Video Analytics Benchmarks Utilizing Gstreamer and Nvidia Deepstream on HPC and Edge Computing Systems

Intern's Name: Calvin Van
Mentor(s): Steven Glandon & Warren Reed Williams

Project# ERDC-ITL-2023-0011

Introduction

| School | University of Texas at Dallas

| Major | Computer Science

| Past Research Experience |

This is my first research experience.

| Future Plans | To continue on with my Bachelor's Degree

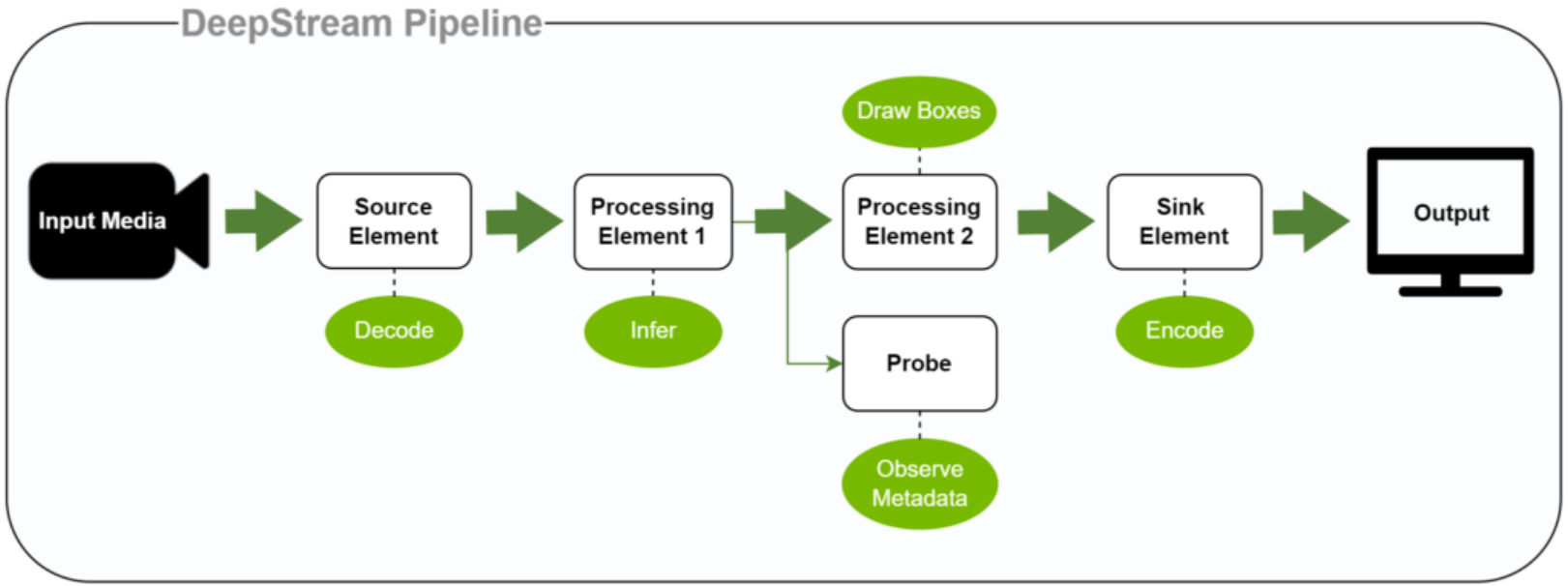
(Rising Sophomore)

| Tell us why you applied for the HIP internship |

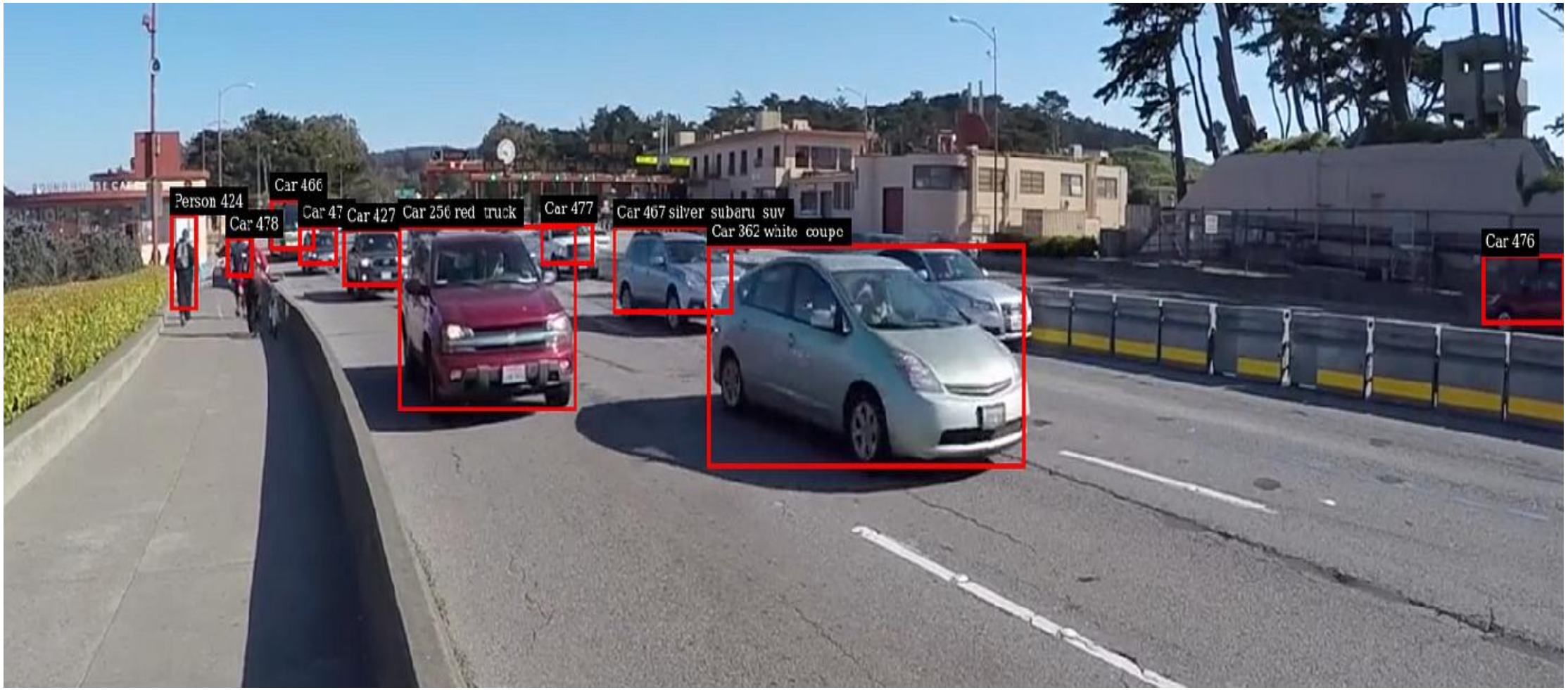
It was a great opportunity for me to get first-hand experience what it’s like to work in the field as a computer scientist!

Project Background

- Nvidia's Deepstream Video Pipelines for video inferencing.
- Implementation of previously used machine learning models into Deepstream pipelines.



Pipeline Example



Project Objectives

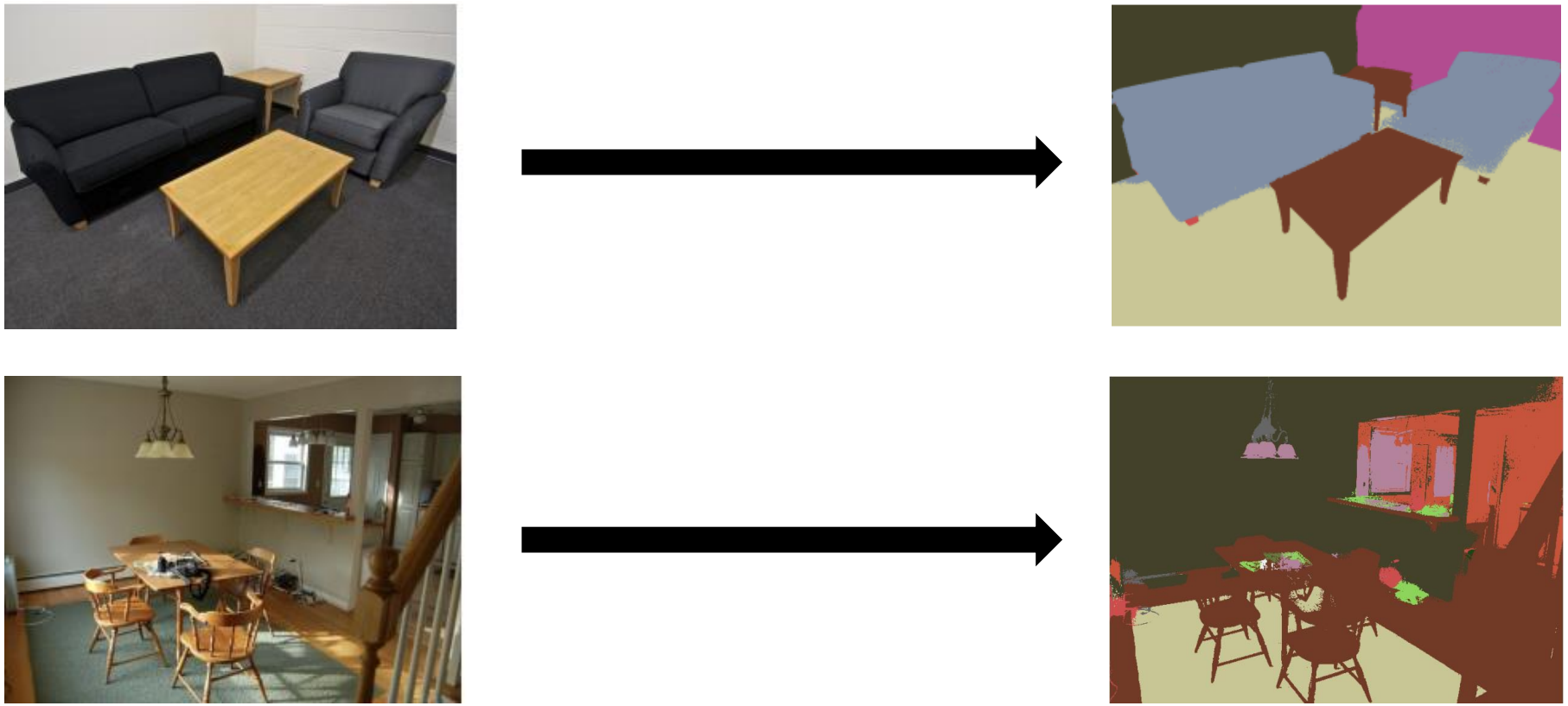
- To explore the capabilities of Nvidia's Deepstream Pipelines.
- To research and experiment with implementing two machine learning models (ESRGAN and Material Segmentation)

The ESRGAN

- **Super-resolution image model**

MINC Segmentation

- **Model that segments materials based on MINC database.**

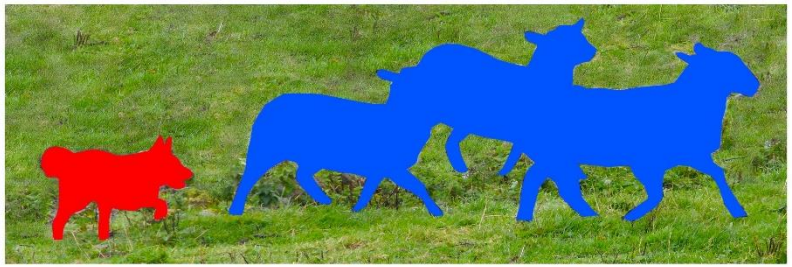


Design and Methods

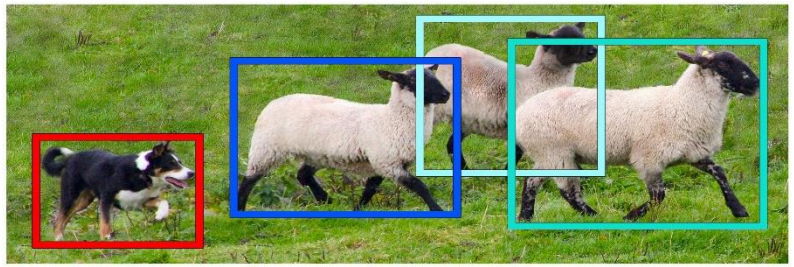
Gst-Nvinfer

- Pipeline inference component
- Four main inference types: object detection, object classification, semantic segmentation and instance segmentation

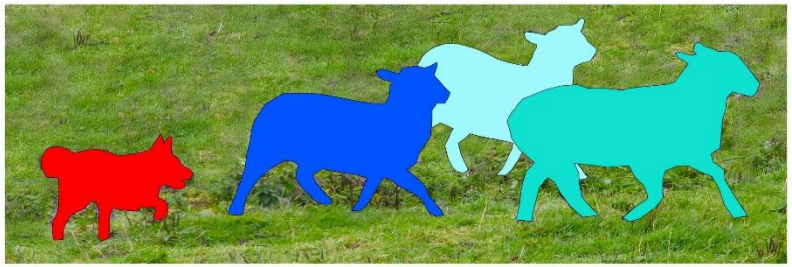
Image Recognition



Semantic Segmentation



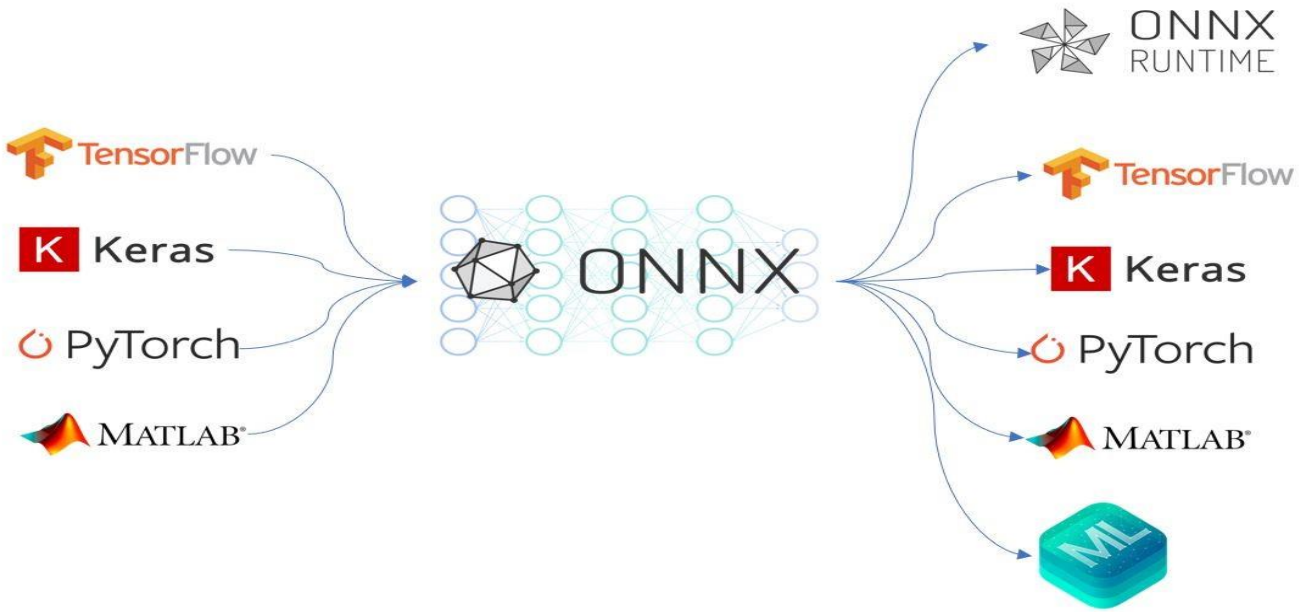
Object Detection



Instance Segmentation

Test Onnx model

- One of the three types of machine learning frameworks used for nvinfer.
- Is theoretically the “easiest” as it requires the least number of configurations to implement.

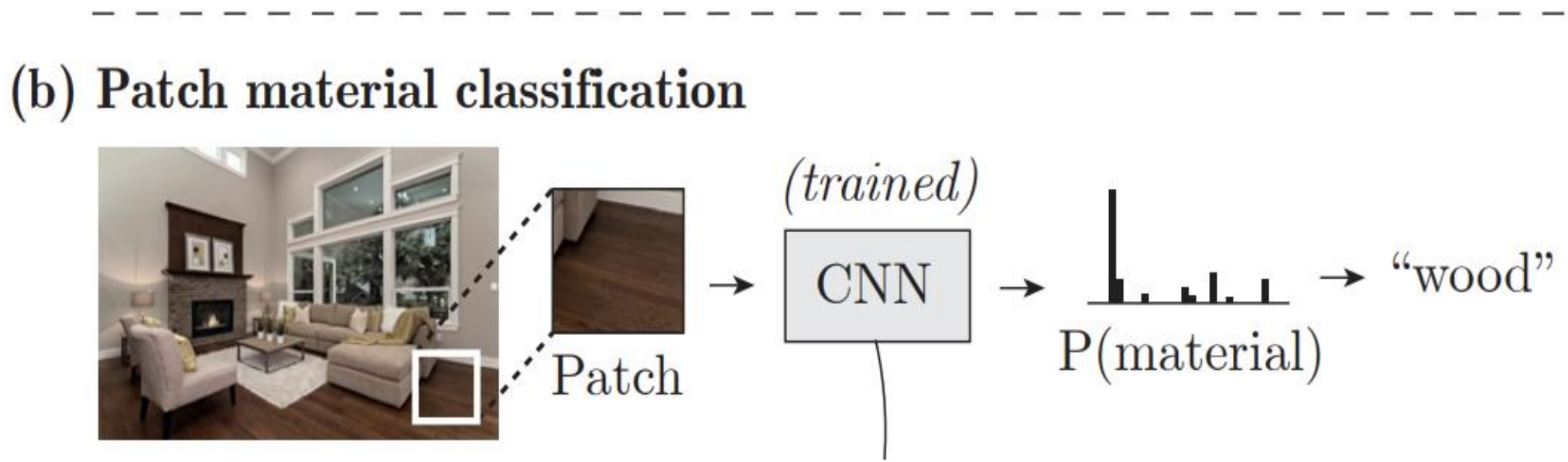


Implement Models into Deepstream

- **Research and experiment with Nvidia Deepstream pipelines to see whether the two target models can be implemented into the pipelines.**

Material Segmentation

- The material “segmentation” model is actually a classification model.
- As a result, a different approach must be done



Nvdspreprocess

Create Custom Segmentation Model

- Code a program that can train a custom multiclass segmentation model.
- Use the MINC database to create a semantic segmentation inference model

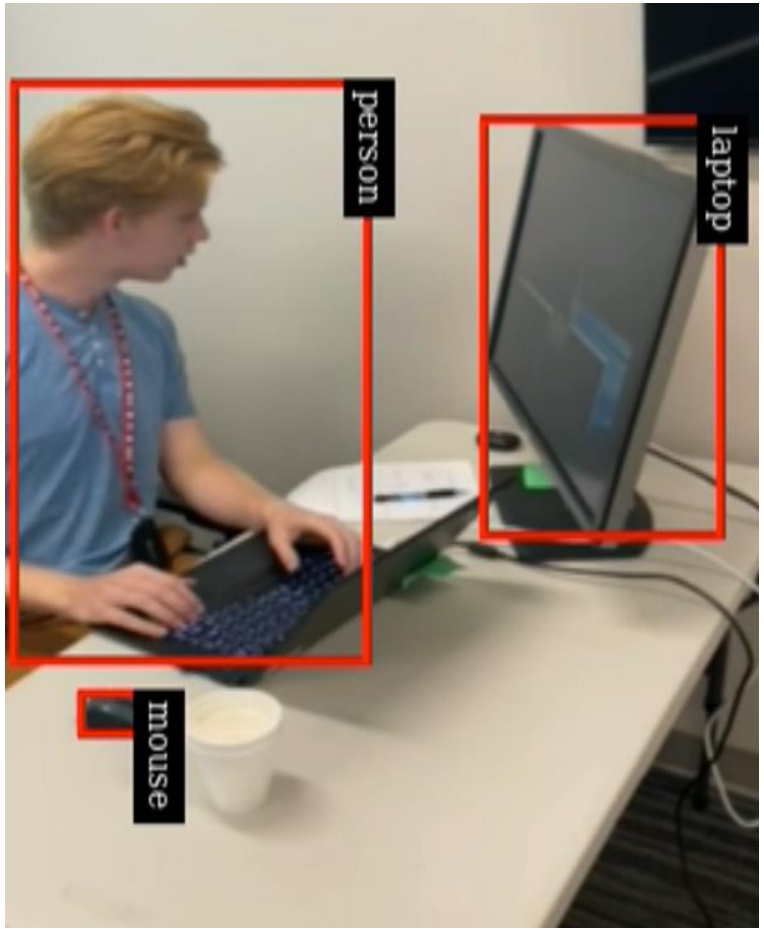
Data / Results

Concerns about Nvinfer

- **Custom libraries are needed for three of the main inference types**
  - object detection and classification, and instant segmentation
- **These inference types look for individual and specific objects and features.**
- **This can pose a challenge as they must be coded.**

Onnx Implementation

- **Was successfully able to convert and deploy an object detection model into the pipeline.**



Concerns with Onnx

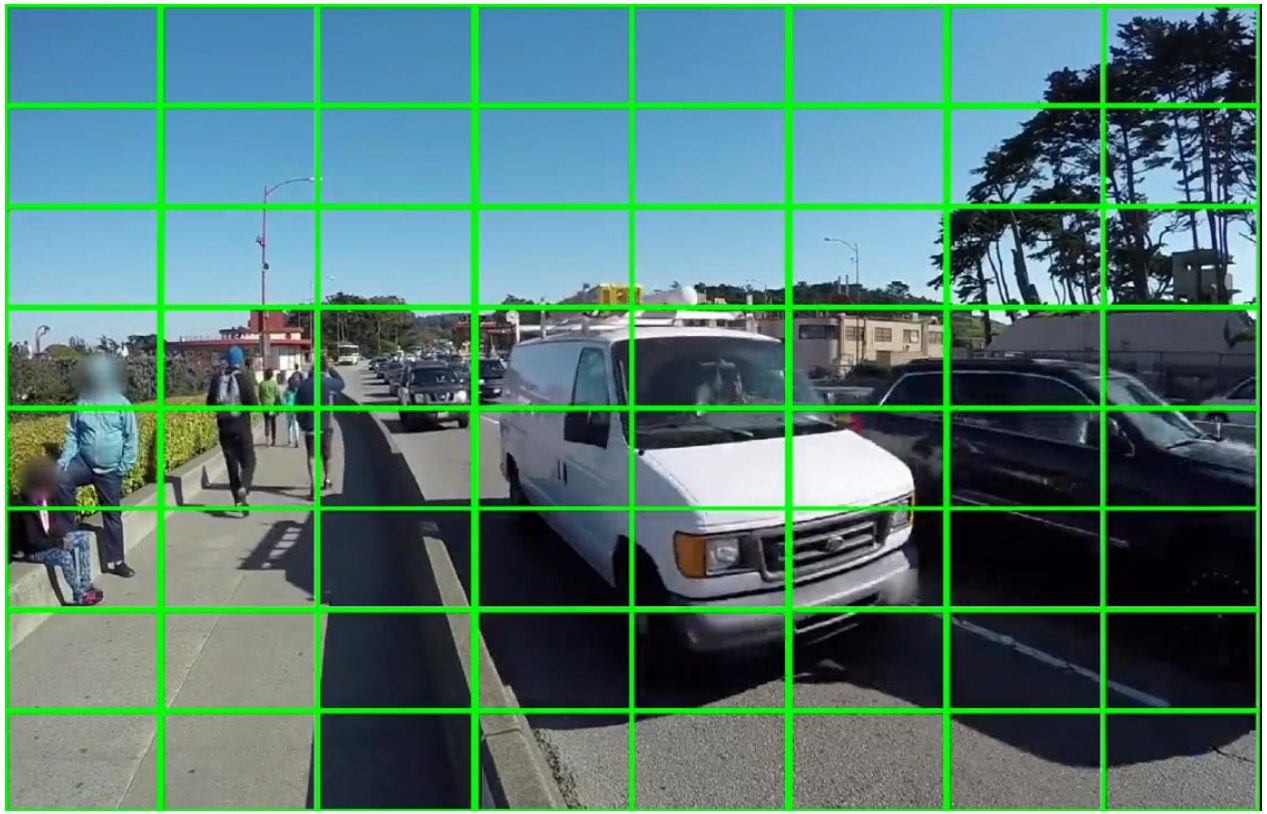
- **Converting onnx can vary in difficulty depending on the inference type.**
  - This entails how many inputs and outputs one may have to specify to convert the model to onnx.
- **It is crucial to know the shape of the input that is sent into the model.**

ESRGAN Implementation

- **Proved to be a difficult task as it is outside the function of nvinfer.**
- **Would possibly require the use of another component called nvdsvideotemplate.**
  - An obstacle to this would be that a custom library is needed to implement this component.

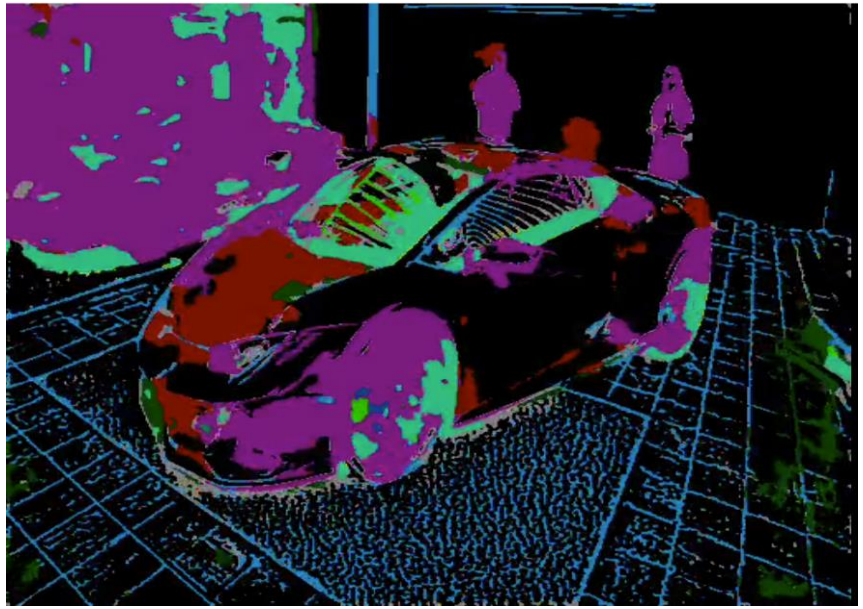
Concerns with Nvdspreprocess

- **Regarding the material segmentation model, it would require too many inputs.**
- **Leads to a stop on the pipeline.**



Segmentation Model Training

- Training the model appears to be the best route.
- Would possibly allow other datasets to be trained through this method.
- Issues would be that said datasets need many images to train from.



Discussion and Conclusions

Nvidia Deepstream Potential

- Deploying current pretrained models that fit within nvinfer inference types can be simple
- Custom models take a lot more time and effort.
- Can have a great use in searching and security with object detection capabilities.

Acknowledgments & References

- Mentors: Steven Glandon & Warren Reed Williams
- Dr. Cheng & Motorny Sergey
- High performance Computing Modernization Program (HPCMP)
- HPC Internship Program (HIP) sponsorship
- Oak Ridge Institute for Science & Education (Orise)
- Fellow Interns: Will Farthing, Raahul Pathak, and Han Saim Jeong.

Any questions?