HCMUTE

fie.hcmute.edu.vn

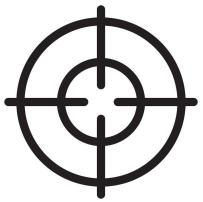
GRADUATION PROJECT

DEVELOPMENT OF A MULTIMODAL 3D OBJECT RECOGNITION AND VISUALIZATION SYSTEM FOR TRAFFIC ENVIRONMENTS USING CAMERA AND LIDAR FUSION

Lê Hoàng Tiến	21145617
Lê Hồ Minh Khoa	21145019
Supervisor	Trần Vũ Hoàng, PhD.

TABLE OF CONTENT

1. INTRODUCTION



1.4. Target

1.1. Problem Statement

Intro

1.3. Scope & Limitations

1.2. Research Objectives

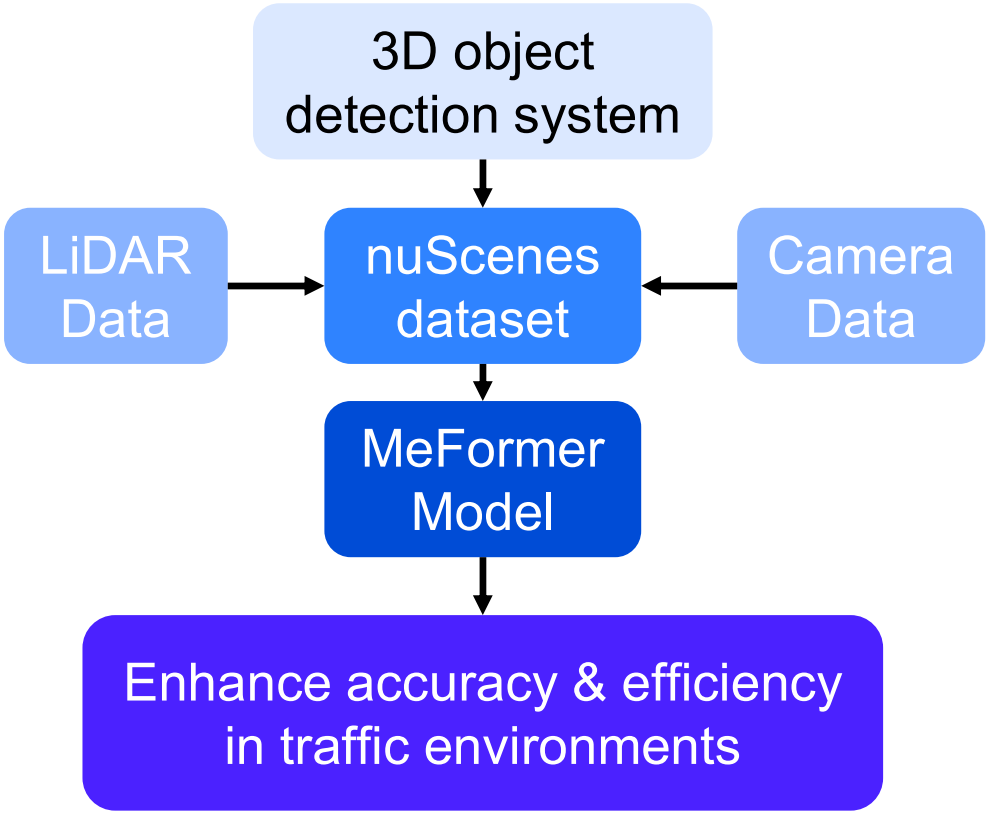
1.1. PROBLEM STATEMENT

1.1. PROBLEM STATEMENT

Low Lighting Scenario

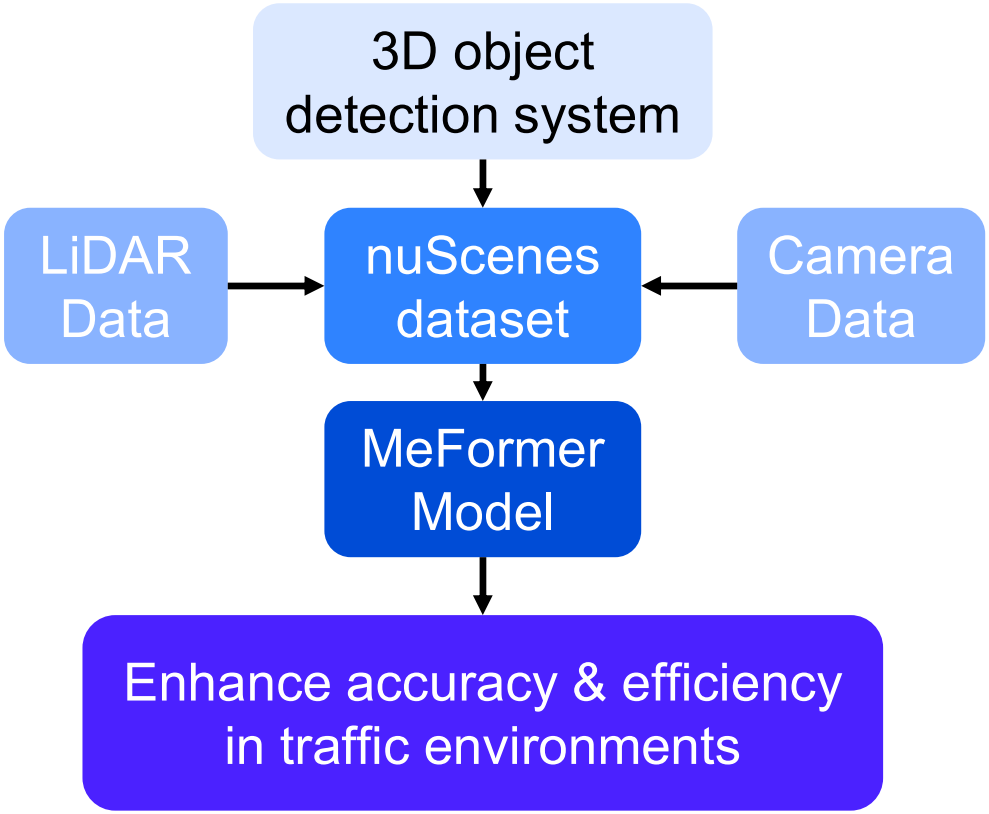
1. INTRODUCTION

1.2. RESEARCH OBJECTIVES



1. INTRODUCTION

1.2. RESEARCH OBJECTIVES



1.3. SCOPE & LIMITATIONS

DATA:
- nuScenes mini (10 scenes)
➔Restricted to specific conditions

HARDWARE:
- No real LiDAR or GPU.
➔Used cloud VPS not real-time.

TIME:
- Only 5 Months
➔No hardware optimization.

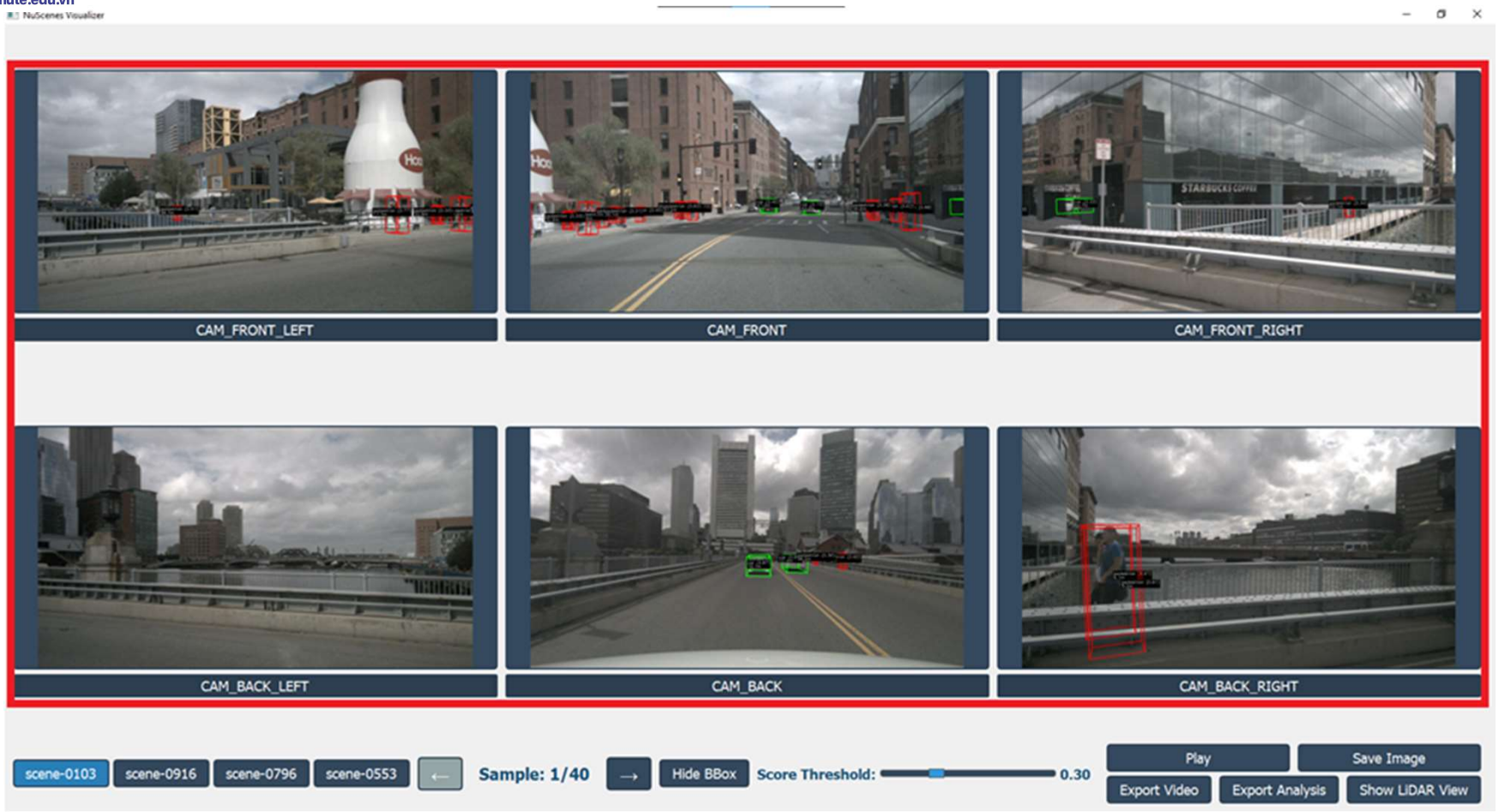
SCOPE:
- Offline 3D detection only.
➔No real-time integration.

1.4. TARGET

Final Product

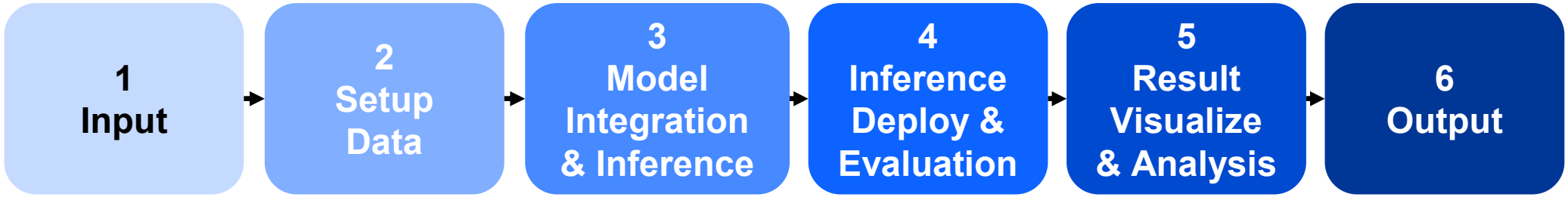
- **Prototype System**
  - An end-to-end multimodal data fusion pipeline
    - 3D object detection
- **Interactive App**
  - Visualizing
  - Evaluating
  - Exporting

1.4. TARGET

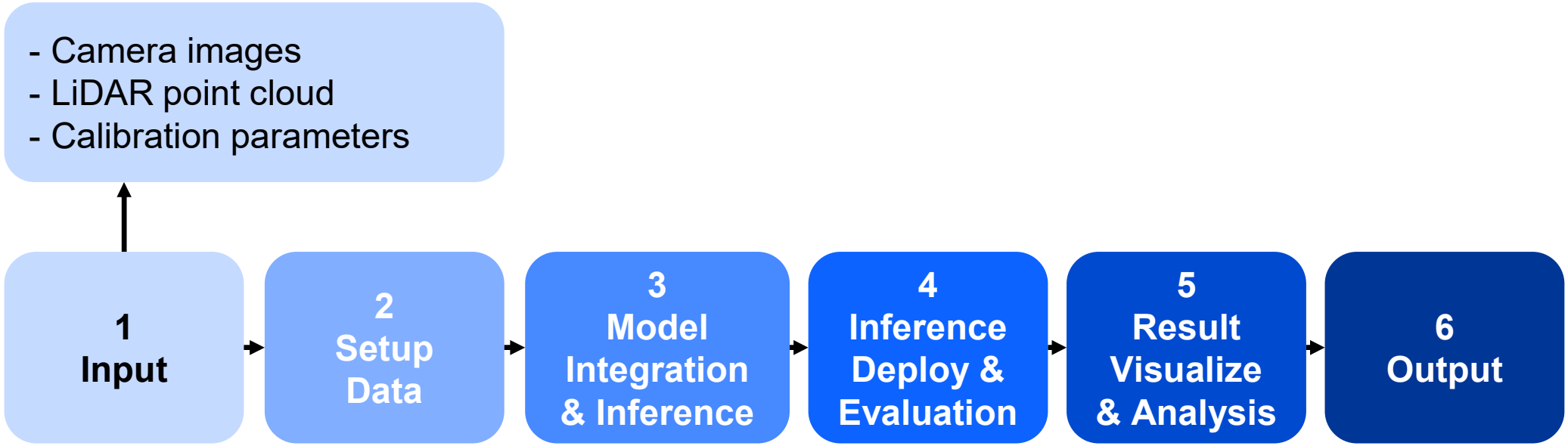


Main User Interface of the Application

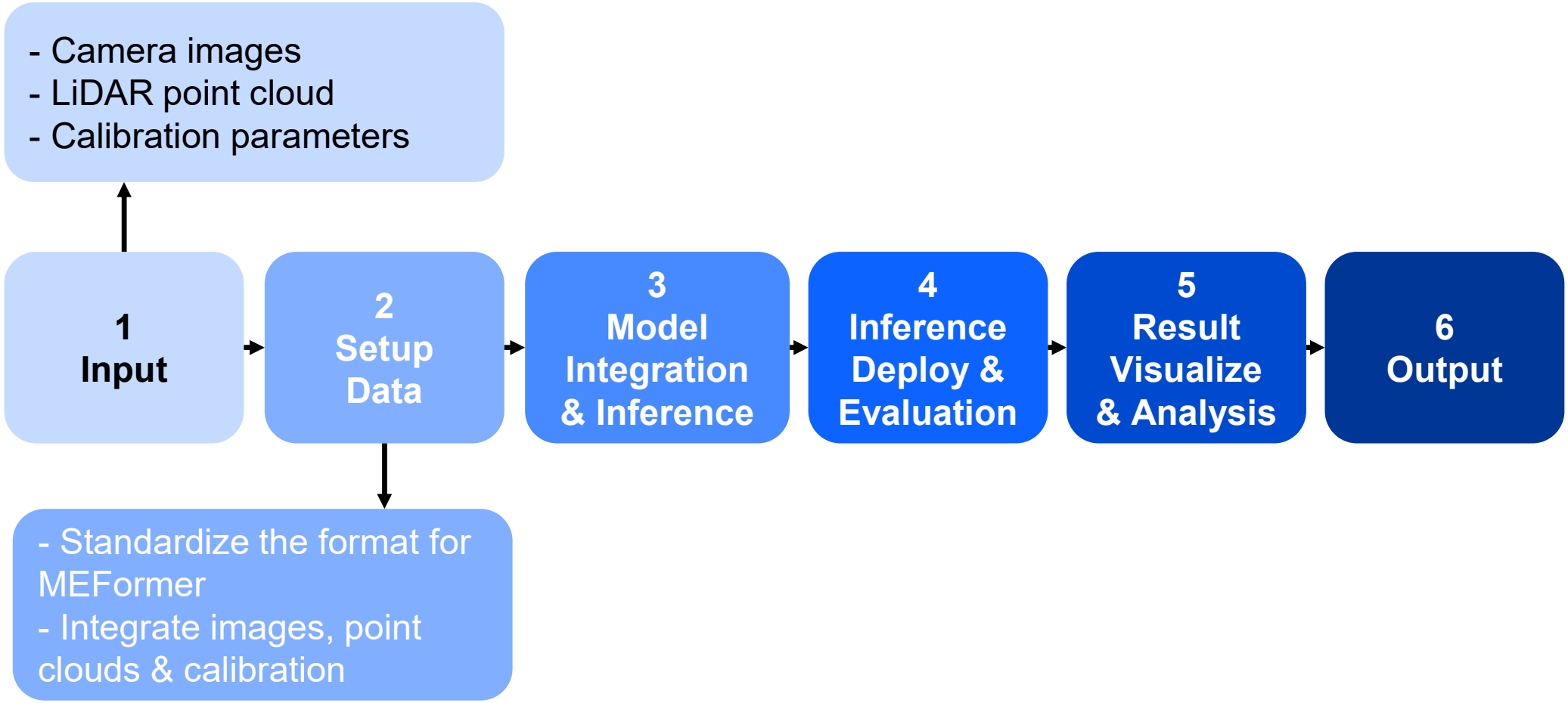
2. OVERVIEW OF PIPELINE



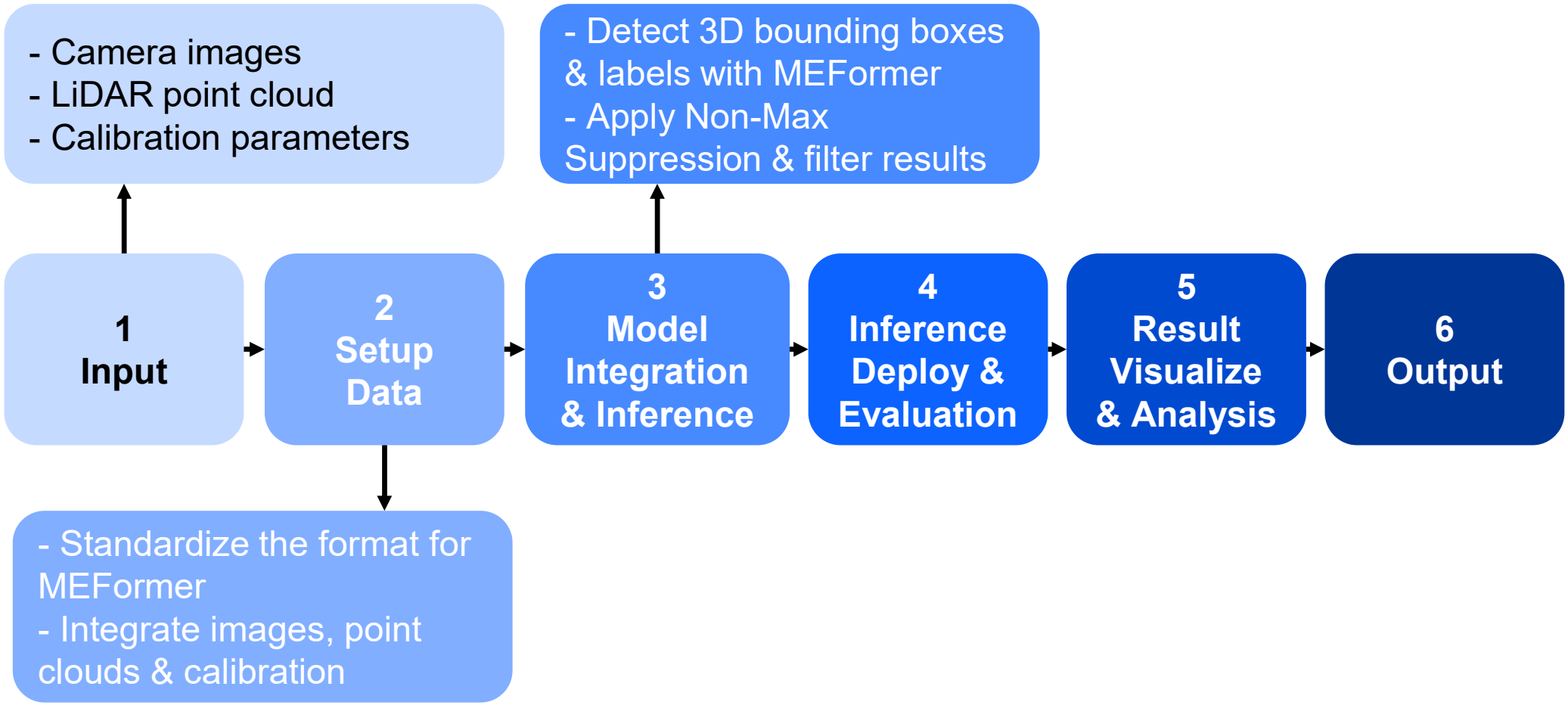
2. OVERVIEW OF PIPELINE



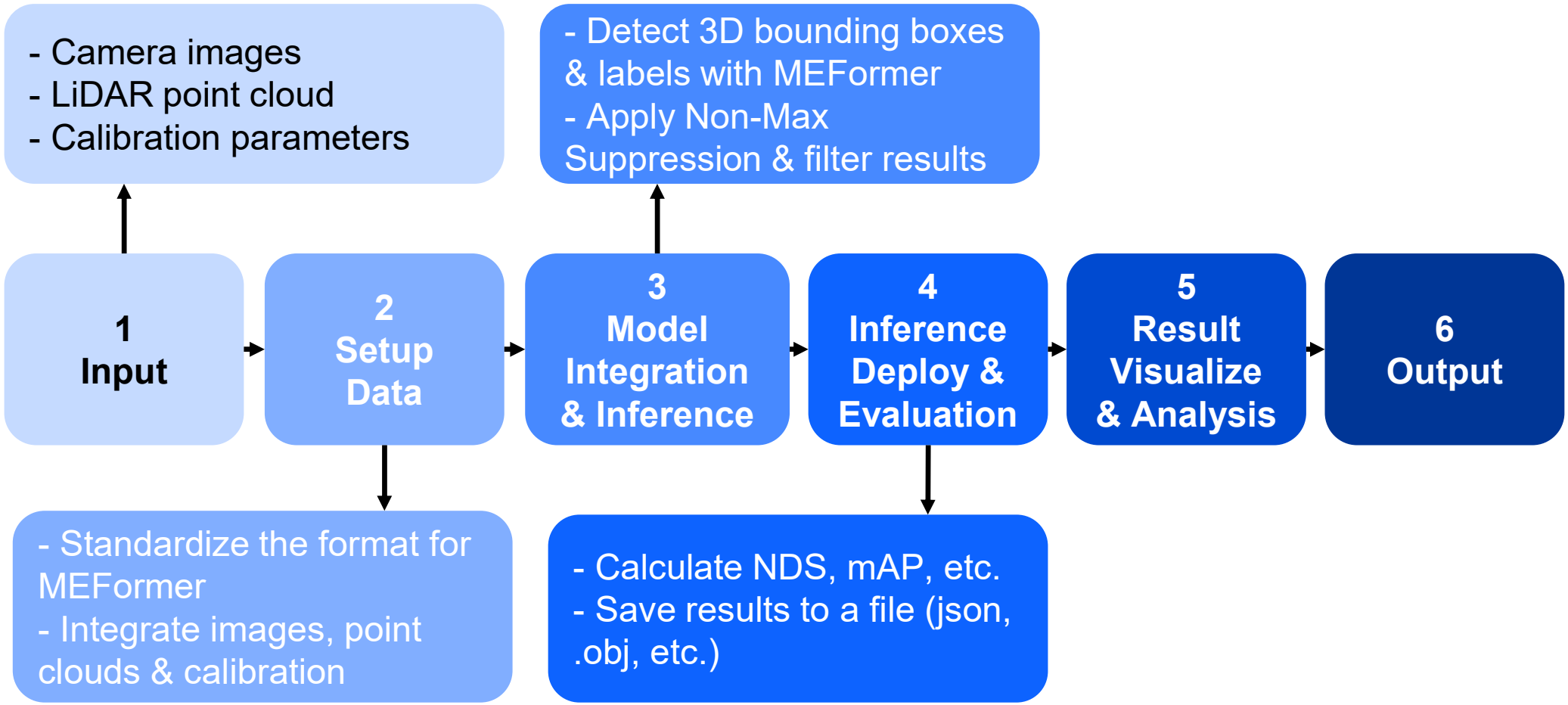
2. OVERVIEW OF PIPELINE



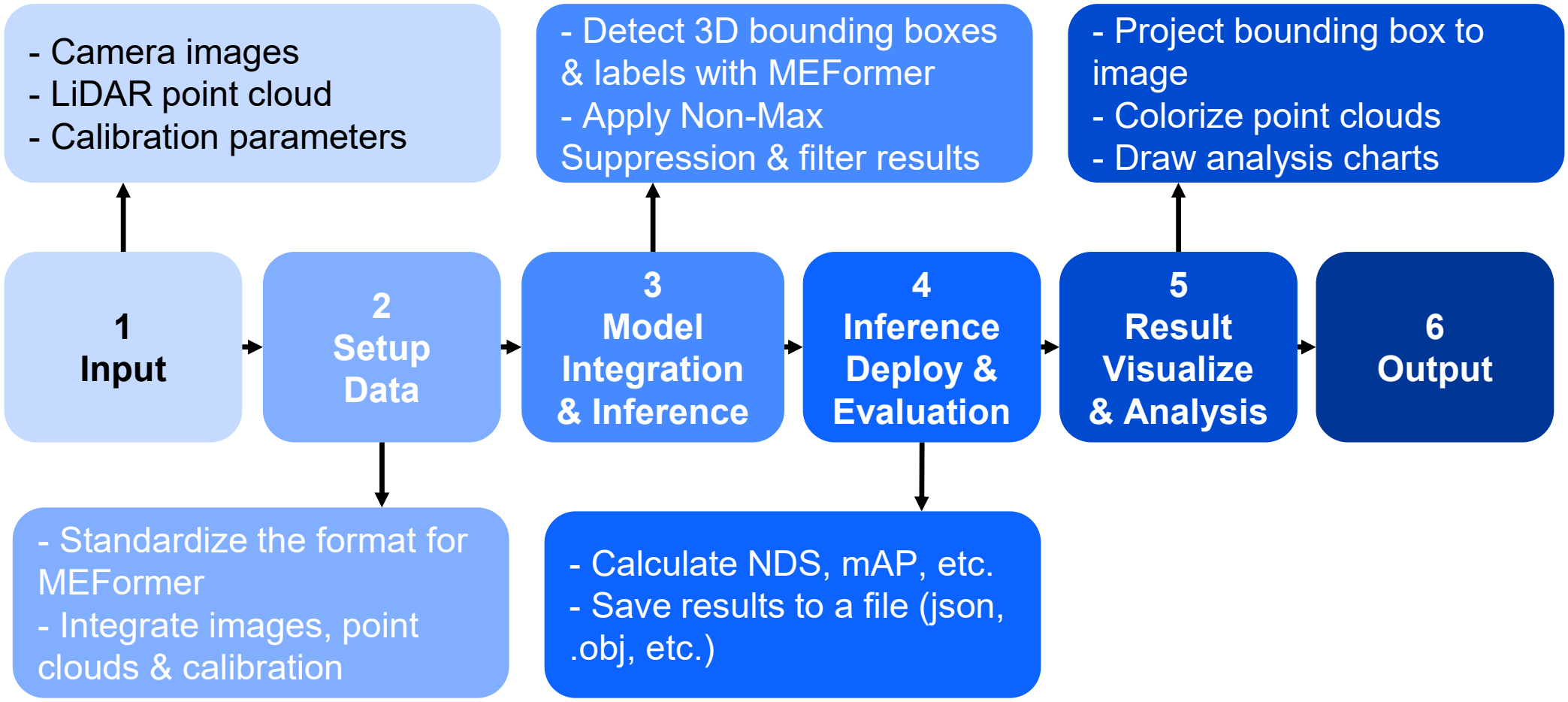
2. OVERVIEW OF PIPELINE



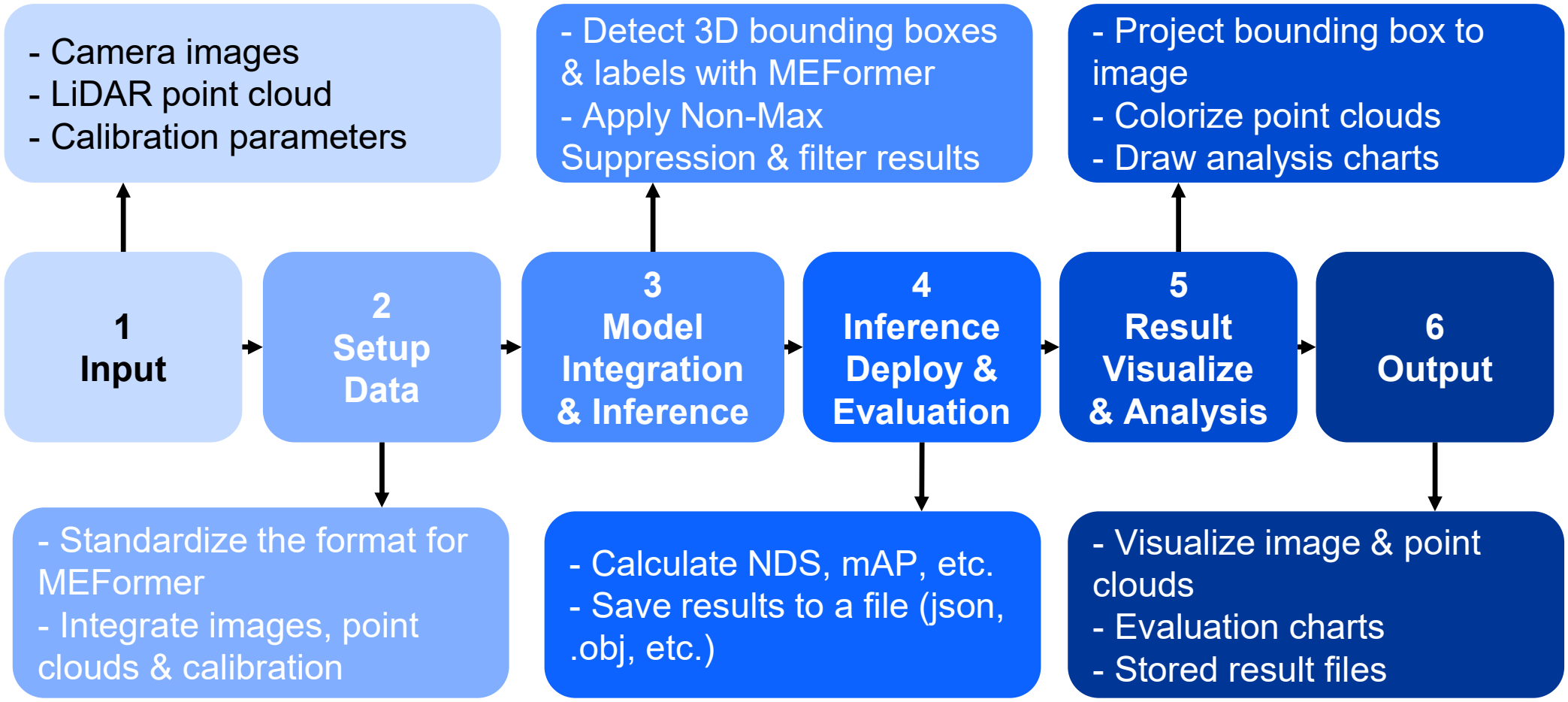
2. OVERVIEW OF PIPELINE



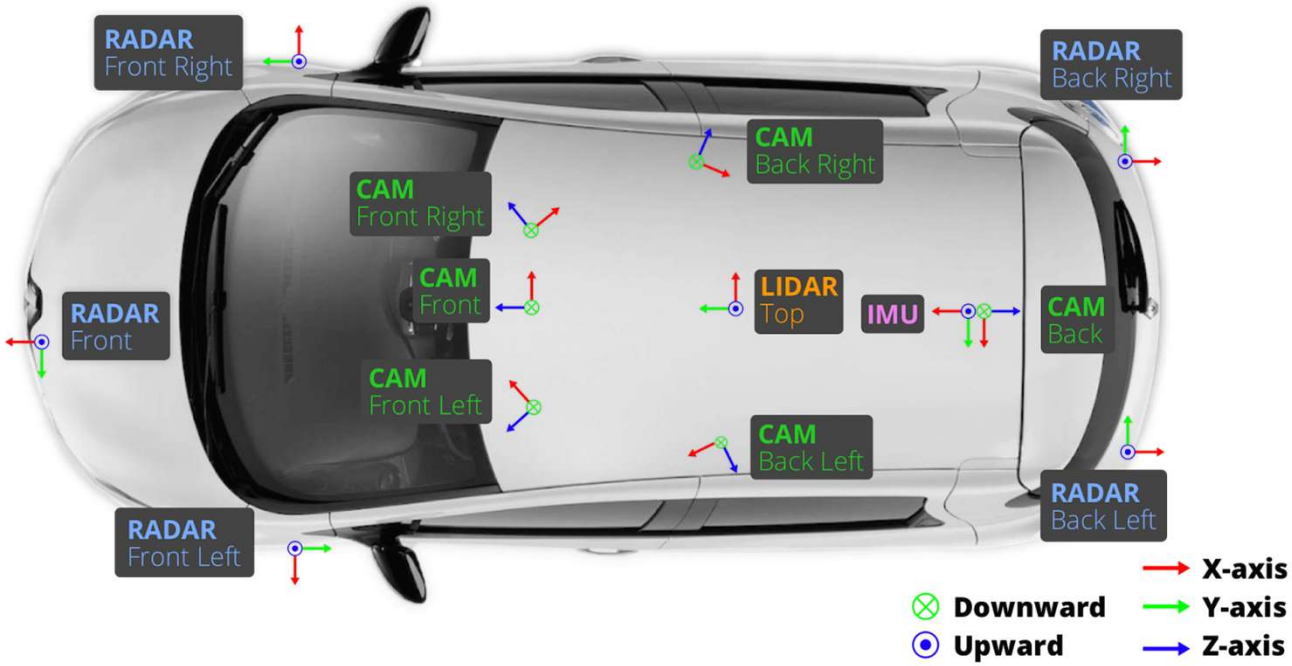
2. OVERVIEW OF PIPELINE



2. OVERVIEW OF PIPELINE



2.1. INPUT & SETUP DATA



Autonomous Vehicle Sensor Layout in nuScenes Dataset

6 CAMERAS

- Front
- Front-left
- Front-right
- Rear
- Rear-left
- Rear-right

1 LIDAR

- A 32-channel device

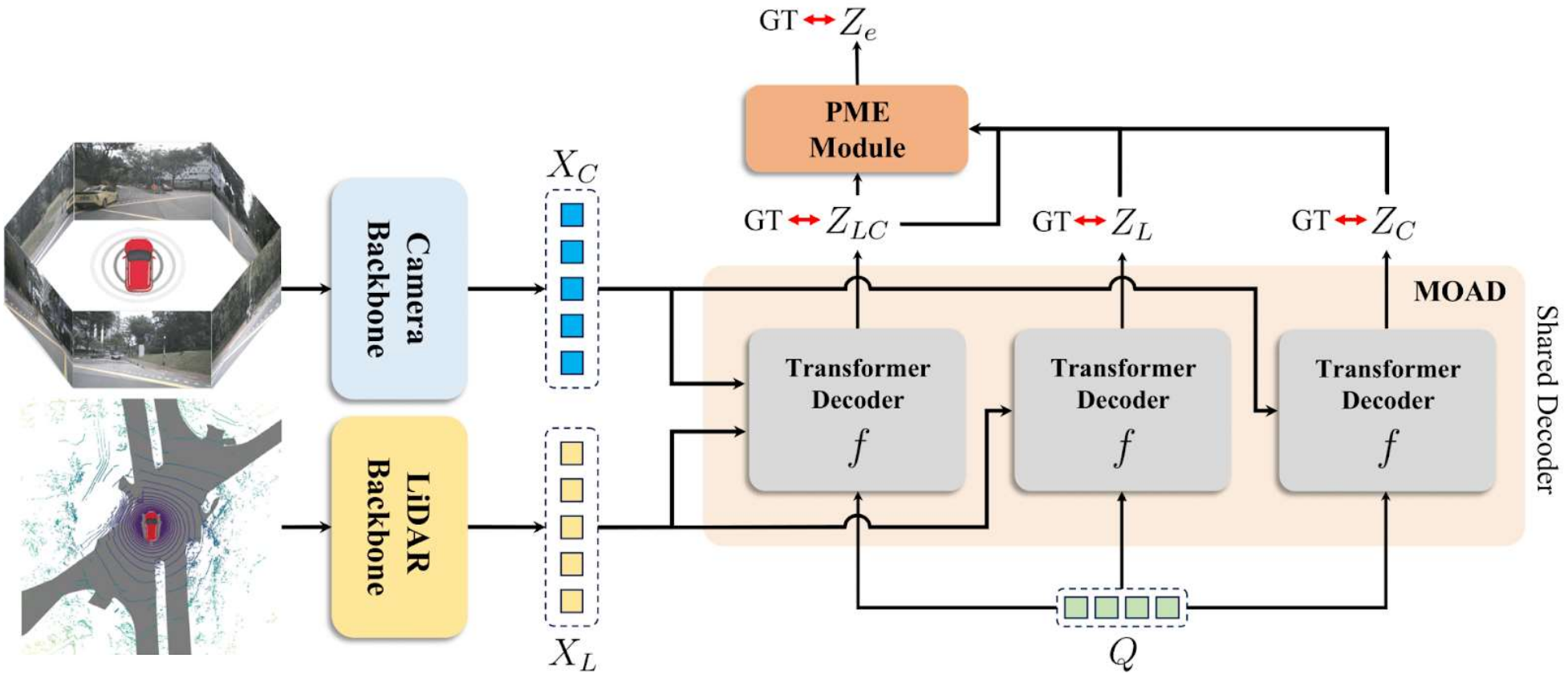
5 RADARS

- Front & 4 Edges

1 IMU

- On the roof

2.2. MODEL INTEGRATION & INFERENCE



The overall architecture of MEFormer

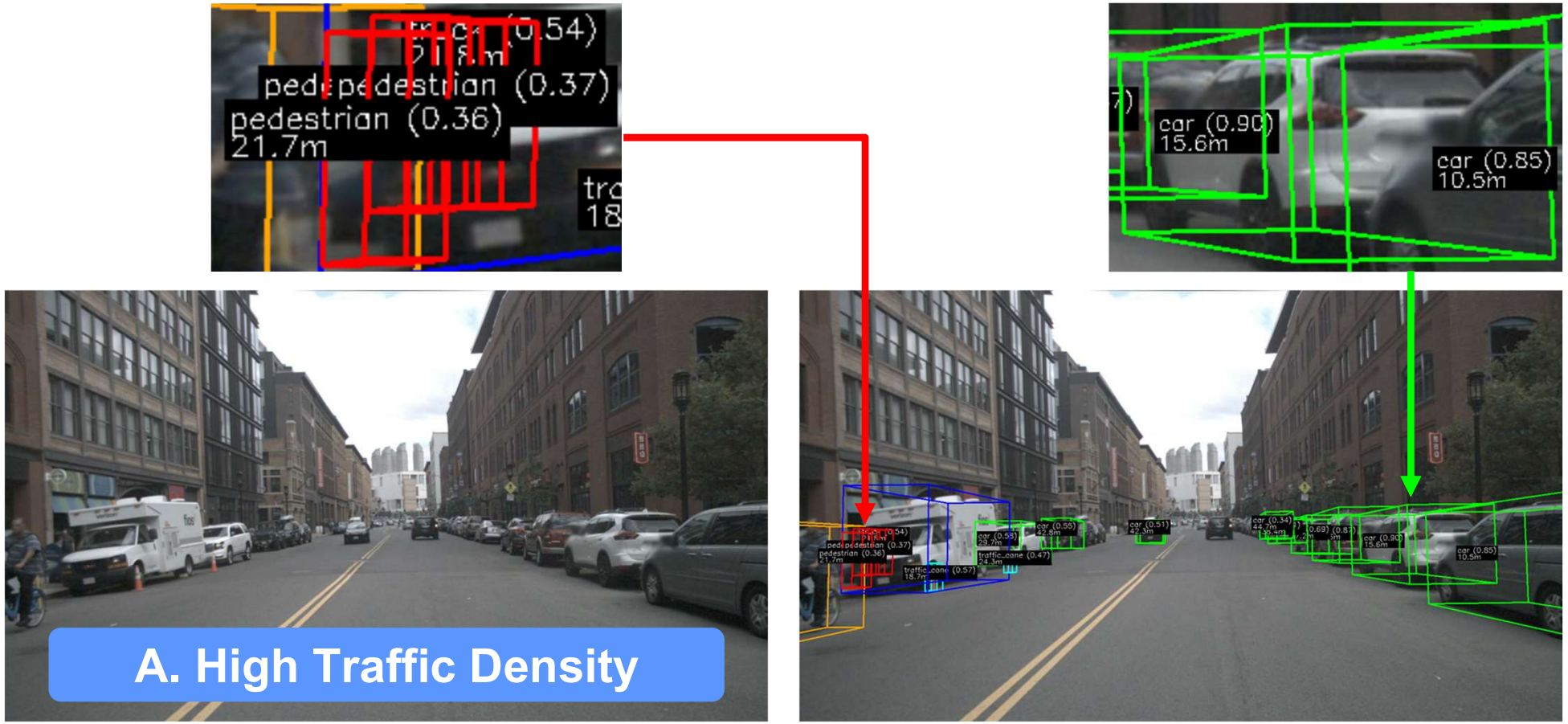
HCMUTE

2.2. MODEL INTEGRATION & INFERENCE

MODELS	mAP (%)	mATE (m)	mAOE (rad)
BEVFormer (C)	44.5	0.58	0.38
DETR3D (C)	41	0.64	0.39
BEVFusion (L+C)	71.3	0.25	0.36
MEFormer (L+C)	**72**	**0.27**	**0.3**
Expected Metrics	≥ 53	≤ 0.49	≤ 0.38
Result	**74**	**0.27**	**0.21**

Performance Comparison of 3D Object Detection Models

2.3. EVALUATION AND VISUALIZATION



Scene-0103, Sample 34 with threshold > 0.3

2.3. EVALUATION AND VISUALIZATION

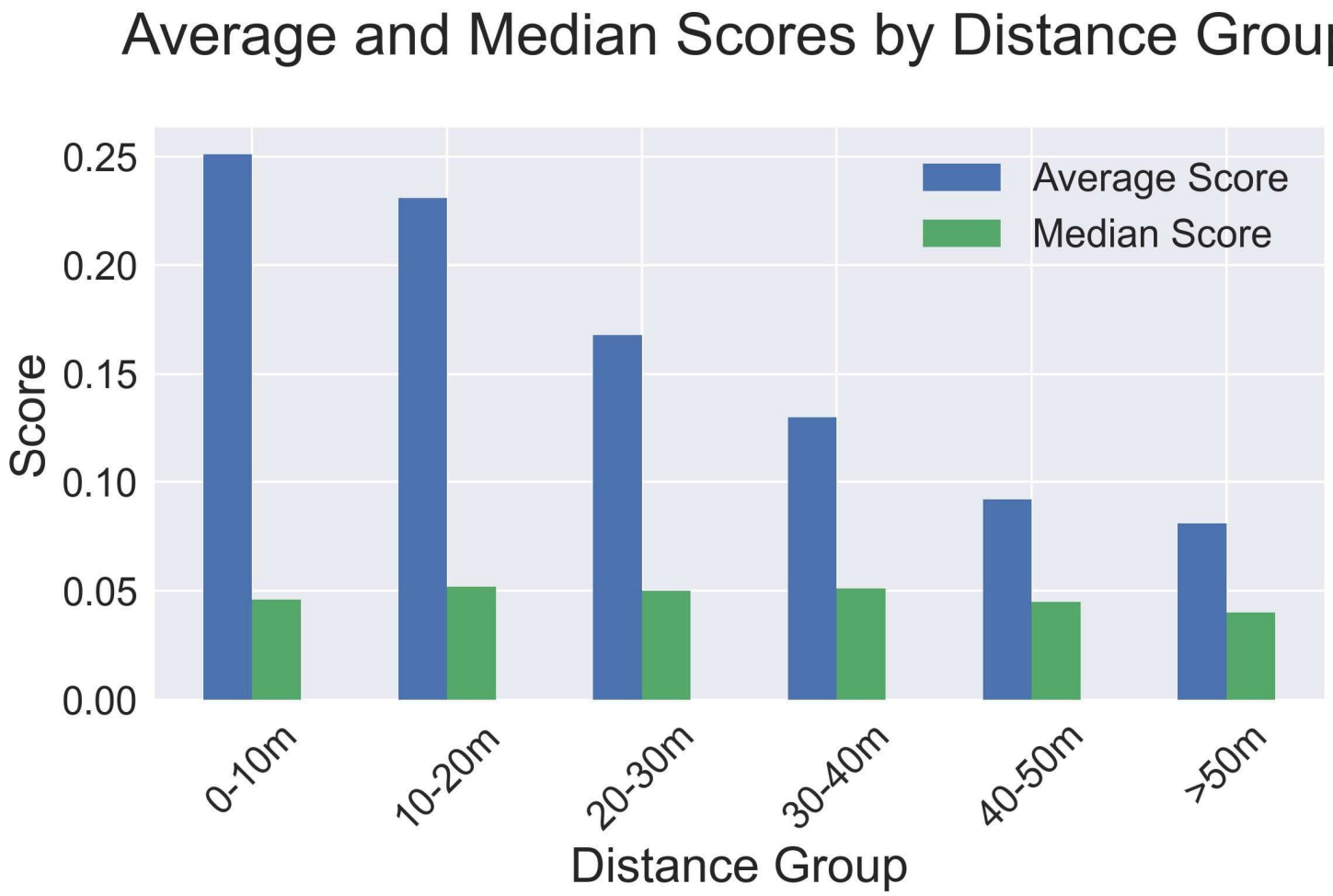
Scene-0916, Sample 8 with threshold > 0.3

2.3. EVALUATION AND VISUALIZATION

C. Faraway Objects

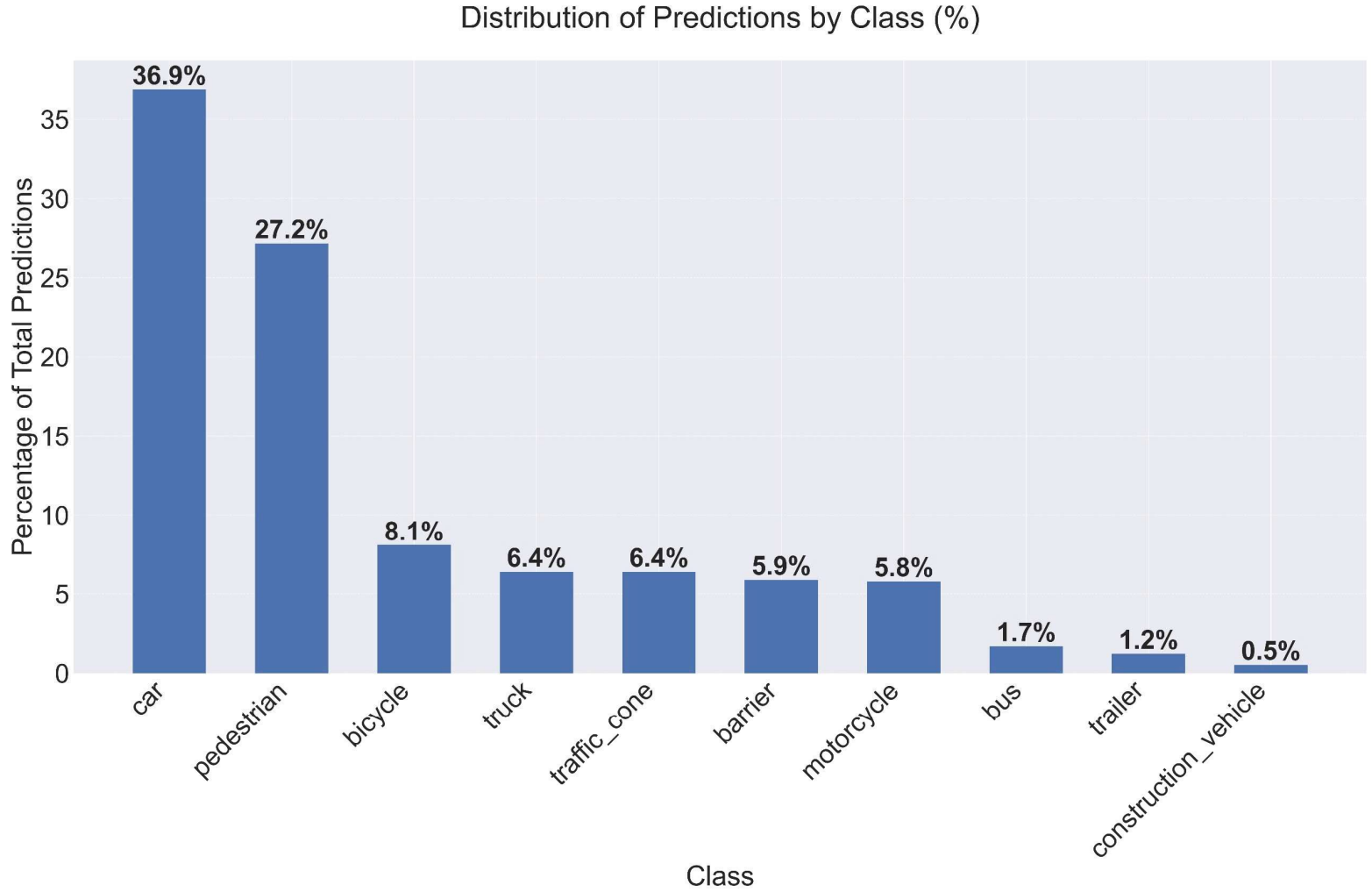
Scene-0796, Sample 11 with threshold > 0.3

2.3. EVALUATION AND VISUALIZATION



HCMUTE

2.3. EVALUATION AND VISUALIZATION



HCMUTE

THE 3-SECOND RULE

A traffic safety guideline for drivers

~2.5-second reaction distance
at 50 km/h (~13.9 m/s)

Recommended Distance = 2.5s x 13.9m/s = 34.75m

Our MEFormer model is most reliable under 40 meters

Our MEFormer model meets the 3-Second Rule

DEMO PRODUCT

3. CONCLUSION

Summary

1. Implemented MeFormer with MMDetection3D on nuScenes.
2. Achieved high mAP & NDS, good for car, bus, pedestrian.
3. Best performance <40m (aligns with **3-Second Rule**).
4. Struggles with faraway, rare, or diverse objects.

3. CONCLUSION

Summary

1. Implemented MeFormer with MMDetection3D on nuScenes.
2. Achieved high mAP & NDS, good for car, bus, pedestrian.
3. Best performance <40m (aligns with **3-Second Rule**).
4. Struggles with faraway, rare, or diverse objects.

Future Development

- **A** **HARDWARE SYSTEM:** Build LiDAR/camera system for collecting real-world data.
- **B** **REAL-WORLD TEST:** Deploy on vehicles/simulators to check performance & latency.
- **C** **RARE CLASS BOOST:** Use augmentation/transfer learning for rare/small classes.
- **D** **DYNAMIC THRESHOLDS:** Tune score thresholds by distance or application.

HCMUTE

fie.hcmute.edu.vn

THANK YOU!

Development of a multimodal 3D object recognition and visualization system for traffic environments using camera and LiDAR fusion

GRADUATION PROJECT