Scene based video segmentation

Submited by : Toukir Sabugar

To : CIPIO.AI

Introduction

- This PPT presents an implementation of scene-based video segmentation using deep learning techniques.
- The goal is to classify scenes as indoor or outdoor and detect the presence of people in each scene.

Dataset Preparation

- Random photos of indoor and outdoor scenes are taken from internet.
- There are total 887 images for training belonging to 2 classes indoor and outdoor.
- for validation 175 images are taken

Scene Classification model

- A pre-trained MobileNetV2 model is used for scene classification.
- The model is fine-tuned on the dataset to classify scenes as indoor or outdoor.
- A threshold of 0.5 is used to determine the scene type based on the model's output probability.
- I run the code for 10 epochs, optimizer=adam,

  loss= binary_crossentropy
- The accuracy of model is **95.83%.**

Why MobileNetv2

- MobileNetV2 is specifically designed for mobile and embedded vision applications, offering a good balance between model size and accuracy.
- It provides a lightweight architecture that allows for efficient inference, making it suitable for real-time applications like scene classification in videos.
- MobileNetV2 is optimized for speed, which is crucial for processing video frames in real-time.

- A pre-trained SSD MobileNetV2 model is used for detecting people in each frame.
- The model is applied to each frame to detect the presence of people.
- Bounding boxes are drawn around detected people for visualization.
- **There are total 231 frames or scene of video clip. My model predict 160 times outdoor and 71 times indoor.**
- Due to less data model is not able to draw bounding box on people but it is showing good result for indoor and outdoor scene detection.

Challenges faced during this work

- The first challenge was to create a quality dataset. I collected images from kaggle dataset and divide it into two classes indoor and outdoor scene.
- Then I searched which pre trained model I can use for this task. The pretrained models are resnet50, vgg16, efficientnet, mobilenetv2 etc.
- The main reason I select mobilenetv2 is it lighter than other model and work best for mobile application.

Solution Implemented

- First, I create dataset of indoor and outdoor scene images. There are total 887 images
- Then, I select and fine-tune a pre-trained MobileNetV2 model for scene classification and a Single Shot MultiBox Detector (SSD) for people detection.
- Load these models and a video file for processing.
- Initialize counters for indoor and outdoor scenes. Process each frame of the video by resizing it to 224x224 for MobileNetV2 input, preprocessing it, and predicting the scene (indoor/outdoor) using the scene classification model.

- Detect people in the frame using the SSD model and draw bounding boxes around them.
- Finally, release the video capture resources and display the counts of indoor and outdoor scenes.
- The model accuracy is **95.83%**.
- The video which I used here have both indoor and outdoor scene **There are total 231 frames or scene of video clip. My model predict 160 times outdoor scene and 71 times indoor scene because outdoor scene has more duration in video compare to indoor Scene.**
- **From the result I can say my model justify the output.**

Potential Improvements

- We can augment the dataset with additional variations of scenes and objects to improve the robustness of the models and enhance their performance in diverse real-world scenarios.
- We can use an ensemble of multiple models for scene classification and object detection to improve overall accuracy and reliability.
- Can apply dynamic thresholding techniques based on scene complexity or frame characteristics to adaptively adjust the threshold for scene classification, improving the accuracy of indoor-outdoor scene classification.

- We can implement a feedback mechanism to refine the models based on user interactions or additional information, improving the adaptability of the system to new environments or scenarios.
- Link of my code:

https://colab.research.google.com/drive/1mKqBRu1f78Q8OTlphzbwUHBmMraepSO0?usp=sharing

Thank you