LUDDY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING
INDIANA UNIVERSITY

Comprehensive Annotation and Object Recognition on Historic Advertisement Videos

Jaswanth Kranthi Boppana, Jagadeesh Kovi, Vishnu Paturu
Final Course Project in Computer Vision, Indiana University Bloomington

B657
Team 13

1. Motivation

- The Indiana University Libraries Moving Image Archive (IULMIA) has a collection of award-winning digitized television ads from the 1960s and 1970s.
- One challenging use case is requests for footage containing a specific object, such as footage containing rings or jewelry or footage containing a TV.
- Manually searching through thousands of video ads to find specific objects can be time-consuming and laborious.

How do we address this problem?
- use object-detection models to identify the objects
- create tags for the videos based on objects
- use those tags to filter the video search

what challenges do we have?
- Unavailability of historical object labelled datasets
- quality of the videos

2. Practical Problem

- It is difficult for the latest models to detect retro images as they are trained on datasets like COCO.
- Basic Object detection models are trained on datasets with limited classes, making it difficult for them to provide detection labels on retro videos.

FIGURE 1 Ice Cream vs Ice Cream (retro Image)

3. Experimental Design

Initial Evaluation: We have evaluated pre-trained object detection models like YOLO and observed that the detection is good but the classes are limited.

Enhancing Performance: To further improve detection Capability on various classes, a labeled dataset is created using SAM+Grouding DINO Arch and trained on tested on Yolo V8.

Future Work: Using Bigger sample sizes to train yolov8 and integration of Optical Character Recognition (OCR) to capture product names from the videos are promising avenues for future exploration.

4. Dataset

- Currently, there are about 1700 add Videos present in IULMIA.
- For Faster processing we are converting Videos into Frames and removing Duplicates.
- We are now using the proposed Auto distill Arch to generate labeled data sets from unlabeled image frames.
- We are now training our target model with the labeled dataset.

FIGURE 2 The Figure Shows the Labelled Dataset preparation from unlabeled image frames.

5. Proposed Method

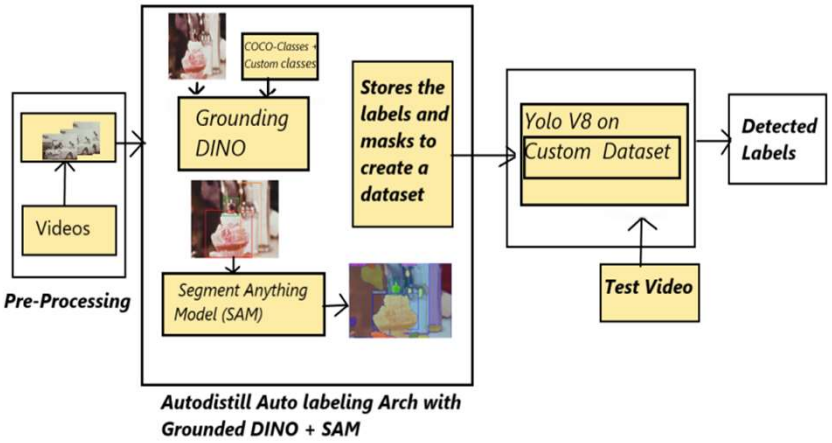

FIGURE 3 The Fig above shows the Implementation of SAM+Grouding DINO Arch and YoloV8 Custom model.

- After trying and testing the object detection using various pre-trained models we felt that it is important to tune these models w.r.t to our use case data for better detection.
- On Working along this thought the Roboflow Auto distill Architecture is identified and used to create image image-labelled datasets using the Grounding DINO + SAM models.
- Later a Yolov8 model is trained on this dataset and its detection capabilities are tested on the images and videos in our use case.
- As part of this model Architecture we also have an option to provide a prompt so that new classes which are not present in COCO can be added and tested.

6. Results So Far

- Managed to extract unique frames from the ad videos
- Experimented with pre-trained object detection models like YOLO.
- Figured out an SAM+DINO and Yolov8 architecture to obtain better detection .

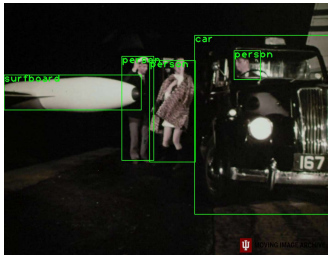

FIGURE i

FIGURE ii

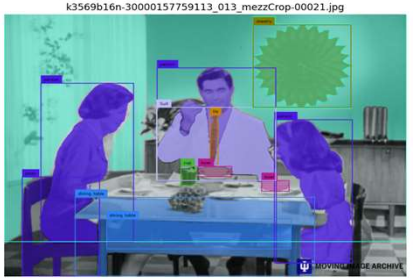

FIGURE .iii

- Fig I and ii show the detection of images on the yolov8 and SSD models pre-trained on COCO.
- The Fig iii shows the detection of the proposed method where there are more no.of classes detected with good accuracy

7. Conclusion & Future Work

- Manually searching for specific objects within these ads can be time-consuming.
- Our project addresses this challenge by leveraging object detection and text recognition.
- This approach creates a labeled dataset out of existing videos by using the proposed DINO+SAM Architecture and the data is later trained and tested using the YoloV8 model.
- Increasing the Size of the Custom Dataset and increasing the Acc.
- Adding OCR on the images to identify text in the images which helps in better search results for the user.

REFERENCES

- Content-based video retrieval in historical collections of the German Broadcasting Archive.
- Semantic video search by automatic video annotation using TensorFlow
- Efficient video annotation with visual interpolation and frame selection guidance.
- RoboFlow Autodistill framework.