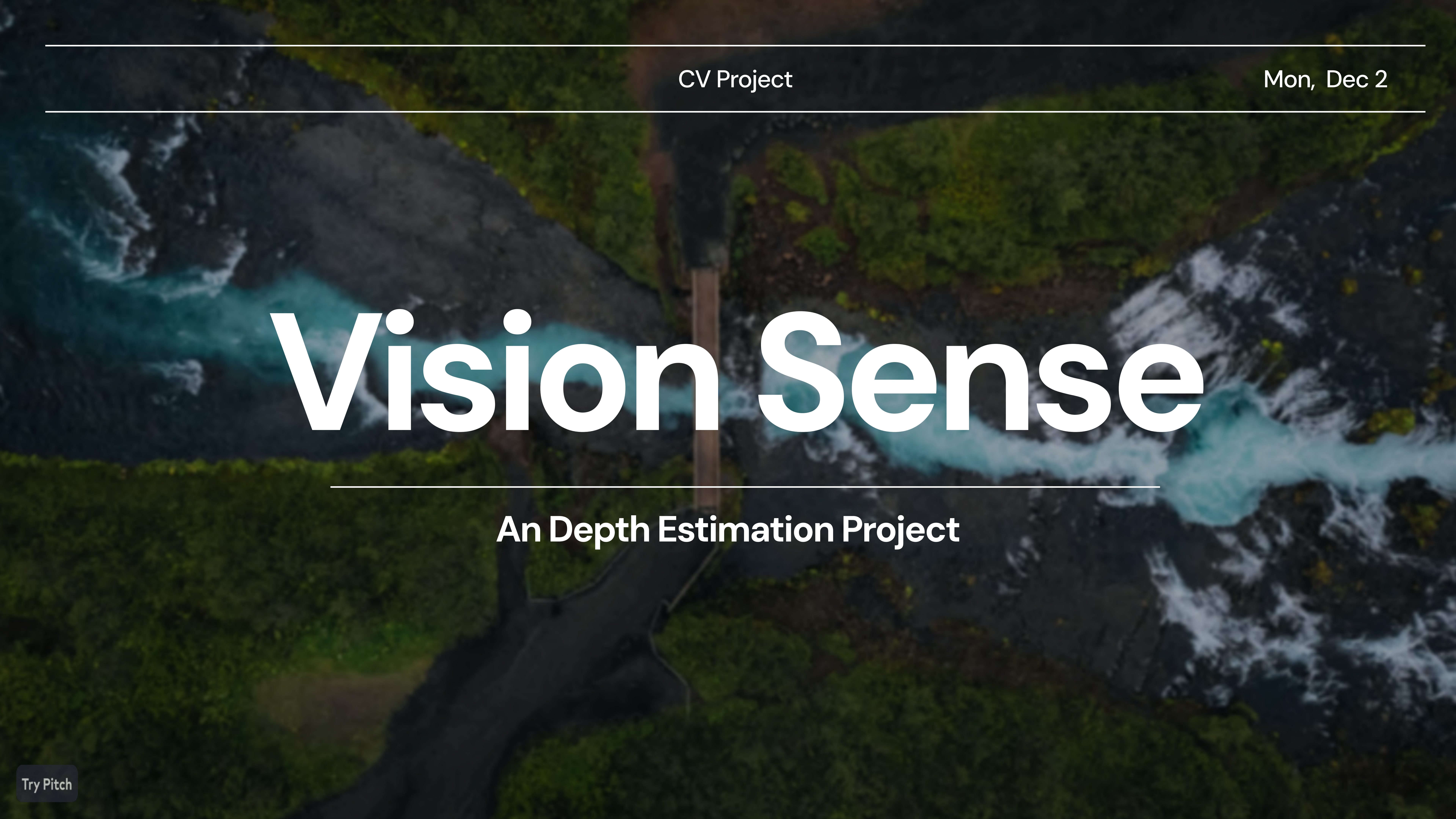


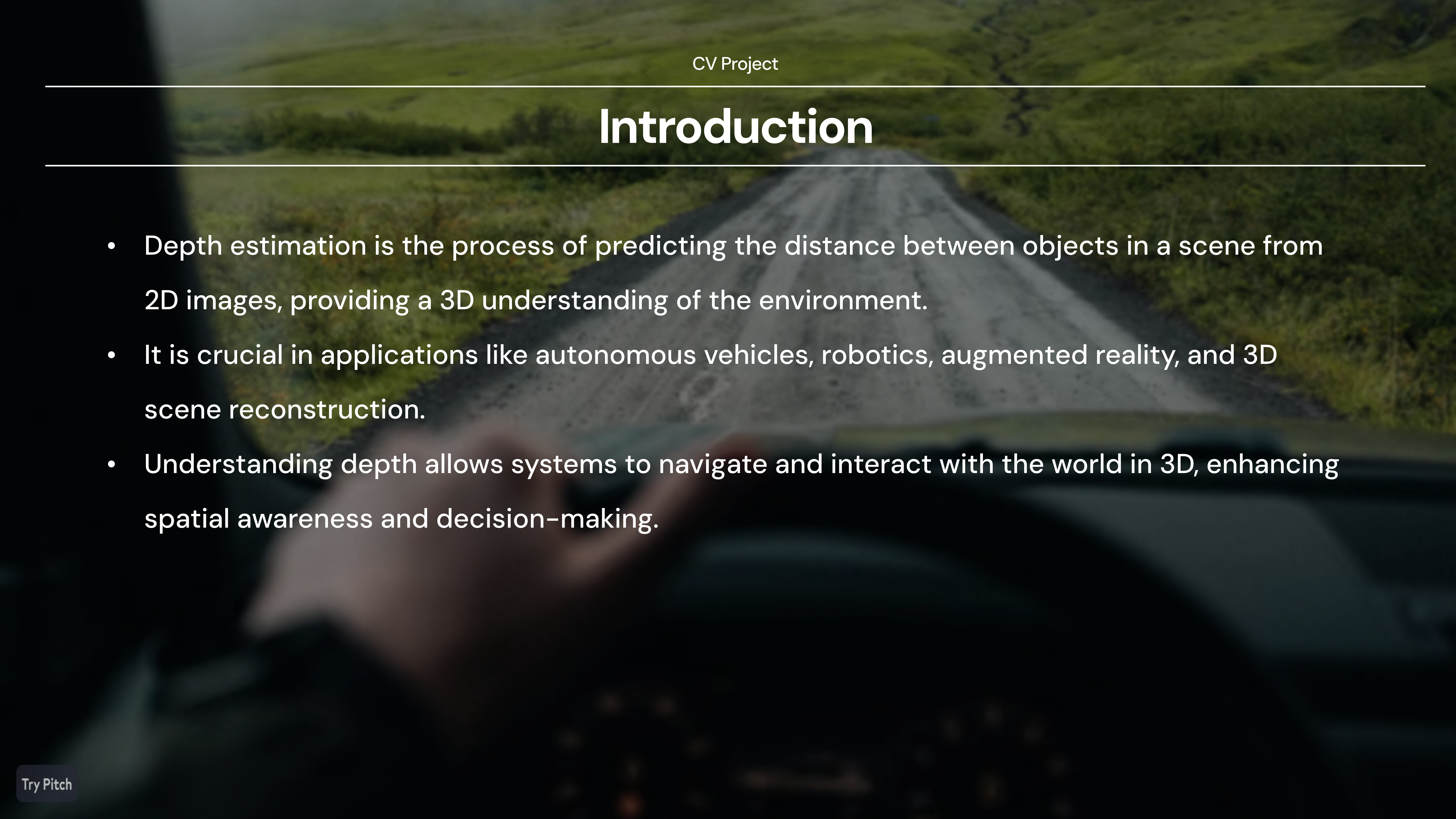
CV Project

Mon, Dec 2

Vision Sense

An Depth Estimation Project

Introduction



- Depth estimation is the process of predicting the distance between objects in a scene from 2D images, providing a 3D understanding of the environment.
- It is crucial in applications like autonomous vehicles, robotics, augmented reality, and 3D scene reconstruction.
- Understanding depth allows systems to navigate and interact with the world in 3D, enhancing spatial awareness and decision-making.

Literature Review

S.No	First Author,Country	Methodology	Highlights	Research Gap	Reference link
1	Shir Gur,Israel	The method estimates depth from focus cues using a differentiable PSF convolutional layer and self-attention in the ASPP module.The model is trained end-to-end to minimize reconstruction loss, using either real or generated data.	The method estimates depth from defocus cues, achieving an Abs Rel of ~0.10 and RMSE of ~4.5, comparable to supervised methods on Nyudepth and Make3D.It reduces overfitting and shows better cross-domain transfer, outperforming traditional supervised approaches.	Limited exploration of deep learning techniques in depth from defocus. Insufficient analysis of varying focus distances in depth estimation	https://doi.org/10.1109/CVPR.2019.00787
2	Geonho Cha,Korea	The ISSL method generates self-samples using random rigid transformations on depth for additional supervision.It improves performance by fully utilizing training images and is evaluated on datasets like KITTI and NYUv2.	The method reduces translation error in dynamic regions, improving dynamic depth error by about 20% compared to the baseline, and slightly enhancing shape error.On the KITTI dataset, it increases accuracy and improves scale-consistency, with the standard deviation of scaling parameters rising from 2.137 to 2.429.	Dynamic regions were often filtered out in photometric loss. Scale ambiguity in monocular depth estimation remains unresolved	https://doi.org/10.1109/LRA.2022.3221871

Literature Review

S.No	First Author,Country	Methodology	Highlights	Research Gap	Reference link
3	Hang Li, China	The method employs a DNET backbone for monocular depth estimation, utilizing dilated convolution.It includes a feature extraction module, a depth perception module that correlates feature maps with depth categories, and an optimization algorithm to address class imbalance in the loss function.	DNET Model Features: DNET is a new CNN that uses dilated convolution and feature fusion to enhance depth estimation, outperforming ResNet50 and DenseNet121 on NYU Depth-v2 and KITTI.Error Reduction: DNET achieves a low RMSE loss of 0.481 and improves accuracy from 0.9094 to 0.9198, capturing detailed object edges effectively.	The paper does not address real-time depth estimation challenges. Insufficient analysis of model performance under varying lighting conditions	https://doi.org/10.3390/app14135833
4	Zonghao Lu,China	The paper introduces 2D image semantic segmentation to improve 3D scene segmentation (SSC) using depth estimation, with 2D features mapped into 3D space through inverse perspective projection.Multiscale 2D plane features facilitate parallel processing with 3D networks, and a Dual-Head Pyramid Pooling module enhances contextual information acquisition	PPMNet demonstrates superior performance on the NYUv2 dataset, achieving a mean Intersection over Union (mIoU) of 60.2%, surpassing previous state-of-the-art methods.On the Semantic KITTI dataset, PPMNet achieves an mIoU of 65.4%, highlighting its effectiveness in 3D Semantic Scene Completion tasks.	Existing models neglect geometric and semantic understanding in 2D. High imbalance in semantic categories within occupied voxels	https://doi.org/10.1016/j.patcog.2024.111030

Literature Review

S.No	First Author,Country	Methodology	Highlights	Research Gap	Reference Link
5	Xin Yang,China	The architecture utilizes a multi-scale cascaded network that includes a Global Understanding Module for capturing global contextual information and a Difference Module for accurate boundary contour estimation. Cascade Modules integrate information from multiple scales, with extensive experiments conducted on the KITTI and NYUv2 datasets.	The proposed model achieved an Absolute Relative Error (Abs. Rel) of 0.057 and a Root Mean Squared Error (RMSE) of 2.415 on the KITTI dataset, indicating high accuracy in depth estimation. For the NYUv2 dataset, the model recorded an Abs. Rel of 0.104 and an RMSE of 0.380, demonstrating its effectiveness in estimating depth information in indoor scene	The model increases time complexity significantly Limited focus on outdoor depth estimation scenarios	10.1109/ACCESS.2021.3076346
6	Shangbin Yu,China	The network adopts an encoder-decoder structure, with the encoder based on the Swin Transformer.Features are fused through interpolation, concatenation, and convolution, with skip connections linking the encoder and decoder.	The proposed depth estimation network achieved a mean absolute error (MAE) of 0.45 on the NYUv2 dataset, improving depth edge accuracy compared to state-of-the-art methods like DPT (MAE 0.55) and BTS (MAE 0.60).The Swin-L model performed better than the Swin-B variant, showing a 10% improvement in depth estimation accuracy, proving its effectiveness in capturing depth features.	The paper does not address real-time depth estimation challenges. Limited exploration of other transformer architectures for depth estimation.	doi:10.1088/1742-6596/2428/1/012019

Literature Review

S.No	First Author,Country	Methodology	Highlights	Research Gap	Reference link
7	Zihang Liu,China	The model features a dual-stream encoder that combines ResNet with the Swin Transformer for enhanced feature extraction. A lightweight decoder using a multi-head Cross-Attention Module, along with edge guidance loss functions and Sobel operator-based edge gradient calculations, improves depth estimation accuracy.	• The edge-enhanced dual-stream method combines ResNet and Swin Transformer, improving depth estimation with a lightweight decoder and edge-guided loss, performing well on NYU Depth V2, KITTI, and SUN RGB-D datasets. • It achieves an AbsRel error of 0.1 and RMSE of 0.9 on NYU Depth V2 and KITTI, outperforming existing methods.	• The paper does not address real-time depth estimation challenges. • Lack of comparison with state-of-the-art methods in depth estimation	https://doi.org/10.3390/electronics13091652

Research Gap

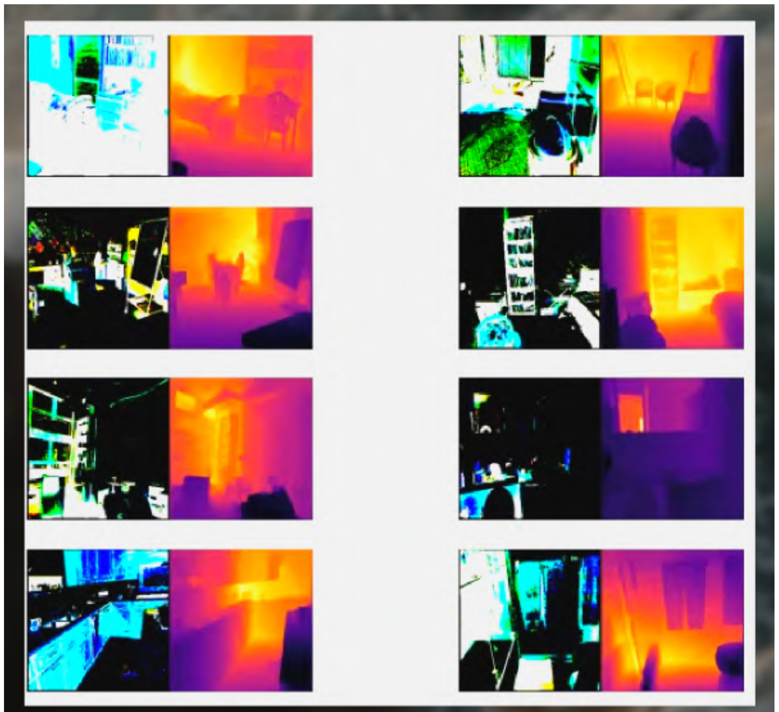
Robust model

Depth Map Visualization

Comparative analysis

CV Project

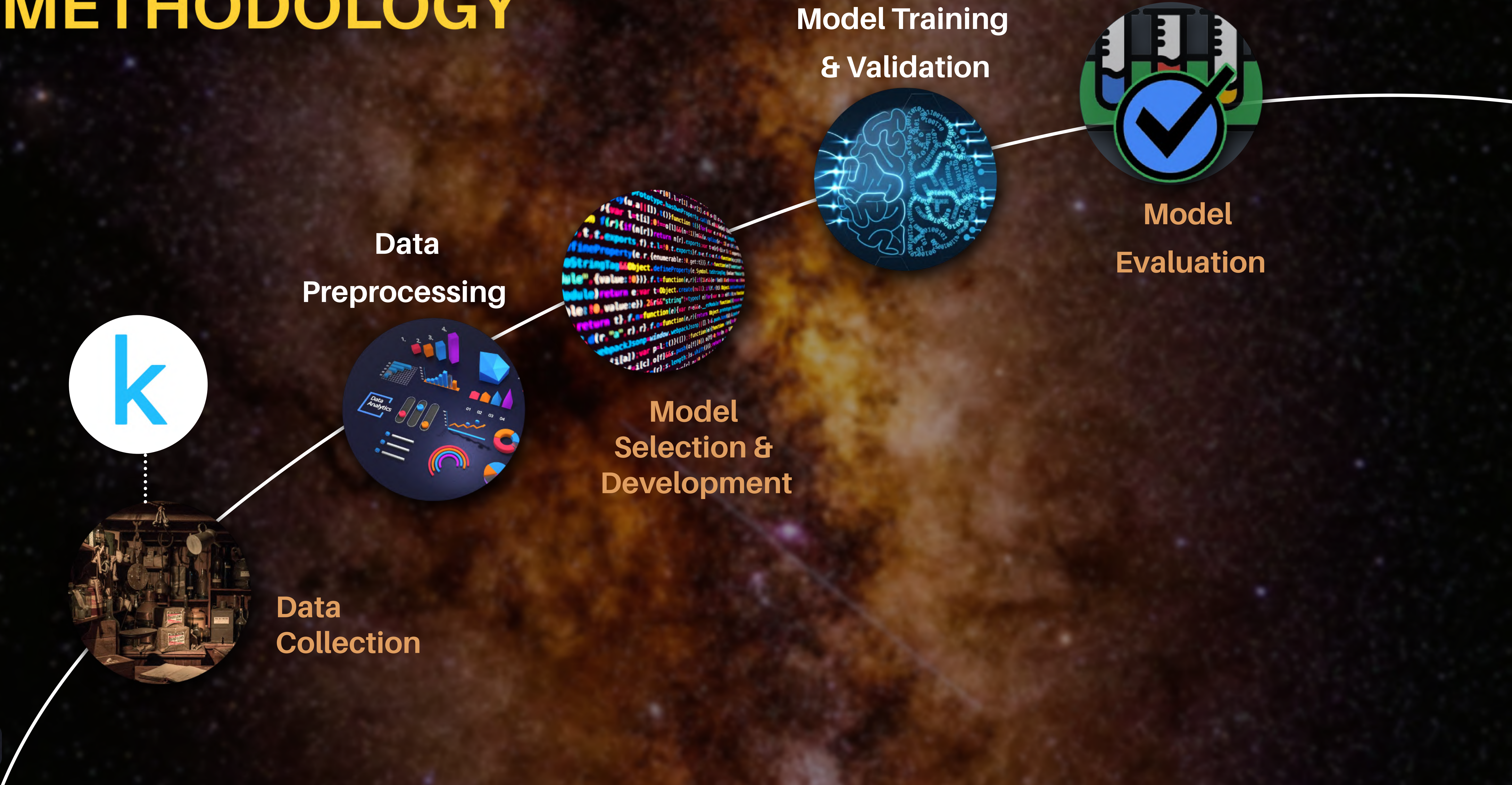
Computer vision concepts



DEPTH ESTIMATION

EDGE DETECTION

METHODOLOGY



Data Collection

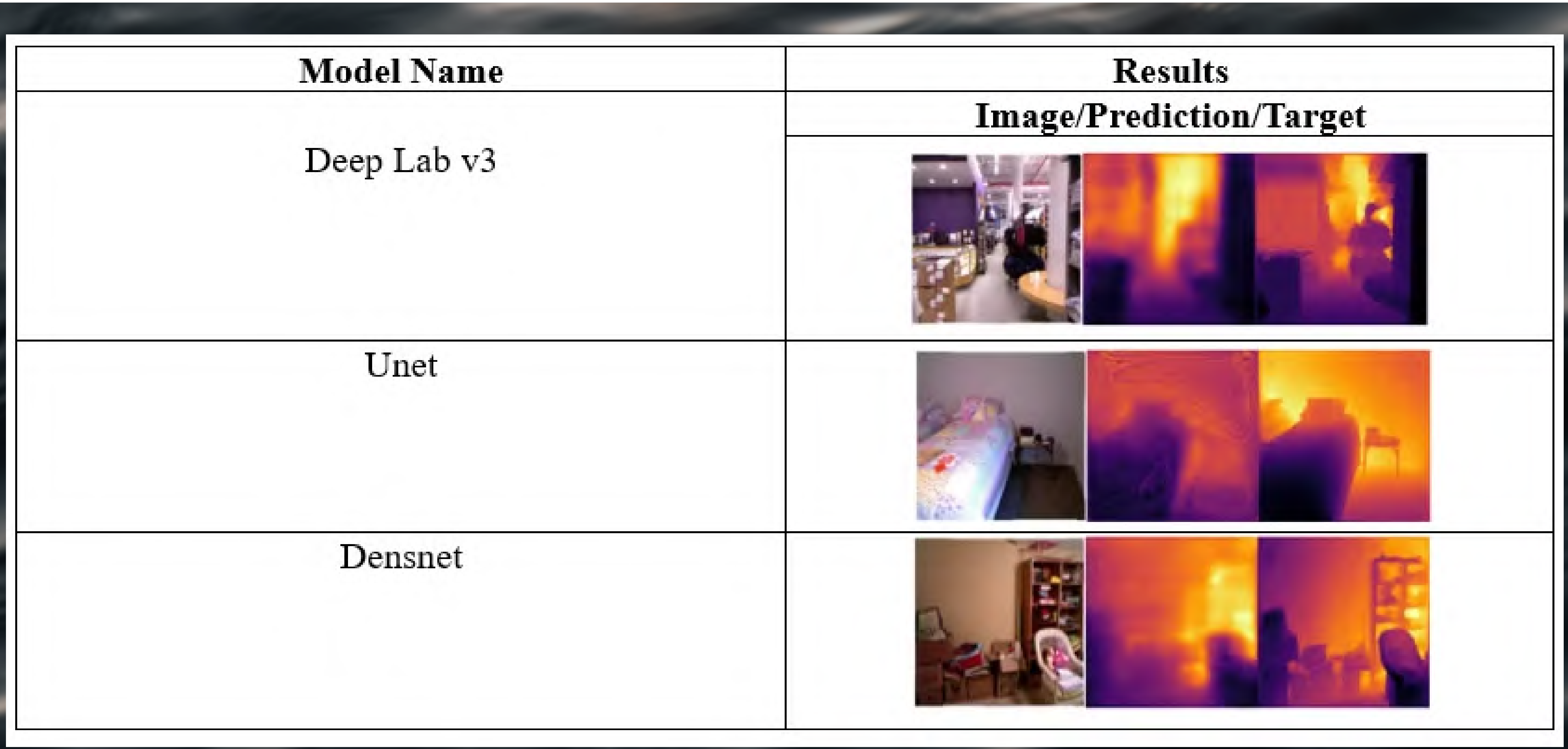
Data Preprocessing

Model Selection & Development

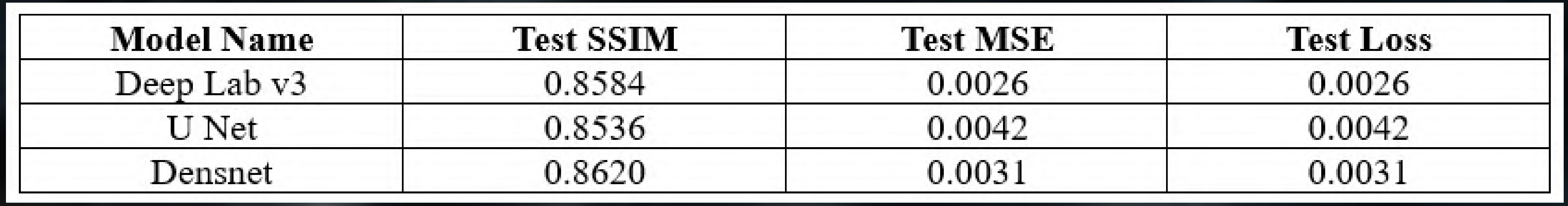
Model Training & Validation

Model Evaluation

RESULTS



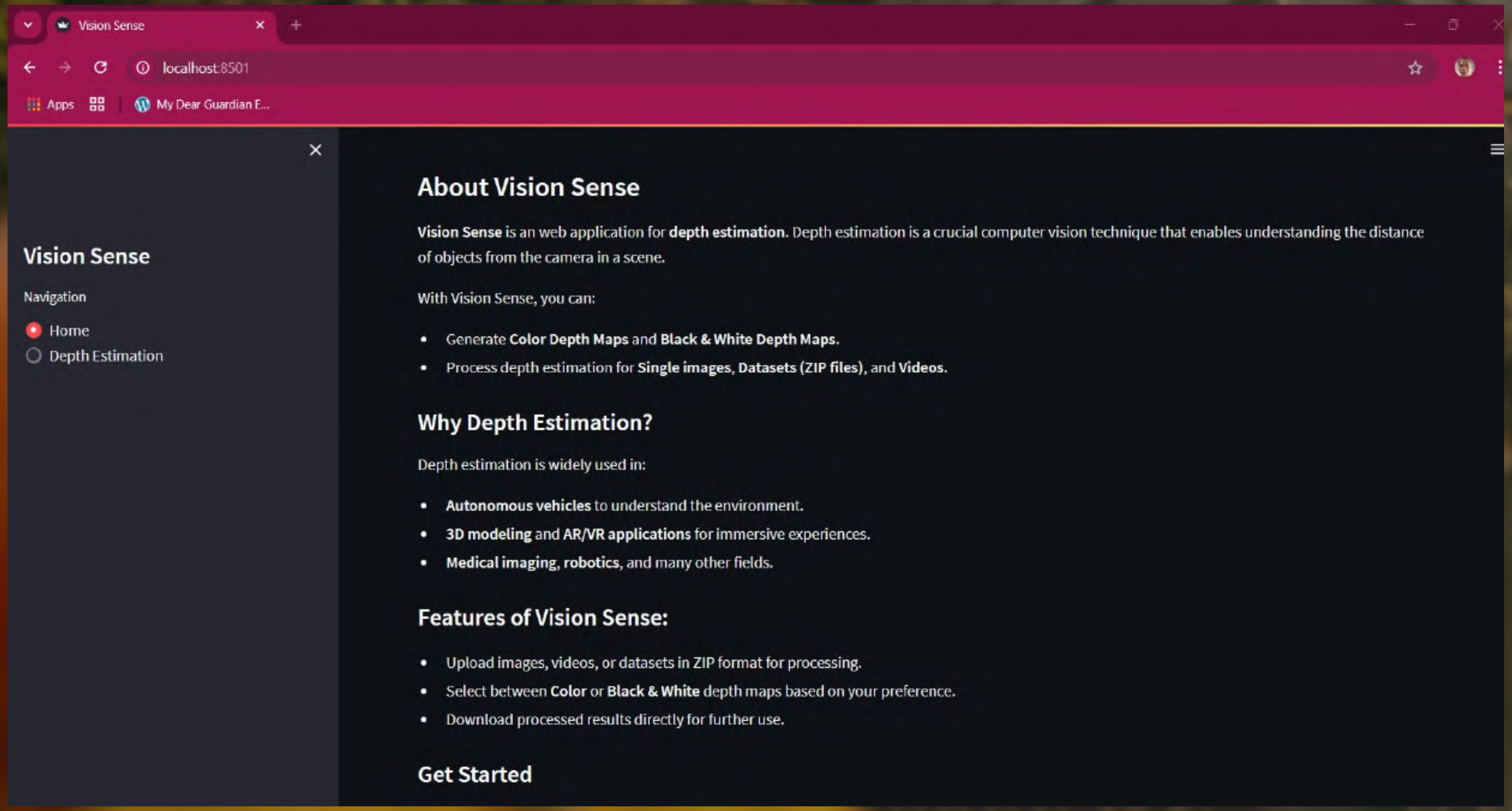
Model Name	Results
	Image/Prediction/Target
Deep Lab v3	
Unet	
Densnet	



Model Name	Test SSIM	Test MSE	Test Loss
Deep Lab v3	0.8584	0.0026	0.0026
U Net	0.8536	0.0042	0.0042
Densnet	0.8620	0.0031	0.0031

CV Project

RESULTS



Vision Sense

Navigation

● Home
○ Depth Estimation

About Vision Sense

Vision Sense is an web application for **depth estimation**. Depth estimation is a crucial computer vision technique that enables understanding the distance of objects from the camera in a scene.

With Vision Sense, you can:

- Generate **Color Depth Maps** and **Black & White Depth Maps**.
- Process depth estimation for **Single images**, **Datasets (ZIP files)**, and **Videos**.

Why Depth Estimation?

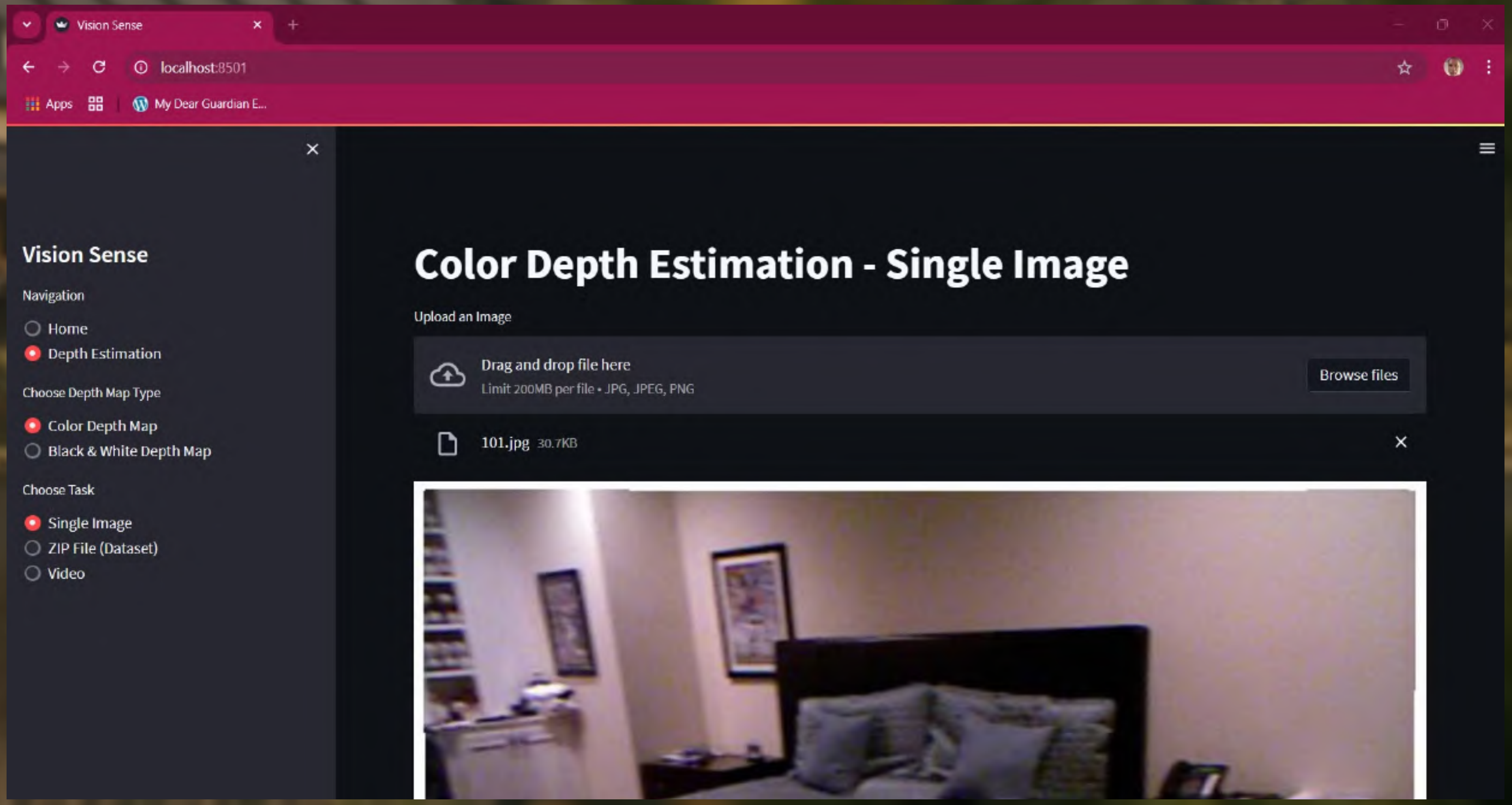
Depth estimation is widely used in:

- **Autonomous vehicles** to understand the environment.
- **3D modeling** and **AR/VR applications** for immersive experiences.
- **Medical imaging**, **robotics**, and many other fields.

Features of Vision Sense:

- Upload images, videos, or datasets in ZIP format for processing.
- Select between **Color** or **Black & White** depth maps based on your preference.
- Download processed results directly for further use.

Get Started



Vision Sense

Navigation

○ Home
● Depth Estimation

Choose Depth Map Type

● Color Depth Map
○ Black & White Depth Map

Choose Task

● Single Image
○ ZIP File (Dataset)
○ Video

Color Depth Estimation - Single Image

Upload an Image

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG

101.jpg 30.7KB

THANK YOU