

Data Preparation

AI Model Development

Time: 60 mins

Introduction

In this class, the student/s will learn to prepare data that can be used to train the ML model to predict the age of the person in real-time.

New Commands Introduced

- `.split()` Splits the string into a list. You can specify the separator, default separator is any whitespace.
- `If img != '.git':` Checks if the img is not in the .git format
- `cv2.cvtColor()` Converts the image color space from one to another
- `np.array()` Creates a numpy array with the specified datatype
- `img[Y:Y+H, X:X+W]` Crops the height and width of an image

Vocabulary

- The **RGB** color model is an additive color model in which the red, green and blue primary colors of light are added together in various ways to reproduce a broad array of colors..
- **Data preparation** is the process of transforming raw data so that data scientists and analysts can run it through machine learning algorithms to uncover insights or make predictions.

Learning Objectives

Student/s should be able to:

- **Recall** how to use the face detector module to detect faces for data preparation.
- **Demonstrate** how to create an AI prediction tool that can predict the age of the person.
- **Explain** the working of trained models to predict the age of a person in the video.

Activities

1. Class Narrative: (2 mins)

- Brief the student/s that you will create an AI tool to predict the age of a person and we need to prepare data which can be used to train the ML model.

2. Concept Introduction Activity: (5 mins)

- Let the student/s play the explore-activity to explore the prepared data with images of people with various age groups.
- Using the slides, explain that the student/s will learn:
 - to feed the labeled image.
 - to store and display age.
 - to preprocess the image.

3. Activity 1: Feed the Labeled Image (14 mins)

Teacher Activity: (7 mins)

- Demonstrate how to add and label the images in the dataset folder as per specified format.
- Iterate through the images in the folder and print their labels.

Student Activity: (7 mins)

- Guide the student/s to feed the sample images and label the images in the dataset folder as per specified format.

4. Activity 2: Store and Display Age (12 mins)

Teacher Activity: (6 mins)

- Explain how labeling an image can be used to define the age category by splitting the label, append the ages and store it in the NumPy array.
- Explain how to read an image, store it in RGB format and further display age over an image.

Student Activity: (6 mins)

- Guide the students to store and display the age by splitting the label and fetching the age.

5. Activity 3: Preprocess the Image (12 mins)

Teacher Activity: (6 mins)

- Explain the student to detect the face and store its coordinates and dimension, normalize the image by cropping and resizing the image.
- Explain appending the resized images and storing them in the NumPy array.

Student Activity: (6 mins)

- Guide the student/s to preprocess the images before model training by detecting the face, storing its coordinates and dimension, cropping and resizing the image.

6. Introduce the Post class project: (2 min)

- Label the training dataset images to distinguish as infected or uninfected with Pneumothorax.

7. Test and Summarize the class learnings: (5 mins)

- Check for understanding through quizzes and summarize learning after respective missions.
- Summarize the overall class learning towards the end of the class.

8. Additional activities:

- Encourage the student/s to rename the labels as per the specified format to train the model.
- Encourage the student/s to add a condition to maximize a dataset to 50 images for each age group.

9. State the Next Class Objective: (1 min)

- In the next class, student/s will learn to create an AI machine learning model which can be used to predict the age.

U.S. Standards:

CSTA: 2-AP-11, 2-AP-12, 2-AP-13, 2-AP-14, 2-AP-19

Links Table		
Activity	Activity Name	Link
Class Presentation	Data Preparation	https://s3-whjr-curriculum-uploads.whjr.online/79c07d7c-3479-4489-a4d1-b04dbb5871ac.html
Explore Activity	Data Preparation	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Activity 1	Feed the Labeled Images	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS-BP
Teacher Reference: Teacher Activity 1 Solution	Feed the Labeled Images	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS
Student Activity 1	Feed the Labeled Images	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Reference: Student Activity 1 Solution	Feed the Labeled Images	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS
Teacher Activity 2	Store and Display Age	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS-BP
Teacher Reference: Teacher Activity 2 Solution	Store and Display Age	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS

Student Activity 2	Store and Display Age	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Reference: Student Activity 2 Solution	Store and Display Age	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS
Teacher Activity 3	Preprocess the Image	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS-BP
Teacher Reference: Teacher Activity 3 Solution	Preprocess the Image	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-TAS
Student Activity 3	Preprocess the Image	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Reference: Student Activity 3 Solution	Preprocess the Image	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS
Student's Additional Activity 1	Label the Training Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Reference: Student's Additional Activity 1 Solution	Label the Training Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS
Student's Additional Activity 2	Maximize a Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS-BP
Teacher Reference: Student's Additional Activity 2 Solution	Maximize a Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-SAS
Post Class Project	Label the Training Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-PCP-BP
Teacher Reference: Post Class Project Solution	Label the Training Dataset	https://github.com/Tynker-Computer-Vision/TNK-M10-PRO-C74-PCP