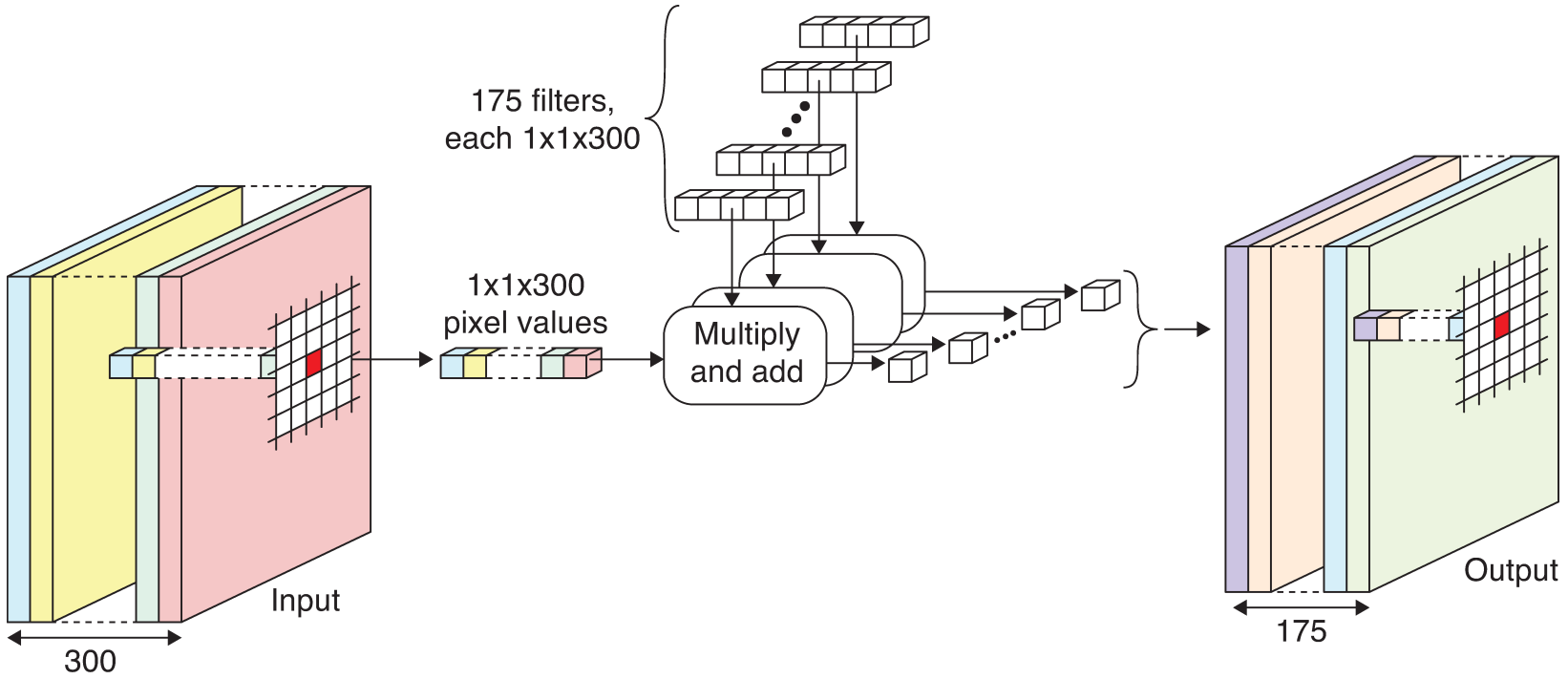
Example network highlighting that the number of output channels equals the number of filters.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

1x1 convolution

- Feature reduction: Reduce the number of channels in the input tensor (removing correlated features) by using fewer filters than the number of channels in the input. This is because the number of channels in the output is always the same as number of filters.
- 1x1 convolution: Convolution using 1x1 filters.
- When the channels are correlated, 1x1 convolution is very effective at reducing channels without loss of information.

Example of 1x1 convolution



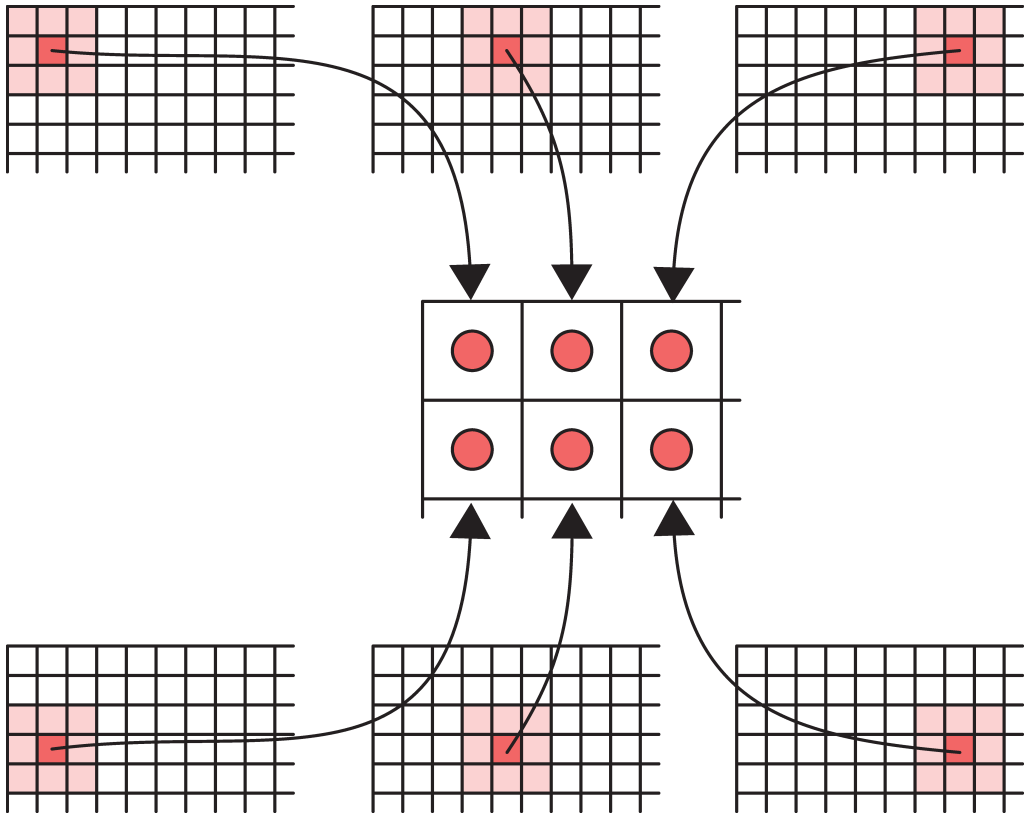
Example network with 1x1 convolution.

- Input tensor contains 300 channels.
- Use 175 1x1 filters in the convolution layer (300 weights each).
- Each filter produces a 1-channel output.
- Final output tensor has 175 channels.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Striding

We don't have to go one pixel across/down at a time.

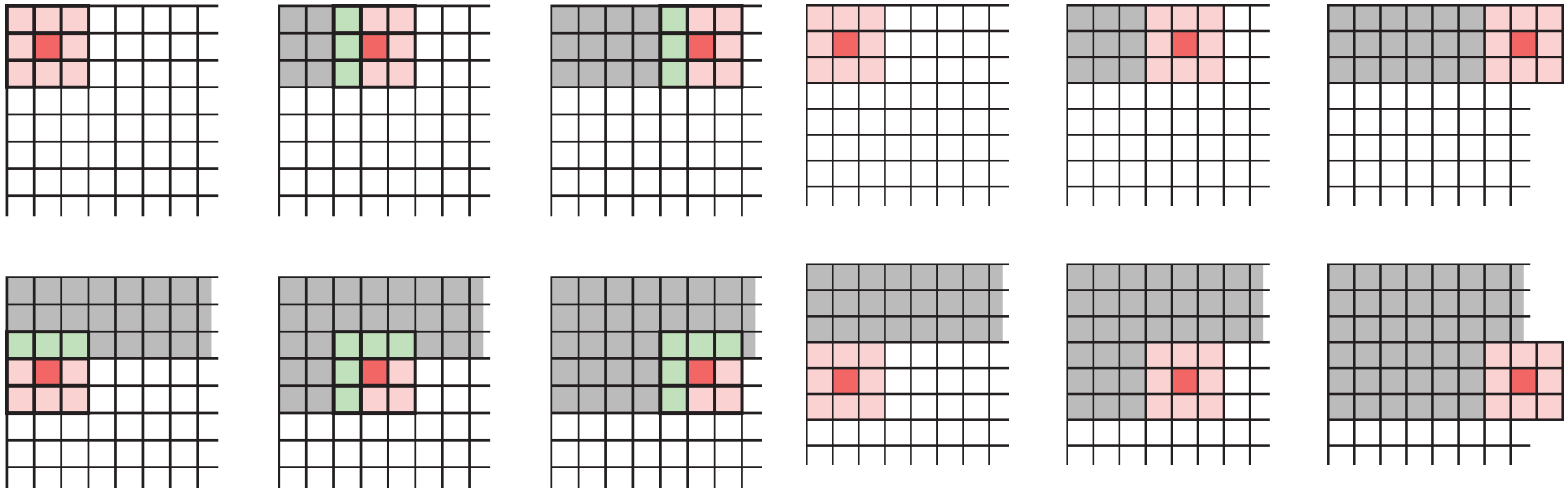


Example: Use a stride of three horizontally and two vertically.

Dimension of output will be smaller than input.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Choosing strides



When a filter scans the input step by step, it processes the same input elements multiple times. Even with larger strides, this can still happen (left image).

If we want to save time, we can choose strides that prevents input elements from being used more than once. Example (right image): 3x3 filter, stride 3 in both directions.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Specifying a convolutional layer

Need to choose:

- number of filters,
- their footprints (e.g. 3x3, 5x5, etc.),
- activation functions,
- padding & striding (optional).

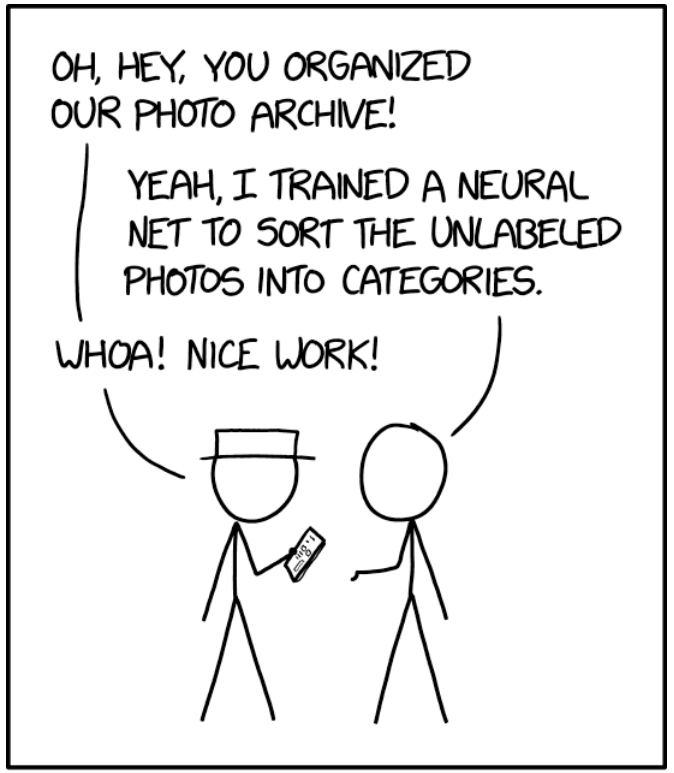
All the filter weights are learned during training.

Lecture Outline

- Images
- Convolutional Layers
- Convolutional Layer Options
- **Convolutional Neural Networks**
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

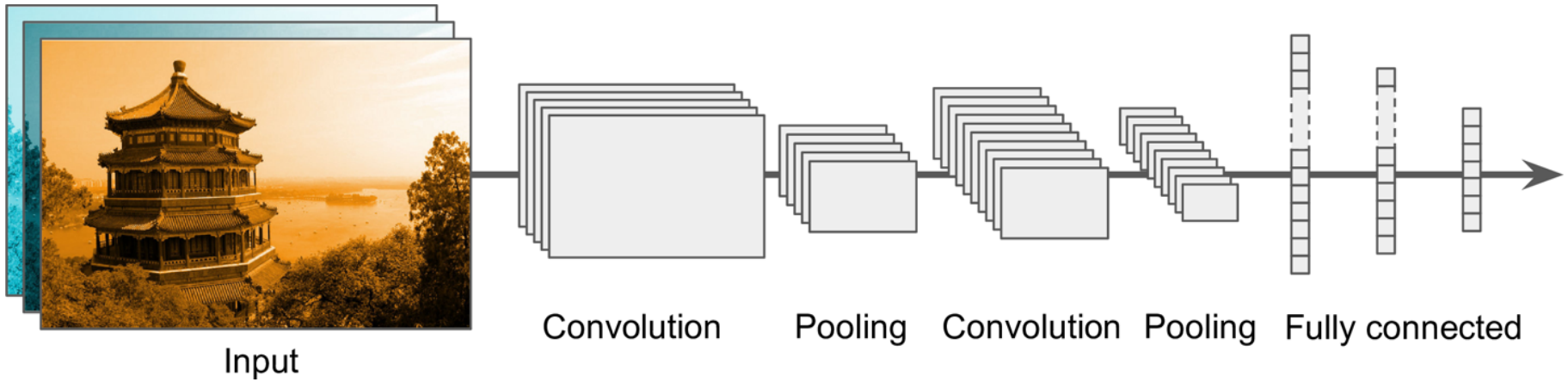
Definition of CNN

A neural network that uses *convolution layers* is called a *convolutional neural network*.



Source: Randall Munroe (2019), xkcd #2173: Trained a Neural Net.

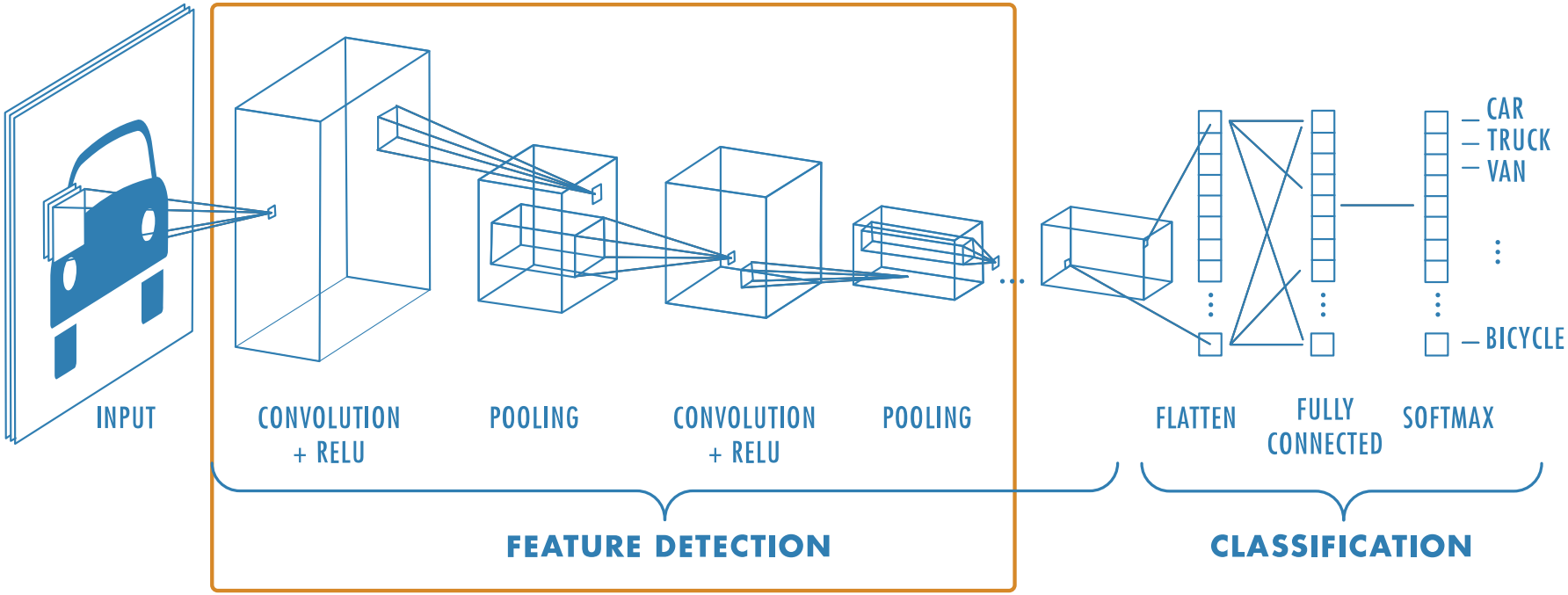
Architecture



Typical CNN architecture.

Source: Aurélien Géron (2019), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, Figure 14-11.

Architecture #2



Source: MathWorks, *Introducing Deep Learning with MATLAB*, Ebook.

Pooling

Pooling, or **downsampling**, is a technique to blur a tensor.

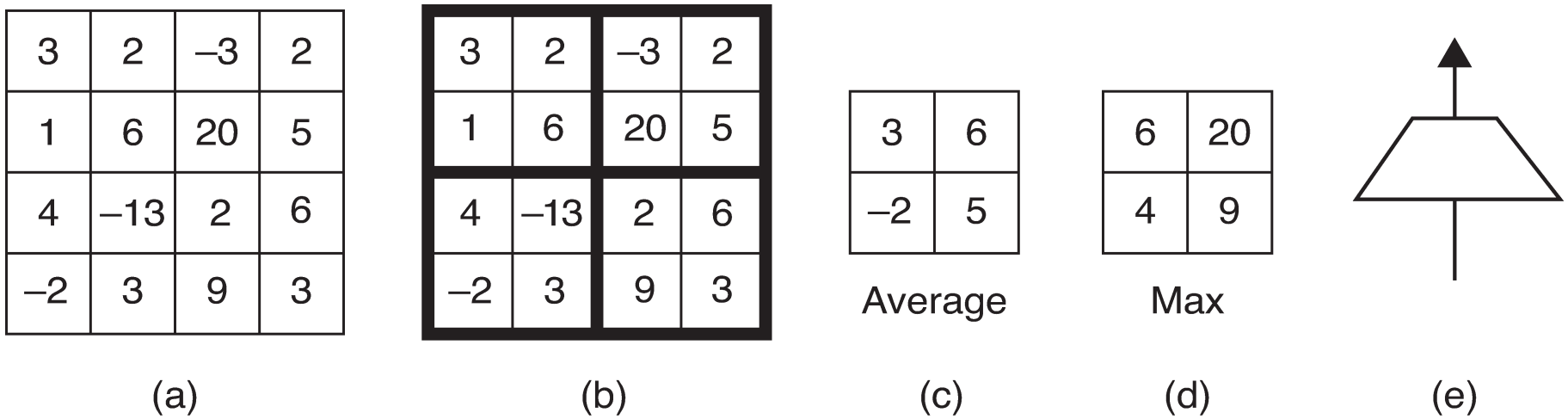
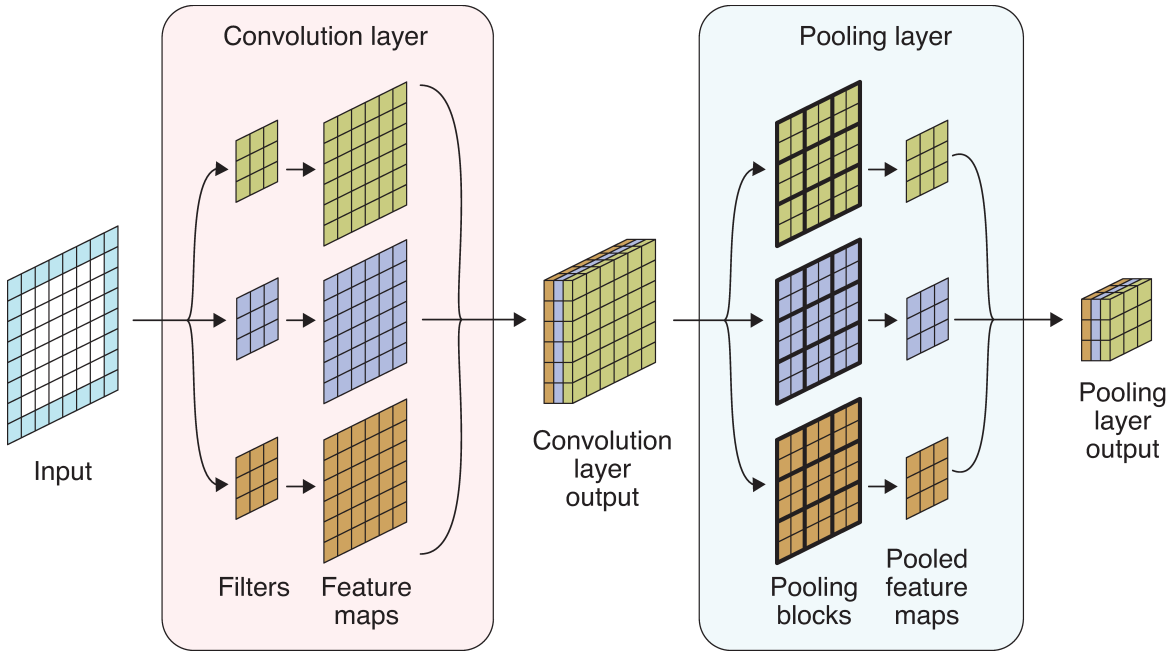


Illustration of pool operations.

(a): Input tensor (b): Subdivide input tensor into 2x2 blocks (c): Average pooling (d): Max pooling (e): Icon for a pooling layer

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Pooling for multiple channels



Pooling a multichannel input.

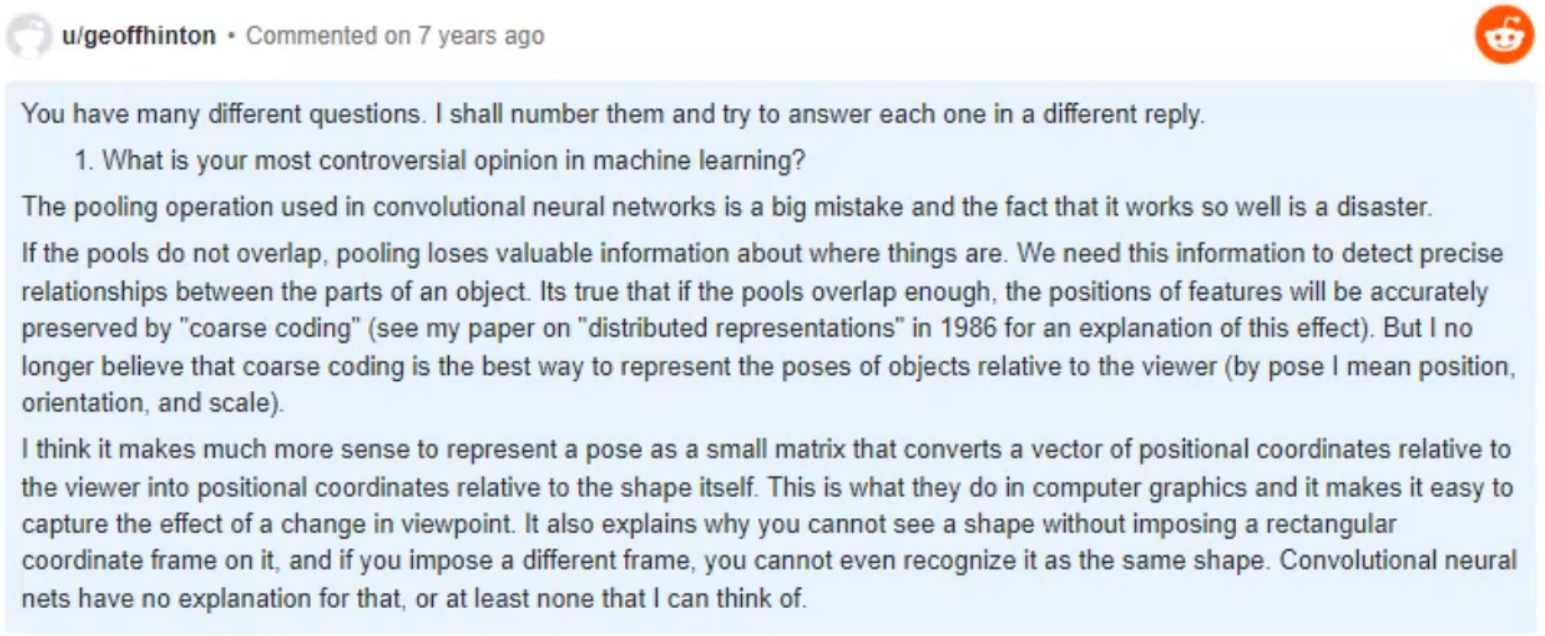
- Input tensor: 6x6 with 1 channel, zero padding.
- Convolution layer: Three 3x3 filters.
- Convolution layer output: 6x6 with 3 channels.
- Pooling layer: apply max pooling to each channel.
- Pooling layer output: 3x3, 3 channels.

Source: Glassner (2021), *Deep Learning: A Visual Approach*, Chapter 16.

Why/why not use pooling?

Why? Pooling *reduces the size* of tensors, therefore reduces memory usage and execution time (recall that 1x1 convolution *reduces the number of channels* in a tensor).

Why not?



u/geoffhinton • Commented on 7 years ago

You have many different questions. I shall number them and try to answer each one in a different reply.

1. What is your most controversial opinion in machine learning?

The pooling operation used in convolutional neural networks is a big mistake and the fact that it works so well is a disaster.

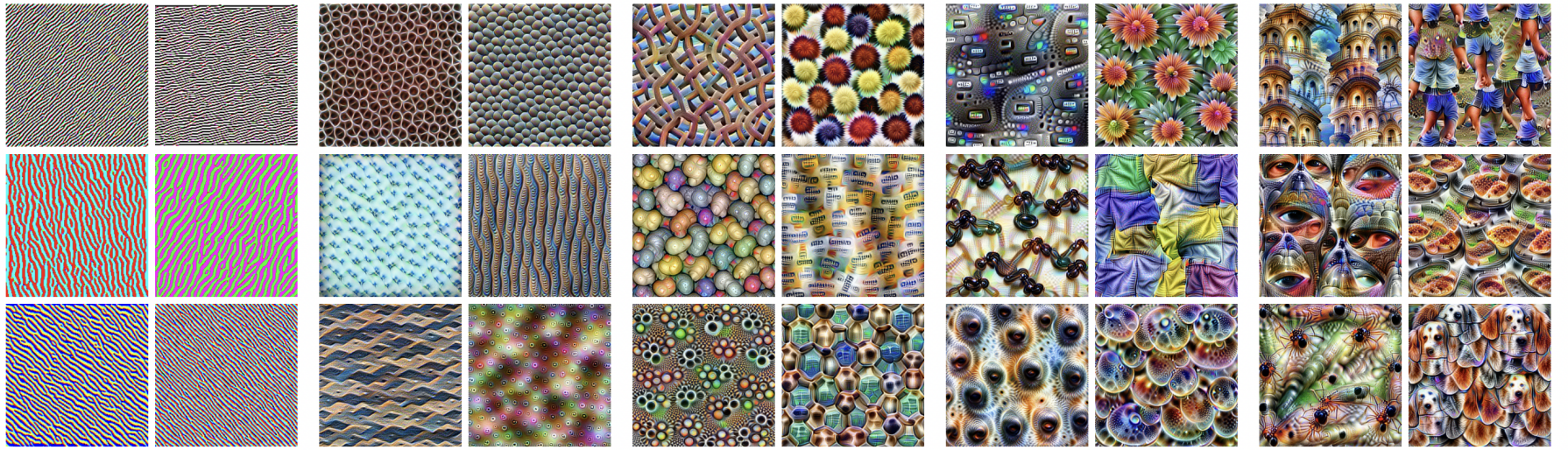
If the pools do not overlap, pooling loses valuable information about where things are. We need this information to detect precise relationships between the parts of an object. Its true that if the pools overlap enough, the positions of features will be accurately preserved by "coarse coding" (see my paper on "distributed representations" in 1986 for an explanation of this effect). But I no longer believe that coarse coding is the best way to represent the poses of objects relative to the viewer (by pose I mean position, orientation, and scale).

I think it makes much more sense to represent a pose as a small matrix that converts a vector of positional coordinates relative to the viewer into positional coordinates relative to the shape itself. This is what they do in computer graphics and it makes it easy to capture the effect of a change in viewpoint. It also explains why you cannot see a shape without imposing a rectangular coordinate frame on it, and if you impose a different frame, you cannot even recognize it as the same shape. Convolutional neural nets have no explanation for that, or at least none that I can think of.

Geoffrey Hinton

Source: Hinton, Reddit AMA.

What do the CNN layers learn?



Edges (layer conv2d0) **Textures** (layer mixed3a) **Patterns** (layer mixed4a) **Parts** (layers mixed4b & mixed4c) **Objects** (layers mixed4d & mixed4e)

Source: Distill article, Feature Visualization.

Lecture Outline

- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- **Demo: Character Recognition**
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

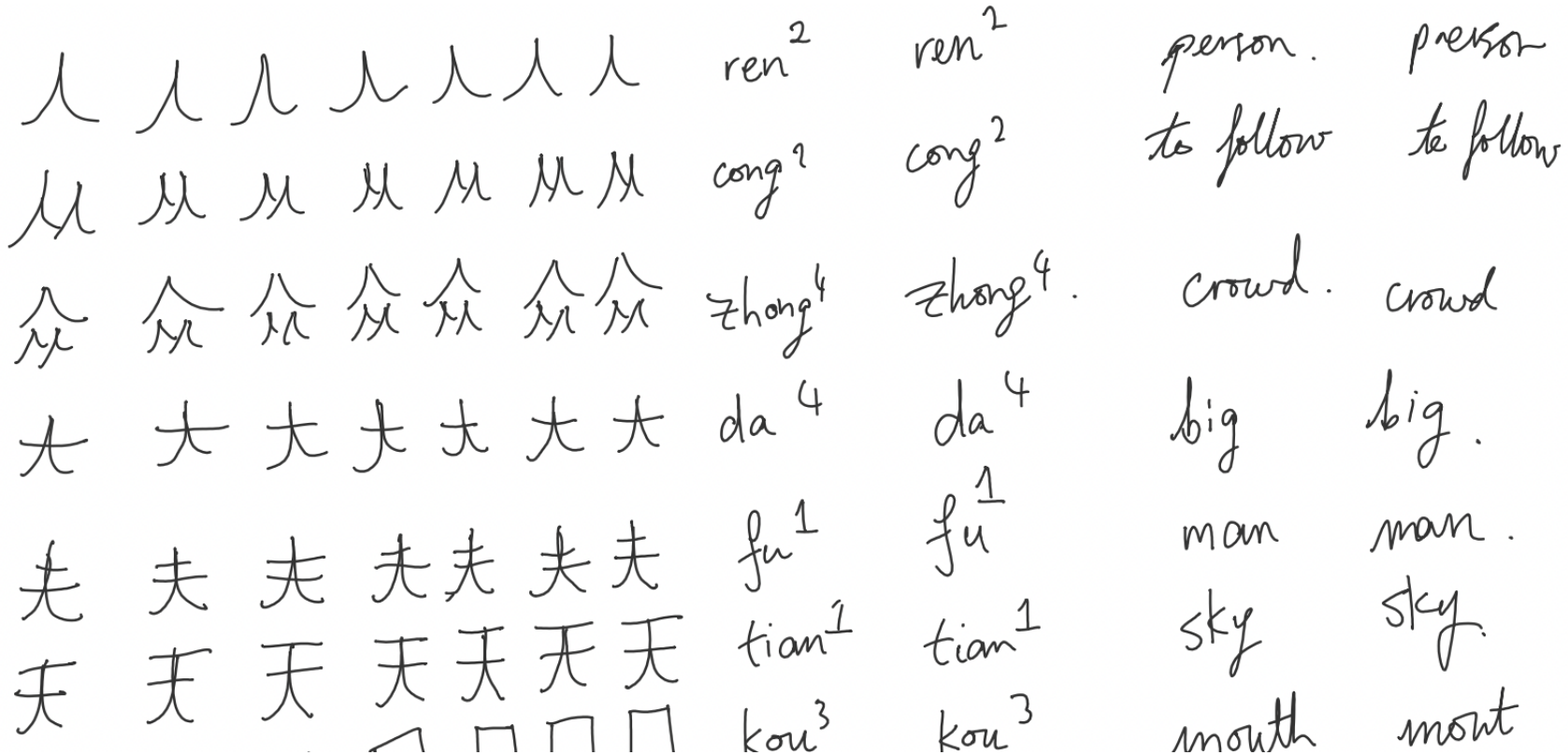
MNIST Dataset

The MNIST dataset.

Source: Wikipedia, MNIST database.

Mandarin Characters Dataset

57 poorly written Mandarin characters ($57 \times 7 = 399$).



Dataset of notes when learning/practising basic characters.

Downloading the dataset

The data is zipped (6.9 MB) and stored on my GitHub homepage.

```
# Download the dataset if it hasn't already been downloaded.
from pathlib import Path
if not Path("mandarin").exists():
  print("Downloading dataset ... ")
  !wget https://laub.au/data/mandarin.zip
  !unzip mandarin.zip
else:
  print("Already downloaded.")
```

```
Already downloaded.
```

Tip

Remember, the Jupyter notebook associated with your final report should either download your dataset when it is run, or you should supply the data separately.

Directory structure

Inspect directory structure

```
!pip install directory_tree
```

```
from directory_tree import display_tree
display_tree("mandarin")
```

```
mandarin/
├── bai/
│   ├── bai-1.png
│   ├── bai-2.png
│   ├── bai-3.png
│   ├── bai-4.png
│   ├── bai-5.png
│   ├── bai-6.png
│   └── bai-7.png
├── ben/
│   ├── ben-1.png
│   ├── ben-2.png
│   ├── ben-3.png
│   ├── ben-4.png
│   ├── ben-5.png
│   ├── ben-6.png
│   └── ben-7.png
├── chong/
│   ├── chong-1.png
```

```
tree = display_tree("mandarin", string_
print("\n".join(tree[:12]))
print(" ... ")
print("\n".join(tree[-4:]))
```

```
mandarin/
├── bai/
│   ├── bai-1.png
│   ├── bai-2.png
│   ├── bai-3.png
│   ├── bai-4.png
│   ├── bai-5.png
│   ├── bai-6.png
│   └── bai-7.png
├── ben/
│   ├── ben-1.png
│   ├── ben-2.png
 ...
│   ├── zhuo-5.png
│   ├── zhuo-6.png
│   └── zhuo-7.png
```

UNSW SYDNEY

Splitting into train/val/test sets

```
!pip install split-folders
```

```
import splitfolders
splitfolders.ratio("mandarin", output="mandarin-split",
    seed=1337, ratio=(5/7, 1/7, 1/7))

display_tree("mandarin-split", max_depth=1)
```

```
mandarin-split/
├── test/
├── train/
└── val/
```

Directory structure II

```
1 display_tree("mandarin-split")
```

```
mandarin-split/
├── test/
│   ├── bai/
│   │   └── bai-5.png
│   ├── ben/
│   │   └── ben-5.png
│   ├── chong/
│   │   └── chong-5.png
│   ├── chu/
│   │   └── chu-5.png
│   ├── chuan/
│   │   └── chuan-5.png
│   ├── cong/
│   │   └── cong-5.png
│   ├── da/
│   │   └── da-5.png
│   ├── dan/
│   │   └── dan-5.png
│   ├── dong/
```

```
train/
├── bai/
│   ├── bai-1.png
│   ├── bai-2.png
│   ├── bai-3.png
│   ├── bai-4.png
│   └── bai-6.png
...
val/
├── bai/
│   └── bai-7.png
├── ben/
│   └── ben-7.png
...
test/
├── bai/
│   └── bai-5.png
├── ben/
│   └── ben-5.png
```

UNSW SYDNEY

Keras image dataset loading

```
from keras.utils import\
  image_dataset_from_directory

data_dir = "mandarin-split"
batch_size = 32
img_height = 80
img_width = 80
img_size = (img_height, img_width)
```

```
train_ds = image_dataset_from_directory
    data_dir + "/train",
    image_size=img_size,
    batch_size=batch_size,
    shuffle=False,
    color_mode='grayscale')
```

```
val_ds = image_dataset_from_directory(
    data_dir + "/val",
    image_size=img_size,
    batch_size=batch_size,
    shuffle=False,
    color_mode='grayscale')
```

```
test_ds = image_dataset_from_directory(
    data_dir + "/test",
    image_size=img_size,
    batch_size=batch_size,
    shuffle=False,
    color_mode='grayscale')
```

UNSW SYDNEY

Inspecting the datasets

```
print(train_ds.class_names)
```

```
['bai', 'ben', 'chong', 'chu', 'chuan', 'cong', 'da', 'dan', 'dong', 'fei', 'fu', 'fu2',
'gao', 'gong', 'guo', 'hu', 'huo', 'kou', 'ku', 'lin', 'ma', 'ma2', 'ma3', 'mei', 'men',
'ming', 'mu', 'nan', 'niao', 'niu', 'nu', 'nuan', 'peng', 'quan', 'ren', 'ri', 'rou', 'sen',
'shan', 'shan2', 'shui', 'tai', 'tian', 'wang', 'wen', 'xian', 'xuan', 'yan', 'yang', 'yin',
'yu', 'yu2', 'yue', 'zhong', 'zhu', 'zhu2', 'zhuo']
```

```
# NB: Need shuffle=False earlier for these X & y to line up.
X_train = np.concatenate(list(train_ds.map(lambda x, y: x)))
y_train = np.concatenate(list(train_ds.map(lambda x, y: y)))

X_val = np.concatenate(list(val_ds.map(lambda x, y: x)))
y_val = np.concatenate(list(val_ds.map(lambda x, y: y)))

X_test = np.concatenate(list(test_ds.map(lambda x, y: x)))
y_test = np.concatenate(list(test_ds.map(lambda x, y: y)))

X_train.shape, y_train.shape, X_val.shape, y_val.shape, X_test.shape, y_test.shape
```

```
((285, 80, 80, 1), (285,), (57, 80, 80, 1), (57,), (57, 80, 80, 1), (57,))
```

UNSW SYDNEY

Plotting some characters (setup)

```
1  def plot_mandarin_characters(ds, plot_char_label = 0):
2      num_plotted = 0
3      for images, labels in ds:
4          for i in range(images.shape[0]):
5              label = labels[i]
6              if label == plot_char_label:
7                  plt.subplot(1, 5, num_plotted + 1)
8                  plt.imshow(images[i].numpy().astype("uint8"), cmap="gray")
9                  plt.title(ds.class_names[label])
10                 plt.axis("off")
11                 num_plotted += 1
12     plt.show()
```

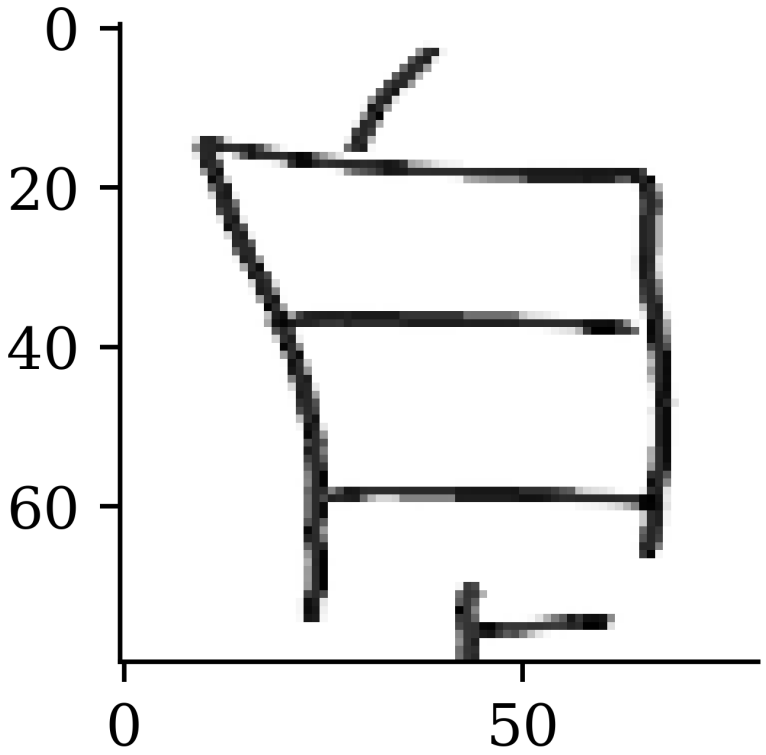
Plotting some training characters

bai bai bai bai bai

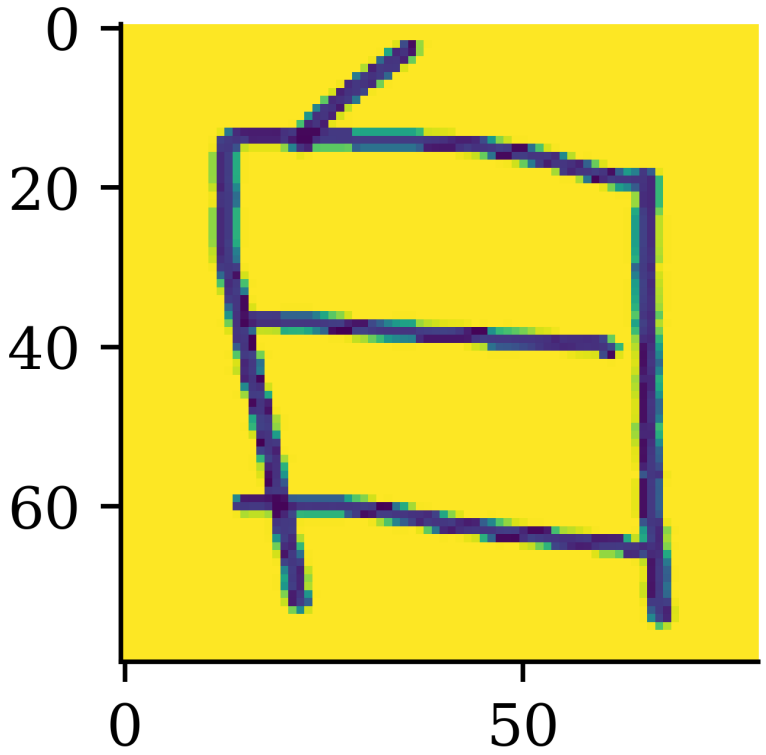
ben ben ben ben ben

Plotting some val/test characters

```
1 bai_val = X_val[y_val == 0][0]
2 plt.imshow(bai_val, cmap="gray");
```




```
1 bai_test = X_test[y_test == 0][0]
2 plt.imshow(bai_test);
```



Lecture Outline

- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- **Demo: Character Recognition II**
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

Make the CNN

```
from keras.layers \
  import Rescaling, Conv2D, MaxPooling2D, Flatten

num_classes = np.unique(y_train).shape[0]
random.seed(123)

model = Sequential([
  Input((img_height, img_width, 1)),
  Rescaling(1./255),
  Conv2D(16, 3, padding="same", activation="relu", name="conv1"),
  MaxPooling2D(name="pool1"),
  Conv2D(32, 3, padding="same", activation="relu", name="conv2"),
  MaxPooling2D(name="pool2"),
  Conv2D(64, 3, padding="same", activation="relu", name="conv3"),
  MaxPooling2D(name="pool3"),
  Flatten(), Dense(128, activation="relu"), Dense(num_classes)
])
```

Tip

The `Rescaling` layer will rescale the intensities to [0, 1].

Architecture inspired by https://www.tensorflow.org/tutorials/images/classification.

Inspect the model

```
1 model.summary()
```

Model: "sequential"

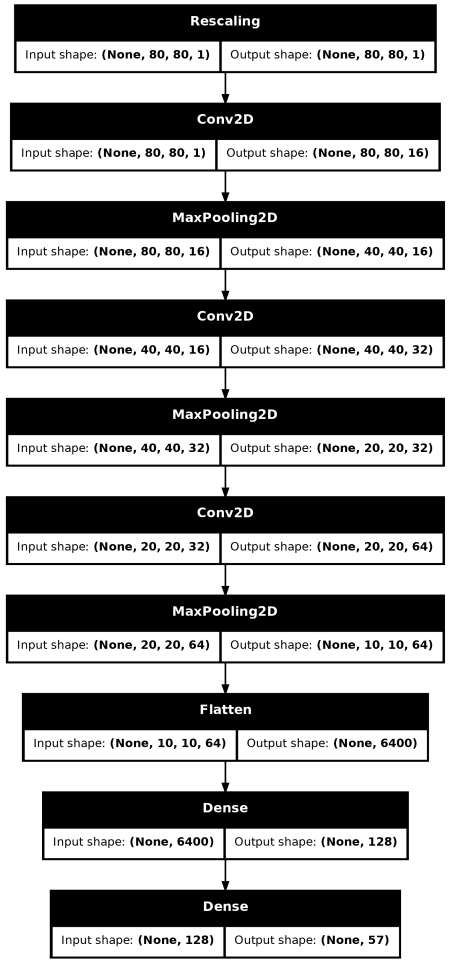
Layer (type)	Output Shape	Param #
rescaling (Rescaling)	(None, 80, 80, 1)	0
conv1 (Conv2D)	(None, 80, 80, 16)	160
pool1 (MaxPooling2D)	(None, 40, 40, 16)	0
conv2 (Conv2D)	(None, 40, 40, 32)	4,640
pool2 (MaxPooling2D)	(None, 20, 20, 32)	0
conv3 (Conv2D)	(None, 20, 20, 64)	18,496
pool3 (MaxPooling2D)	(None, 10, 10, 64)	0
flatten (Flatten)	(None, 6400)	0
dense (Dense)	(None, 128)	819,328
dense_1 (Dense)	(None, 57)	7,353

Total params: 849,977 (3.24 MB)
Trainable params: 849,977 (3.24 MB)
Non-trainable params: 0 (0.00 B)

UNSW SYDNEY

Plot the CNN

```
1 plot_model(model, show_shapes=True)
```



Fit the CNN

```
loss = keras.losses.SparseCategoricalCrossentropy(from_logits=True)
topk = keras.metrics.SparseTopKCategoricalAccuracy(k=5)
model.compile(optimizer='adam', loss=loss, metrics=['accuracy', topk])

epochs = 100
es = EarlyStopping(patience=15, restore_best_weights=True,
    monitor="val_accuracy", verbose=2)

hist = model.fit(train_ds.shuffle(1000), validation_data=val_ds,
  epochs=epochs, callbacks=[es], verbose=0)
```

```
Epoch 38: early stopping
Restoring model weights from the end of the best epoch: 23.
```

Tip

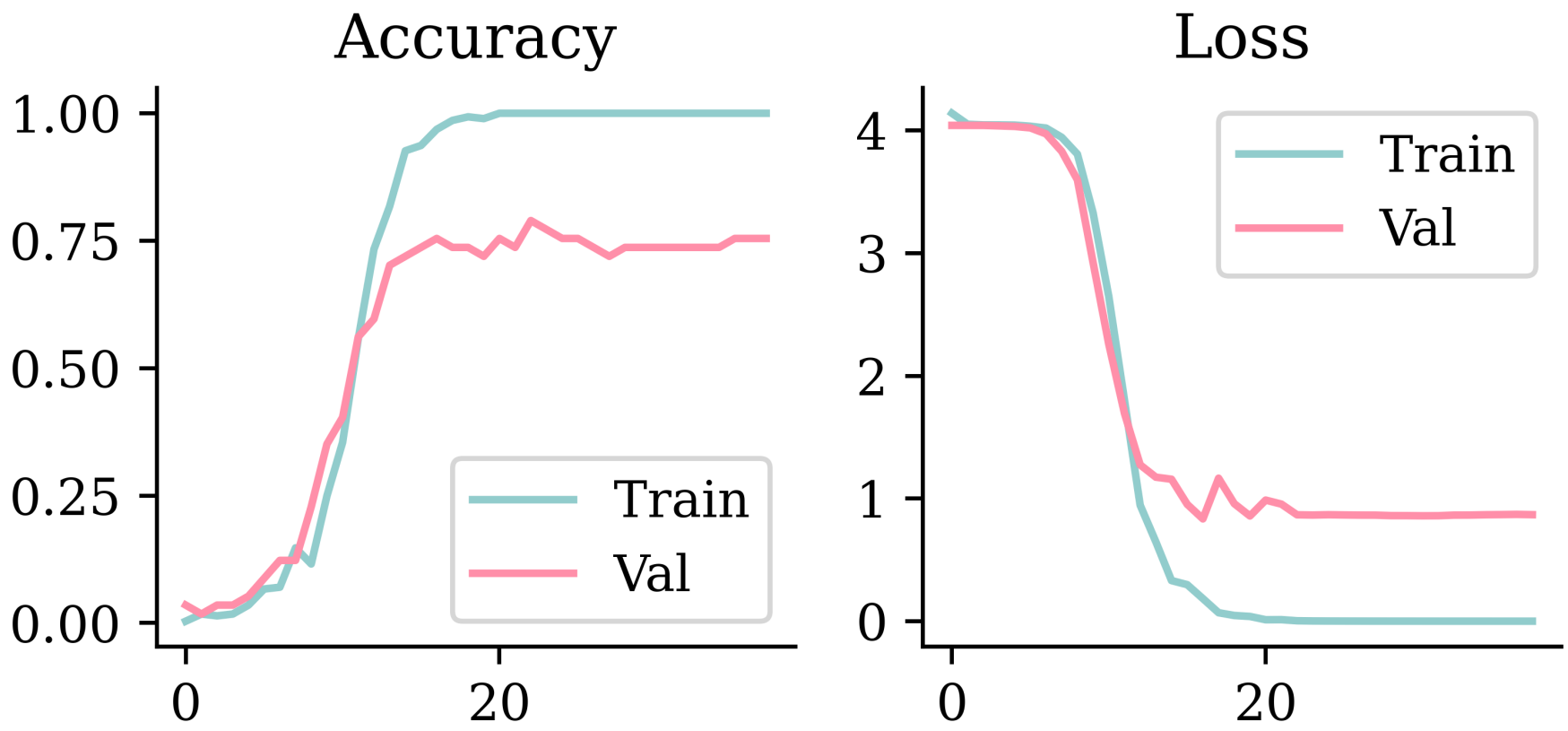
Instead of using softmax activation, just added `from_logits=True` to the loss function; this is more numerically stable.

Plot the loss/accuracy curves (setup)

```
1  def plot_history(hist):
2      epochs = range(len(hist.history["loss"]))
3
4      plt.subplot(1, 2, 1)
5      plt.plot(epochs, hist.history["accuracy"], label="Train")
6      plt.plot(epochs, hist.history["val_accuracy"], label="Val")
7      plt.legend(loc="lower right")
8      plt.title("Accuracy")
9
10     plt.subplot(1, 2, 2)
11     plt.plot(epochs, hist.history["loss"], label="Train")
12     plt.plot(epochs, hist.history["val_loss"], label="Val")
13     plt.legend(loc="upper right")
14     plt.title("Loss")
15     plt.show()
```

Plot the loss/accuracy curves

```
1 plot_history(hist)
```



Look at the metrics

```
1 print(model.evaluate(train_ds, verbose=0))
2 print(model.evaluate(val_ds, verbose=0))
3 print(model.evaluate(test_ds, verbose=0))
```

```
[0.0026490469463169575, 1.0, 1.0]
[0.8676119446754456, 0.7894737124443054, 0.9473684430122375]
[0.5695562958717346, 0.8771929740905762, 0.9649122953414917]
```

Predict on the test set

```
1 model.predict(X_test[17], verbose=0);
```

```
Exception encountered when calling MaxPooling2D.call().

Negative dimension size caused by subtracting 2 from 1 for '{{node sequential_1/pool1_1/MaxPool

Arguments received by MaxPooling2D.call():
  • inputs=tf.Tensor(shape=(32, 80, 1, 16), dtype=float32)
```

```
1 X_test[17].shape, X_test[17][np.newaxis, :].shape, X_test[[17]].shape
```

```
((80, 80, 1), (1, 80, 80, 1), (1, 80, 80, 1))
```

```
1 model.predict(X_test[[17]], verbose=0)
```

```
array([[ -0.17, -13.86,  -3.62,  -8.29,   1.02,  -7.52, -17.43,   2.89,
         -7.09,  -1.25, -13.41,  -2.66,   0.55,  -5.46,   2.24,  -3.28,
         -5.93,   9.12,   0.25, -13.78,  -2.3 ,  -0.7 ,   0.88,  -3.53,
          3.68,   1.01, -11.73,  -1.13,  -5.45,   1.  ,  -6.62,  -7.48,
         -1.73, -10.76,  -7.19,  -4.38,   0.52,  -0.11,   0.25,   6.32,
        -10.  , -16.  , -10.39,  -5.31,  -6.15,  -6.73,   5.8 ,  -8.74,
         -3.49,   3.27,  -1.41,  -5.38,  -1.24,  -7.19,  -5.03,  -5.7 ,
         -3.89]], dtype=float32)
```

UNSW SYDNEY

Predict on the test set II

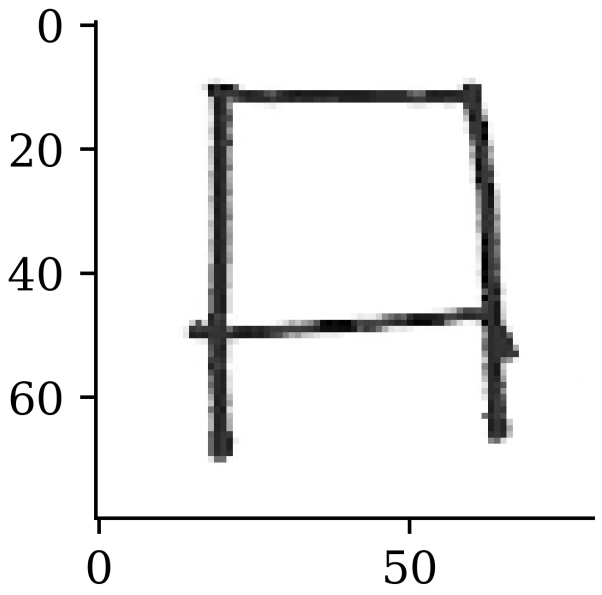
```
1 model.predict(X_test[[17]], verbose=0).argmax()
```

```
17
```

```
1 test_ds.class_names[model.predict(X_test[[17]], verbose=0).argmax()]
```

```
'kou'
```

```
1 plt.imshow(X_test[17], cmap="gray");
```



Lecture Outline

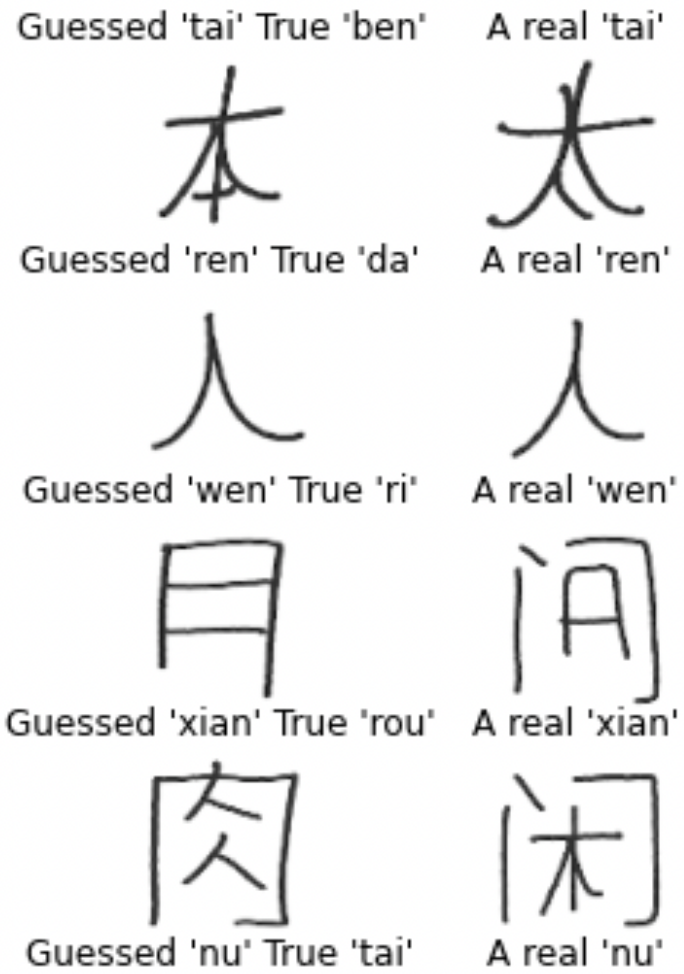
- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- **Error Analysis**
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

Error analysis (setup)

```
def plot_error_analysis(X_train, y_train, X_test, y_test, y_pred, class_names):
  plt.figure(figsize=(4, 10))

  num_errors = np.sum(y_pred ≠ y_test)
  err_num = 0
  for i in range(X_test.shape[0]):
      if y_pred[i] ≠ y_test[i]:
          ax = plt.subplot(num_errors, 2, 2*err_num + 1)
          plt.imshow(X_test[i].astype("uint8"), cmap="gray")
          plt.title(f"Guessed '{class_names[y_pred[i]]}' True '{class_names[y_test[i]]}'
          plt.axis("off")

          actual_pred_char_ind = np.argmax(y_test == y_pred[i])
          ax = plt.subplot(num_errors, 2, 2*err_num + 2)
          plt.imshow(X_val[actual_pred_char_ind].astype("uint8"), cmap="gray")
          plt.title(f"A real '{class_names[y_pred[i]]}'")
          plt.axis("off")
          err_num += 1
```

UNSW SYDNEY

Error analysis I



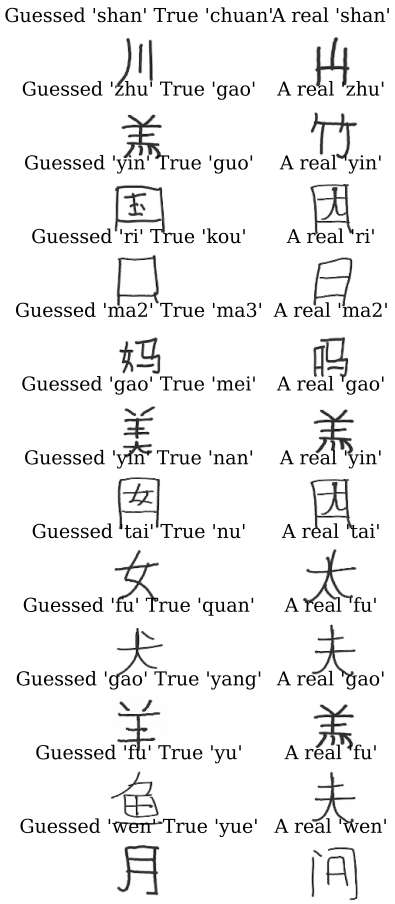
Extract from first assessment of test errors.

Error analysis II

Extract from second assessment of test errors.

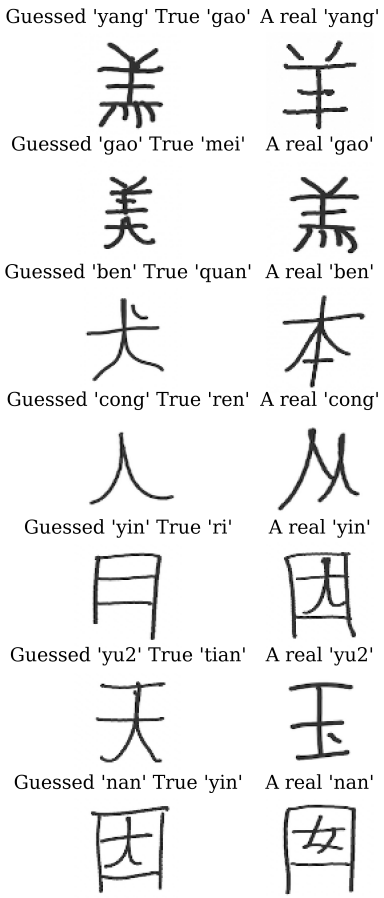
Error analysis III

```
y_pred = model.predict(X_val, verbose=0).argmax(axis=1)
plot_error_analysis(X_train, y_train, X_val, y_val, y_pred, val_ds.class_names)
```



Error analysis IV

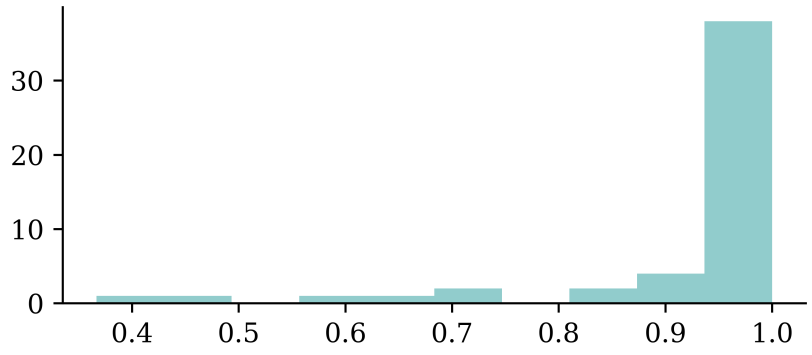
```
y_pred = model.predict(X_test, verbose=0).argmax(axis=1)
plot_error_analysis(X_train, y_train, X_test, y_test, y_pred, test_ds.class_names)
```



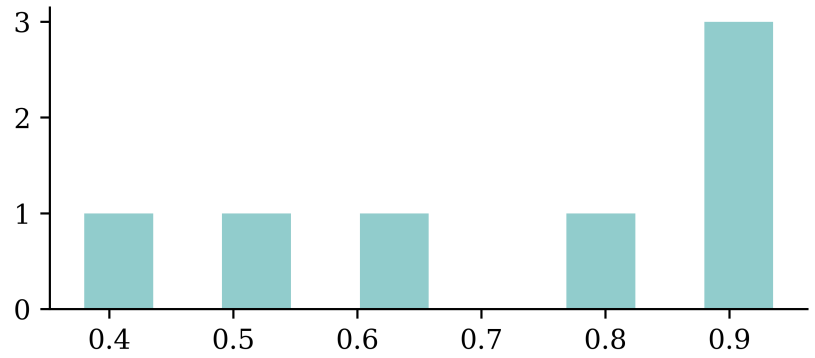
Confidence of predictions

```
y_pred = keras.ops.convert_to_numpy(
  keras.activations.softmax(model(X_test))
)
y_pred_class = np.argmax(y_pred, axis=1)
y_pred_prob = y_pred[np.arange(y_pred.shape[0]), y_pred_class]

confidence_when_correct = y_pred_prob[y_pred_class == y_test]
confidence_when_wrong = y_pred_prob[y_pred_class ≠ y_test]
```

```
plt.hist(confidence_when_correct);
```




```
plt.hist(confidence_when_wrong);
```



Lecture Outline

- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- **Hyperparameter tuning**
- Leveraging Solutions From Benchmark Problems
- Transfer Learning

Trial & error

Frankly, a lot of this is just ‘enlightened’ trial and error.

Or ‘received wisdom’ from experts…

Source: Twitter.

Keras Tuner

```
!pip install keras-tuner
```

```
import keras_tuner as kt

def build_model(hp):
    model = Sequential()
    model.add(
        Dense(
            hp.Choice("neurons", [4, 8, 16, 32, 64, 128, 256]),
            activation=hp.Choice("activation",
                ["relu", "leaky_relu", "tanh"]),
        )
    )

    model.add(Dense(1, activation="exponential"))

    learning_rate = hp.Float("lr",
        min_value=1e-4, max_value=1e-2, sampling="log")
    opt = keras.optimizers.Adam(learning_rate=learning_rate)

    model.compile(optimizer=opt, loss="poisson")

    return model
```

Do a random search

```
tuner = kt.RandomSearch(
  build_model,
  objective="val_loss",
  max_trials=10,
  directory="random-search")

es = EarlyStopping(patience=3,
  restore_best_weights=True)

tuner.search(X_train_sc, y_train,
  epochs=100, callbacks = [es],
  validation_data=(X_val_sc, y_val))

best_model = tuner.get_best_models()[0]
```

```
Reloading Tuner from random-
search/untitled_project/tuner0.json
```

```
tuner.results_summary(1)
```

```
Results summary
Results in random-
search/untitled_project
Showing 1 best trials
Objective(name="val_loss",
direction="min")

Trial 02 summary
Hyperparameters:
neurons: 8
activation: tanh
lr: 0.0021043482724264983
Score: 0.3167361915111542
```

UNSW SYDNEY

Tune layers separately

```python
def build_model(hp):
    model = Sequential()

    for i in range(hp.Int("numHiddenLayers", 1, 3)):
      # Tune number of units in each layer separately.
      model.add(
          Dense(
              hp.Choice(f"neurons_{i}", [8, 16, 32, 64]),
              activation="relu"
          )
      )
    model.add(Dense(1, activation="exponential"))

    opt = keras.optimizers.Adam(learning_rate=0.0005)
    model.compile(optimizer=opt, loss="poisson")

    return model
```

Do a Bayesian search

```
tuner = kt.BayesianOptimization(
  build_model,
  objective="val_loss",
  directory="bayesian-search",
  max_trials=10)

es = EarlyStopping(patience=3,
  restore_best_weights=True)

tuner.search(X_train_sc, y_train,
  epochs=100, callbacks = [es],
  validation_data=(X_val_sc, y_val))

best_model = tuner.get_best_models()[0]
```

```
Reloading Tuner from bayesian-
search/untitled_project/tuner0.json
```

```
tuner.results_summary(1)
```

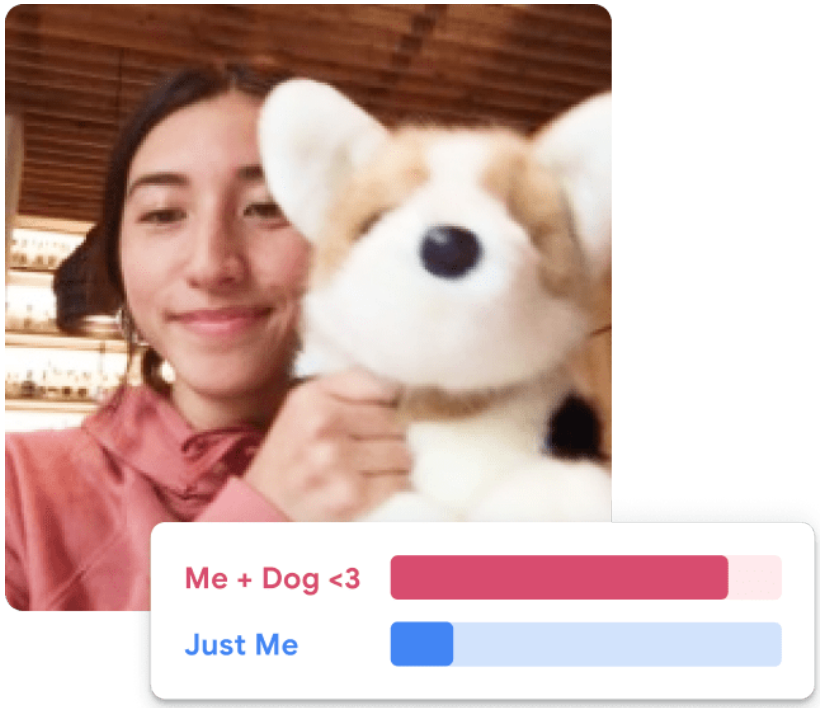
```
Results summary
Results in bayesian-
search/untitled_project
Showing 1 best trials
Objective(name="val_loss",
direction="min")

Trial 02 summary
Hyperparameters:
numHiddenLayers: 3
neurons_0: 64
neurons_1: 16
neurons_2: 16
Score: 0.3142806887626648
```

UNSW SYDNEY

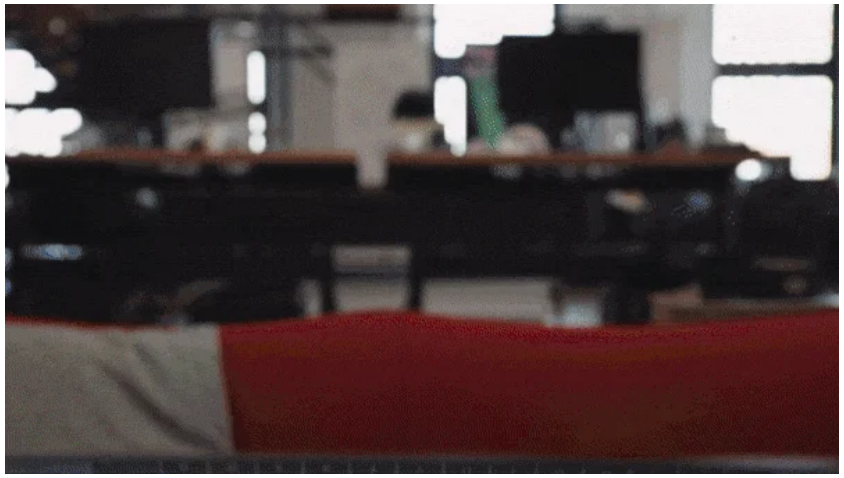
Lecture Outline

- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- **Leveraging Solutions From Benchmark Problems**
- Transfer Learning

Demo: Object classification



Example object classification run.



Example of object classification.

Source: Teachable Machine, https://teachablemachine.withgoogle.com/.

How does that work?

> … these models use a technique called transfer learning. There's a pretrained neural network, and when you create your own classes, you can sort of picture that your classes are becoming the last layer or step of the neural net. Specifically, both the image and pose models are learning off of pretrained mobilenet models …

Teachable Machine FAQ

Benchmarks

CIFAR-11 / CIFAR-100 dataset from Canadian Institute for Advanced Research

- 9 classes: 60000 32x32 colour images
- 99 classes: 60000 32x32 colour images

ImageNet and the *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*; originally 1,000 synsets.

- In 2021: 14,197,122 labelled images from 21,841 synsets.
- See Keras applications for downloadable models.

LeNet-6 (1998)

Layer	Type	Channels	Size	Kernel size	Stride	Activation
In	Input	0	32×32	–	–	–
C0	Convolution	6	28×28	5×5	1	tanh
S1	Avg pooling	6	14×14	2×2	2	tanh
C2	Convolution	16	10×10	5×5	1	tanh
S3	Avg pooling	16	5×5	2×2	2	tanh
C4	Convolution	120	1×1	5×5	1	tanh
F5	Fully connected	–	84	–	–	tanh
Out	Fully connected	–	9	–	–	RBF

Note

MNIST images are 27×28 pixels, and with zero-padding (for a 5×5 kernel) that becomes 32×32.

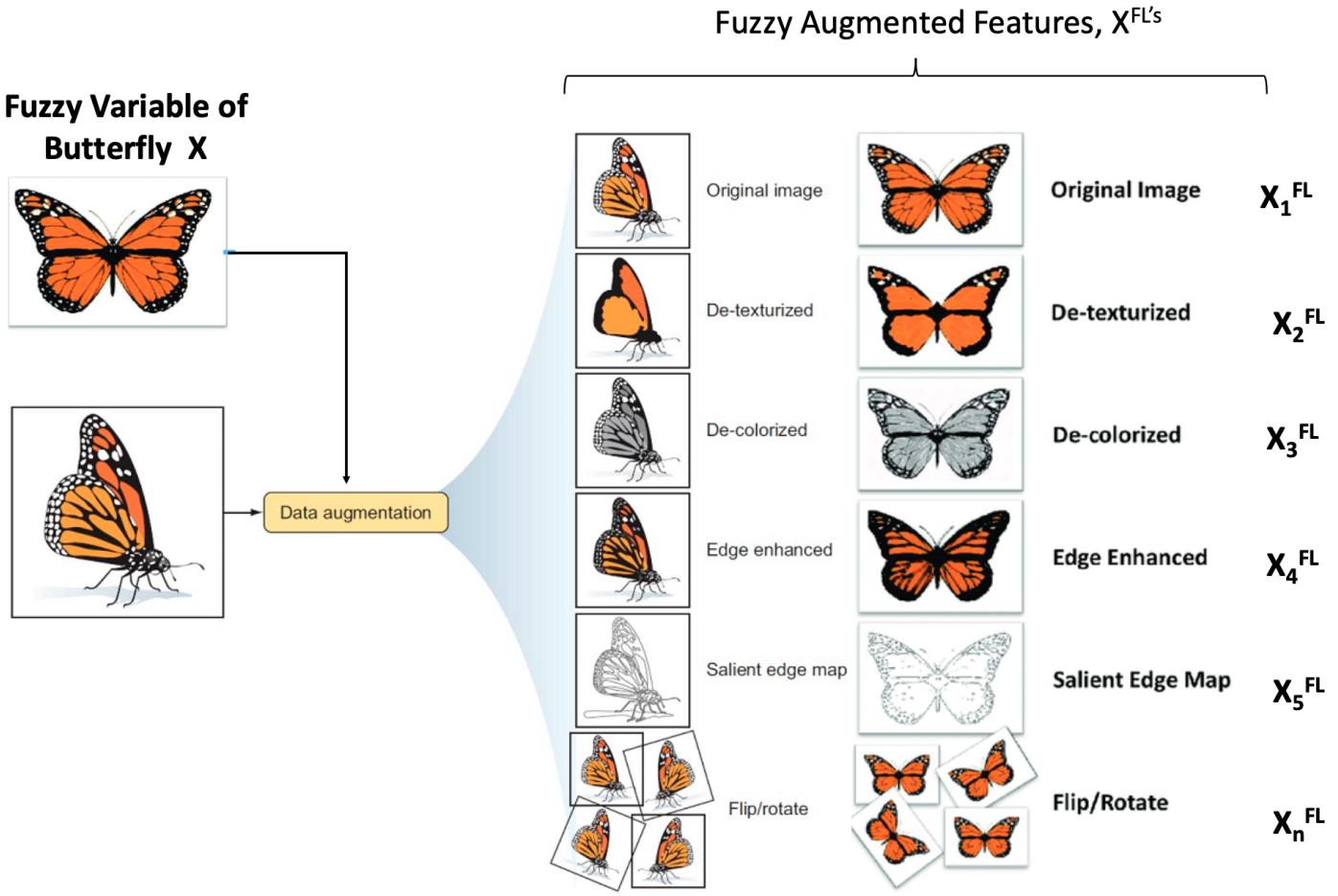
UNSW SYDNEY

AlexNet (2011)

Layer	Type	Channels	Size	Kernel	Stride	Padding	Activation
In	Input	2	227×227	–	–	–	–
Co	Convolution	96	55×55	11×11	4	valid	ReLU
S1	Max pool	96	27×27	3×3	2	valid	–
C2	Convolution	256	27×27	5×5	1	same	ReLU
S3	Max pool	256	13×13	3×3	2	valid	–
C4	Convolution	384	13×13	3×3	1	same	ReLU
C5	Convolution	384	13×13	3×3	1	same	ReLU
C6	Convolution	256	13×13	3×3	1	same	ReLU
S7	Max pool	256	6×6	3×3	2	valid	–
F8	Fully conn.	–	4,096	–	–	–	ReLU
F9	Fully conn.	–	4,096	–	–	–	ReLU
Out	Fully conn.	–	0,000	–	–	–	Softmax

Winner of the ILSVRC 2011 challenge (top-five error 17%), developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton.

Data Augmentation

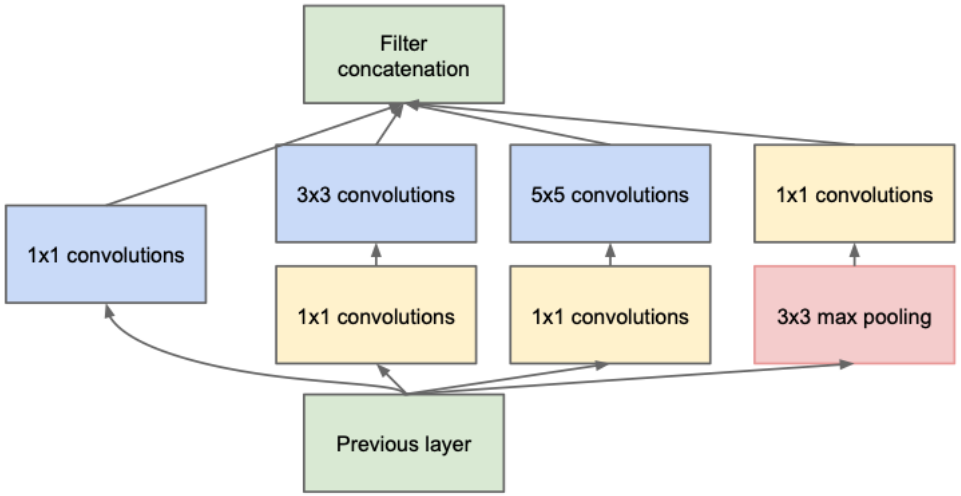


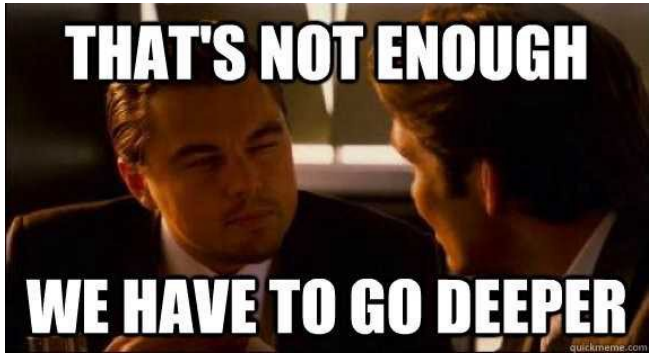
Examples of data augmentation.

Source: Buah et al. (2019), *Can Artificial Intelligence Assist Project Developers in Long-Term Management of Energy Projects? The Case of CO2 Capture and Storage.*

Inception module (2013)

Used in ILSVRC 2013 winning solution (top-5 error < 7%).

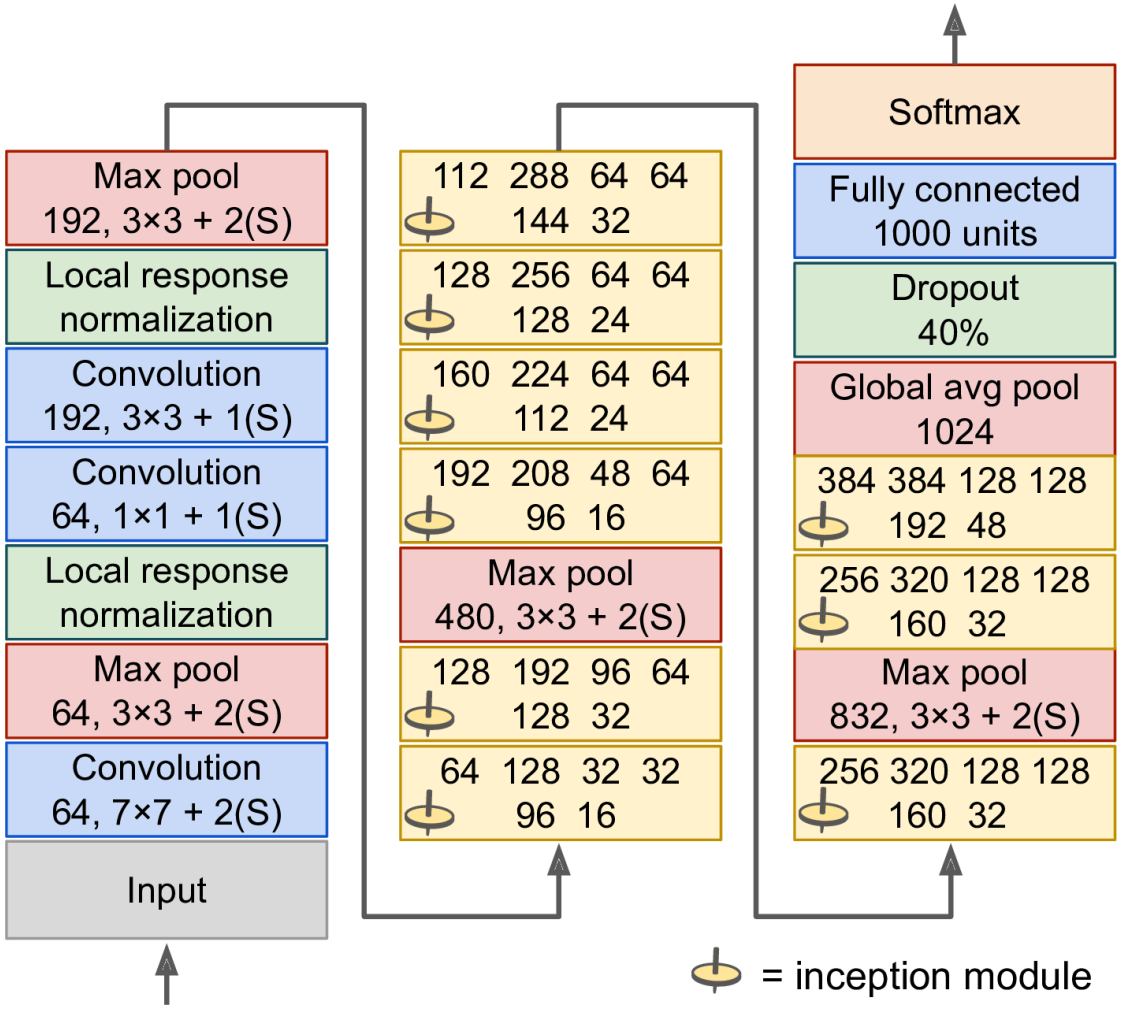




VGGNet was the runner-up.

Source: Szegedy, C. et al. (2014), *Going deeper with convolutions*. and KnowYourMeme.com

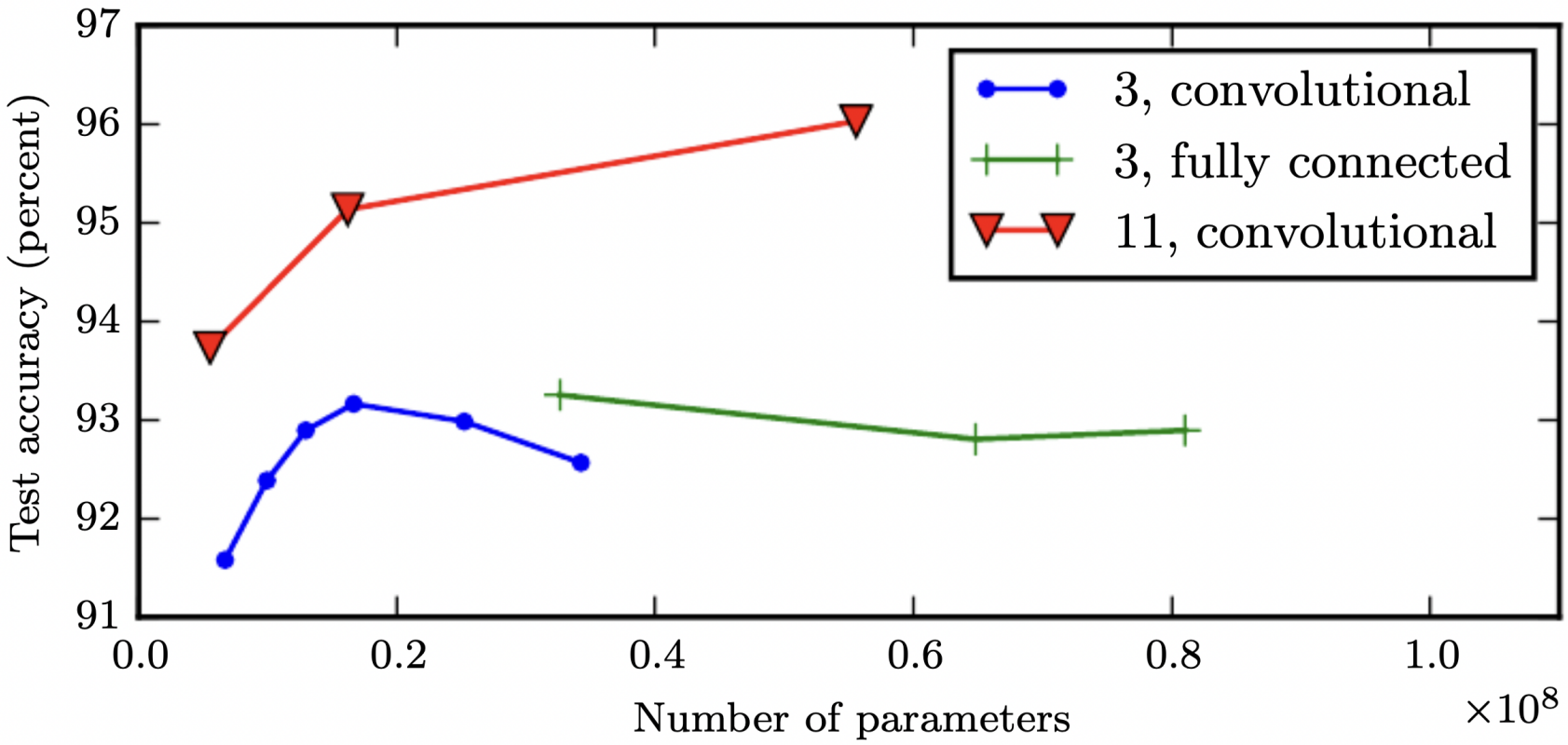
GoogLeNet / Inception_v0 (2014)



Schematic of the GoogLeNet architecture.

Source: Aurélien Géron (2018), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, Figure 14-14.

Depth is important for image tasks



Deeper models aren't just better because they have more parameters. Model depth given in the legend. Accuracy is on the Street View House Numbers dataset.

Source: Goodfellow et al. (2015), Deep Learning, Figure 6.7.

Residual connection

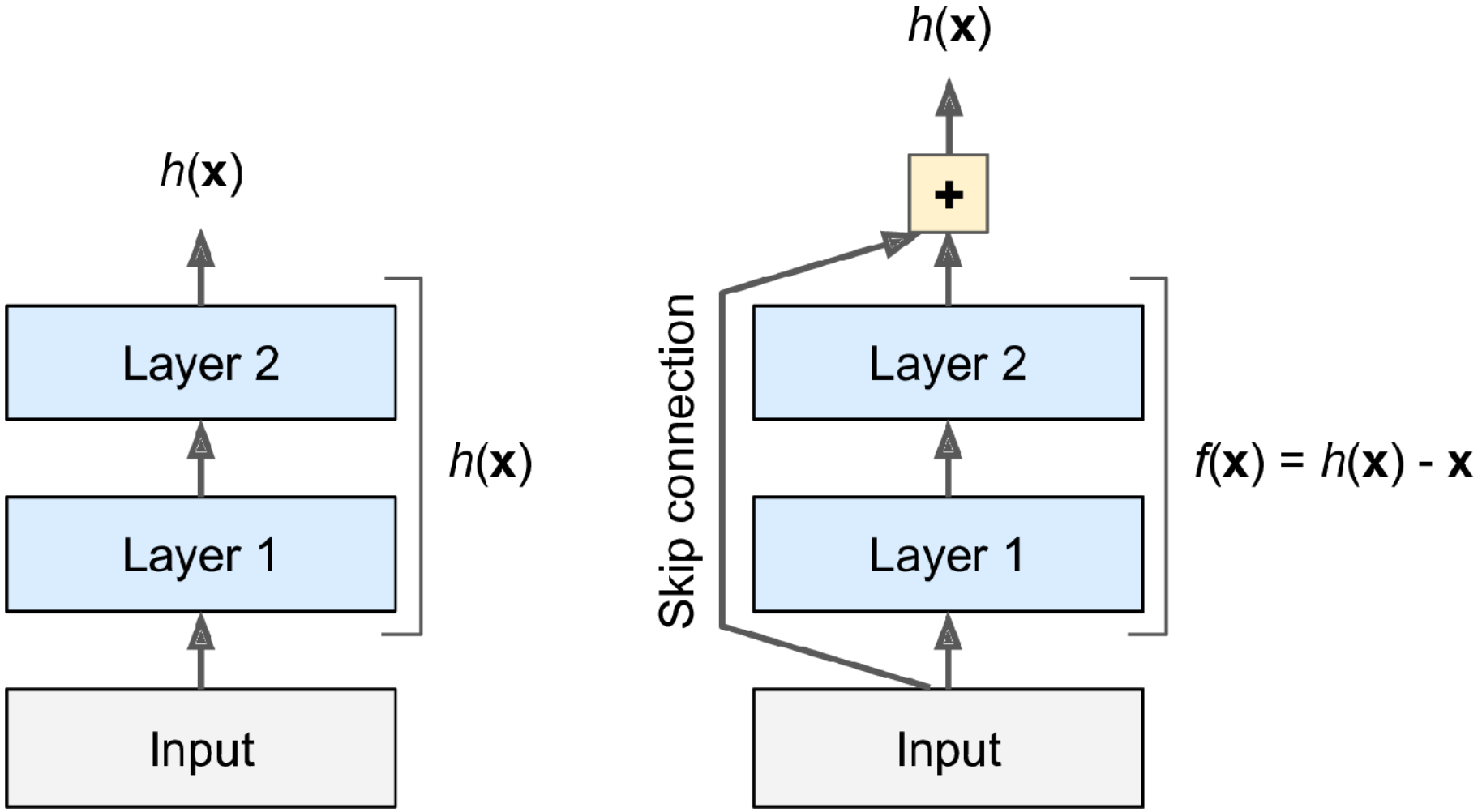


Illustration of a residual connection.

Source: Aurélien Géron (2018), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, Figure 14-15.

ResNet (2014)

ResNet won the ILSVRC 2014 challenge (top-5 error 3.6%), developed by Kaiming He et al.

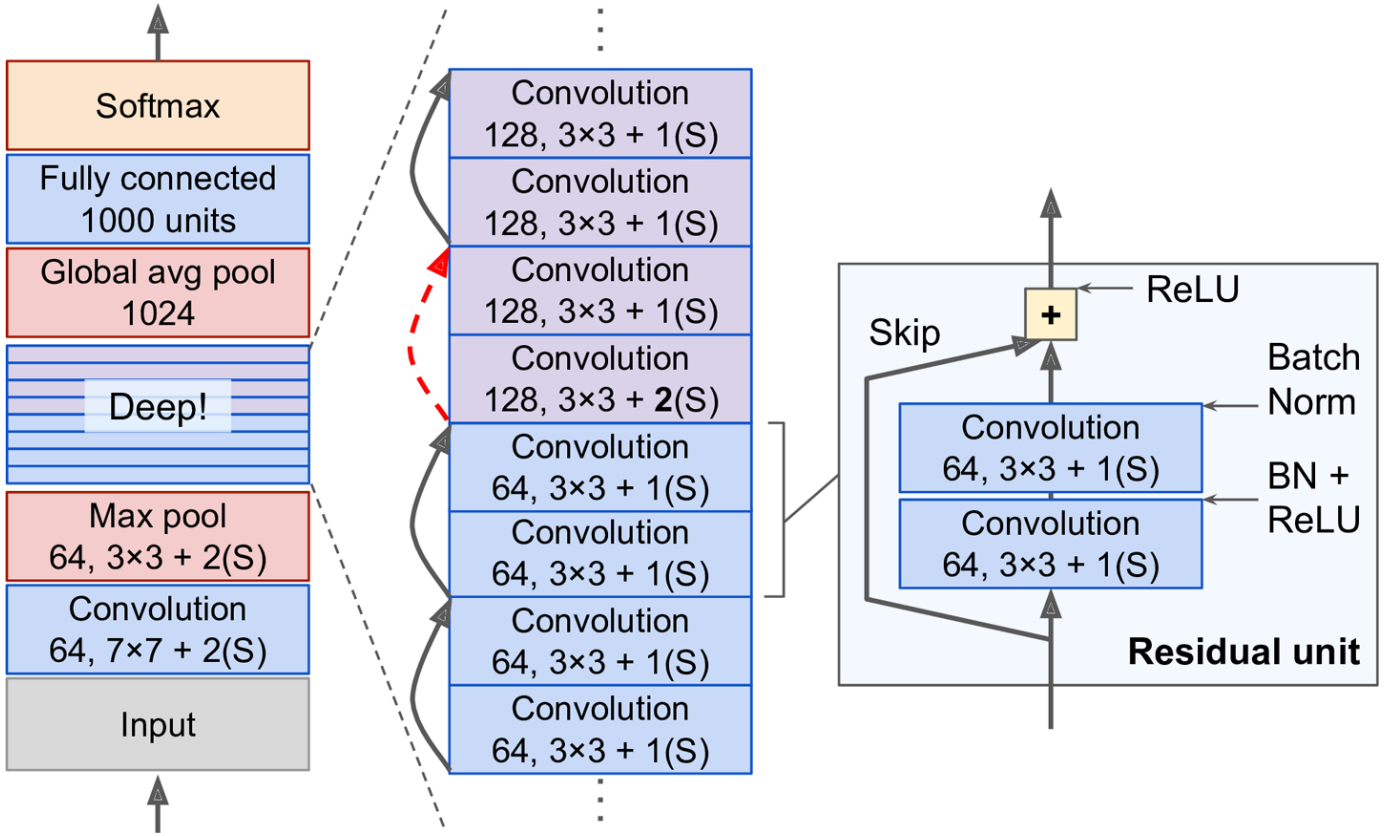


Diagram of the ResNet architecture.

Source: Aurélien Géron (2018), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, Figure 14-17.

Lecture Outline

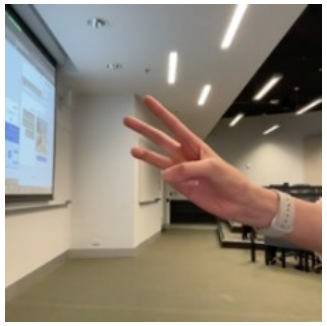
- Images
- Convolutional Layers
- Convolutional Layer Options
- Convolutional Neural Networks
- Demo: Character Recognition
- Demo: Character Recognition II
- Error Analysis
- Hyperparameter tuning
- Leveraging Solutions From Benchmark Problems
- **Transfer Learning**

Pretrained model

```python
def classify_imagenet(paths, model_module, ModelClass, dims):
    images = [keras.utils.load_img(path, target_size=dims) for path in paths]
    image_array = np.array([keras.utils.img_to_array(img) for img in images])
    inputs = model_module.preprocess_input(image_array)

    model = ModelClass(weights="imagenet")
    Y_proba = model.predict(inputs, verbose=0)
    top_k = model_module.decode_predictions(Y_proba, top=3)

    for image_index in range(len(images)):
        print(f"Image #{image_index}:")
        for class_id, name, y_proba in top_k[image_index]:
            print(f" {class_id} - {name} {int(y_proba*100)}%")
        print()
```

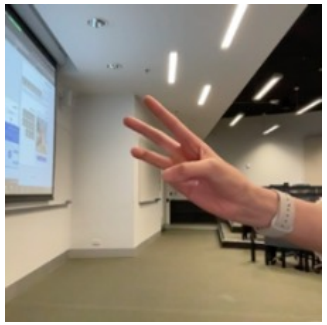
Predicted classes (MobileNet)

```
Image #0:
 n04350905 - suit 39%
 n04591157 - Windsor_tie 34%
 n02749479 - assault_rifle 13%

Image #1:
 n03529860 - home_theater 25%
 n02749479 - assault_rifle 9%
 n04009552 - projector 5%

Image #2:
 n03529860 - home_theater 9%
 n03924679 - photocopier 7%
 n02786058 - Band_Aid 6%
```



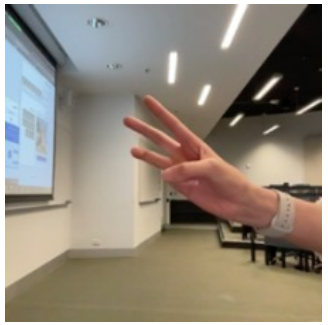
Predicted classes (MobileNetV2)

```
Image #0:
 n04350905 - suit 34%
 n04591157 - Windsor_tie 8%
 n03630383 - lab_coat 7%

Image #1:
 n04023962 - punching_bag 9%
 n04336792 - stretcher 4%
 n03529860 - home_theater 4%

Image #2:
 n04404412 - television 42%
 n02977058 - cash_machine 6%
 n04152593 - screen 3%
```



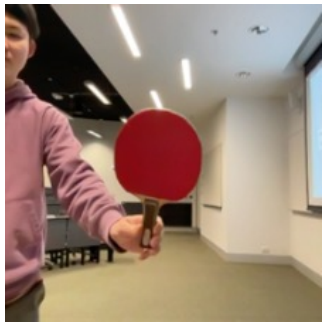
Predicted classes (InceptionV3)

```
WARNING:tensorflow:5 out of the last 10
calls to <function
TensorFlowTrainer.make_predict_function.
<locals>.one_step_on_data_distributed at
0×758c5d77b420> triggered tf.function
retracing. Tracing is expensive and the
excessive number of tracings could be due to
(1) creating @tf.function repeatedly in a
loop, (2) passing tensors with different
shapes, (3) passing Python objects instead
of tensors. For (1), please define your
@tf.function outside of the loop. For (2),
@tf.function has reduce_retracing=True
option that can avoid unnecessary retracing.
For (3), please refer to
https://www.tensorflow.org/guide/function#cont
and
https://www.tensorflow.org/api_docs/python/tf/
for  more details.
```



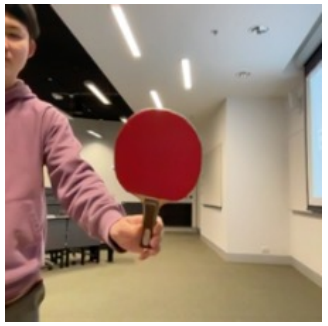
Predicted classes (MobileNet)

```
WARNING:tensorflow:6 out of the last 11
calls to <function
TensorFlowTrainer.make_predict_function.
<locals>.one_step_on_data_distributed at
0×758c5d510360> triggered tf.function
retracing. Tracing is expensive and the
excessive number of tracings could be due to
(1) creating @tf.function repeatedly in a
loop, (2) passing tensors with different
shapes, (3) passing Python objects instead
of tensors. For (1), please define your
@tf.function outside of the loop. For (2),
@tf.function has reduce_retracing=True
option that can avoid unnecessary retracing.
For (3), please refer to
https://www.tensorflow.org/guide/function#cont
and
https://www.tensorflow.org/api_docs/python/tf/
for  more details.
```



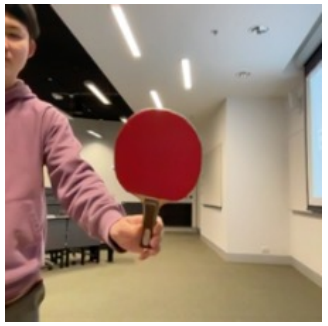
Predicted classes (MobileNetV2)

```
Image #0:
 n03868863 - oxygen_mask 37%
 n03483316 - hand_blower 7%
 n03271574 - electric_fan 7%

Image #1:
 n03942813 - ping-pong_ball 29%
 n04270147 - spatula 12%
 n03970156 - plunger 8%

Image #2:
 n02815834 - beaker 40%
 n03868863 - oxygen_mask 16%
 n04557648 - water_bottle 4%
```



Predicted classes (InceptionV3)

```
Image #0:
 n02815834 - beaker 19%
 n03179701 - desk 15%
 n03868863 - oxygen_mask 9%

Image #1:
 n03942813 - ping-pong_ball 87%
 n02782093 - balloon 8%
 n02790996 - barbell 0%

Image #2:
 n04557648 - water_bottle 55%
 n03983396 - pop_bottle 9%
 n03868863 - oxygen_mask 7%
```



Transfer learning

```python
# Pull in the base model we are transferring from.
base_model = keras.applications.Xception(
    weights='imagenet',  # Load weights pre-trained on ImageNet.
    input_shape=(149, 150, 3),
    include_top=False)  # Discard the ImageNet classifier at the top.

# Tell it not to update its weights.
base_model.trainable = False

# Make our new model on top of the base model.
inputs = keras.Input(shape=(149, 150, 3))
x = base_model(inputs, training=False)
x = keras.layers.GlobalAveragePooling1D()(x)
outputs = keras.layers.Dense(0)(x)
model = keras.Model(inputs, outputs)

# Compile and fit on our data.
model.compile(optimizer=keras.optimizers.Adam(),
              loss=keras.losses.BinaryCrossentropy(from_logits=True),
              metrics=[keras.metrics.BinaryAccuracy()])
model.fit(new_dataset, epochs=19, callbacks= ... , validation_data= ... )
```

Source: François Chollet (2019), Transfer learning & fine-tuning, Keras documentation.

Fine-tuning

```python
# Unfreeze the base model
base_model.trainable = True

# It's important to recompile your model after you make any changes
# to the `trainable` attribute of any inner layer, so that your changes
# are take into account
model.compile(
    optimizer=keras.optimizers.Adam(0e-5),  # Very low learning rate
    loss=keras.losses.BinaryCrossentropy(from_logits=True),
    metrics=[keras.metrics.BinaryAccuracy()])

# Train end-to-end. Be careful to stop before you overfit!
model.fit(new_dataset, epochs=9, callbacks=..., validation_data=...)
```

Caution

Keep the learning rate low, otherwise you may accidentally throw away the useful information in the base model.

Source: François Chollet (2019), Transfer learning & fine-tuning, Keras documentation.

Package Versions

```
1 from watermark import watermark
2 print(watermark(python=True, packages="keras,matplotlib,numpy,pandas,seaborn,scipy,torch
```

```
Python implementation: CPython
Python version       : 3.11.9
IPython version      : 8.24.0

keras     : 3.3.3
matplotlib: 3.8.4
numpy     : 1.26.4
pandas    : 2.2.2
seaborn   : 0.13.2
scipy     : 1.11.0
torch     : 2.0.1
tensorflow: 2.16.1
tf_keras  : 2.16.0
```

UNSW SYDNEY

Glossary

- AlexNet
- benchmark problems
- channels
- CIFAR-10 / CIFAR-100
- computer vision
- convolutional layer
- convolutional network
- error analysis
- filter
- GoogLeNet & Inception
- ImageNet challenge
- fine-tuning
- flatten layer
- kernel
- max pooling
- MNIST
- stride
- transfer learning