Automatic Segmentation And Labelling of Objects in Video

Main Project Presentation: Zeroth Review

Guided by:

Dr. Deepa V.

Presented by:

Batch 4

Ajay T Shaju, SJC20AD004
Emil Saj Abraham, SJC20AD028
Justin Thomas Jo, SJC20AD046
Vishnuprasad KG, SJC20AD063

Outline

- Introduction
- Motivation
- Literature Review
- Objectives
- Block diagram
- Data Description
- Future Scope
- Conclusion
- References

Introduction

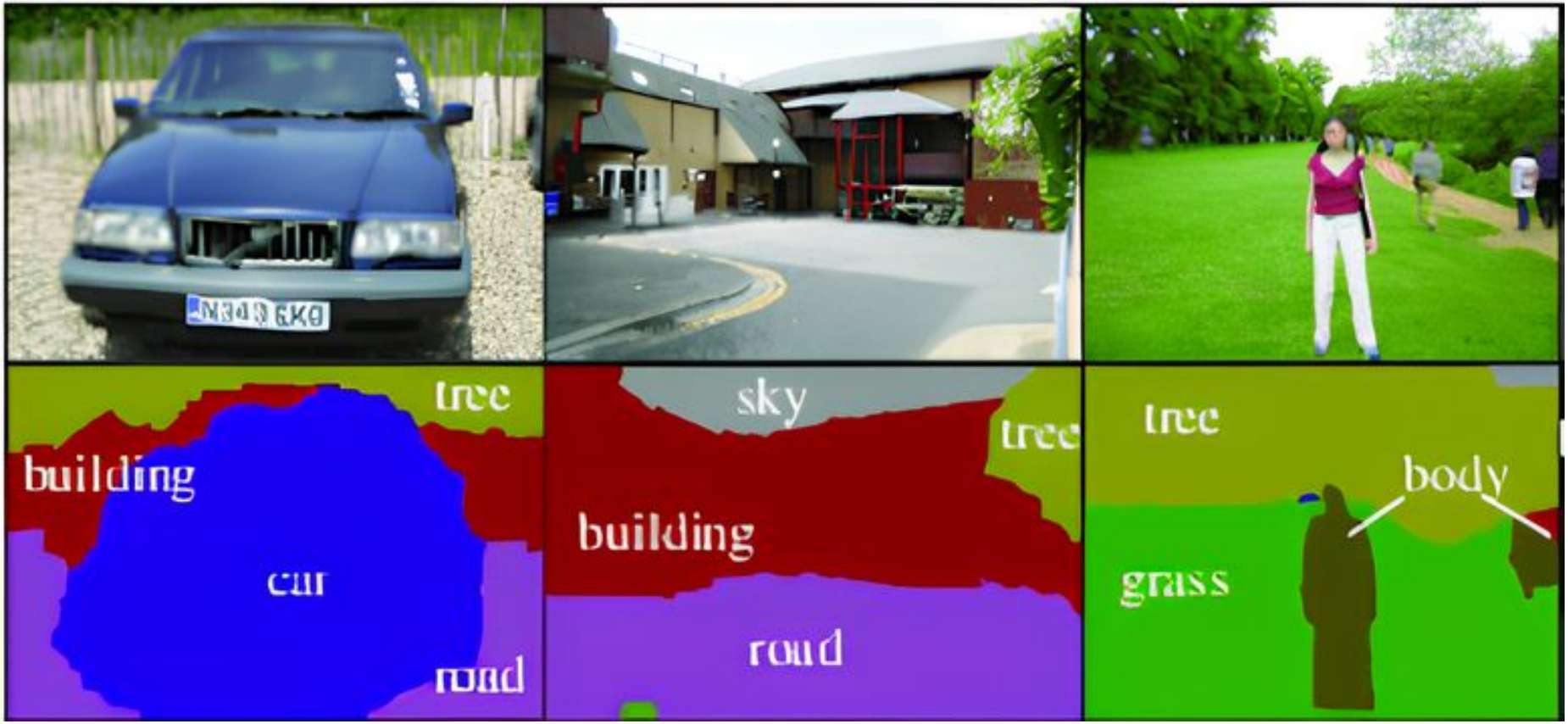
- **How does a Self Driving Car study the environment.**
- **How does a wheeled-robot navigates through it's surroundings?**
- We need to train them on **huge sets of labelled data, a huge manual work.**

→ **Automating Segmentation And Labelling of Video Data.**

- Object segmentation: partitioning an image into multiple regions or segments, where each segment corresponds to a distinct object or region of interest.
- Labeling: assigning a category or label to each segmented object.

Raw Video

After Segmentation



Labelling on segmented Image

Motivation

- Self-driving vehicle development is an exciting challenge, but **achieving full autonomy** (level 5) is hindered by the need for vast amounts of **labeled data**.

- Efficient labeling tools are crucial for AI algorithms, yet current methods require substantial **human intervention**.

- The goal is to develop **automatic labeling algorithms** to expedite progress towards **environment-aware navigating objects**.

Literature Review

[1] Yao, Rui, et al. **"Video object segmentation and tracking: A survey."** ACM Transactions on Intelligent Systems and Technology (TIST) 11.4 (2020): 1-47.

- Solved the difficulties in handling fast motion, out-of-view, and real-time processing by video object segmentation and tracking(VOST)
- Tried with different learning methods like unsupervised VOS, semi-supervised VOS, interactive VOS, and segmentation-based tracking methods.
- **Gap:** Affected by low resolution, motion blur. Segmentation is not flexible with object shape.
- **Future Scope**: Multi-camera video object segmentation and tracking, 3D video object segmentation and tracking.

Literature Review

[2] Yiwen Wang; Ye Lyu; Yanpeng Cao; Michael Ying Yang **“Deep Learning for Semantic Segmentation of UAV Videos”** IEEE International Symposium on Geoscience and Remote Sensing (2019)

- Proposed model combines FCN & LSTM for segmentation
- FCN segments each frame individually
- LSTM acts as the post processing method that uses temporal information of consecutive frames
- **Gap:** Noise in the frame reduces accuracy. Resized images shows low performance
- **Future Scope:** Noise reduction algorithms and better hardware equipment could significantly increase Accuracy

Literature Review

[3] Yang, Linjie, Yuchen Fan, and Ning Xu. "**Video instance segmentation.**" Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

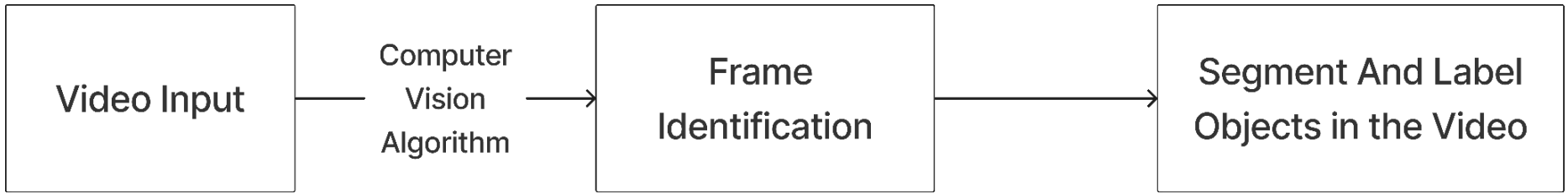
- Presented the first large scale dataset, **YouTube-VIS**, for video instance segmentation
- Proposed **MaskTrack R-CNN** for video instance segmentation.
- **Gap:** unable to associate objects due to object occlusions and fast motion.
- **Future Scope:** object detection with spatial-temporal features, end-to-end trainable matching criterion, and incorporating motion information for better recognition.

Objectives

- Focus on robustness.
- Achieve real-time performance on real-world videos.
- Evaluate on benchmark datasets.
- Create user-friendly tool.

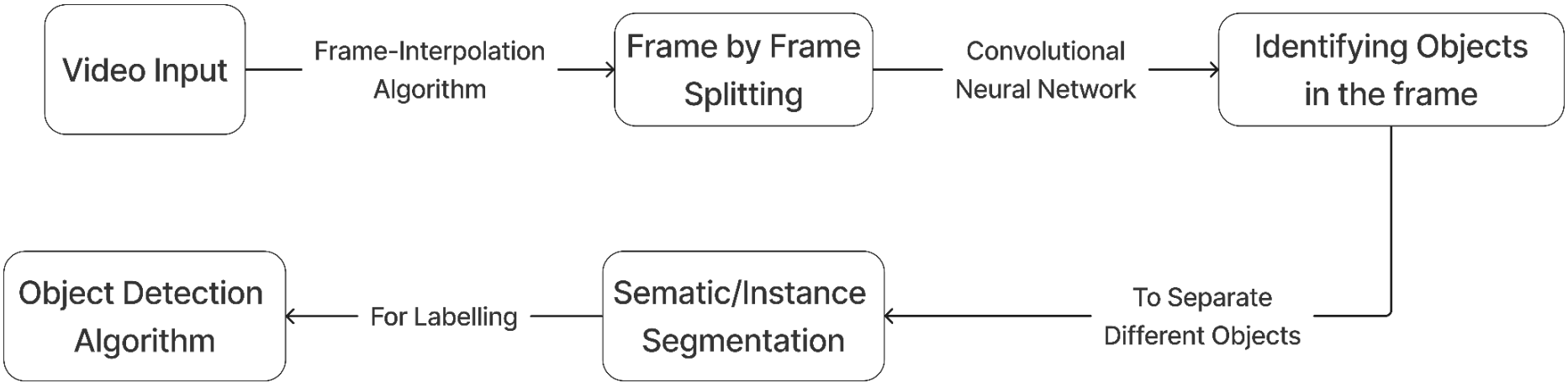
Block Diagram

High-Level Overview Of The System



Block Diagram

Detailed View Of The System



Data Description

- Primary data - Video files of any formats.
- Data Collection Methods:
  - Manual data collection by taking videos ourselves.
  - From major datasets like CamVid(Cambridge-Driving Labelled Video Database) or DAVIS 2016.
    - Real-World Urban Videos – highly relevant for autonomous driving tasks.
    - Diverse Environmental Conditions – lighting and colors.
    - Benchmark dataset for video segmentation algorithms.

Future scope

- **Integrate labeling algorithms** into **autonomous vehicles** for easy and immediate access to enhanced labeling tools.
- Upgrade the labeling algorithms to **label objects in real-time**, helping self-driving cars **adapt to evolving road conditions**.
- Customize for uses like **robotics**, **industrial automation**, and **healthcare**, where precise labeling is crucial.
- Extend the labeling algorithms to **handle multimodal data fusion**, combining information from various sensors to generate **more accurate** object annotations.

Conclusion

- The project **targets a major challenge** in **self-driving** vehicle development.
- Development and optimization of state-of-the-art algorithms to **automate labeling**.
- Enhancing accuracy, speed, and scalability for **efficient labeling**.
- Acceleration of progress towards **fully autonomous** vehicles by revolutionizing the data labeling process.

References

[1] Yao, Rui, et al. **"Video object segmentation and tracking: A survey."** ACM Transactions on Intelligent Systems and Technology (TIST) 11.4 (2020): 1-47.

[2] Yiwen Wang; Ye Lyu; Yanpeng Cao; Michael Ying Yang **“Deep Learning for Semantic Segmentation of UAV Videos”** IEEE International Symposium on Geoscience and Remote Sensing (2019)

[3] Yang, Linjie, Yuchen Fan, and Ning Xu. **"Video instance segmentation."** Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[4] Caelles, Sergi, et al. **"The 2019 davis challenge on vos: Unsupervised multi-object segmentation."** arXiv preprint arXiv:1905.00737 (2019).

[5] J. Usón, J. Cabrera, D. Corregidor and N. García, "**Analysing Foreground Segmentation in Deep Learning Based Depth Estimation on Free-Viewpoint Video Systems"** *2022 IEEE 12th International Conference on Consumer Electronics (ICCE-Berlin)*, Berlin, Germany, (2022)

[6] H. Liang, L. Liu, Y. Bo and C. Zuo, **"Semi-Supervised Video Object Segmentation Based on Local and Global Consistency Learning,"** in IEEE Access, vol. 9, pp. 127293-127304, (2021)

[7] Kai Xu, Longyin Wen, **“Self-Supervised Deep TripleNet for Video Object Segmentation”**, IEEE Transactions on Multimedia, pp: 1-11, 2020.

[8] J.Park, C.Lee, Chang-Su Kim, **“Asymmetric Bilateral Motion Estimation for Video Frame Interpolation”**, arXiv, pp:1-10, Aug 2021.

[9] W.Choi, Y.Jun Koh, Chang-Su Kim, **“Video Frame Interpolation Based on Symmetric and Asymmetric Motions”**, IEEE Access, Vol. 11, pp: 22394-22403, Feb. 2023.

Questions?

Thank You

Presentation Setting

Opening, Intro, Motivation - Ajay

Review of Lit - Justin

Objectives, Block diagram - Emil

Data Description, Future Scope - Vishnu

Conclusion, References - Ajay

Don’t Present this

Don’t Present this