INTEGRATING MIDAS FOR 3D ROAD SAFETY MEASUREMENT

NELDA DIXON MUT22CS107
Roll no 48
Muthoot Institute of Technology and Science,
Varikoli

Project Guide: Nikita Pinheiro
Assistant Professor
Department of Computer Science and Engineering

TABLE OF CONTENTS

MOTIVATION

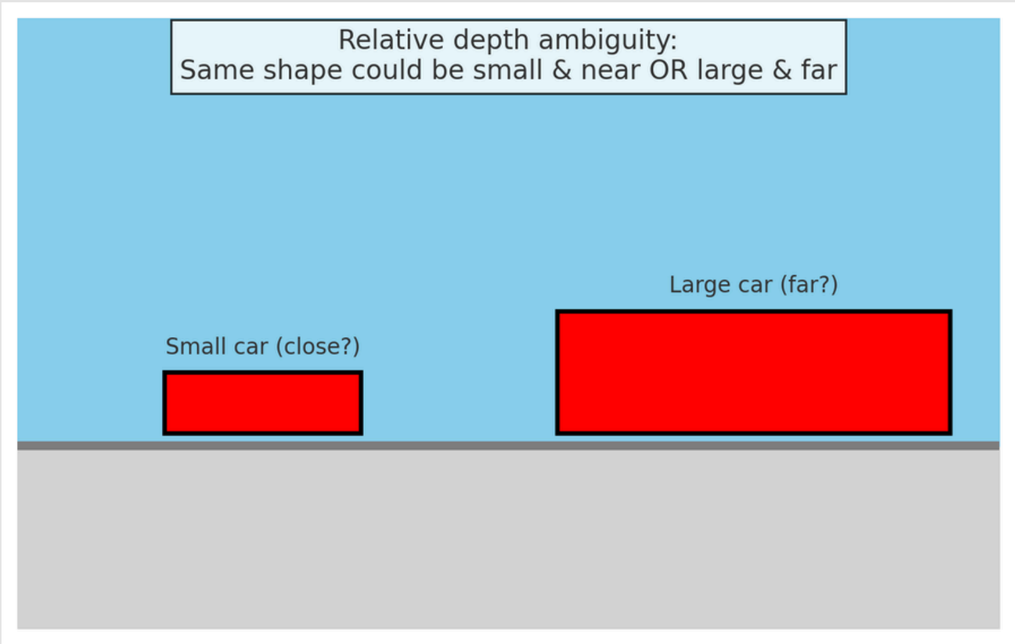
Why accurate 3D measurement on roads matters:

- Accident prevention:
  accurate distance to pedestrians/vehicles
- Asset management:
  pothole sizing, pavement deformation measurement
- Low-cost alternative:
  monocular cameras vs LiDAR (cost, weight)

FUNDAMENTAL LIMITATION:
SCALE AMBIGUITY

WHAT DOES IT MEAN?

- A single image cannot tell absolute size or distance.
- Deep learning models (like MiDaS, FastDepth) predict relative depth maps meaning they can rank which pixels are closer or farther but they don't know the true scale.
- For Example: A toy car close to the camera and a real car far away can look the same in 2D.



FUNDAMENTAL LIMITATION:

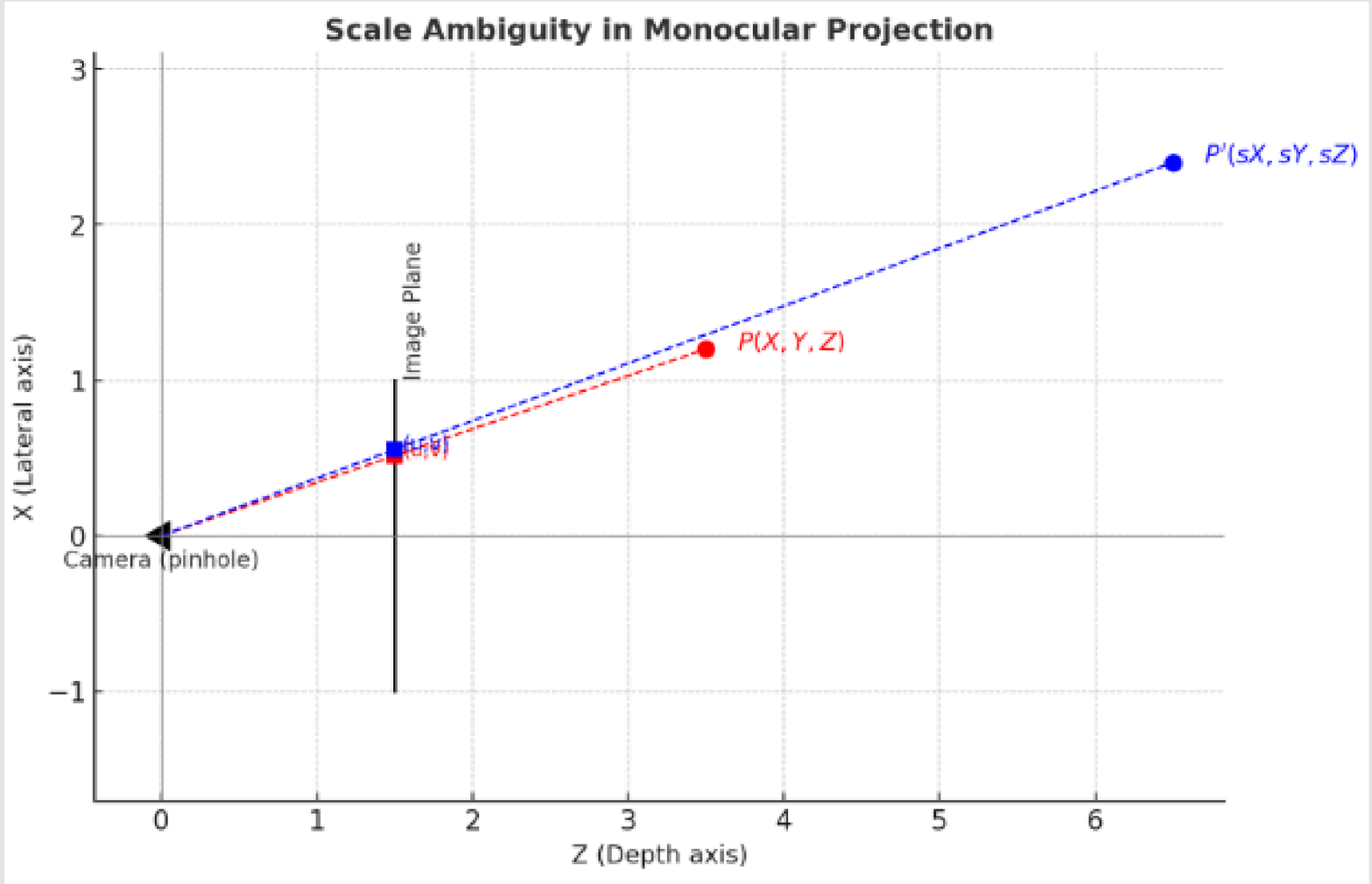
SCALE AMBIGUITY ... CONT.

WHY DOES IT HAPPEN?

- Cameras capture projections of 3D scenes into 2D.
- Without knowing the camera's zoom (focal length) or reference sizes, you lose the scale factor when going from 3D → 2D.
- Mathematically:
- $(X,Y,Z) \xrightarrow{\text{Projection}} (u,v)=(f_x/Z, f_y/Z)$
- Depth Z shrinks everything proportionally, so doubling all distances produces the same image.

FUNDAMENTAL LIMITATION:

SCALE AMBIGUITY ... CONT.



FUNDAMENTAL LIMITATION:

SCALE AMBIGUITY ... CONT.

WHY IS IT A PROBLEM?

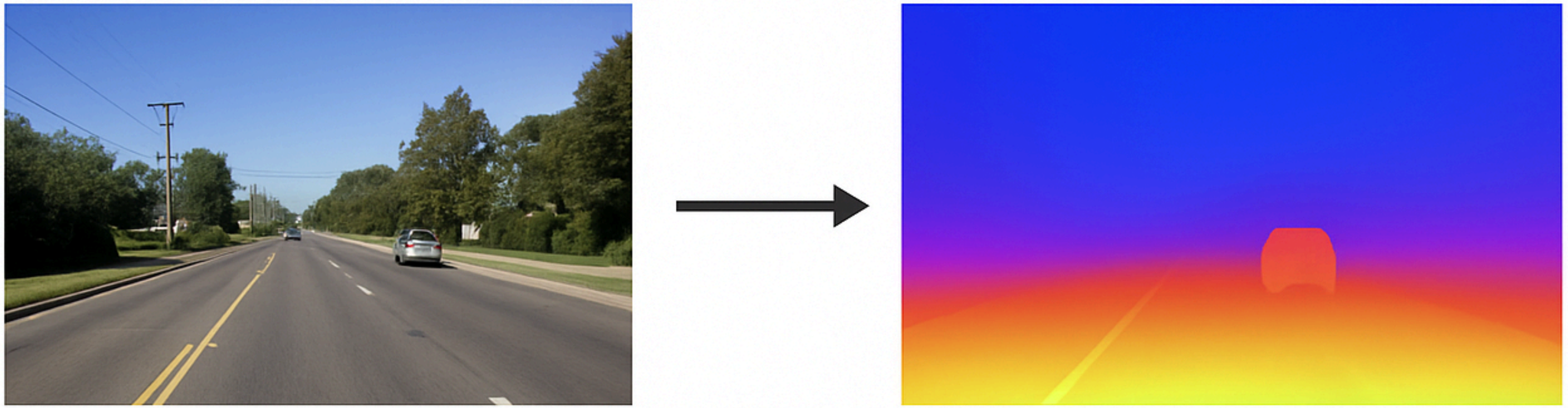
- Road Safety Applications need matric depth:
  - Braking & Collision Avoidance → Must know if pedestrian is 3 m or 30 m away to apply correct braking force.
  - Pothole Detection → Requires actual width, depth, and volume for repair prioritization and cost estimation.
  - Lane Changes & Warnings → Safety depends on exact distances to vehicles, not just relative closeness.
  - Relative Depth ≠ Usable → Engineering decisions demand real-world units (m, cm), not just depth ordering.

MONOCULAR DEPTH ESTIMATION (MDE)

What is Monocular Depth Estimation (MDE)?

- Predicts 3D depth from a single RGB image
- Eliminates need for stereo cameras or LiDAR
- Useful for low-cost, large-scale deployment
- Challenges:
  - Scale ambiguity → Hard to know absolute distances
  - Occlusion handling → Hidden objects not visible
  - Lighting & weather → Shadows, glare, fog affect accuracy

MONOCULAR DEPTH ESTIMATION (MDE)

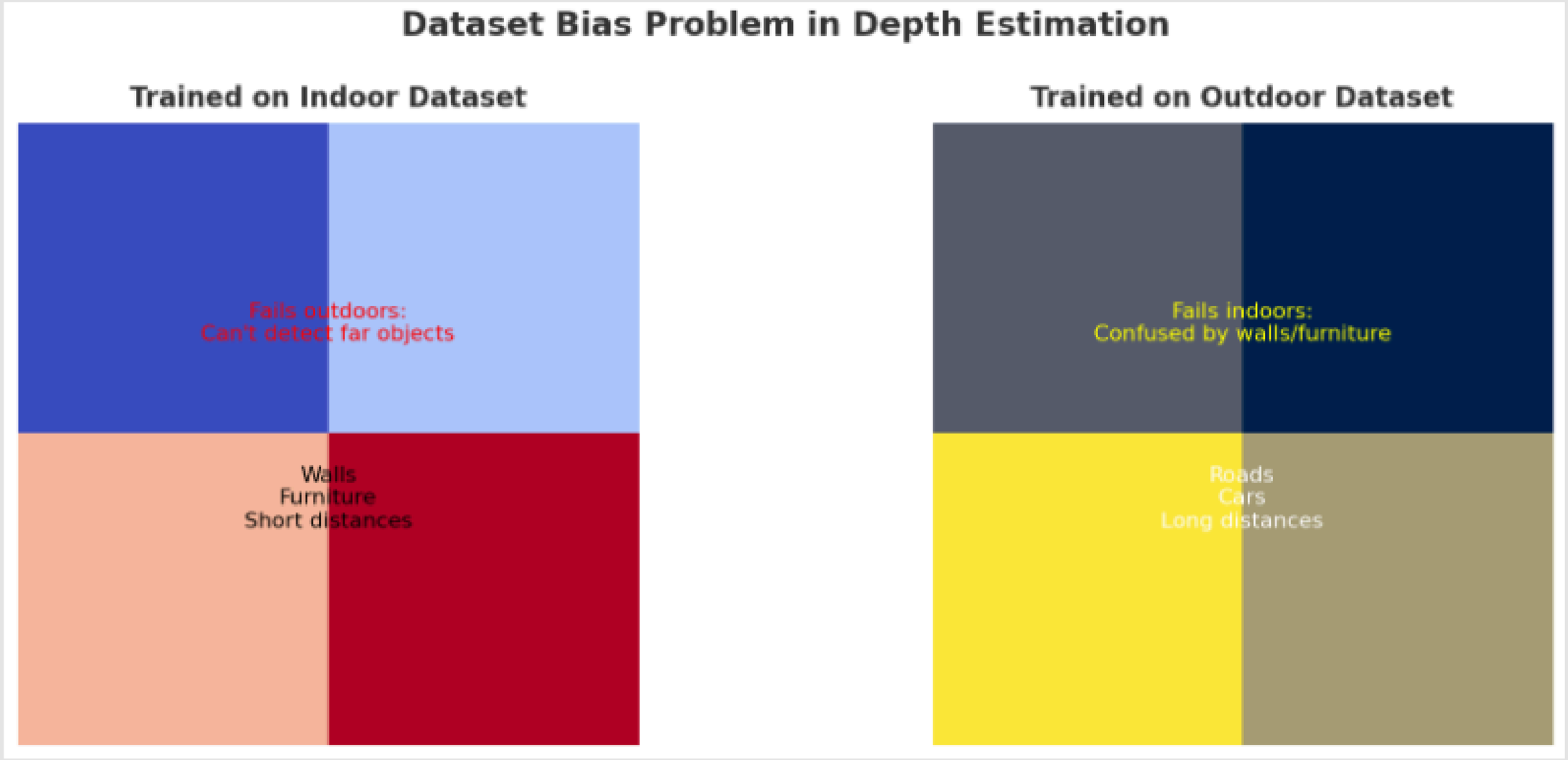


MIDAS

MiDaS Overview:

- Developed by Intel & University of Tübingen.
- Trained on multiple diverse datasets → strong generalization.
- Predicts relative depth maps (scaled depth).
- Works across a wide range of scenes: indoors, outdoors, urban, natural.
- Key strength: transfer learning from large-scale image tasks

MIDAS


Dataset Bias Problem in Depth Estimation
Trained on Indoor Dataset
Fails outdoors:
Can't detect far objects
Walls
Furniture
Short distances
Trained on Outdoor Dataset
Fails indoors:
Confused by walls/furniture
Roads
Cars
Long distances

MIDAS

Typical Use Cases:

- Offline depth estimation (non real-time):
  - MiDaS is relatively heavy, so it's more common in research, content creation, and offline 3D reconstruction rather than in embedded, real-time systems.

- High-accuracy dense depth maps:
  - Produces pixel-level dense predictions rather than sparse points (traditional stereo methods).

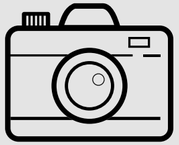
- Cross-domain applications:
  - Works well for AR/VR, 3D photography, background replacement, and robotics mapping where no dataset-specific tuning is possible.

MIDAS
ARCHITECTURE

Encoder–Decoder Backbone:

- Encoder: Pre-trained CNN/Transformer (e.g., ResNet, EfficientNet, Vision Transformers) extracts features
- Feature Fusion: Combines multi-scale features for context
- Decoder: Upsampling network reconstructs dense depth map
- Output: Relative depth map (pixel-wise depth prediction)

MIDAS
ARCHITECTURE



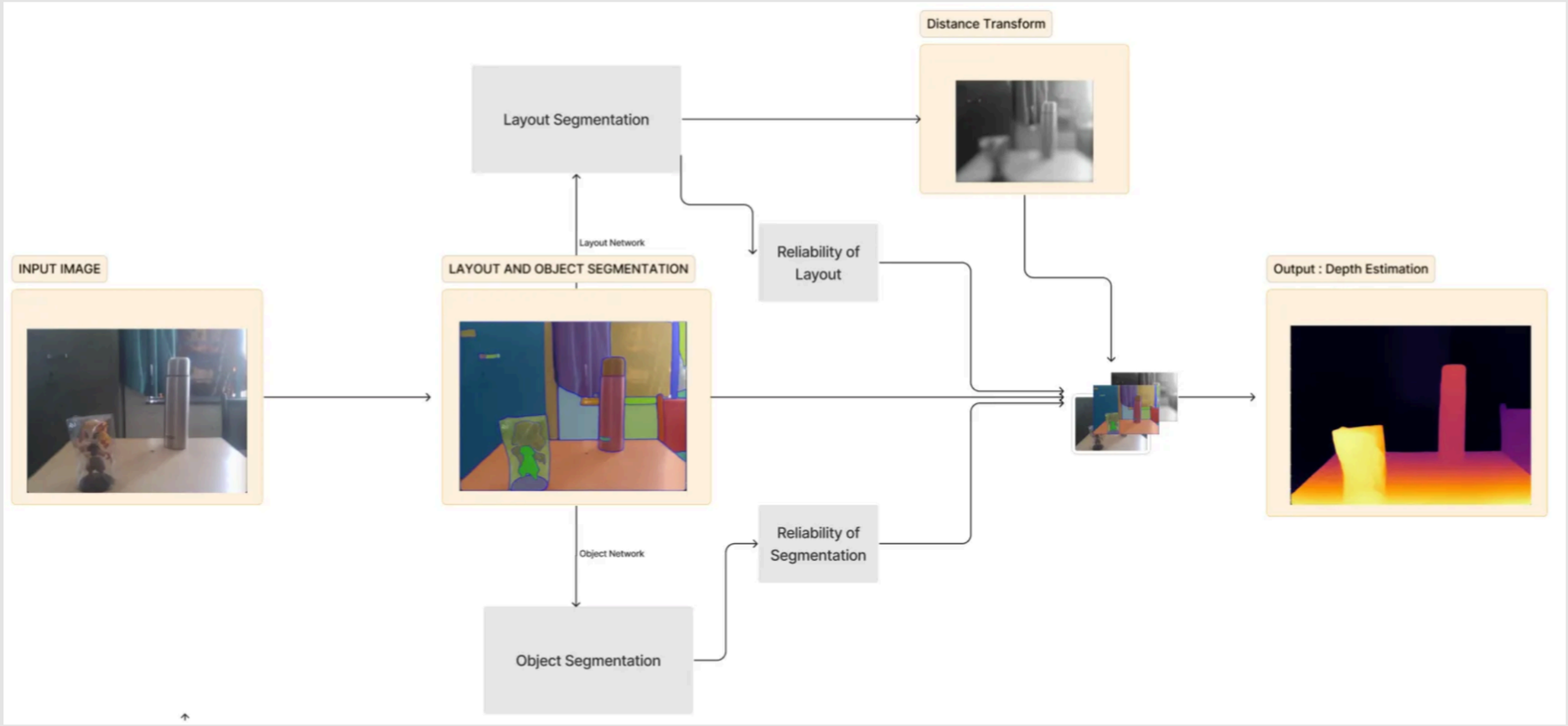
Input Image

Encoder (CNN/Transformer)

Multi-Scale Feature Fusion

Decoder (Depth Map)

Predicted Depth Map



MIDAS

Evolution of MiDaS Versions:

2019 — MiDaS v1:
- Introduced by Intel Intelligent Systems Lab (ISL).
- Based on ResNet encoder–decoder.
- First multi-dataset training approach for robust monocular depth.

2020 — MiDaS v2
- Improved accuracy and generalization.
- Extended dataset coverage (KITTI, NYU Depth, TUM, etc.).

2021 — MiDaS v3 (DPT Models)
- Introduced Dense Prediction Transformer (DPT) backbone.
- Variants: DPT-Large, DPT-Hybrid, DPT-Small (speed vs. accuracy trade-off).

2022–2023 — Optimization for Edge Devices
- Lightweight MiDaS models released for mobile and embedded devices.
- Deployment via ONNX, TensorRT, PyTorch Mobile.

2024–2025 — Wider Applications
- Adopted in autonomous driving, AR/VR, road safety, robotics.
- Community-trained custom versions for local datasets (e.g., Indian road conditions).

MIDAS

Feature	MiDaS v2.0	MiDaS v3.0	MiDaS v3.1
Architecture	Classic	DPT	Multi-Transformer
Accuracy	Base	21%	28%
Speed	Moderate	Improved	Up to 4x faster
Real-Time Ready	❌	❌	✅
Best For	Static Images	Precision Tasks	Real-Time Applications

MIDAS

COMPARATIVE STUDY INSIGHTS:

Model	Accuracy	Generalization	Speed	Notes
Monodepth2	Medium	Limited	Fast	Works only on trained dataset
FastDepth	Lower	Limited	Very Fast	Mobile-focused
DPT	High	Good	Slower	Transformer-based
MiDaS v3.1	Very High	Excellent	Fast	Best balance for deployment

MIDAS
NEED?

Road Safety Need; Pothole Detection:

- Major cause of accidents, especially for two-wheelers.
- Traditional detection: manual surveys, vibration sensors → costly & slow.
- MiDaS can estimate depth variation in road surface.
- Enables:
  - Early detection of potholes
  - Prioritization of repairs
  - Real-time driver alerts

MIDAS
NEED?

Why MiDaS for Pothole Detection?

- Single Camera Setup → Low-cost, easy to install roadside or on vehicles
- Dense Depth Maps → Captures fine surface variations (potholes, cracks)
- Robust Across Conditions → Works in urban, rural, and highway scenes
- Scalable → Deployable on large road networks without expensive sensors
- Better than 2D Detection → Goes beyond just spotting dark patches; estimates depth & severity

CALIBRATION FOR ABSOLUTE DEPTH

- Relative Depth Output
  - MiDaS provides depth up to scale (i.e., relative depth, not real-world units).
  - Values indicate “closer vs farther,” but not meters directly.

- Conversion to Absolute Depth
  - Camera Intrinsics: Focal length, sensor size, and principal point are used to scale relative depth to actual distances.
  - Known Objects / Reference Points: Measure objects of known size in the scene to compute a scale factor.
  - Formula (simplified):

  - $D_{absolute} = s \times D_{relative}$

  - where s is the scale factor from intrinsics or known objects.

CALIBRATION FOR ABSOLUTE DEPTH

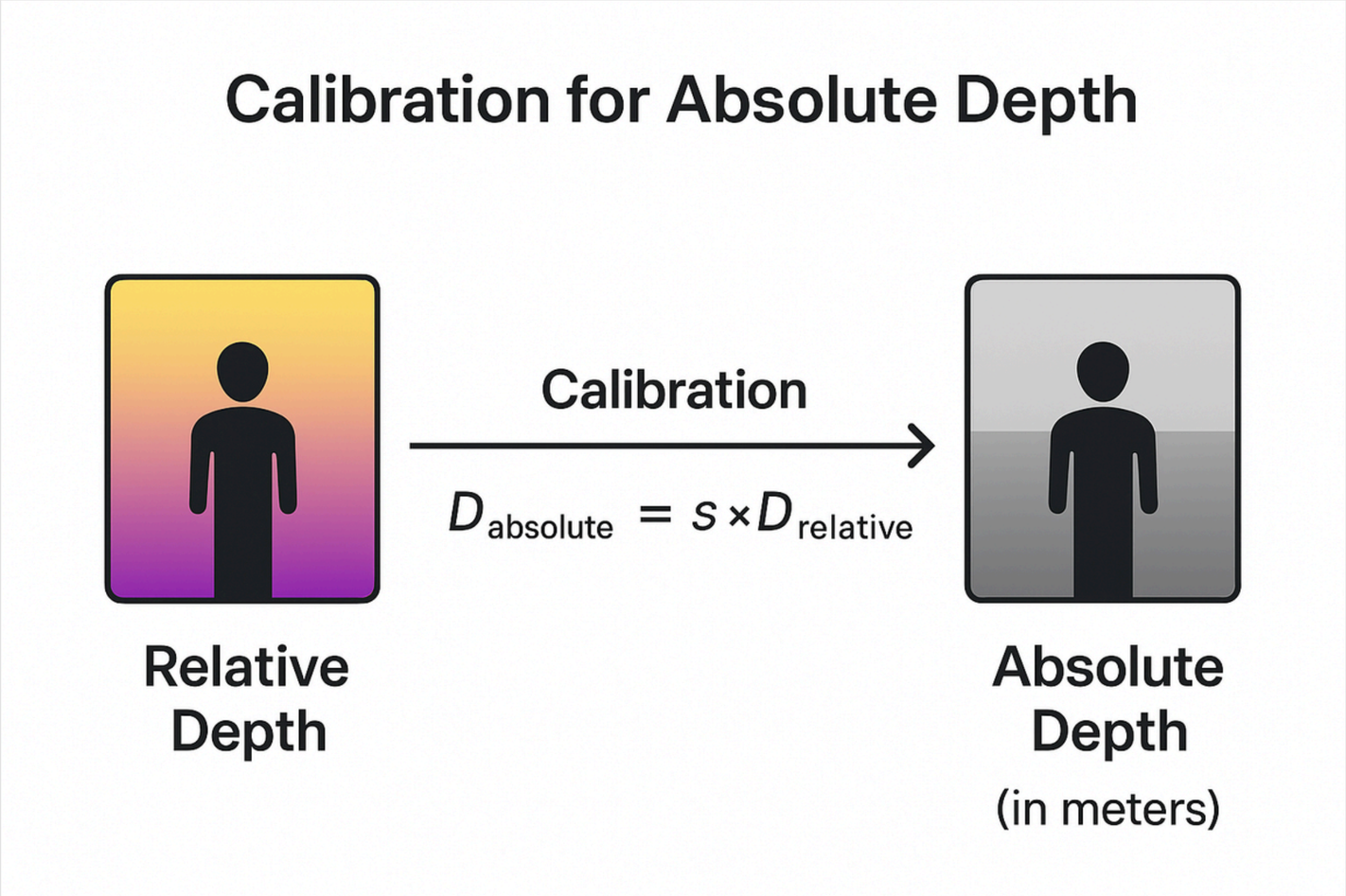
Optional Sensor Fusion:

- Combine with GPS for global positioning of objects.
- IMU (Inertial Measurement Unit) helps refine camera pose and orientation.
- Sensor fusion improves accuracy and consistency across frames.

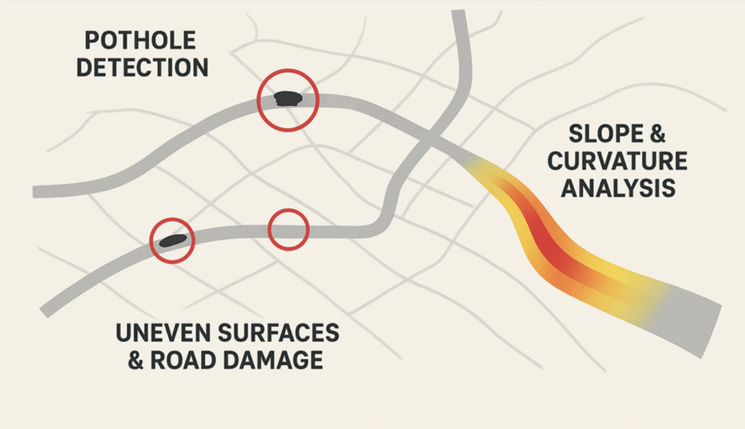
Challenges:

- Scale ambiguity: Relative depth maps don’t know actual distances without calibration.
- Dynamic scenes: Moving objects or changing camera angles complicate calibration.
- Lighting and occlusions can affect depth estimation reliability.

CALIBRATION FOR ABSOLUTE DEPTH



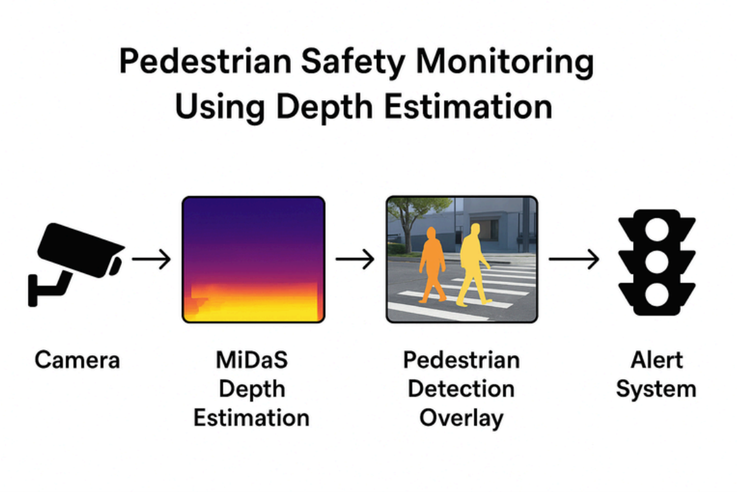
USE CASE 1



Road Infrastructure Monitoring:

- Pothole Detection:
  - Automatic identification of potholes using camera images or LiDAR/point cloud data.
  - Alerts maintenance teams in real-time for quicker repairs.
  - Enables prioritization of high-traffic or hazardous areas.
- Uneven Surfaces & Road Damage:
  - Detect cracks, rutting, and surface irregularities.
  - Track deterioration over time using periodic surveys.
  - Integrate with predictive maintenance systems to prevent accidents.
- Slope & Curvature Analysis:
  - Assess road inclines, banking angles, and sharp curves.
  - Identify potentially dangerous sections, especially for highways and mountain roads.
  - Support planning for new roads or upgrades to existing infrastructure.

USE CASE 2



Pedestrian Safety Monitoring Using Depth Estimation:

- Detect Pedestrians with Depth:
  - MiDaS predicts the distance of objects from the camera in real time.
  - Helps differentiate between pedestrians on sidewalks versus those stepping onto the road.
- Crosswalk Monitoring:
  - Focus on pedestrian-heavy zones like crosswalks, school zones, and intersections.
  - Detect unusual behavior: e.g., jaywalking or pedestrians waiting at the curb.
- Alert Systems & Roadside Integration:
  - Cameras feed into a roadside computing unit running MiDaS.
  - Real-time alerts can be sent to vehicles (V2X), traffic lights, or local authorities.
  - Enables proactive measures—slowing down vehicles or signaling pedestrians.

IMPLEMENTATION

```
# Import dependencies
import cv2
import torch
import matplotlib.pyplot as plt

# Download the MiDaS
#midas = torch.hub.load('intel-isl/MiDaS', 'MiDaS_small')
midas = torch.hub.load('intel-isl/MiDaS', 'DPT_Large')

# Use GPU if available
device = torch.device("cuda")
if torch.cuda.is_available()
else torch.device("cpu")
midas.to(device)
midas.eval()

# Use transforms to resize and normalize the image
midas_transforms = torch.hub.load("intel-isl/MiDaS", "transforms")

if midas == "DPT_Large" or midas == "DPT_Hybrid":
    transform = midas_transforms.dpt_transform
else:
    transform = midas_transforms.small_transform

# Hook into OpenCV
cap = cv2.VideoCapture(0)
while cap.isOpened():
    ret, frame = cap.read()
```

```
# Transform input for midas
 img = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
 imgbatch = transform(img).to('cpu')

 # Make a prediction
 with torch.no_grad():
 prediction = midas(imgbatch)
 prediction = torch.nn.functional.interpolate(
 prediction.unsqueeze(1),
 size = img.shape[:2],
 mode='bicubic',
 align_corners=False
 ).squeeze()

 output = prediction.cpu().numpy()

 print(output)
 plt.imshow(output)
 cv2.imshow('CV2Frame', frame)
 plt.pause(0.00001)

 if cv2.waitKey(10) & 0xFF == ord('q'):
 cap.release()
 cv2.destroyAllWindows()

plt.show()
```

LIMITATIONS OF MIDAS

Technical Limitations:

- Relative Depth Only → MiDaS predicts relative depth (closer/farther), not absolute distances in meters.
- Calibration Needed → Needs scaling or sensor fusion (e.g., GPS, LiDAR) for real-world measurements.
- High Computation → Heavy models like DPT_Large demand GPUs or optimized edge deployment.

Environmental Limitations:

- Lighting Sensitivity → Accuracy drops in low light, night driving, rain, fog, and glare.
- False Positives → Shadows, water puddles, and oil stains sometimes misclassified as depth variations.
- Dataset Bias → Pretrained mostly on global datasets (KITTI, NYU, ReDWeb). Local road conditions (e.g., Indian rural roads) may reduce performance.

Practical Limitations:

- Real-Time Challenges → Running on mobile or embedded systems requires pruning/quantization.
- No Temporal Awareness → Works frame by frame, so sudden glitches or flickers can occur in video streams.

CONCLUSION

- MiDaS enables 3D road safety measurement using just a single RGB camera.
- Effective in detecting potholes, road irregularities, and hazards with depth precision.
- Provides a low-cost, scalable alternative to LiDAR and stereo systems.
- Still faces challenges: relative depth scaling, lighting issues, computational load.
- Future directions: sensor fusion, edge deployment, smart city integration.
- A step towards safer, smarter, and more sustainable roads.

REFERENCES

- A Comparative Study on Monocular Depth Estimation (AtharvMalusare 2023)
- Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer
- MiDaS v3.1 – A Model Zoo for Robust Monocular Relative Depth Estimation

THANK YOU