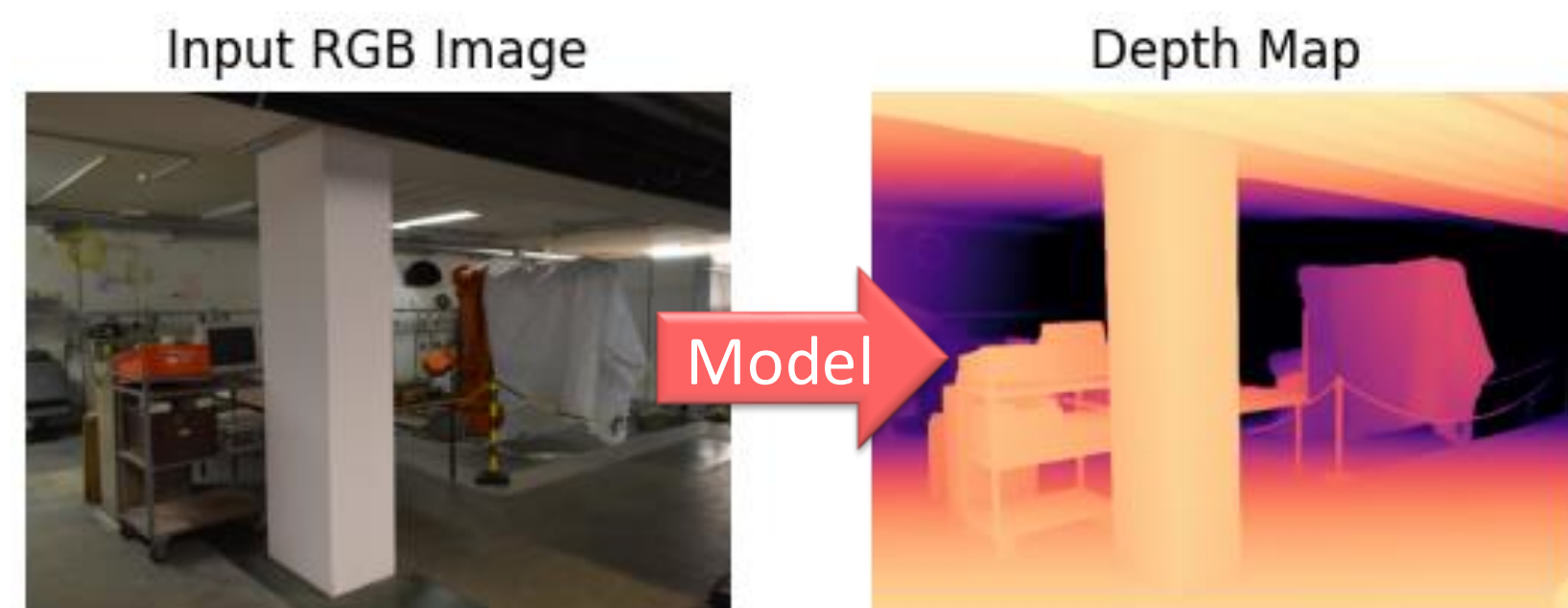# Benchmarking Monocular Depth Estimation Models

Nikolas Spyropoulos, el21202, el21202@mail.ntua.gr
School of Electrical and Computer Engineering, National Technical University of Athens

Introduction & Motivation

- Monocular Depth Estimation (MDE) is critical for **3D Reconstruction**, **AR**, and **Autonomous Driving**

- **The Problem:** Models struggle with **Domain Shift**. Models trained on indoor scenes fail to generalize to outdoor environments

- **The Goal**: Assess the **Zero-Shot Domain Generalization** of modern Foundation Models across diverse environments
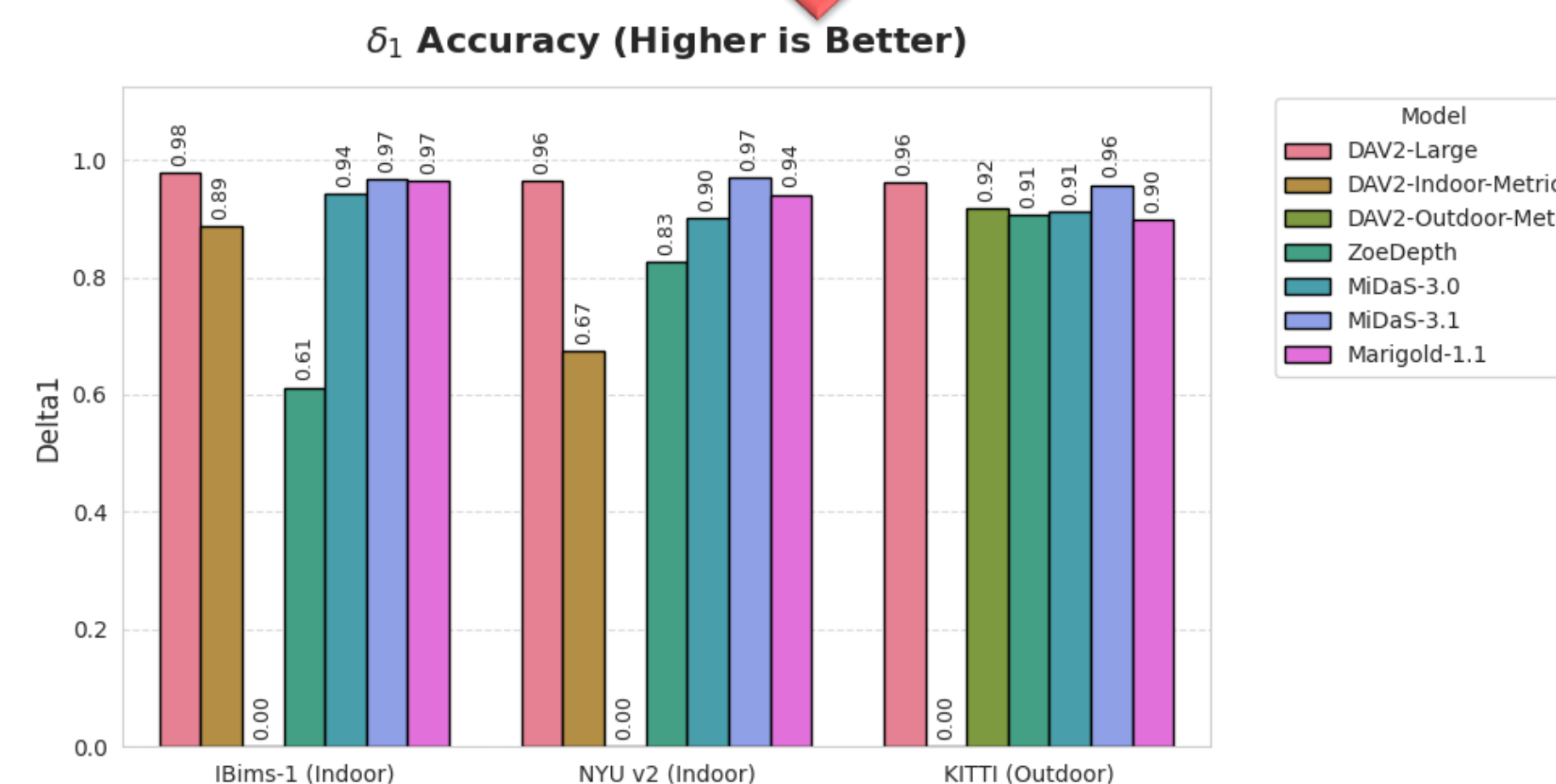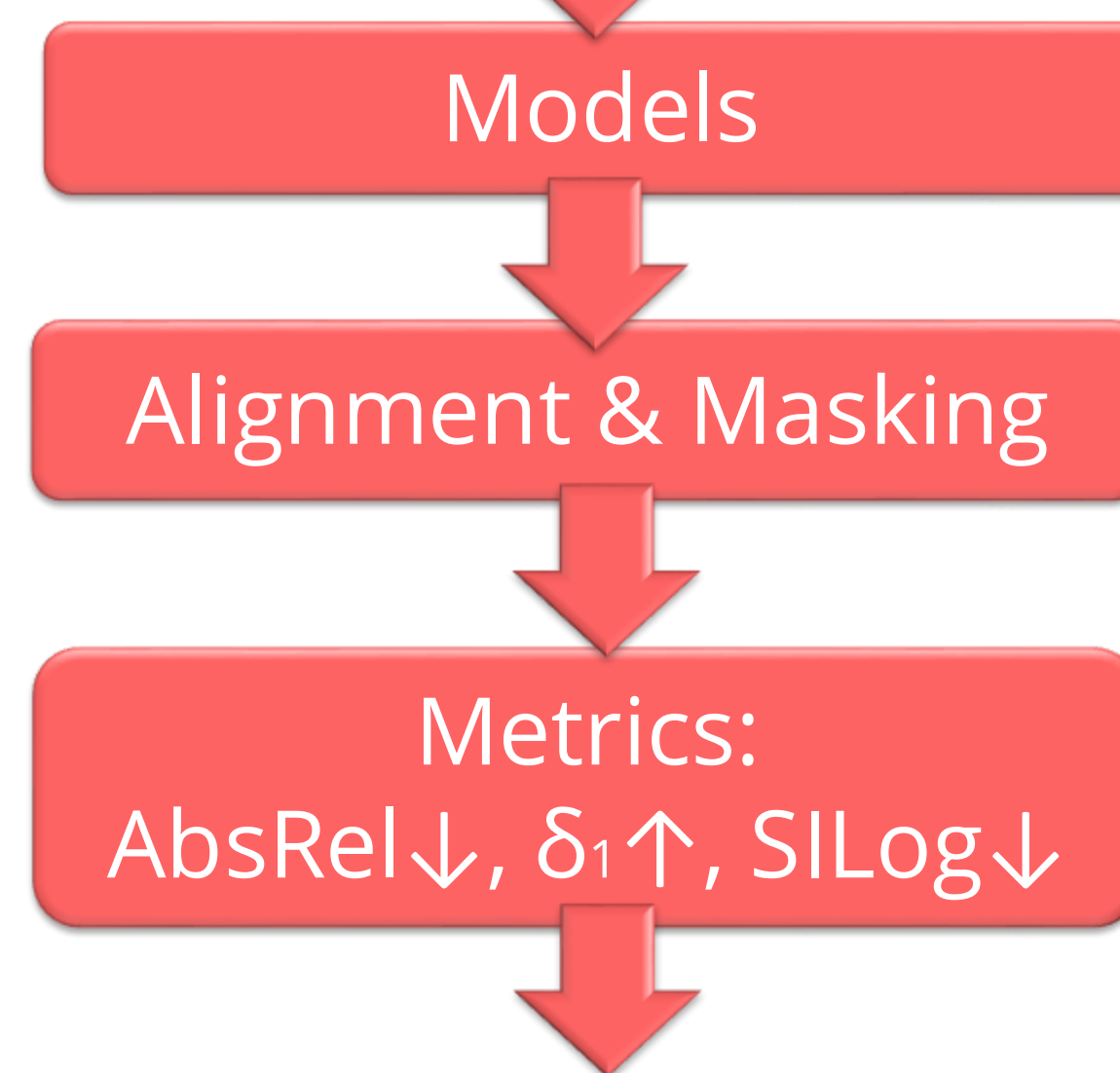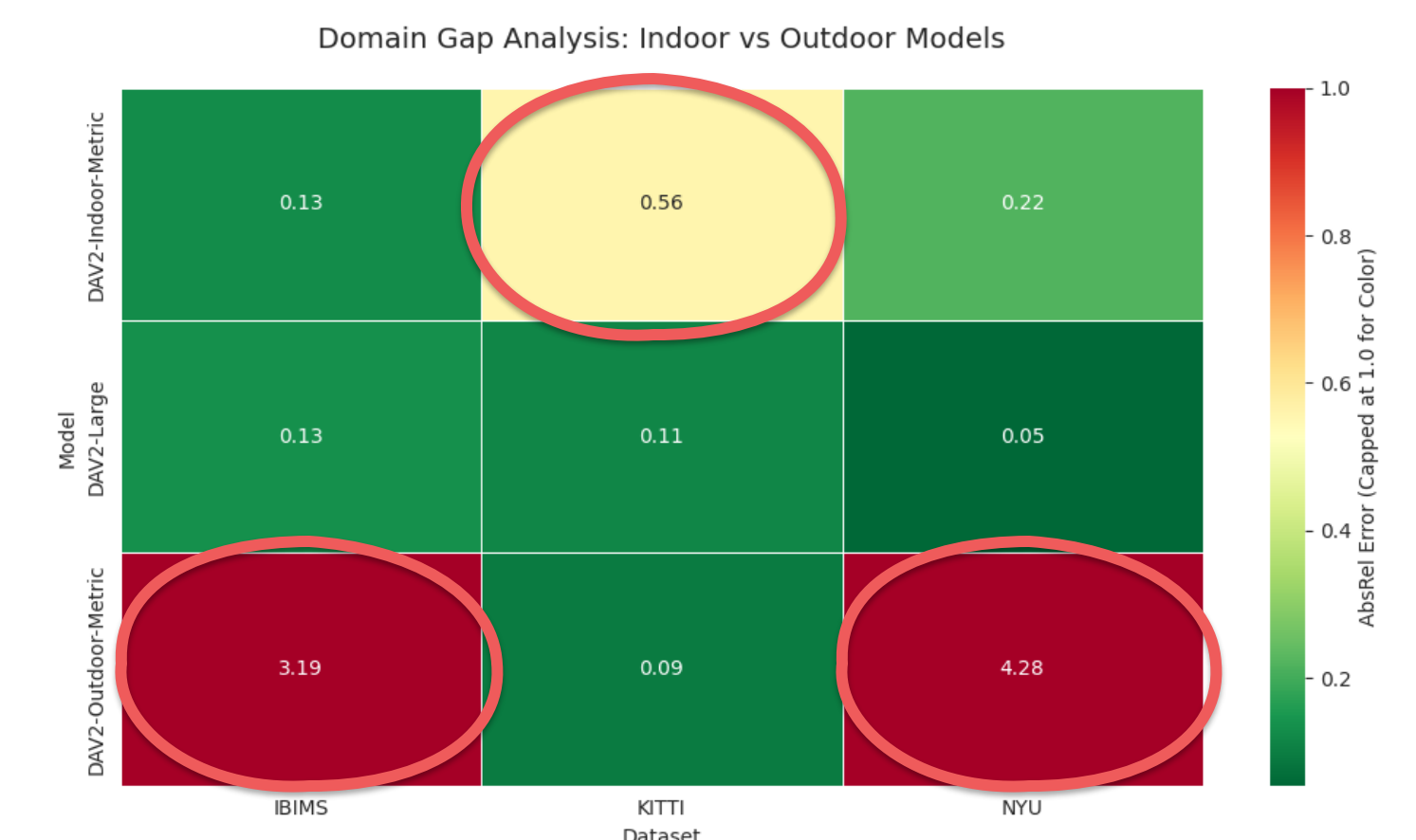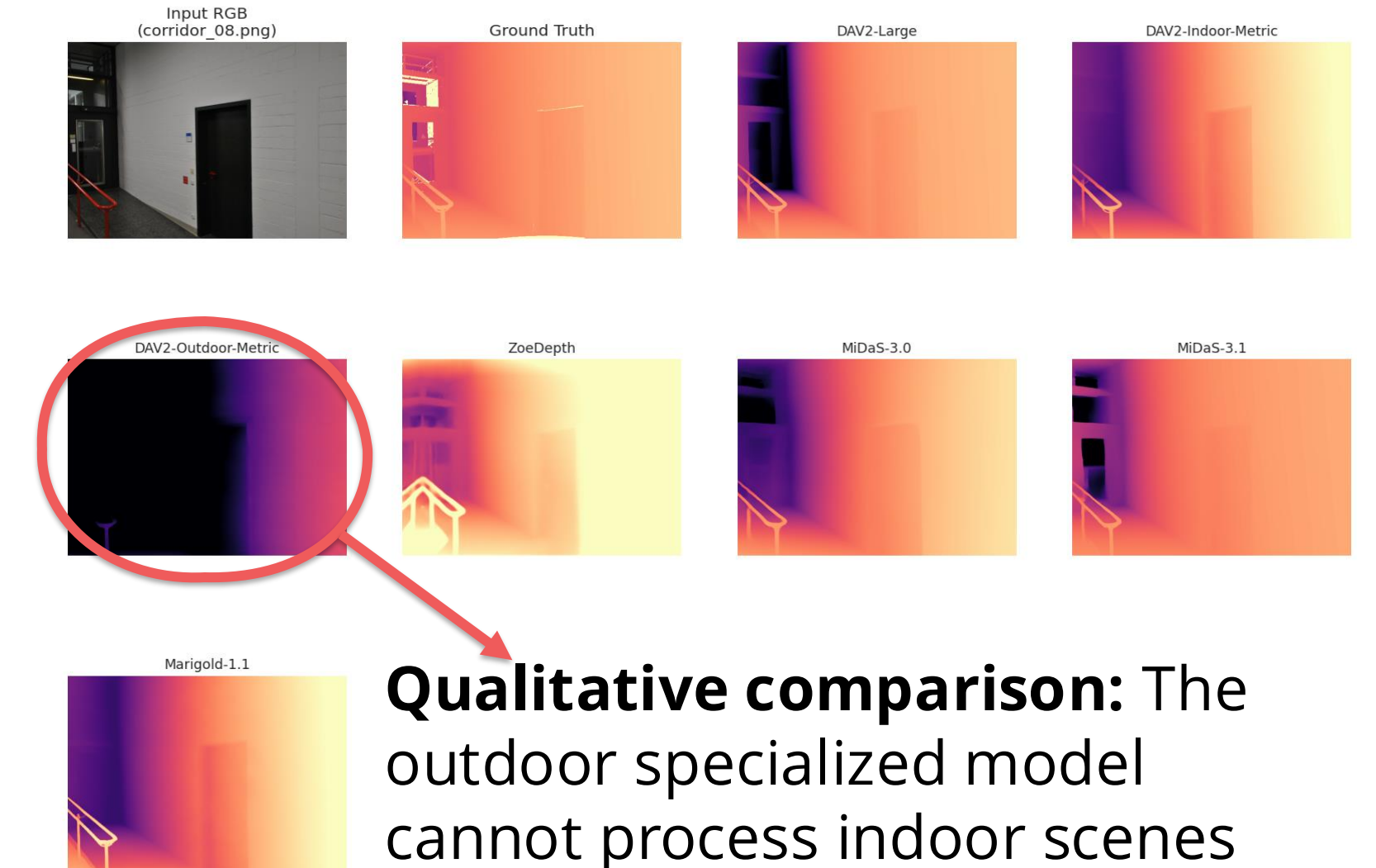

The Models

- **Depth Anything V2**: Relative Depth, Transformer

  - **DAV2-Large**: basic large model

  - **DAV2-Indoor-Metric**: fine-tuned for indoor scenes and metric depth

  - **DAV2-Outdoor-Metric**: fine-tuned for outdoor scenes and metric depth

- **ZoeDepth**: Metric Depth, Transformer

- **MiDaS v3.0 & v3.1**: Relative Depth, Transformer

- **Marigold v1.1**: Relative Depth, Stable Diffusion

The Pipeline

Datasets: **IBims-1**: Indoor/Challenging, **NYU:** Indoor, **KITTI**: Outdoor

Models → Alignment & Masking → Metrics: AbsRel↓, δ_1↑, SILog↓

δ_1 **Accuracy (Higher is Better)**

Domain Gap

Qualitative comparison: The outdoor specialized model cannot process indoor scenes

Conclusions

- **DAV2** is the **best all-rounder** model with **real-time** inference time

- **Marigold** captures the **finest details**, but it cannot be used in real-time

- **Relative depth models** are superior for **"in-the-wild"** scenarios

- **Specialized** models **lack domain generalization** abilities