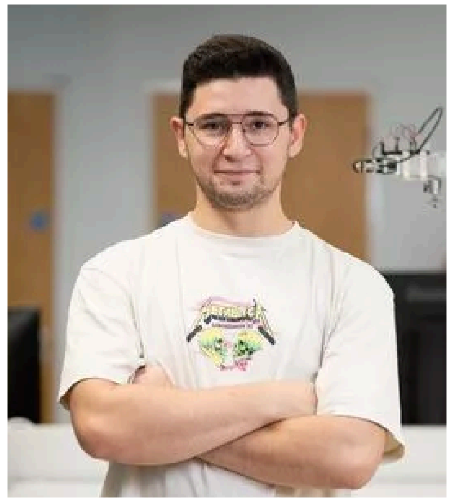
ORIENTATION PREDICTION FOR ROBOTIC MANIPULATION:
ANGLE ENCODING STRATEGIES FOR LINEAR REGRESSION

Faculty of Computing and Engineering – Intelligent Systems Research Centre
University of Ulster – Magee Campus

Authors: Antonio Gambale
Professor Sonya Coleman, Dr Emmett Kerr, Dr Dermot Kerr, Dr Philip Vance, Dr Cornelia Fermuller and Professor Yiannis Aloimonos

ADVCOMP2025



Antonio Gambale

gambale-a@ulster.ac.uk

Professional Experience
- PhD Researcher, School of Computing, Engineering & Intelligent Systems, Ulster University
- PhD Researcher Representative, School of Computing, Engineering & Intelligent Systems, Ulster University

Research & Activities
- Specialises in automation and computer vision for industrial robotics
- Focus areas include orientation detection for components in assembly systems
- Additional research ongoing in autonomous hyperspectral weed identification

List of Publications
- "Computing the Orientation of Hardware Components from Images using Traditional Computer Vision Methods."

2023 The 39th International Manufacturing Conference (IMC39)
- "A Comparative Study of Hough Transform and PCA for Bolt Orientation Detection."

2024 IEEE 22nd International Conference on Industrial Informatics (INDIN)
- "Orientation Prediction for Robotic Manipulation: Angle Encoding Strategies for Linear Regression."

2025 Irish Machine Vision and Image Processing Conference (IMVIP)

Highlights & Initiatives
- **Active representative for PhD researchers on university committees, promoting PhD community interests**
- **Shares research experiences and advancements in automated assembly through university channels**

Background

Challenge in Robotic Manipulation

Robotic grasping pipelines often focus on predicting the gripper pose, typically using complex, multi-parameter representations like grasp rectangles, which are not easily generalisable and can be computationally demanding.

Post-Grasp Tasks

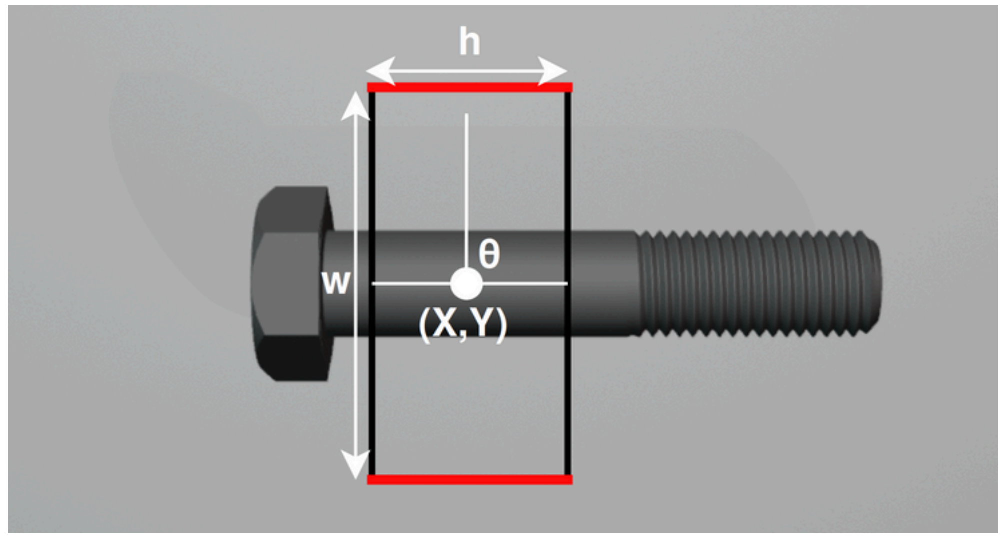
There is a growing need for efficient, object-centric orientation prediction methods that can support diverse manipulation tasks and work with various types of robotic end-effectors. Moving beyond just stable grasping toward robust post-grasp manipulation.

Datasets

Existing datasets and annotation protocols rarely provide the required fine-grained orientation labels, limiting the ability to develop and evaluate object pose centric methods in real-world scenarios.

Gripper Angle 143.6°

Example Multi-Parameter Grasp Rectangle



Object Angle 193.3°

Aims and Contributions

In our paper, we aimed to:

- Develop and benchmark efficient, robust methods for planar object orientation prediction using shallow learning models and a single-angle 360° representation.
- Design a practical annotation pipeline to enrich existing datasets with precise object orientation labels, enabling training and evaluation on both synthetic and real-world data

The contributions of our study are:

- A comprehensive, systematic comparison of encoding schemes, integration strategies and shallow regression models for planar orientation prediction.
- Actionable guidance for deploying reliable, interpretable orientation predictors in robotic manipulation, **identifying XGBoost 1.7 with vector integration and quadrant encoding as the optimal solution for real-world applications.**

Methodology Overview

Patch extraction and pre-processing

- Segment each object from the greater image.
- Extract object patch; normalize background and pad for uniformity.
- Resize to standard input size (224×224), preserving aspect ratio.

CNN Feature Extraction

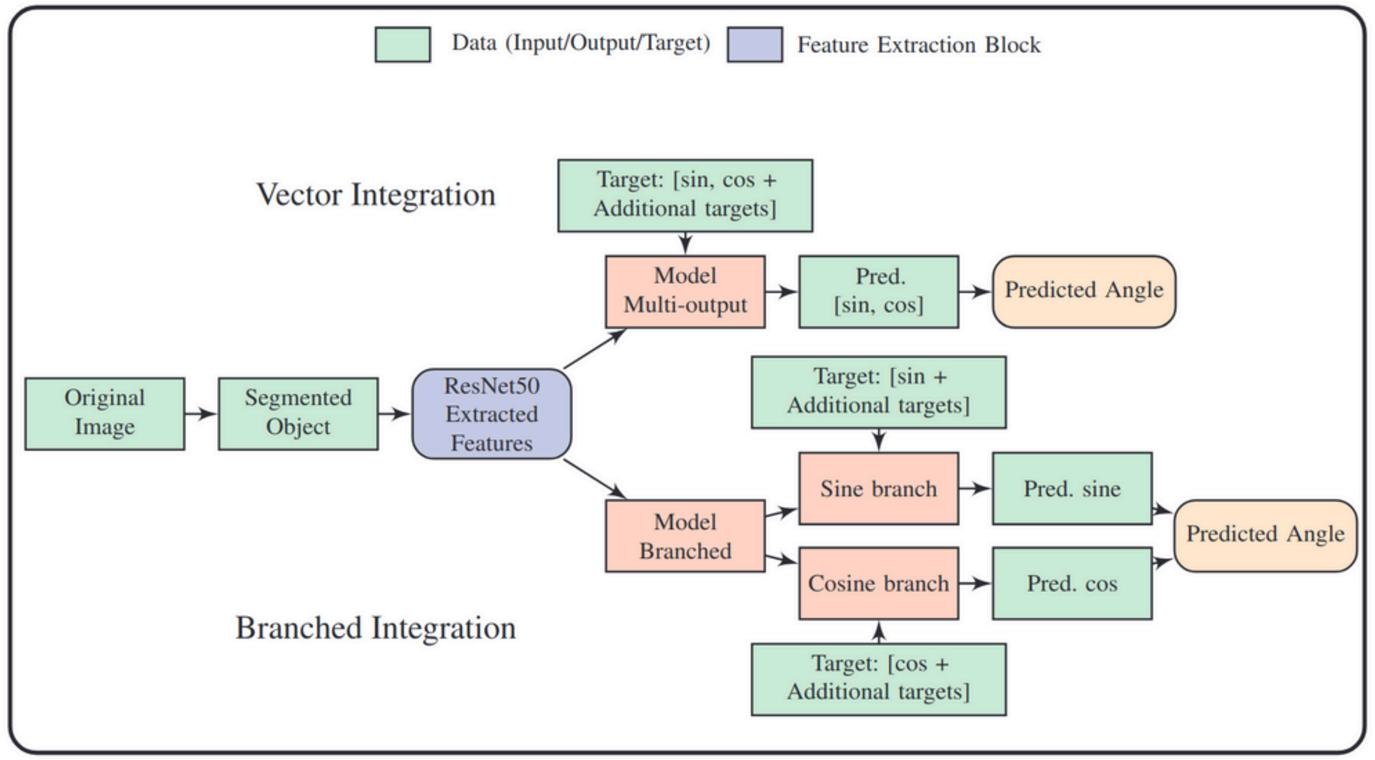
- Use pre-trained ResNet50 (ImageNet, no classification head).
- Extract a 2,048-dimensional feature vector for each patch.

Regression (Orientation Prediction)

- Input extracted features to shallow regressor model.
- Model outputs predictions according to the chosen encoding.

Outputs

- Model outputs are encoded
- Recovered by decoding:
  - Use the inverse tangent function to combine sine/cosine predictions.
  - Angle is normalised to the standard range (0°, 360°).



Angle Encoding Strategies

Encoding Approaches

$$\tan^{-1}\left(\frac{\sin(\theta)}{\cos(\theta)}\right)$$

- **Base Encoding:**

Uses fundamental trigonometric components, encoding angles as ***[sin(θ),cos(θ)]***.

- **Quadrant Encoding:**

Extends base encoding with one-hot encoding for angular quadrants ***[sin(θ),cos(θ),Q1,Q2,Q3,Q4]***.

- **Polar Encoding:**

Adds the angle in radians to the trigonometric components ***[sin(θ),cos(θ),θrad]***.

- **Full Encoding:**

Combines all components for a comprehensive representation ***[sin(θ),cos(θ),Q1,Q2,Q3,Q4,θrad]***.

Model Architectures Tested

Model	Supports Native Multi-Output	Uses Wrapper for Multi-Output	Unified Loss for Multi-Target
Random Forest (RF)	✓		✓
SVR		✓	
M-SVR	✓		✓
XGBoost 1.7		✓	
XGBoost 2.0	✓		✓

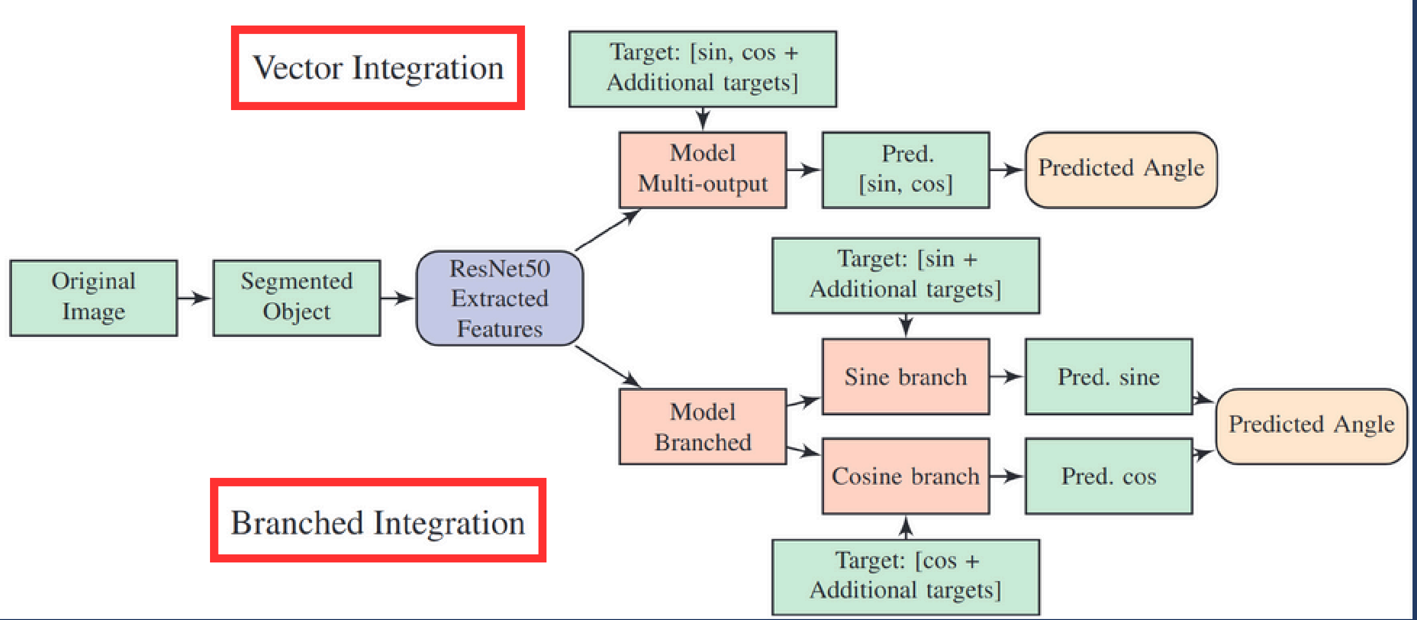
Integration Strategies

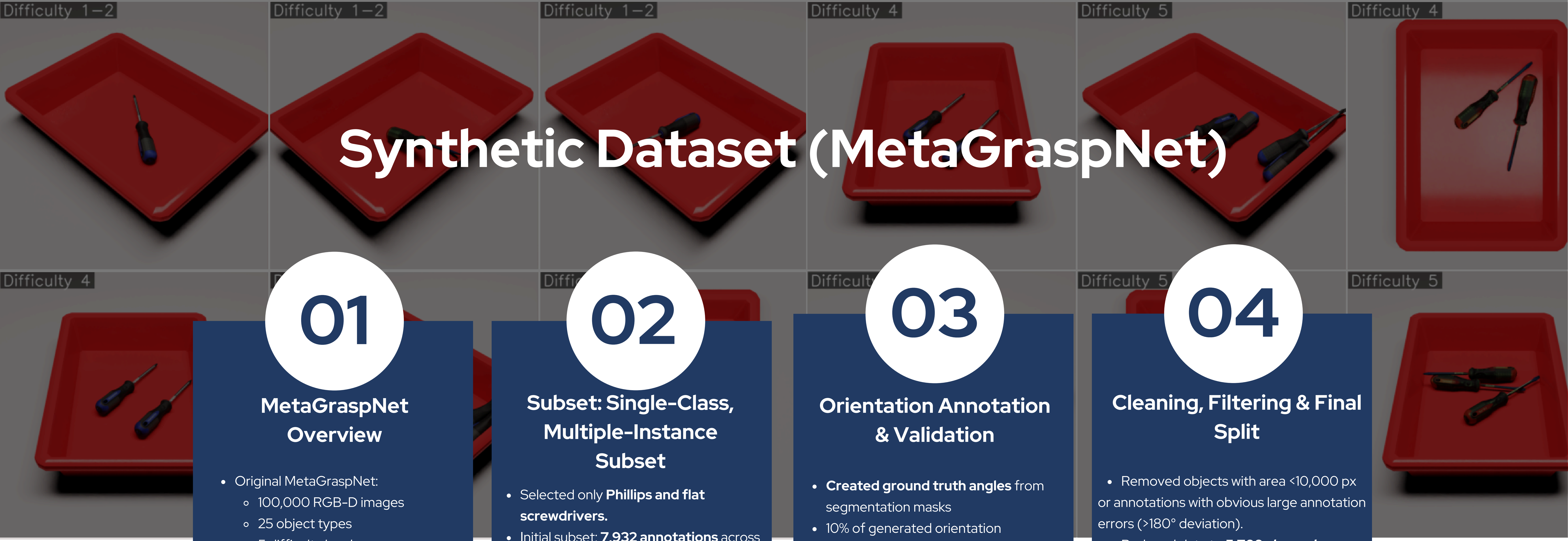
- **Branched Integration:**

Splits model outputs into separate branches for sine and cosine, training them in parallel and combining outputs to recover the angle.

- **Vector Integration:**

Predicts all target variables using a single multi-output model, capturing relationships among outputs and enabling joint error minimisation.





Synthetic Dataset (MetaGraspNet)

01 MetaGraspNet Overview

- Original MetaGraspNet:
  - 100,000 RGB-D images
  - 25 object types
  - 5 difficulty levels
- Designed for evaluating object detection, segmentation, and grasping in varied scenarios.

02 Subset: Single-Class, Multiple-Instance Subset

- Selected only **Phillips and flat screwdrivers.**
- Initial subset: **7,932 annotations** across **2,691 images** from **9 camera poses**

03 Orientation Annotation & Validation

- **Created ground truth angles** from segmentation masks
- 10% of generated orientation annotations manually validated (allowed error ±10°).
- Ensured data integrity for use as ground truth.

04 Cleaning, Filtering & Final Split

- Removed objects with area <10,000 px or annotations with obvious large annotation errors (>180° deviation).
- Reduced data to **5,709 cleaned annotations.**
- Final split: **4,567 training (80%) / 1,142 testing (20%);** all checked for distribution and label quality.

Annotation Creation Pipeline:

Synthetic Dataset (MetaGraspNet) Examples

Real-World Dataset

01 Overview

- Custom dataset of real RGB images of screwdrivers captured under varied lighting and perspectives.
- Comprises **27 images and 81 annotations:** includes different difficulty levels (single, cluttered, and occluded objects)

02 Image capture

- Images captured at 3072 × 4080 pixels resolution using a Samsung ISOCELL GN9 sensor.
- Lens aperture: f/1.9, exposure time: 1/100 second, ISO sensitivity: 386.
- Images collected under uncontrolled ambient daylight, from both overhead and 45° angles, with three screwdriver variants arranged on a white backdrop.

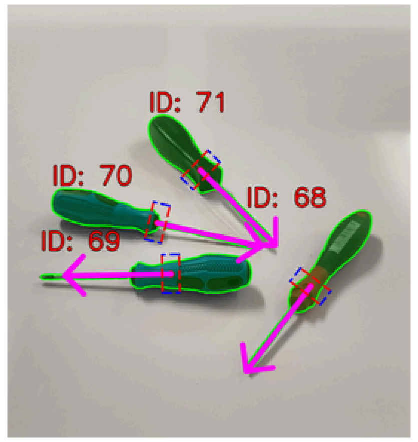
03 Segmentation

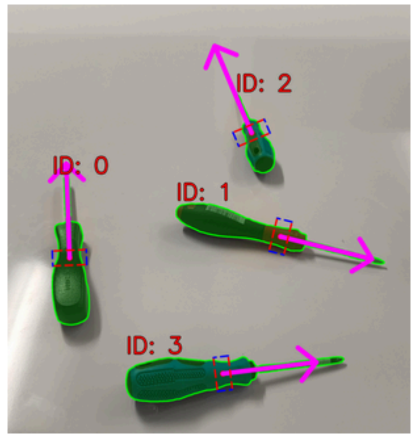
- Segment Anything Model (SAM) used for mask generation.
- Segmentation masks validated by three domain experts (high inter-annotator IoU: 0.95).

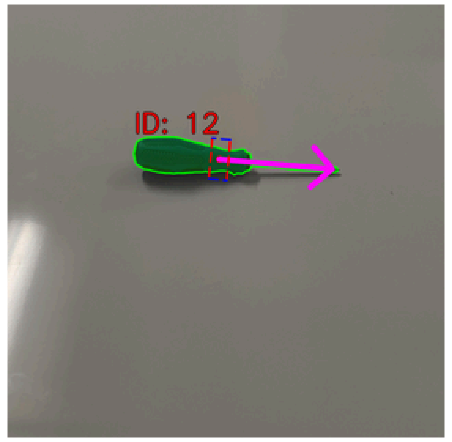
Parameter	Value
Model Type	vit_h
Checkpoint	sam_vit_h_4b8939.pth
Device	cuda
Points Per Side	32
Pred IoU Threshold	0.88
Stability Score Threshold	0.95
Crop N Layers	0
Crop Overlap Ratio	0.3413

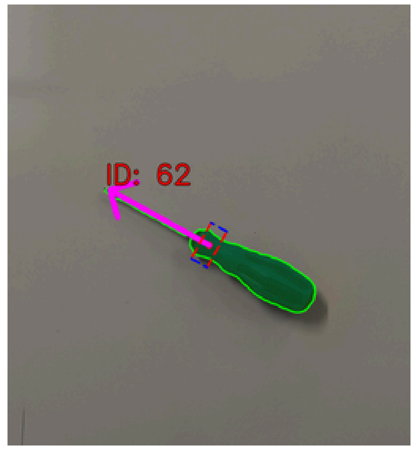
04 Orientation Annotation & Validation

Orientation angles annotated by three domain experts (orientation agreement ±1.8°).

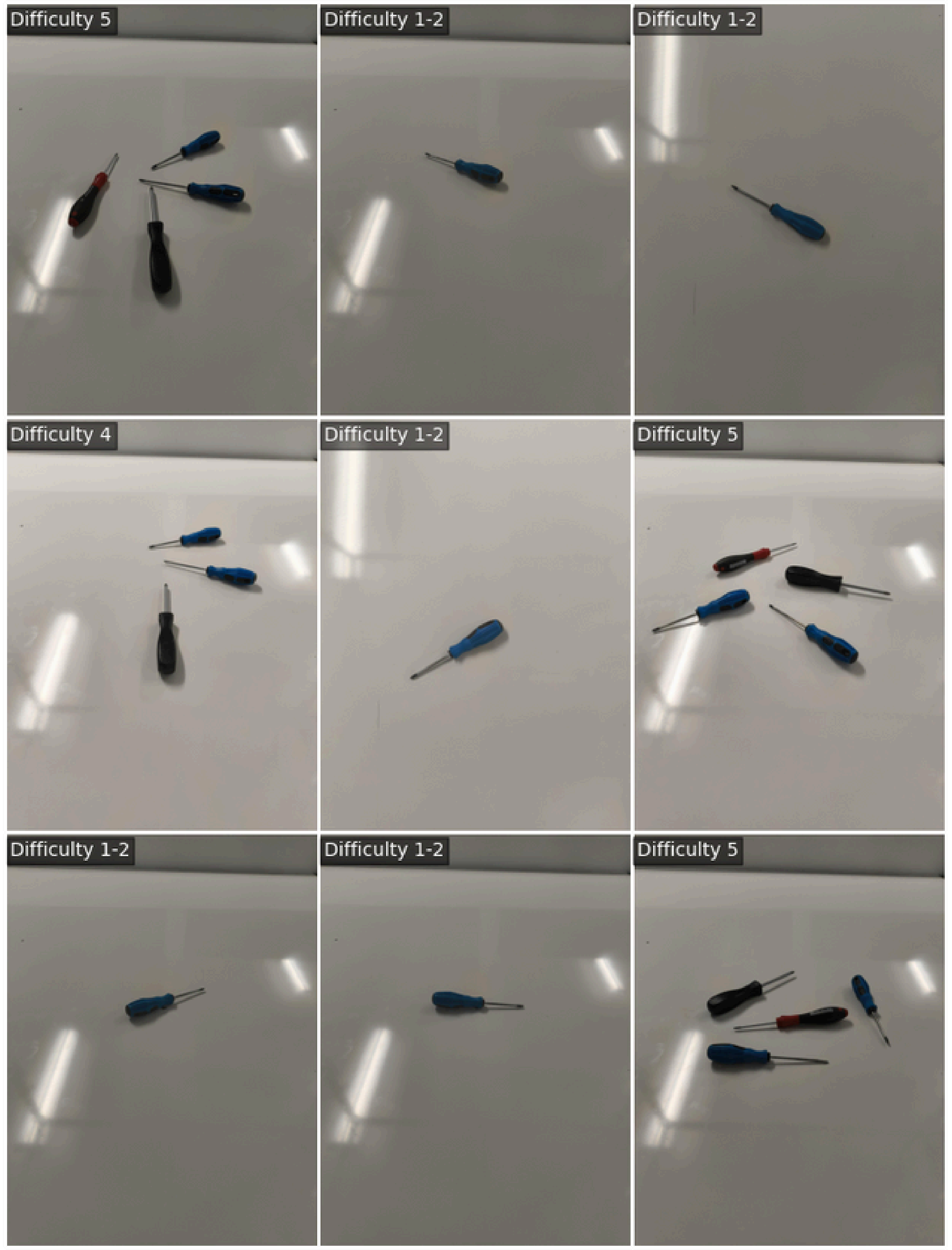


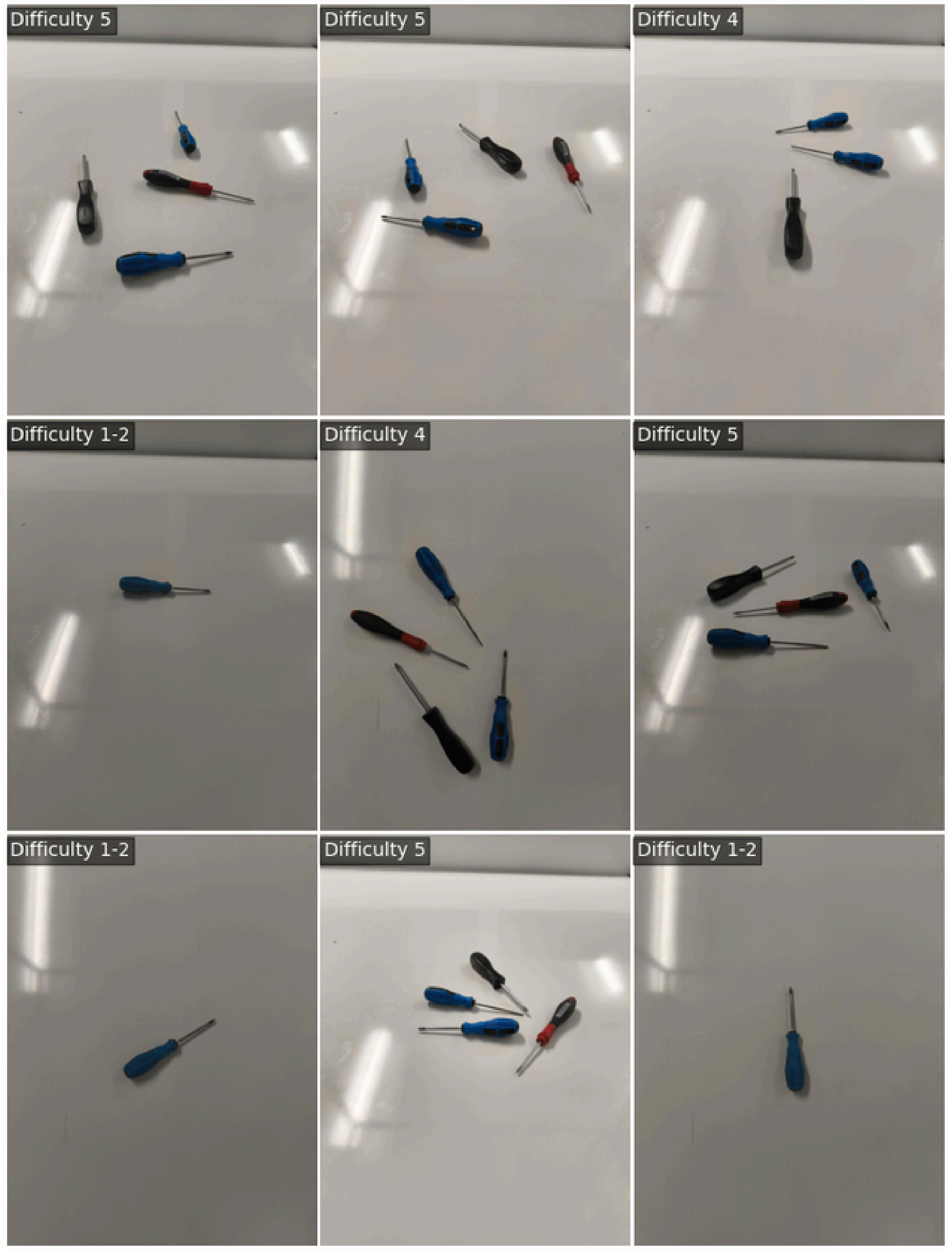






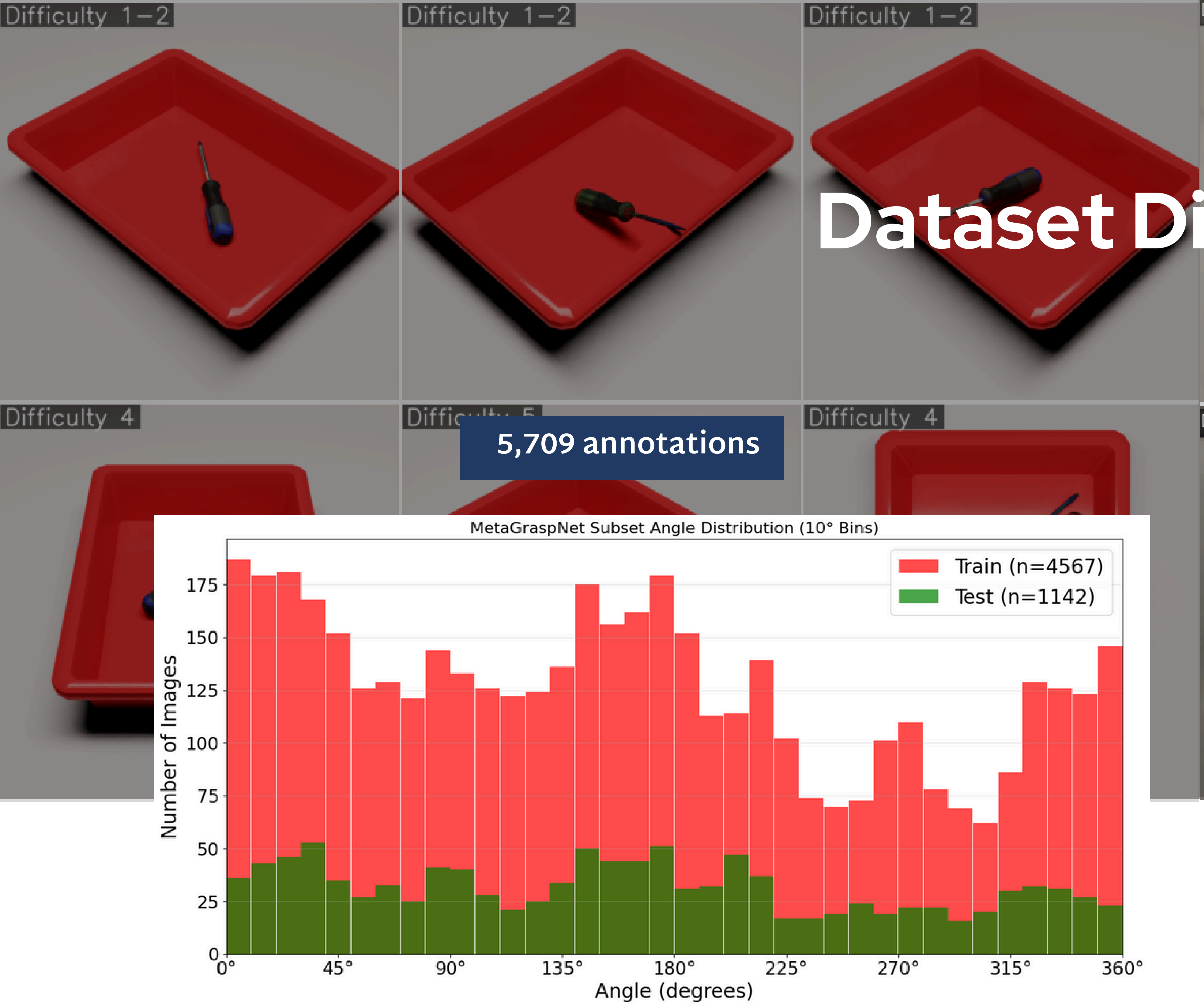
Real-World Dataset Examples



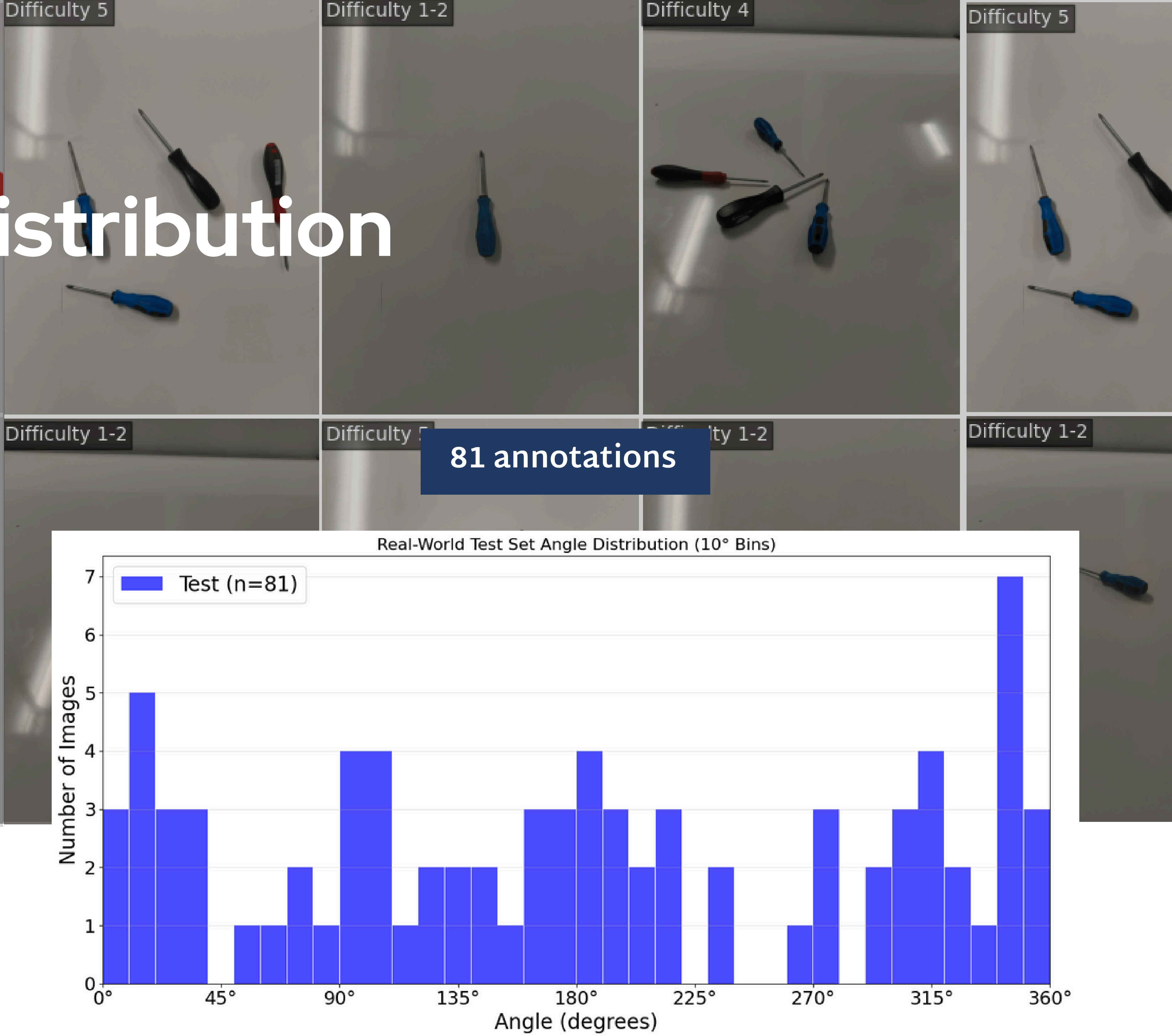


Dataset Distribution

5,709 annotations



81 annotations



Experimental Results

Key Focus

- Metric: Mean Absolute Angular Error (MAAE) in degrees—lower is better.
- Compared: Models, encoding strategies, integration methods (vector vs. branched).
- Datasets: MetaGraspNet (synthetic) and real-world (real-world).
- Speed: Inference time per patch (milliseconds).

Domain Gap Impact:

- **Synthetic → Real-world performance degradation significant**
- **XGBoost 1.7: 5.15° → 8.15° (+58% error increase)**
- **M-SVR: 8.04° → 28.86° (+259% error increase)**

MetaGraspNet Dataset (MAAE in degrees)

Vector Integration Results

Model	Base	Quadrant	Polar	Full
XGBoost 1.7	5.15	5.15	5.15	5.15
XGBoost 2	4.92	5.01	5.46	5.17
M-SVR	8.04	8.04	8.04	8.04
SVR	5.01	5.12	5.01	5.01
Random Forest	5.48	5.08	5.87	5.67

Branched Integration Results

Model	Base	Quadrant	Polar	Full
XGBoost 1.7	5.15	5.15	5.15	5.15
SVR	5.01	5.12	5.01	5.01
Random Forest	7.43	5.44	5.93	5.88

Vector Integration Inference Time Results

Model Type	Base	Quadrant	Polar	Full
XGBoost 1.7	0.76	1.86	0.91	1.73
XGBoost 2	0.76	1.86	0.91	0.29
M-SVR	17.76	17.78	17.8	17.76
SVR	13.48	41.94	20.78	51.25
Random Forest	59.27	56.5	57.9	45.42

Branched Integration Inference Time Results

Model Type	Base	Quadrant	Polar	Full
XGBoost 1.7	0.5	0.83	1.2	3.61
SVR	13.55	70.23	27.68	75.49
Random Forest	119.15	117.08	117.83	91.04

Real-World Dataset (MAAE in degrees)

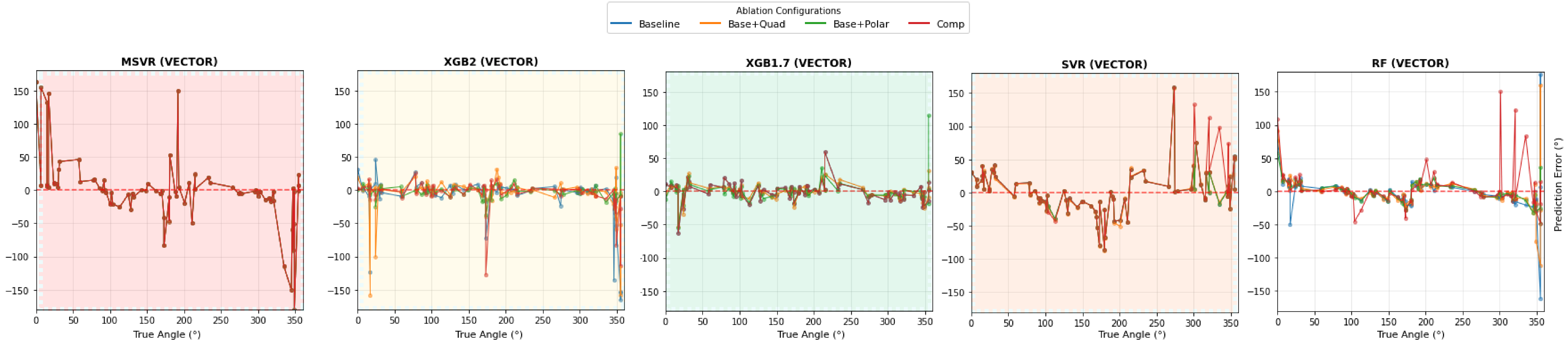
Vector Integration MAAE Results

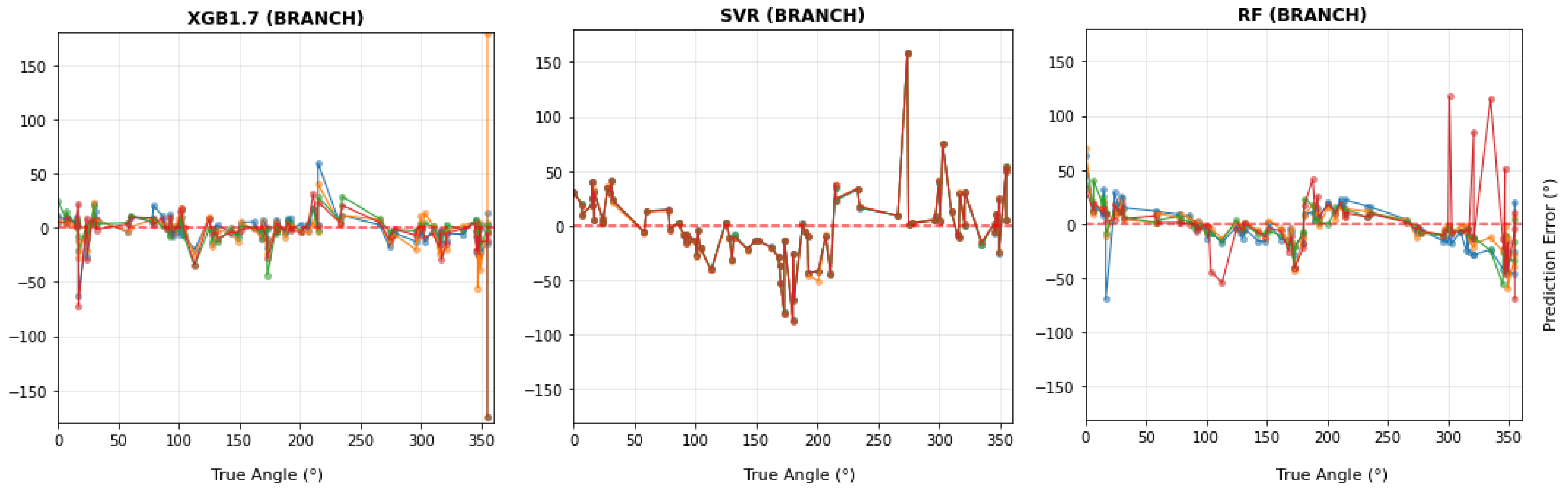
Model	Base	Quadrant	Polar	Full
XGBoost 1.7	9.61	8.15	8.96	9.61
XGBoost 2	17.09	13.91	7.18	10.46
M-SVR	28.86	28.86	28.82	28.82
SVR	23.13	23.54	23.14	28.03
Random Forest	14.81	14.49	11.84	17.43

Branched Integration Results

Model	Base	Quadrant	Polar	Full
XGBoost 1.7	9.61	11.66	9.15	11.14
SVR	23.15	23.54	23.14	23.28
Random Forest	16.45	11.62	12.9	17.5

Experimental Results (Real-world plots)





Boundary Discontinuity Analysis:

- XGBoost 1.7: Most stable across angular range, minimal boundary spikes
- M-SVR/SVR: Severe errors at 0°/359° boundaries, erratic behaviour
- XGBoost 2: Volatile at boundaries AND 180°, despite good MAAE

Integration Strategy Impact:

- Vector preferred for XGBoost 1.7 + Quadrant encoding
- Branched optimal for Random Forest + Quadrant encoding
- Encoding sensitivity varies significantly by model architecture

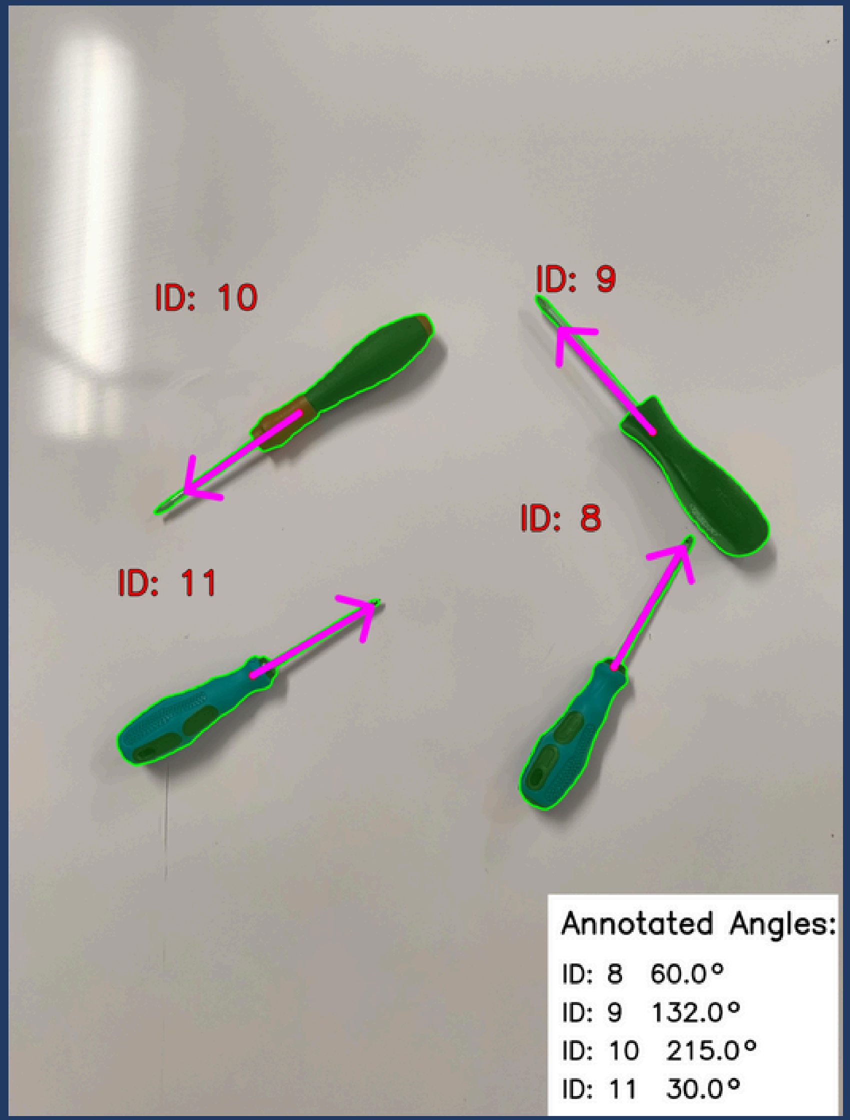
Conclusion & Future Work

01

Conclusion:
- XGBoost 1.7 (Vector + Quadrant) = Best configuration
- Lowest real-world MAAE: 8.15°
- <2ms inference → real-time capable
- Stable predictions, mitigates boundary errors
- Complex encodings (polar/full) = marginal gains, risk instability
- SVR / M-SVR / RF → slower, less reliable, or erratic

Broader Insights
- Synthetic → Real transfer gap is significant
- Integration strategy matters (Vector > Branched for XGBoost)

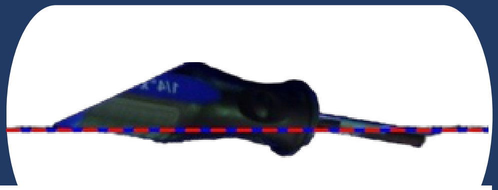
Future Work
- Deploy in real robotic grasping
- Apply domain adaptation to close synthetic–real gap
- Develop hybrid encodings (Quadrant + Polar)
- Add temporal consistency metrics for sequential tasks
- Benchmark vs deep learning for competitiveness

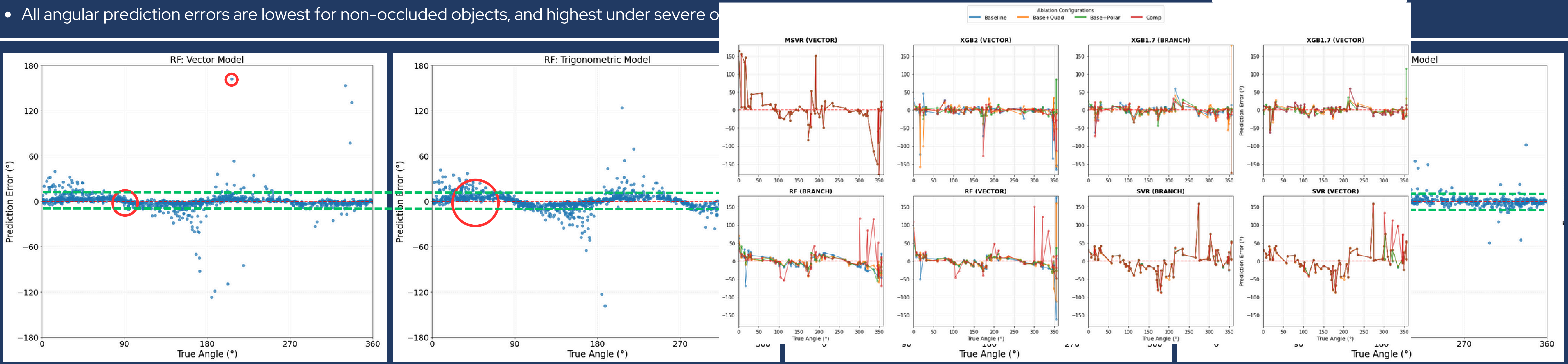


Discussion & Analysis

Key Performance Metrics

- SVR consistently gives high accuracy regardless of encoding, with errors tightly clustered and larger but less frequent outliers.
- RF is clearly improved by vector encoding, but less stable than SVR, with occasional large errors.
- All angular prediction errors are lowest for non-occluded objects, and highest under severe o





- **RF** "snaps" to cardinals: Tree-based models excel at axis-aligned splits. Sin/cos encoding produces extreme values (1,0) or (0,1) at cardinal directions, making these much easier for decision trees to partition precisely. Non-cardinal angles have intermediate values that are harder to split cleanly.
- **SVR** smooths predictions: Support Vector Regression creates continuous, smooth prediction surfaces that interpolate evenly across the angle space. This reduces the accuracy spikes at cardinals but maintains more consistent performance across all orientations.