AI Bootcamp

Preprocessing Image Data

Module 19 Day 1

Class Objectives

By the end of class, you will be able to:

1. Describe the Pillow library.
2. Import an image.
3. Import an image dataset.
4. Prepare image data for a model.
5. Normalize image data.
6. Label image data for supervised learning.

Welcome

Instructor **Demonstration**

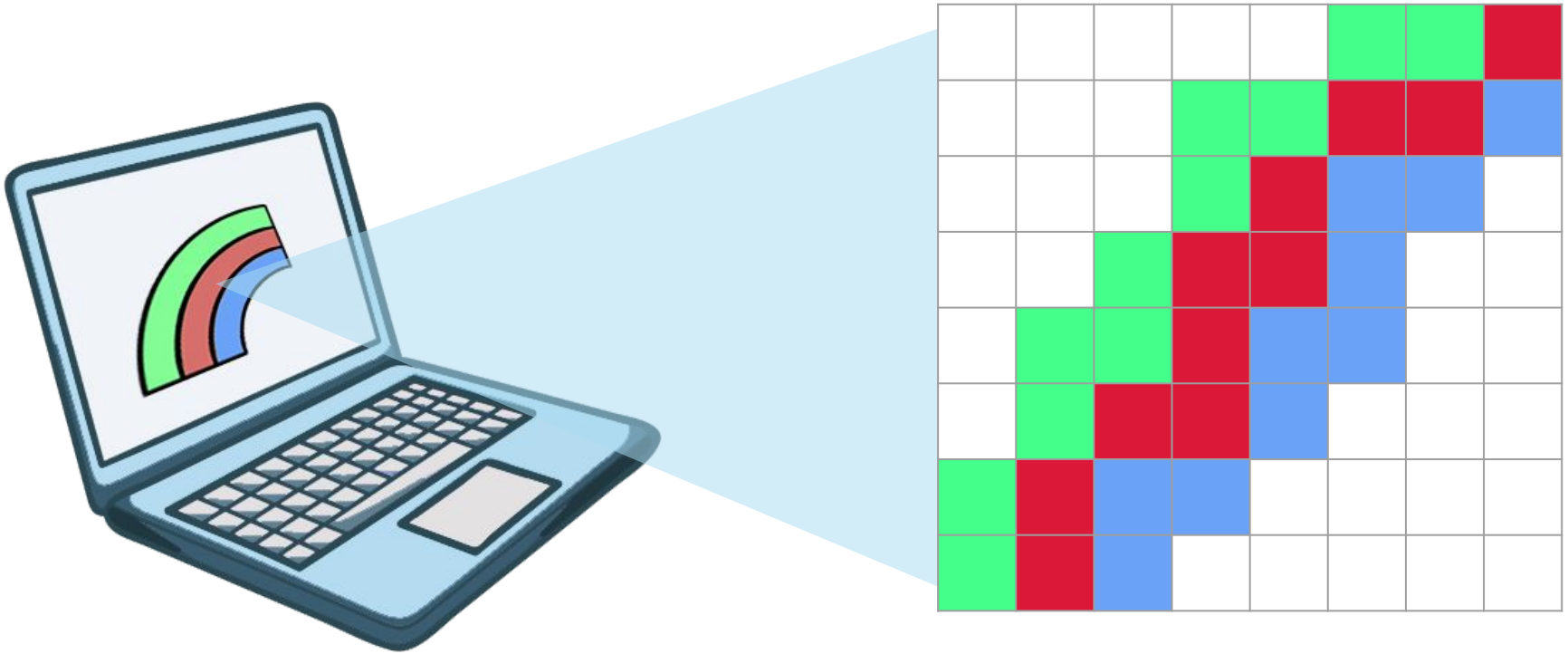
Introduction to Image Data

What kind of data have we used to train and validate our models?

What challenges might we face in processing image data that we didn’t face with data in traditional rows and columns?

Representing Image Data

Any image is made up of a number of pixels:

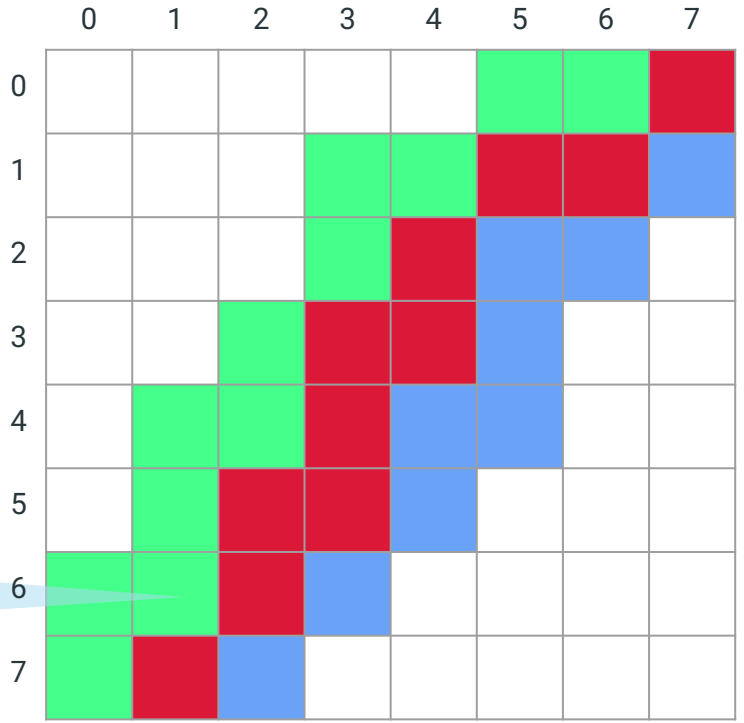


Representing Image Data

Pixel position

This pixel is positioned in row 6 and column 1.

(6, 1)



Representing Image Data

Combinations of RGB values

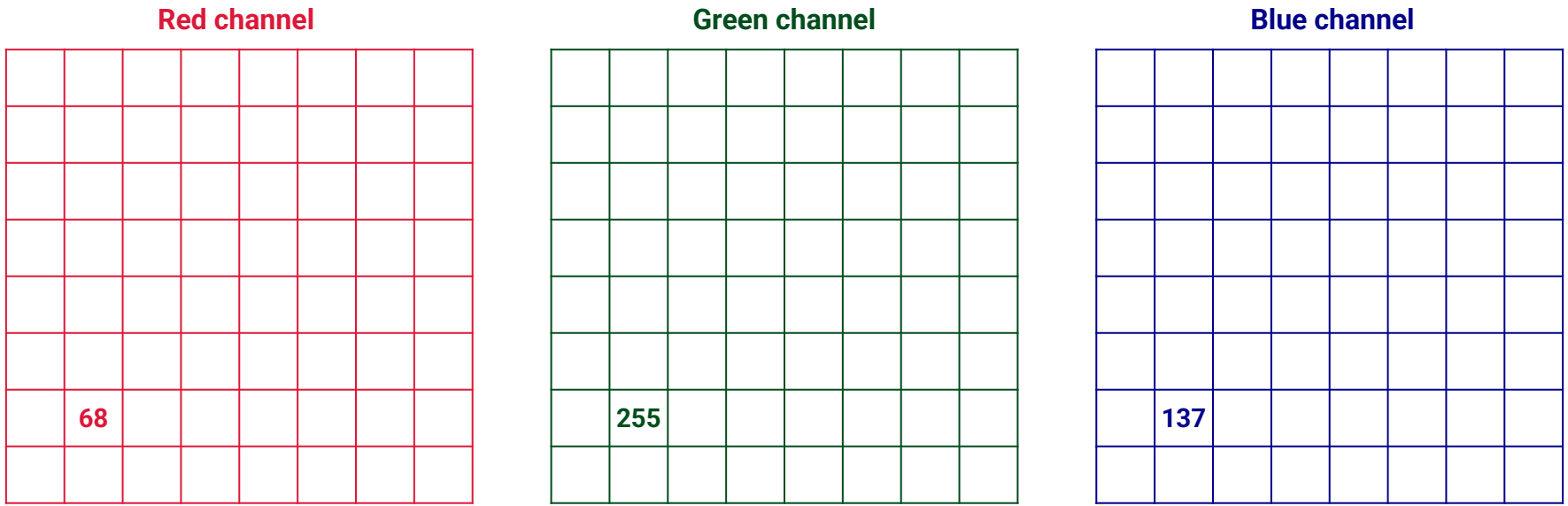
	Red	Green	Blue
Crimson	220	24	57
Burgundy	147	3	46
Carnation pink	253	175	199
Mountbatten pink	162	112	138

Seemingly similar colors have significantly different RGB values. The pixel we singled out earlier at position (6, 1) has RGB values (68, 255, 137).

	Red	Green	Blue
Spring green (68, 255, 137)	68	255	137

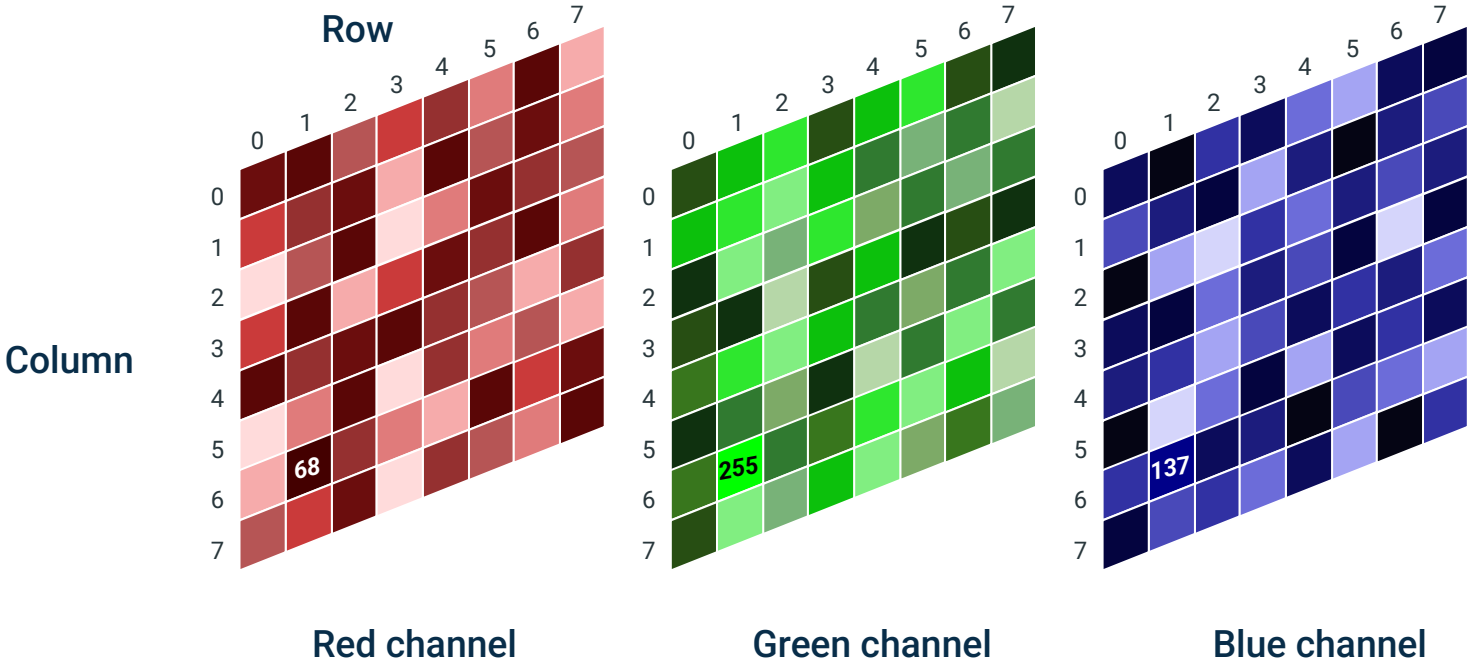
Representing Image Data: Arrays

Each RGB channel has its own array.



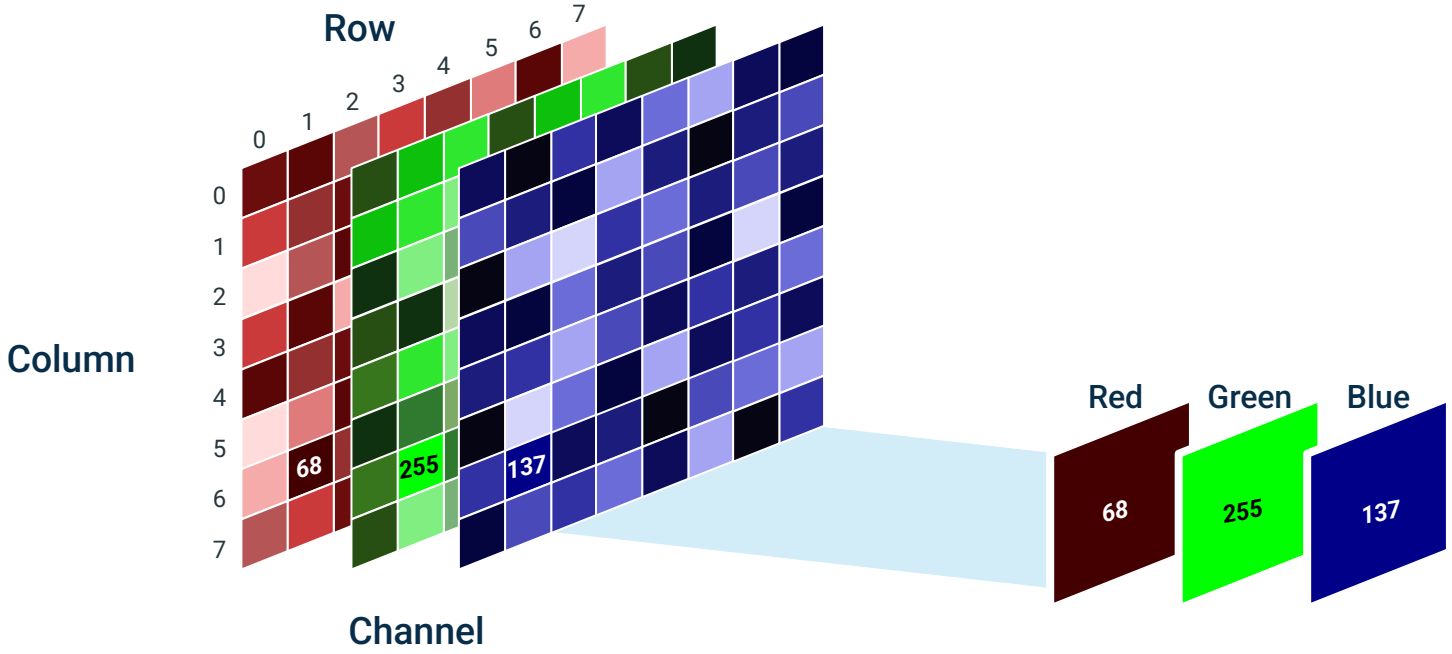
Representing Image Data: Stacked RGB Channels

Stacked RGB channels



Representing Image Data: Stacked RGB Channels

These arrays are stacked to create a representation of each pixel in the image.



If we were to pack all the information describing the single pixel we've focused on into a tuple, what would it look like?

Instructor **Demonstration**

Importing an Image

Activity:

Practice Importing

In this activity, you will use Pillow and the **`requests`** library to import an image, view the image, and view or modify various attributes of the image.

Suggested Time:
10 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Pickling Images

Pickles could hold unsafe and malicious data. Only unpickle data you know is from a **trusted source**.

Activity:

Practice Pickling

In this activity, you will return to the DeFungi dataset and use Pillow and the `requests` library to import 20 images into a list. Then you will store that list as a pickle file.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?

Break

15 mins

Instructor **Demonstration**

Preprocessing Images

Activity:

Preprocessing Fungi

In this activity, you will resize, convert to float, and normalize all images in the DeFungi set.

Suggested Time:
20 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Image Labels

Activity:

Labeling Images

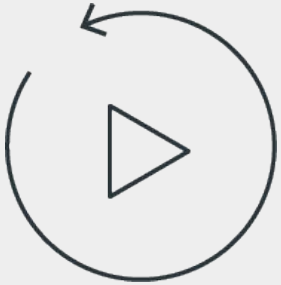
In this activity, you will import a .csv file containing the file names of the 250 fungi images. Using pandas, you will manipulate the file names to extract the class labels for each image.

Suggested Time:
10 Minutes

Time’s up!

Let’s review

Questions?



Let’s **recap**

Recap

After today's lesson, you are able to:

1. Describe the Pillow library.
2. Import an image.
3. Import an image dataset.
4. Prepare image data for a model.
5. Normalize image data.
6. Label image data for supervised learning.

Next

In the next lesson, you'll begin to explore the ins and outs of the CNN architecture.

Questions?

The End