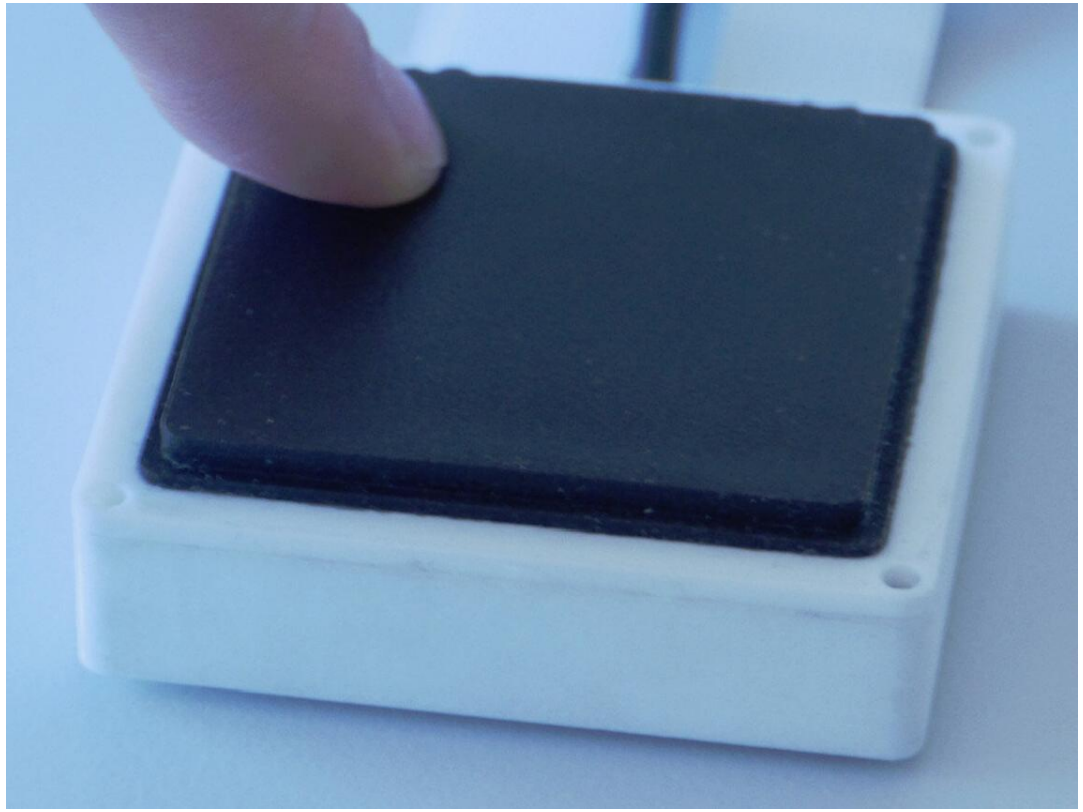
Performance Evaluation Guideline for Vision-Based Tactile Sensors

- Current Research Gap



Lack of standardized performance evaluation criteria

Existing evaluation metrics are often incomplete and do not fully capture the multimodal complexity of tactile sensing

The highly diverse application scenarios of tactile sensors make it difficult to select universal and broadly applicable evaluation metrics.

Main Contribution

- Systematic Evaluation Framework
- Standardized Metrics and Protocols
- Unified Evaluation Pipeline
- Addressing Existing Evaluation Gaps

A IMM: Gelsight (Without Marker) / Digit (Original) RESNET

B MDM Tactip (Black Skin)/ Digit (C-TAC Black) Marker Detection GNN

C IMM+MDM Gelsight (With Marker) / Digit (Reflective layer) RESNET

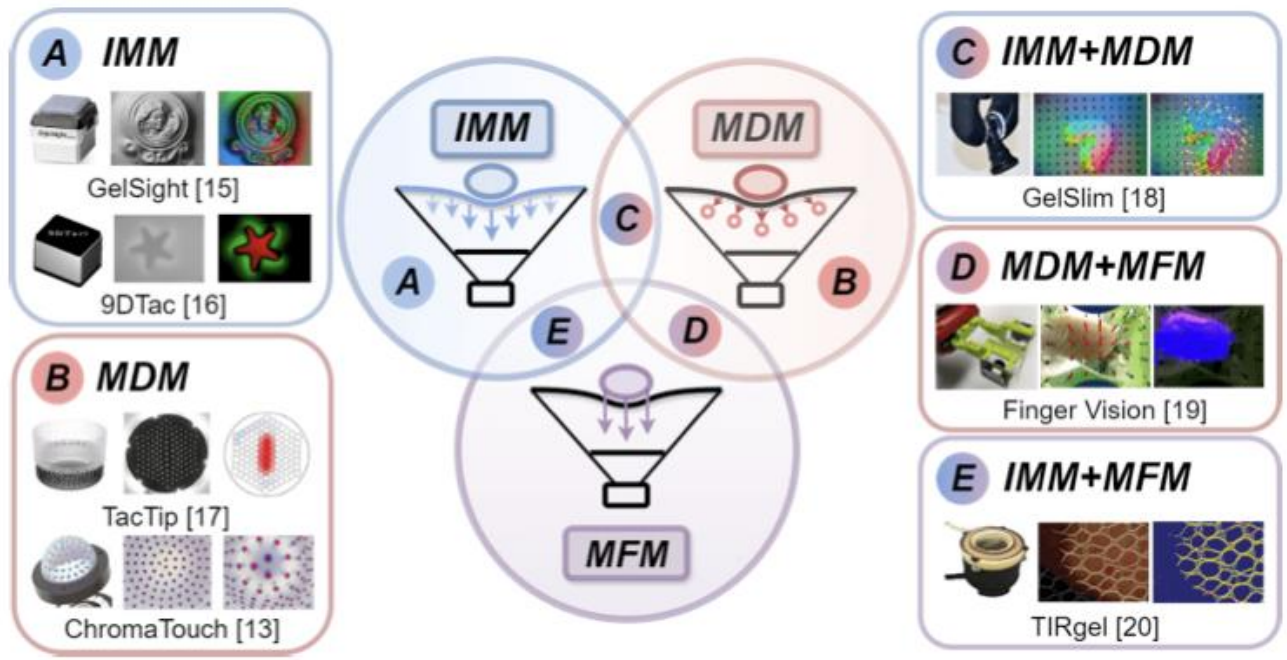
D MDM+MFM Vitactip RESNET

A: RGB, Reflective Layer, No marker

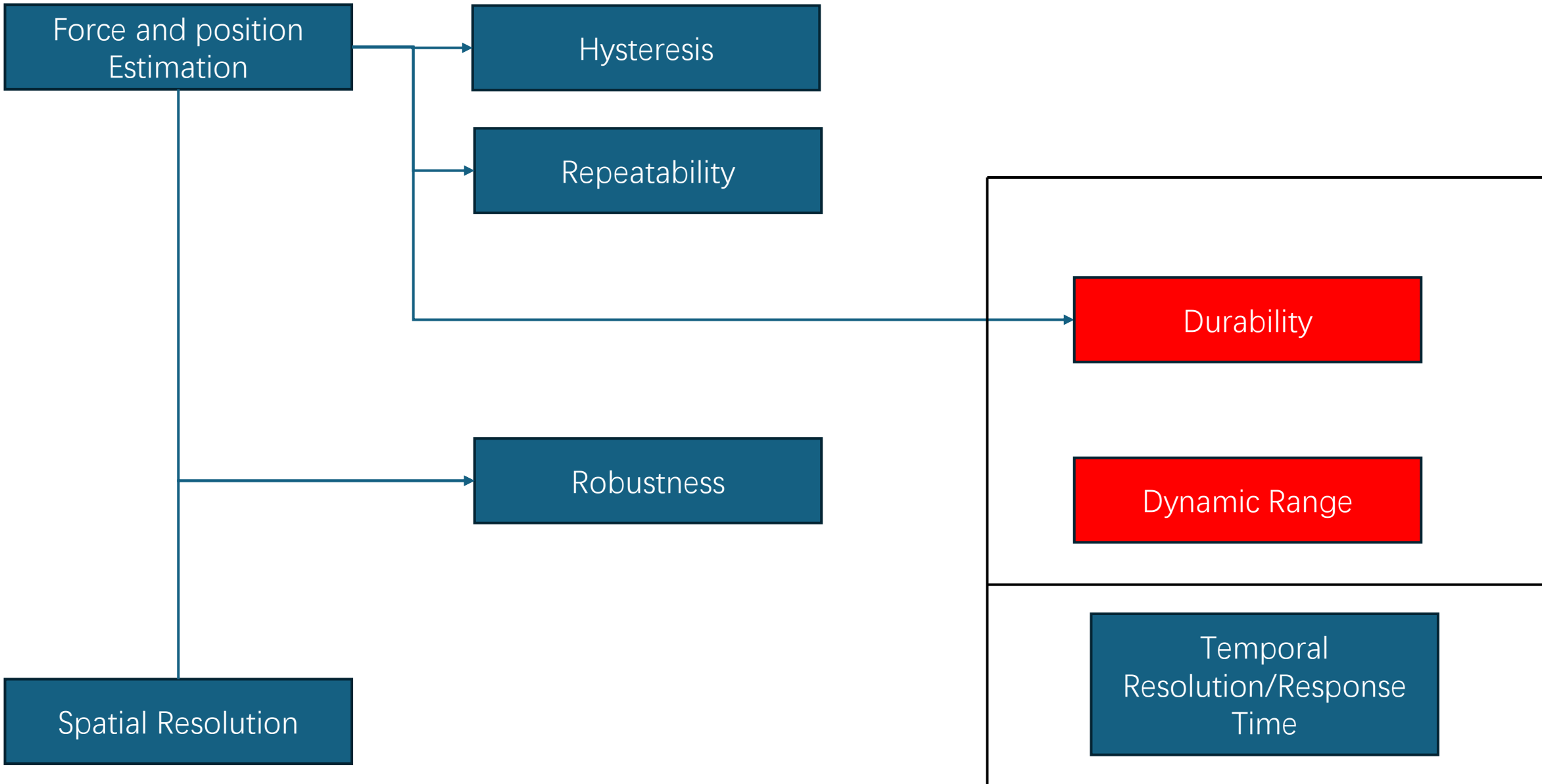
B: ***Black*** Skin, **marker**

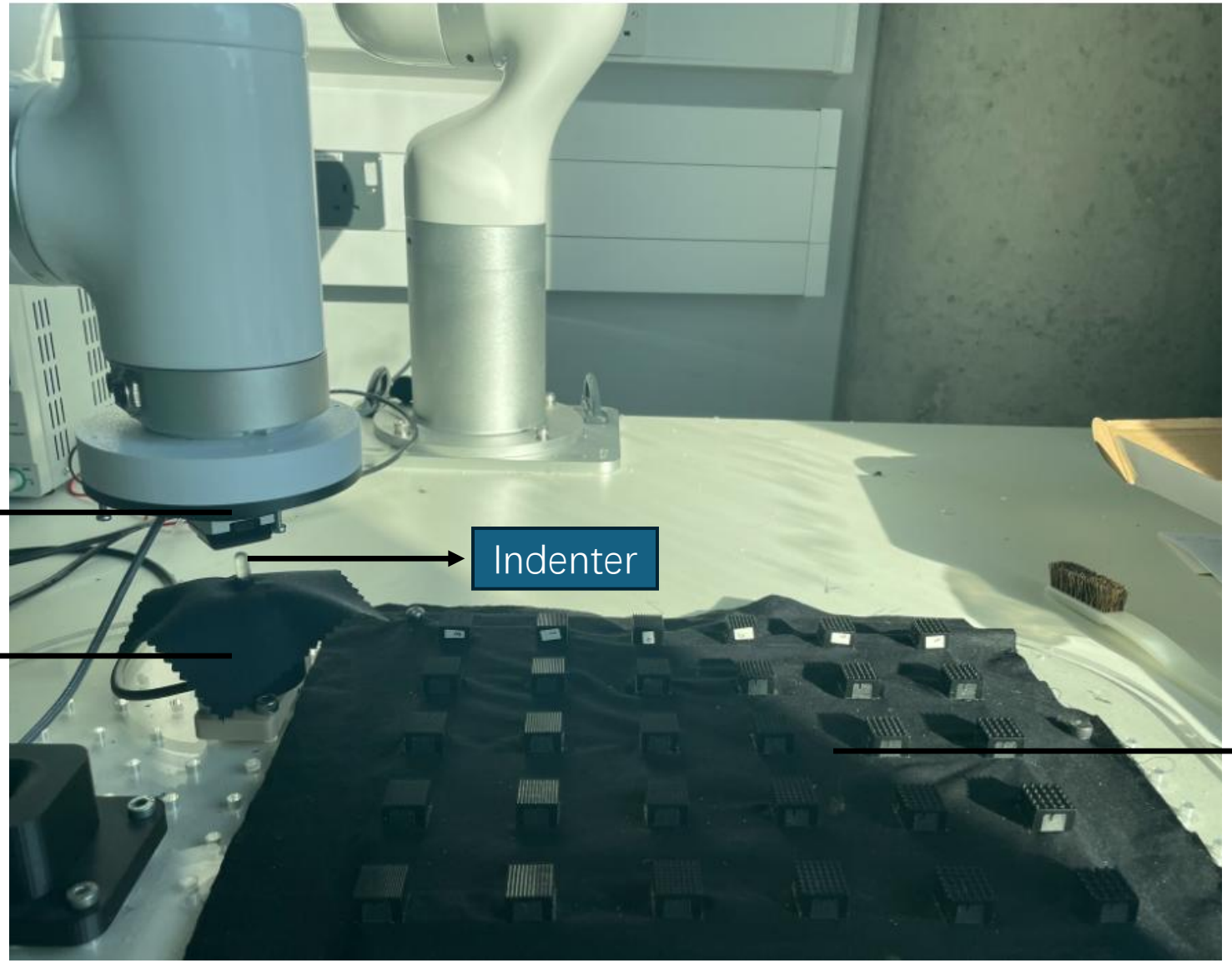
C: RGB, Reflective Layer, **marker**

C: ***Transparent*** Skin, **marker**



Performance Evaluation for tactile sensor

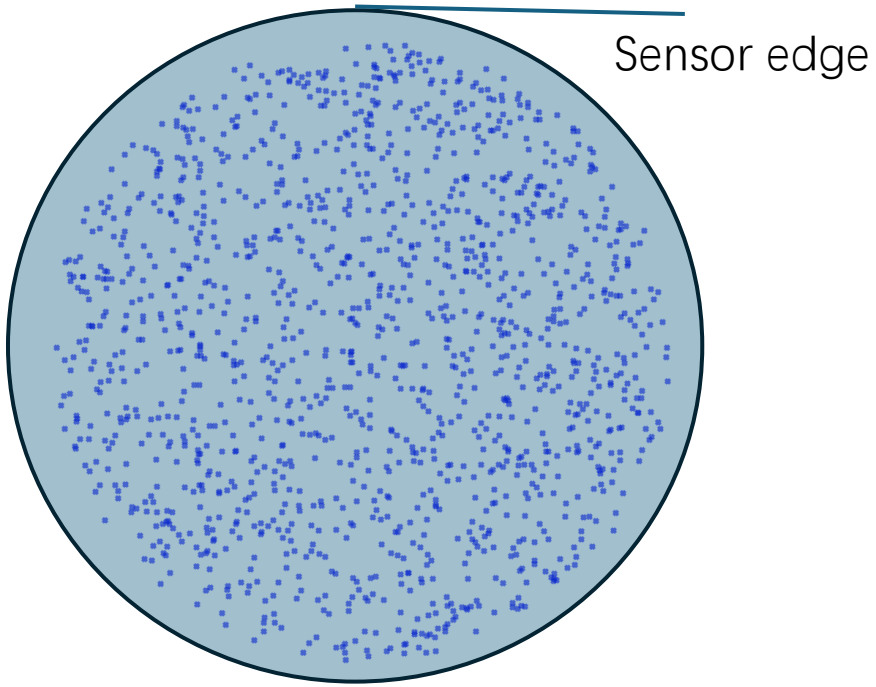



VBTS
Indenter
6-axis force sensor
Spatial Resolution Sample

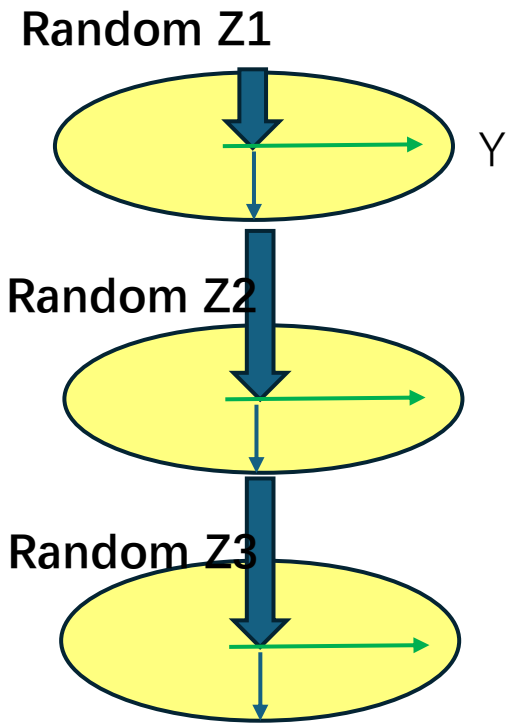
Force estimation and localisation

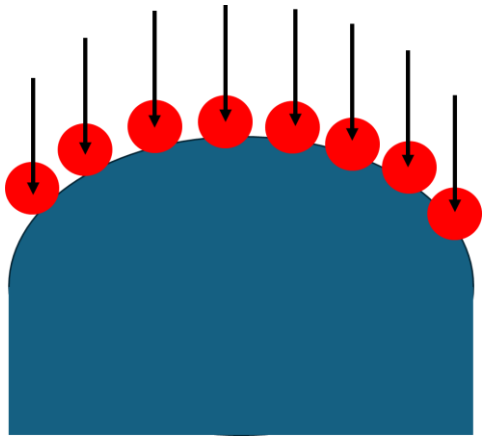
(When force sensor is beyond 0.025N, we assumed it is touched)

1. Find the center of the Sensor
2. Generate a group of random points based on its center
3. Z axis random displacement and X Y random displacement * 3 on the random contact point

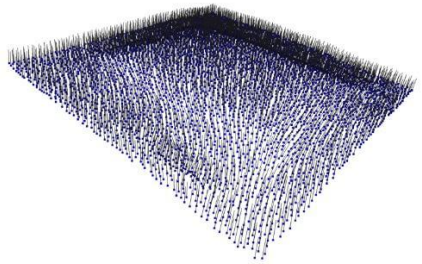


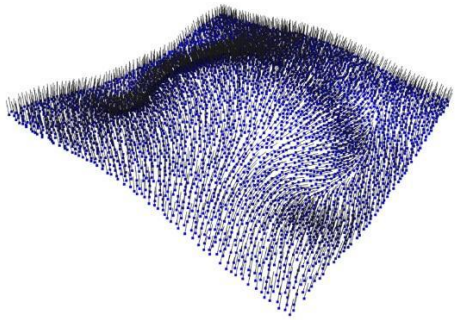
Random Generated points

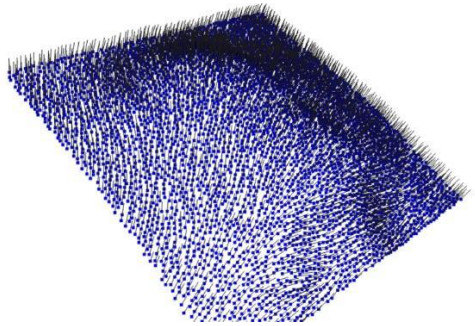




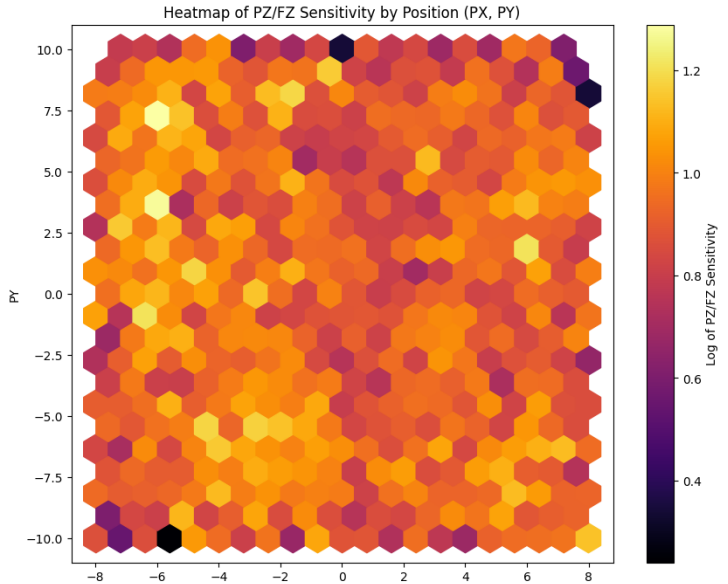
Surface 3D Reconstruction

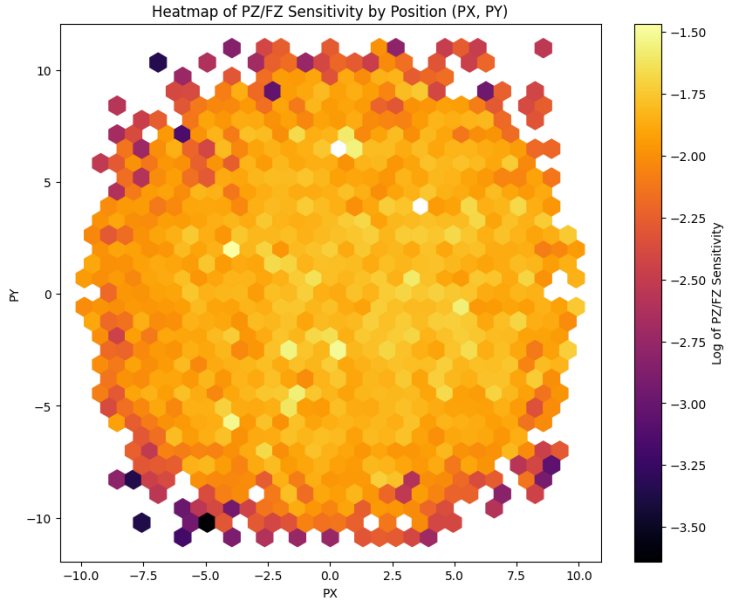






Sensitivity of Sensor in Normal direction



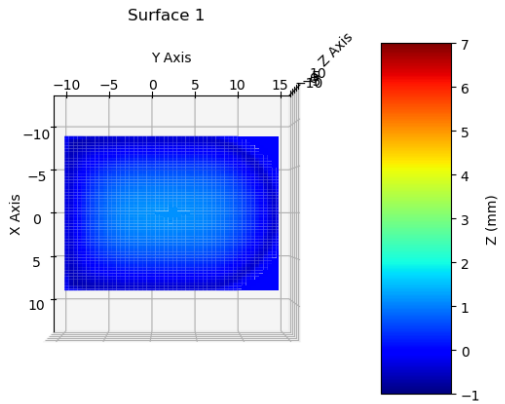


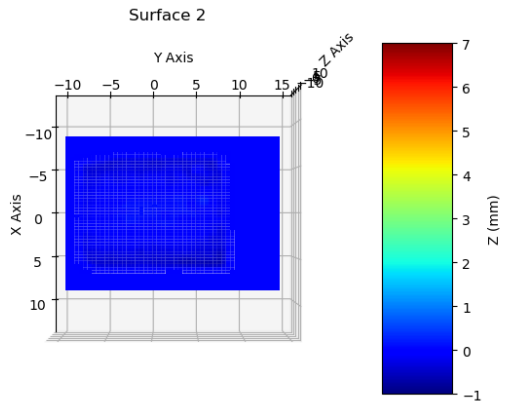
Surface 3D Reconstruction

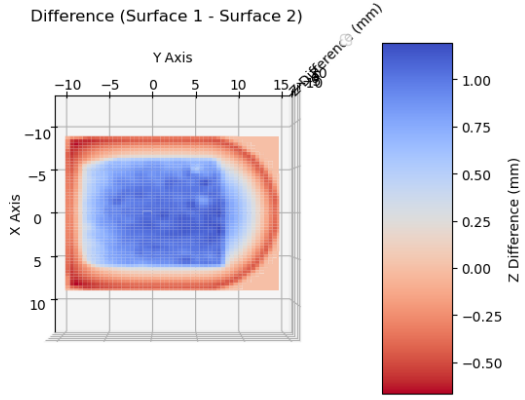
CAD file

Sensor 1

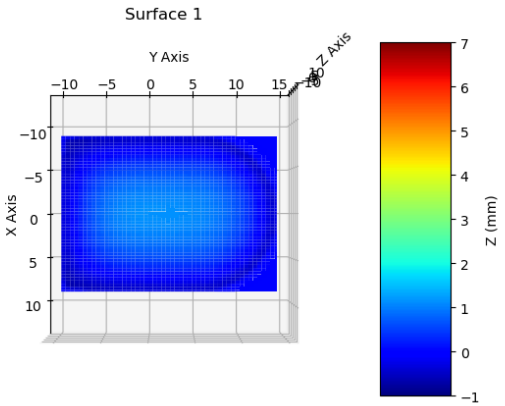
Sensor 2

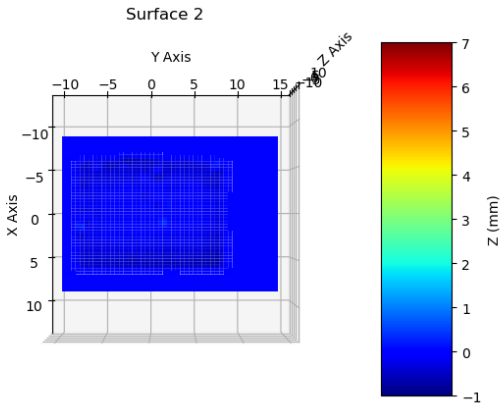


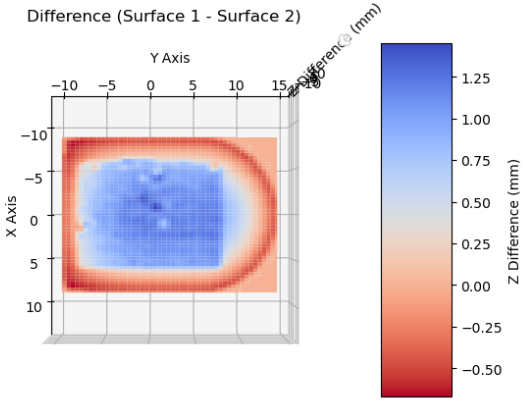




Sensor 1
MSE:**2.94**

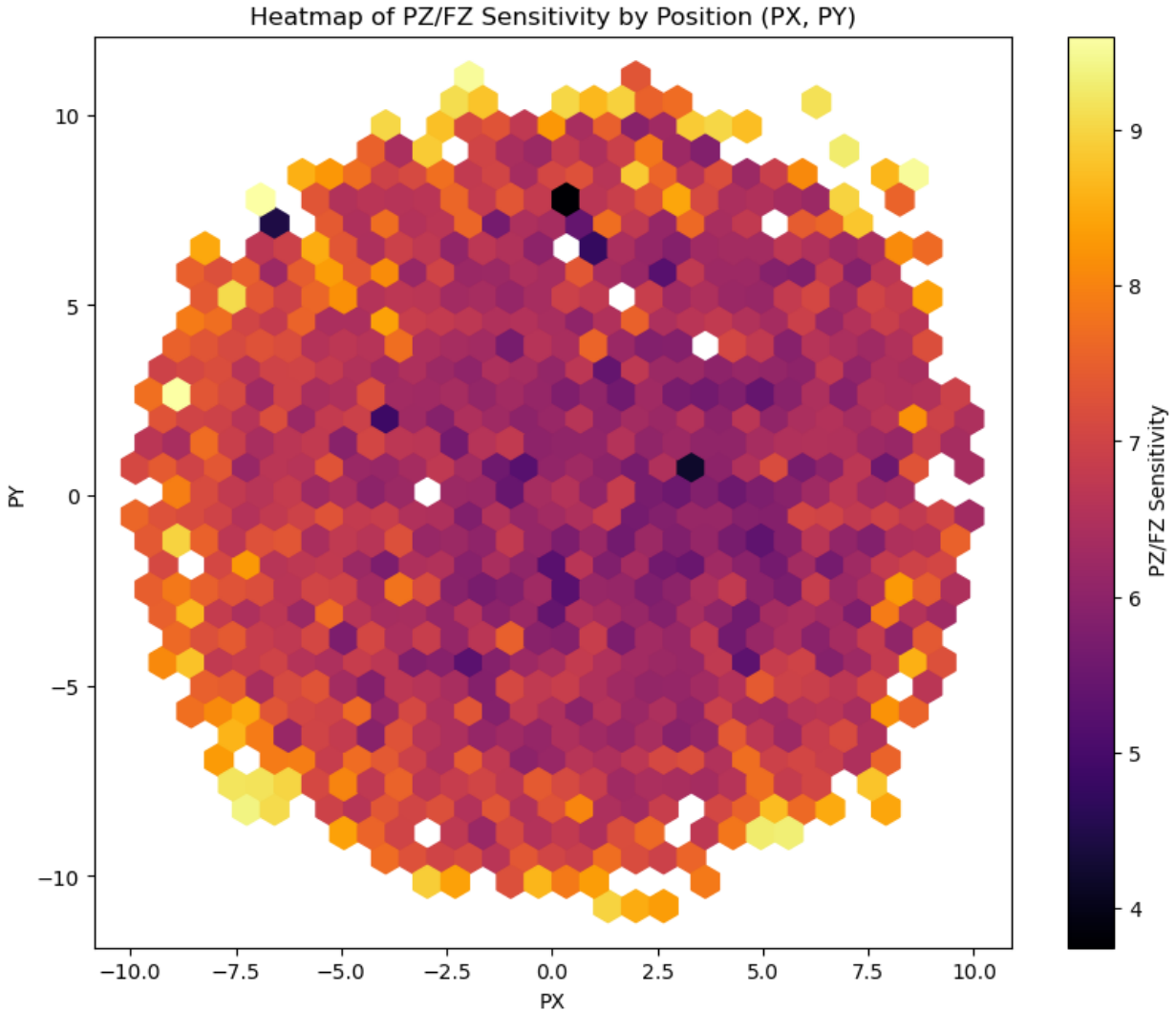




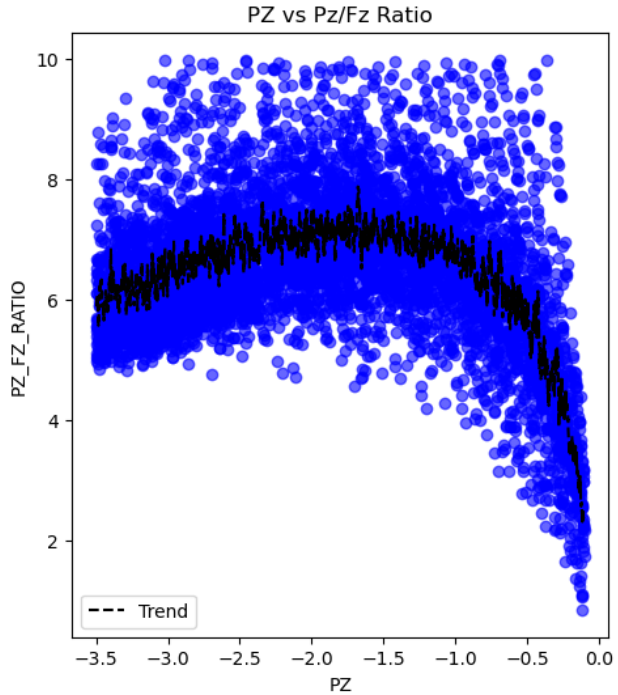


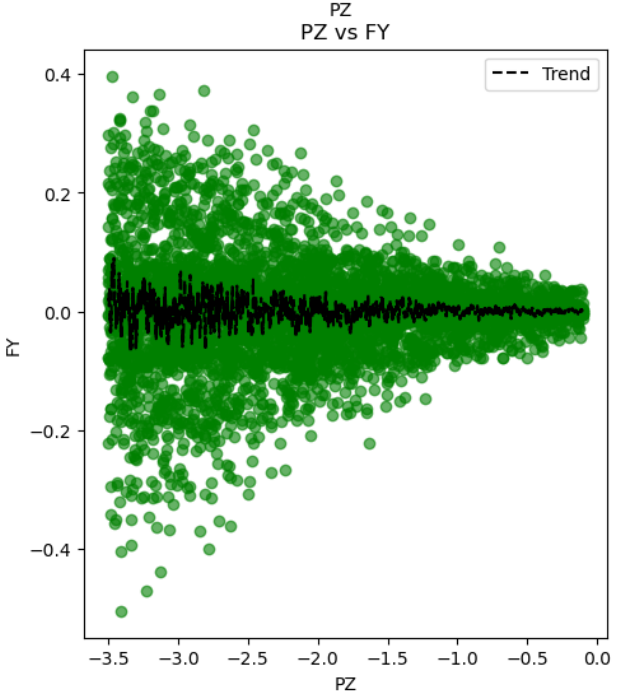
Sensor 2
MSE:**2.927**

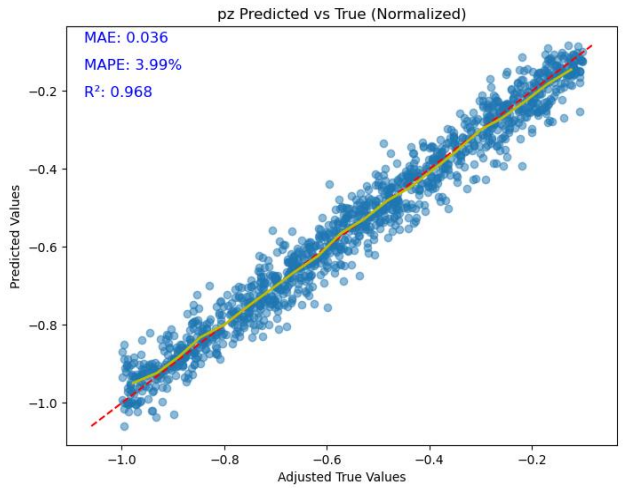
(a)

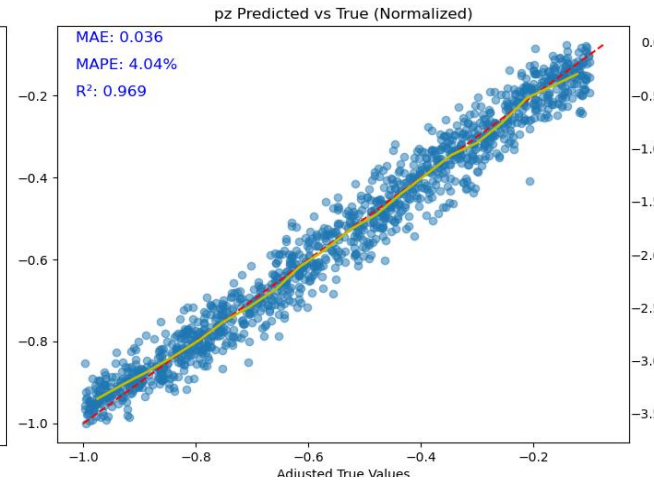

Heatmap of PZ/FZ Sensitivity by Position (PX, PY)
PX
PY
PZ/FZ Sensitivity
10
5
0
−5
−10
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
9
8
7
6
5
4

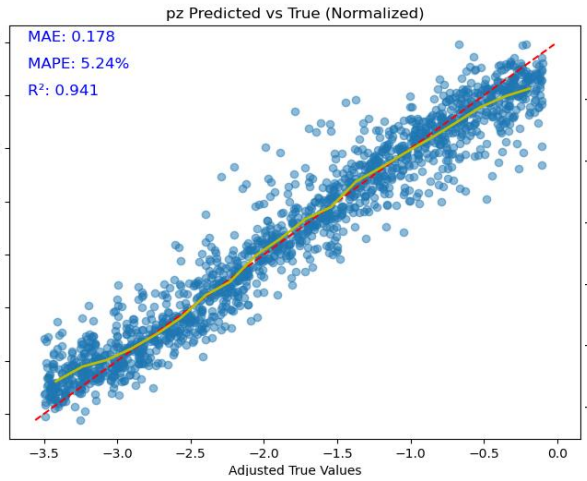
(b)

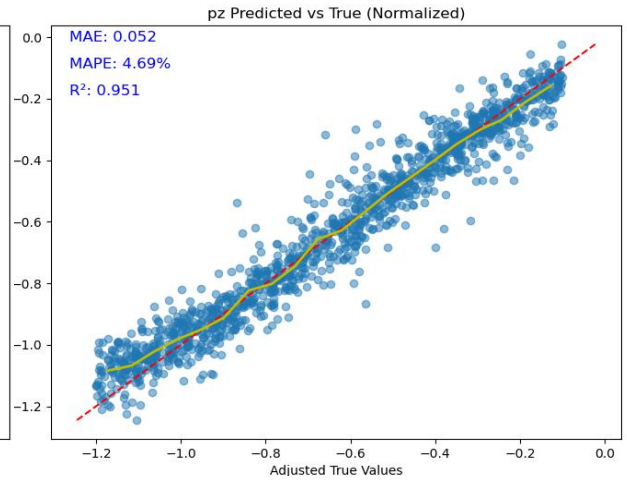

PZ vs Pz/Fz Ratio
PZ_FZ_RATIO
PZ
Trend
10
8
6
4
2
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0

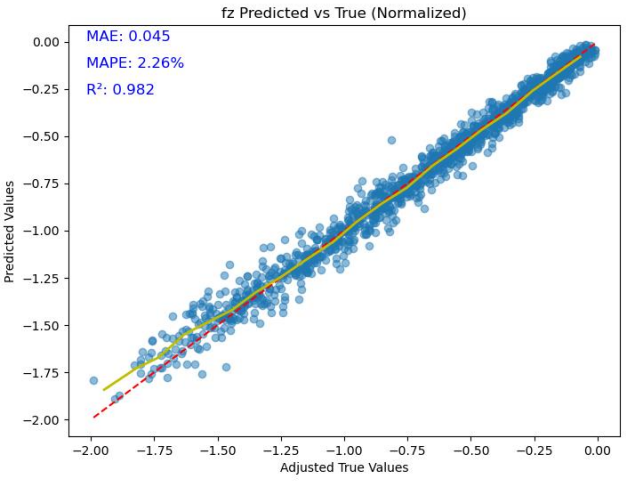

PZ vs FY
FY
PZ
Trend
0.4
0.2
0.0
−0.2
−0.4
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0

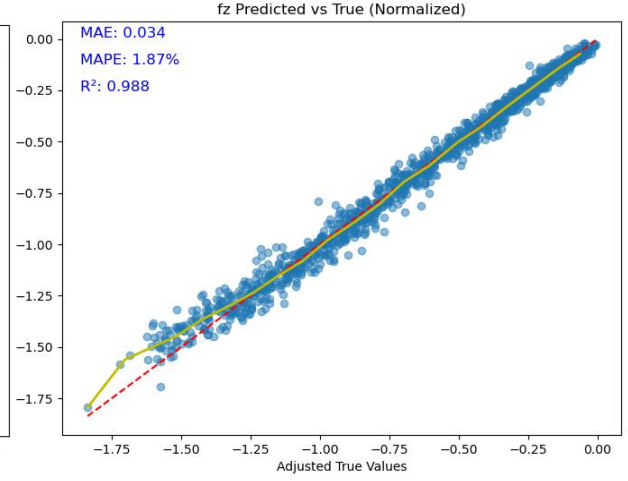


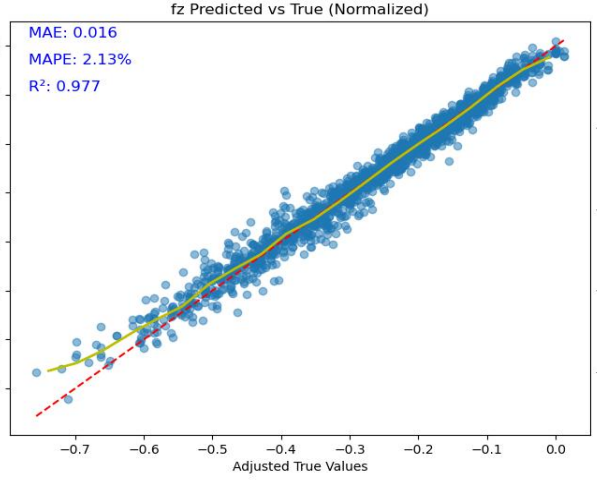


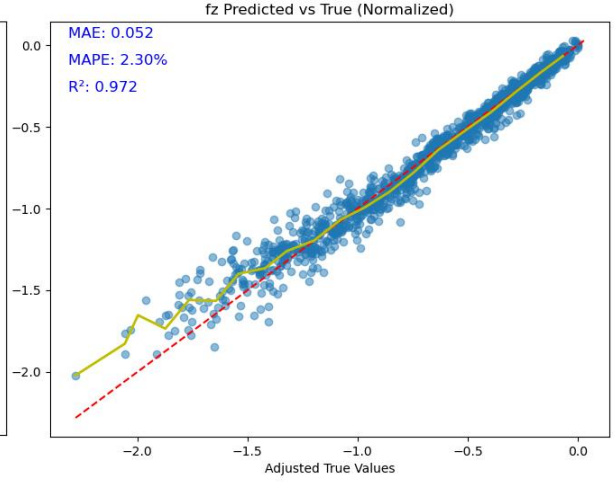




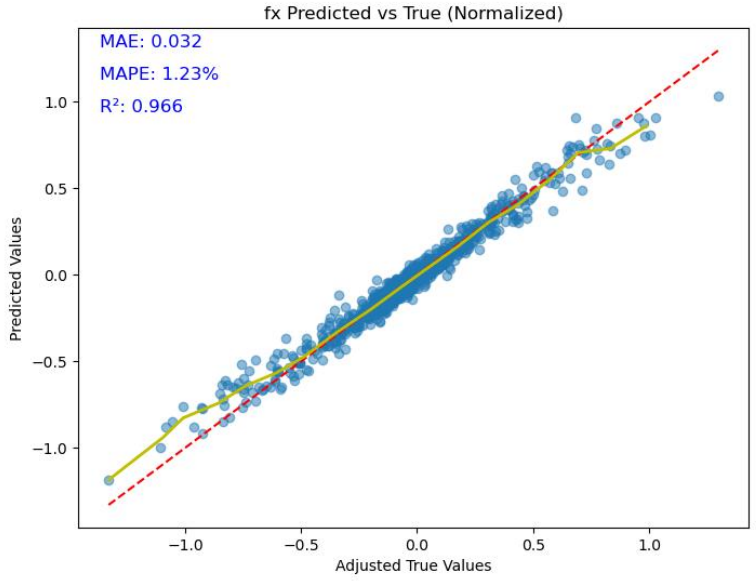


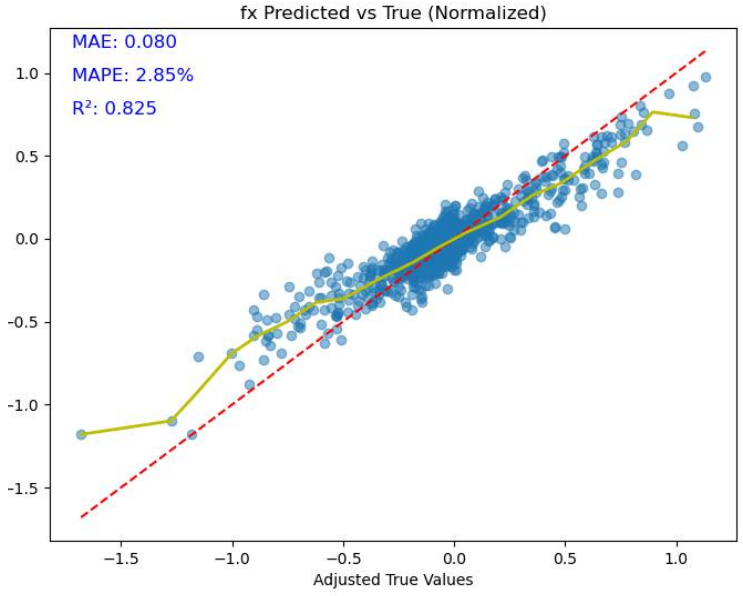


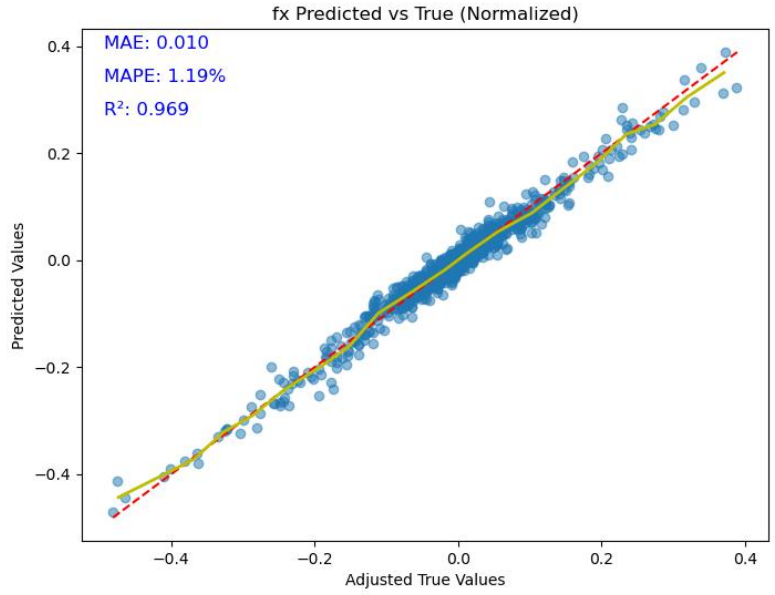


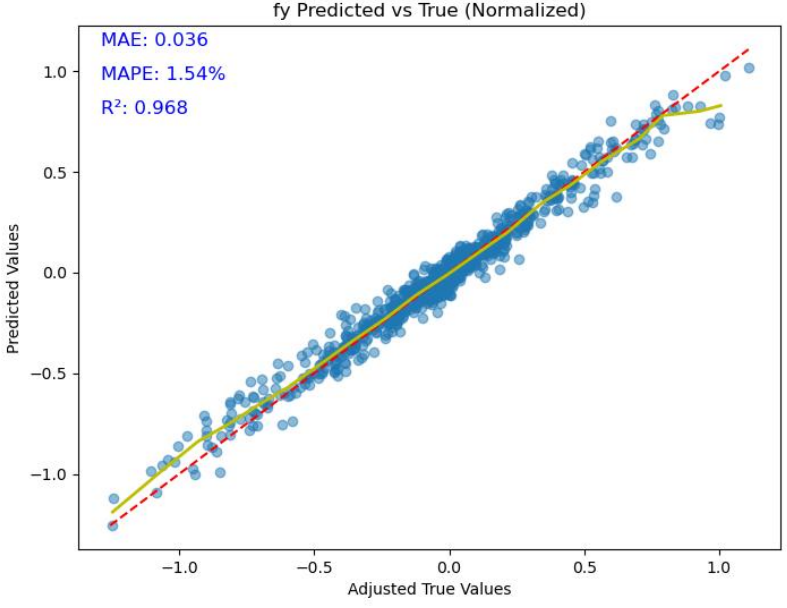


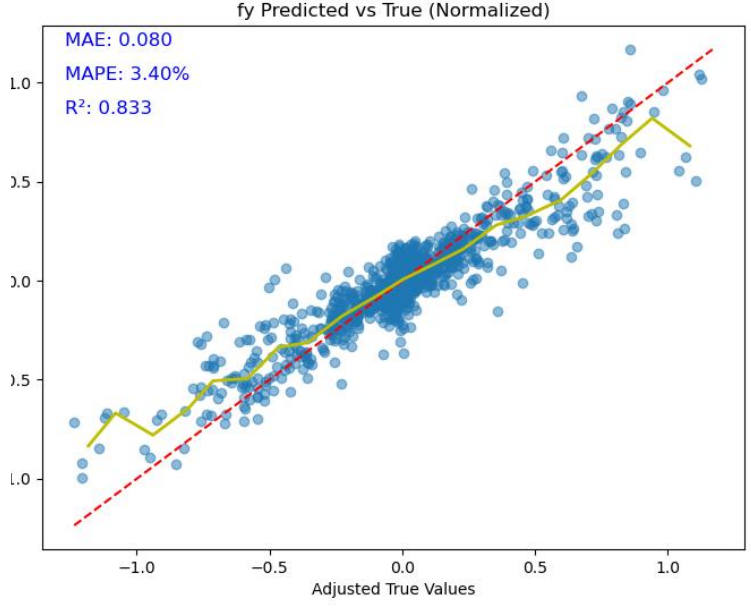
Gelmini Gelmini(Nomarker) Vitactip MagicTac

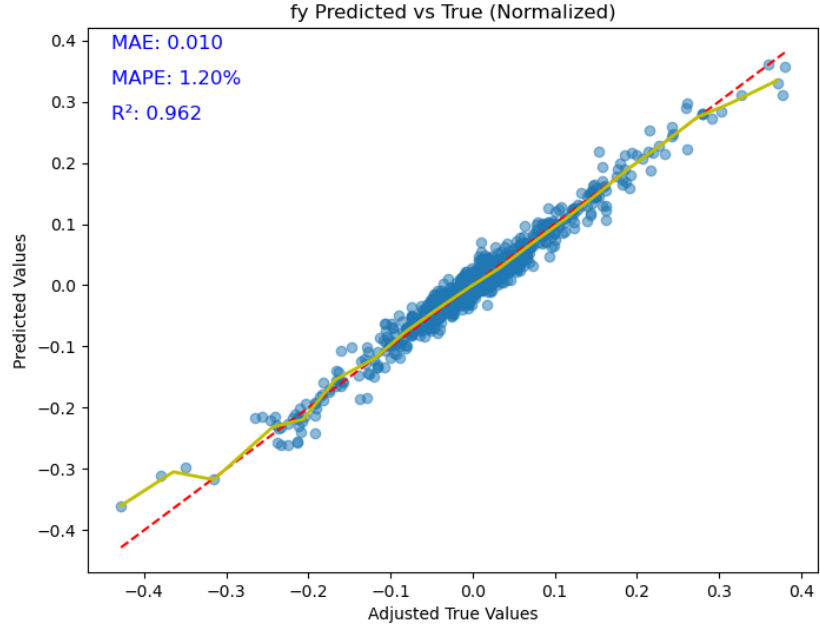








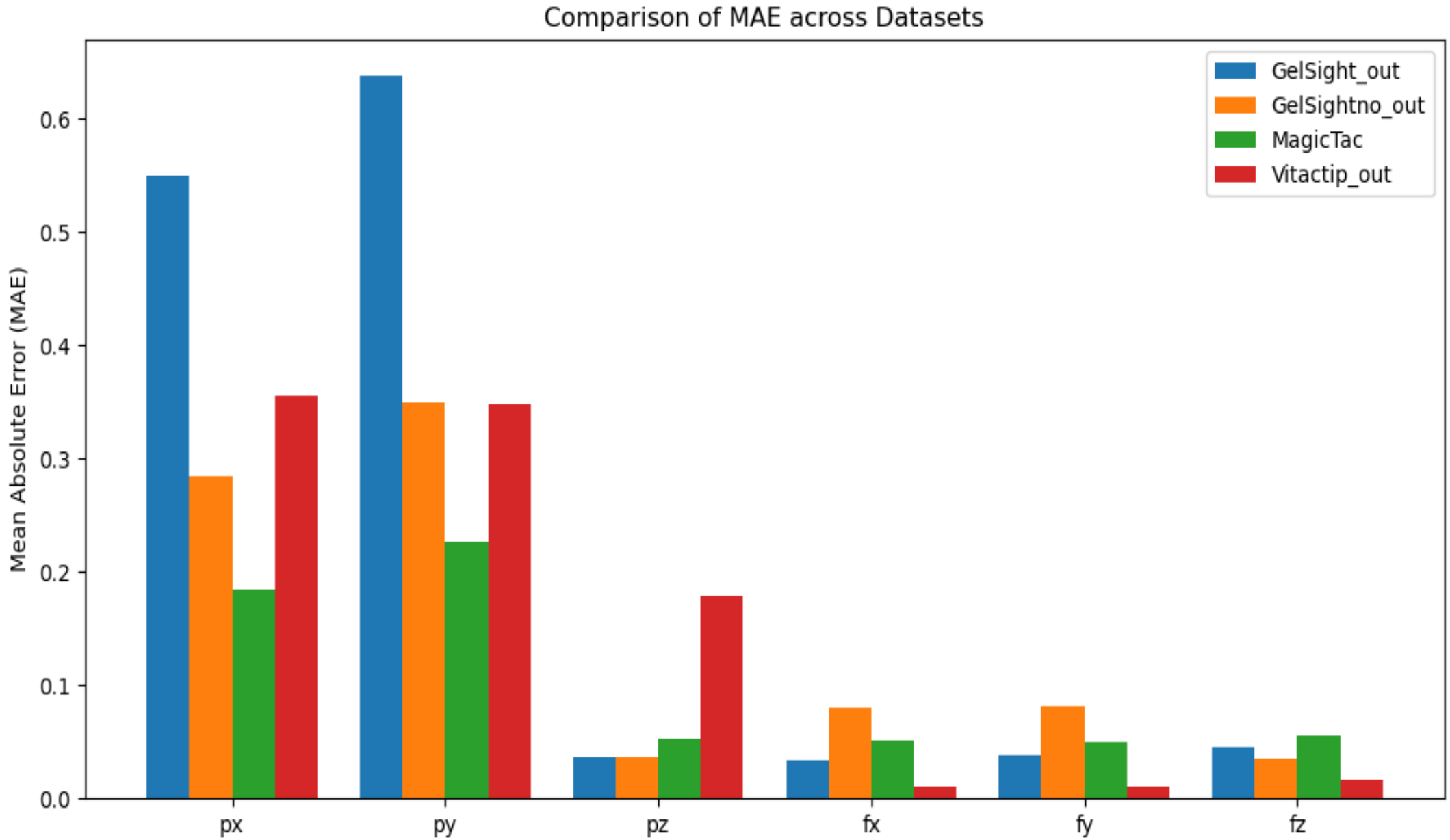




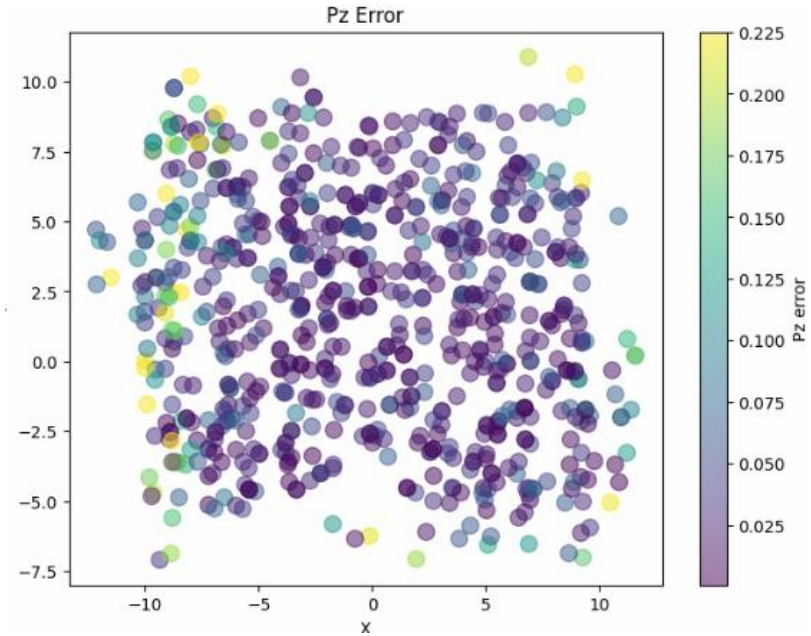
Gelmini

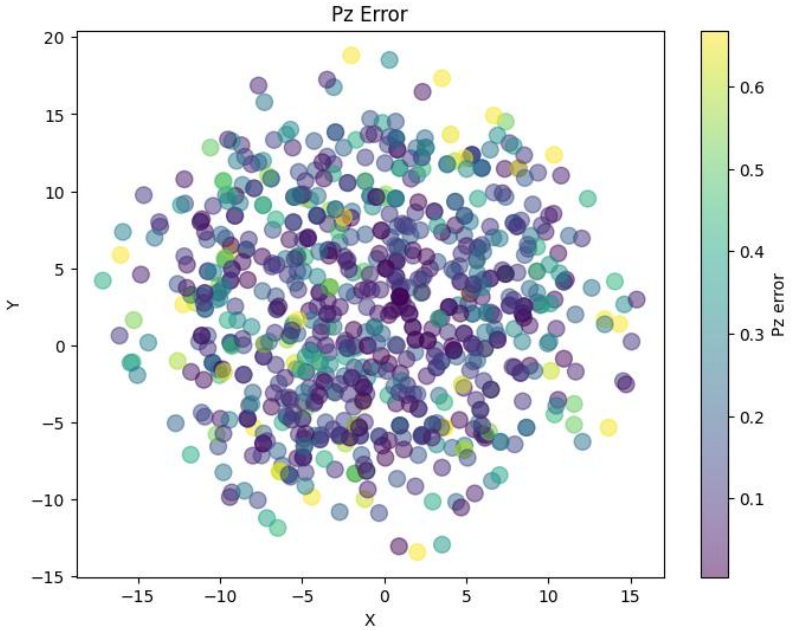
Gelmini (Nomarker)

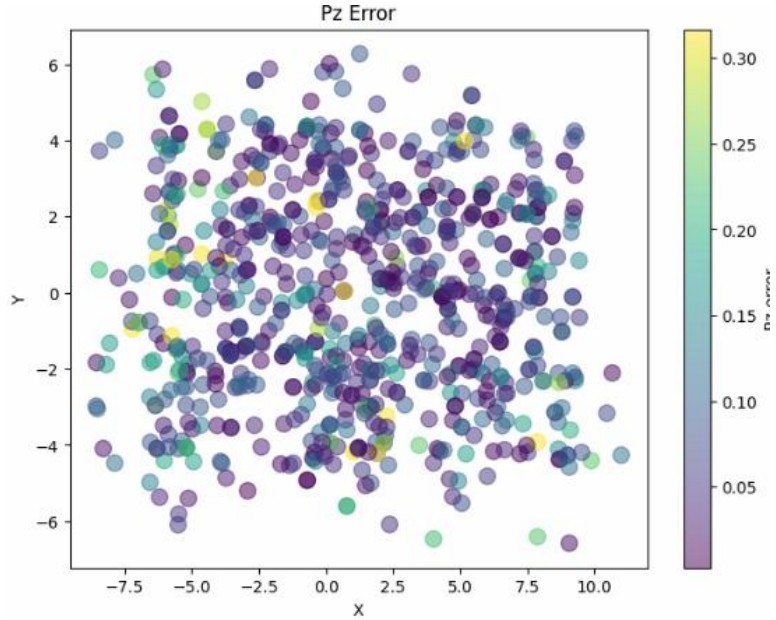
Vitactip

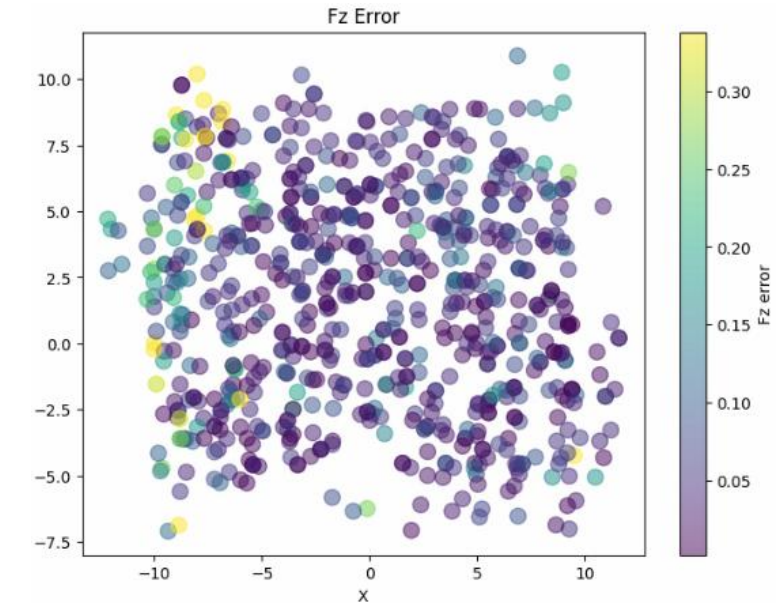

Comparison of MAE across Datasets
Mean Absolute Error (MAE)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
px
py
pz
fx
fy
fz
GelSight_out
GelSightno_out
MagicTac
Vitactip_out

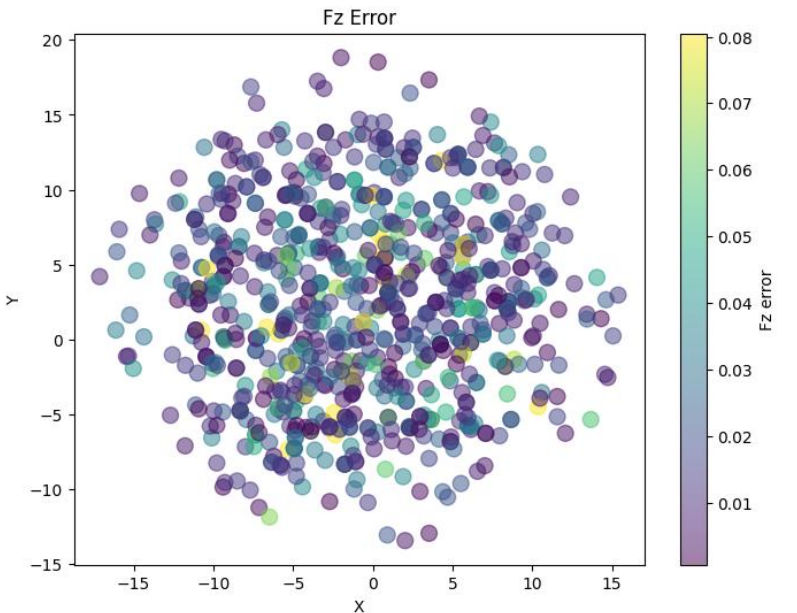
Error Map

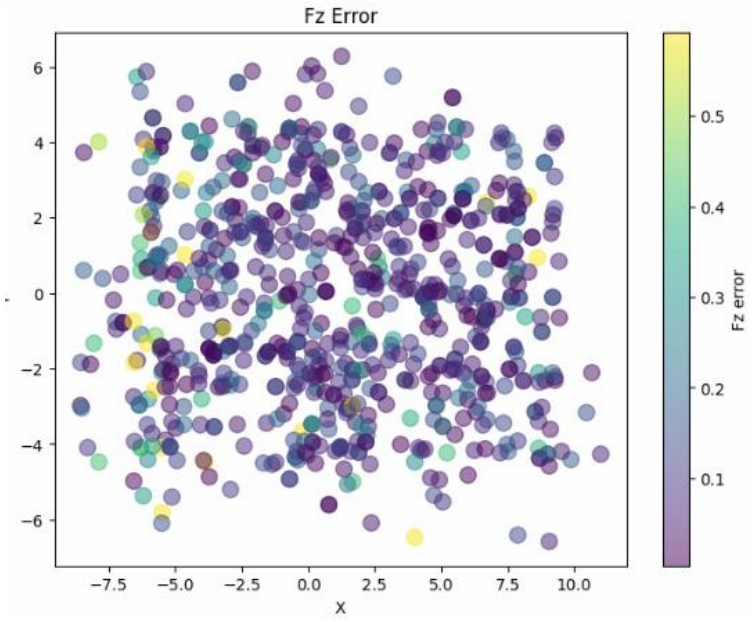

Pz Error
X
Pz error

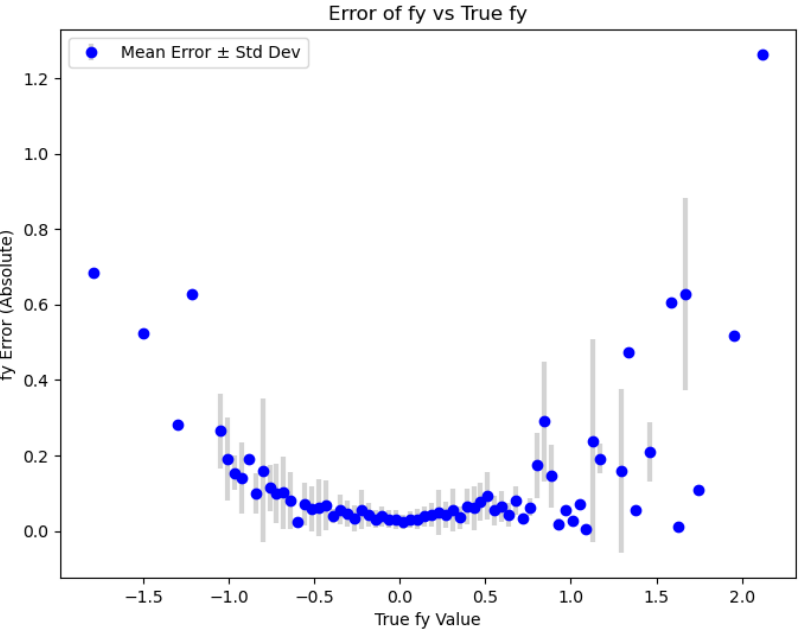

Pz Error
X
Y
Pz error

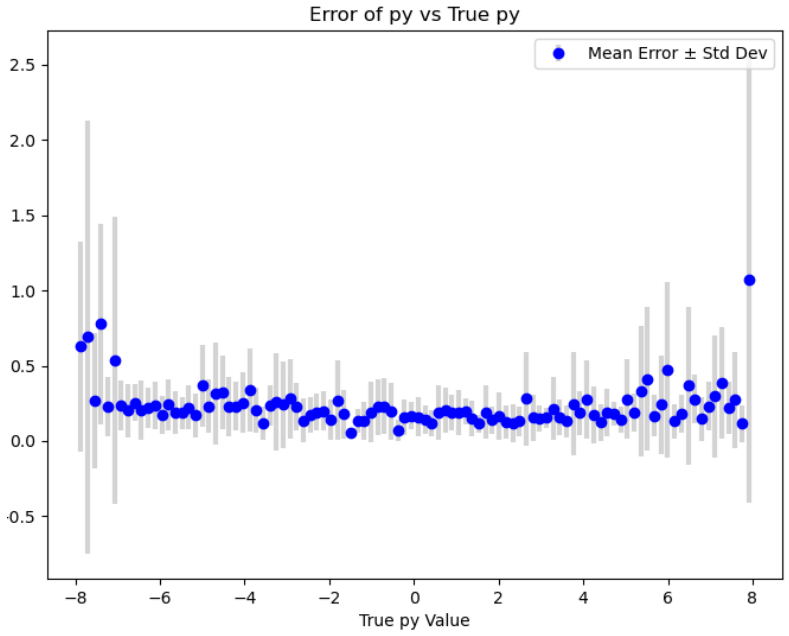

Pz Error
X
Y

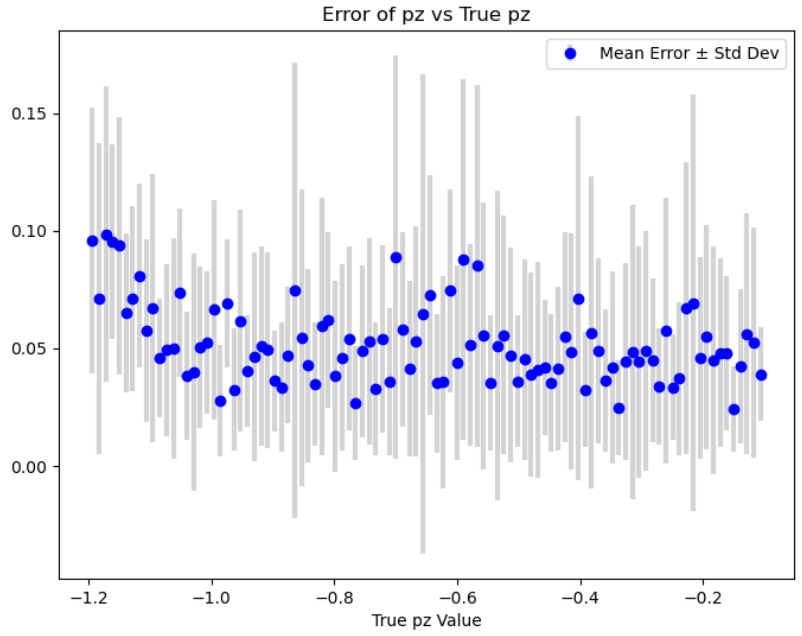

Fz Error
X
Fz error

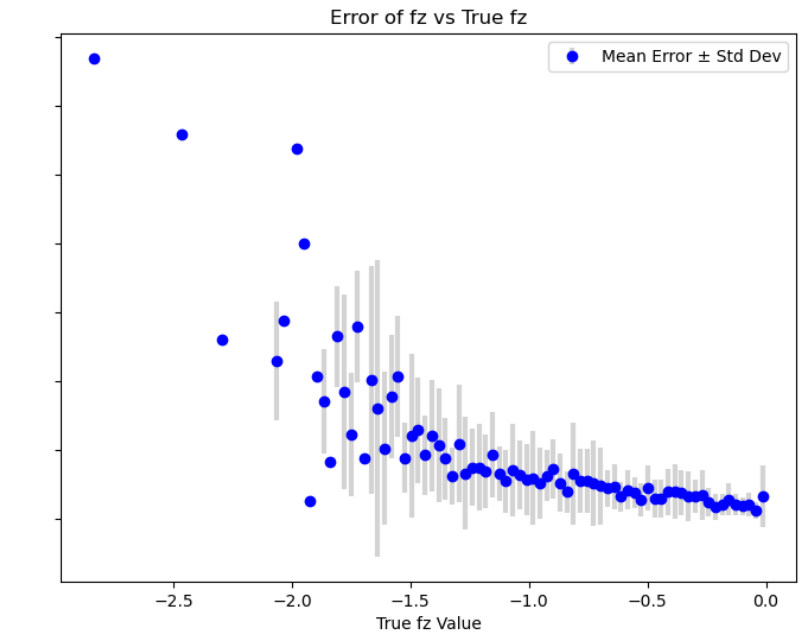

Fz Error
X
Y
Fz error

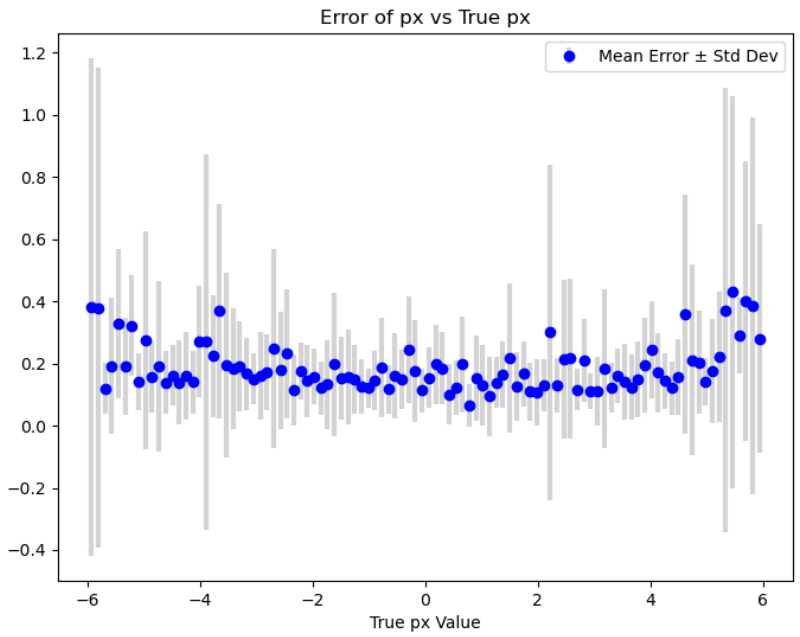

Fz Error
X
Fz error

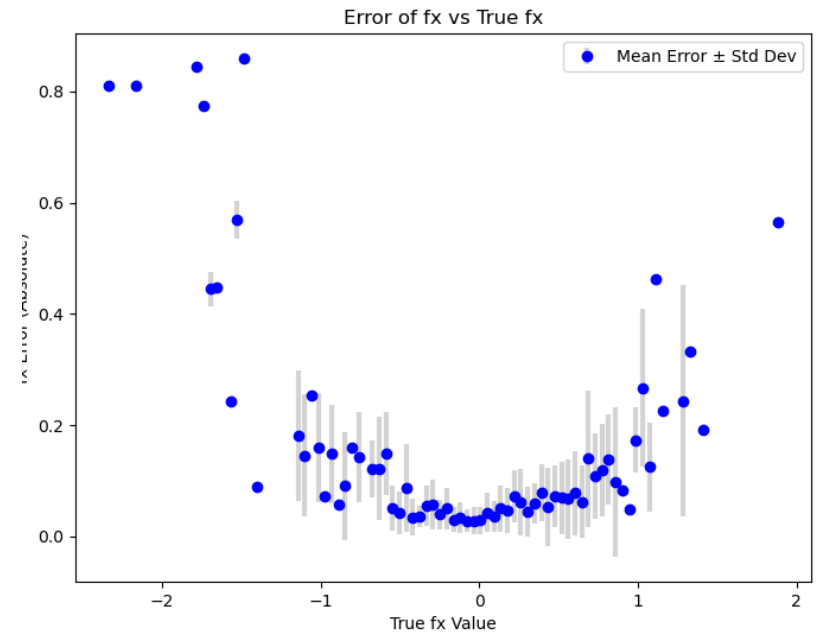

Error of fy vs True fy
Mean Error ± Std Dev
fy Error (Absolute)
True fy Value


Error of py vs True py
Mean Error ± Std Dev
True py Value


Error of pz vs True pz
Mean Error ± Std Dev
True pz Value

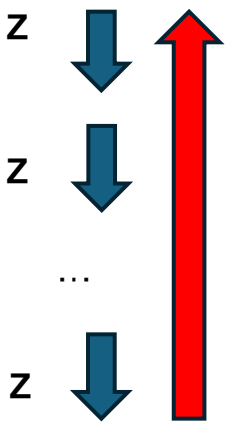

Error of fz vs True fz
Mean Error ± Std Dev
True fz Value

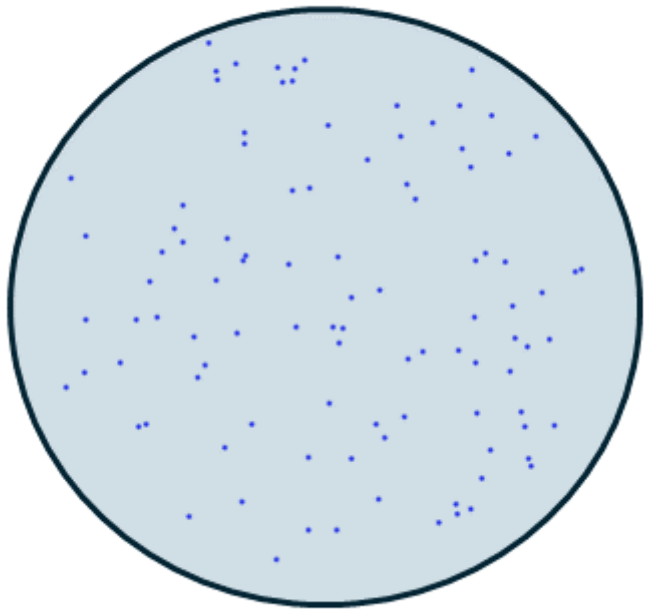

Error of px vs True px
Mean Error ± Std Dev
True px Value

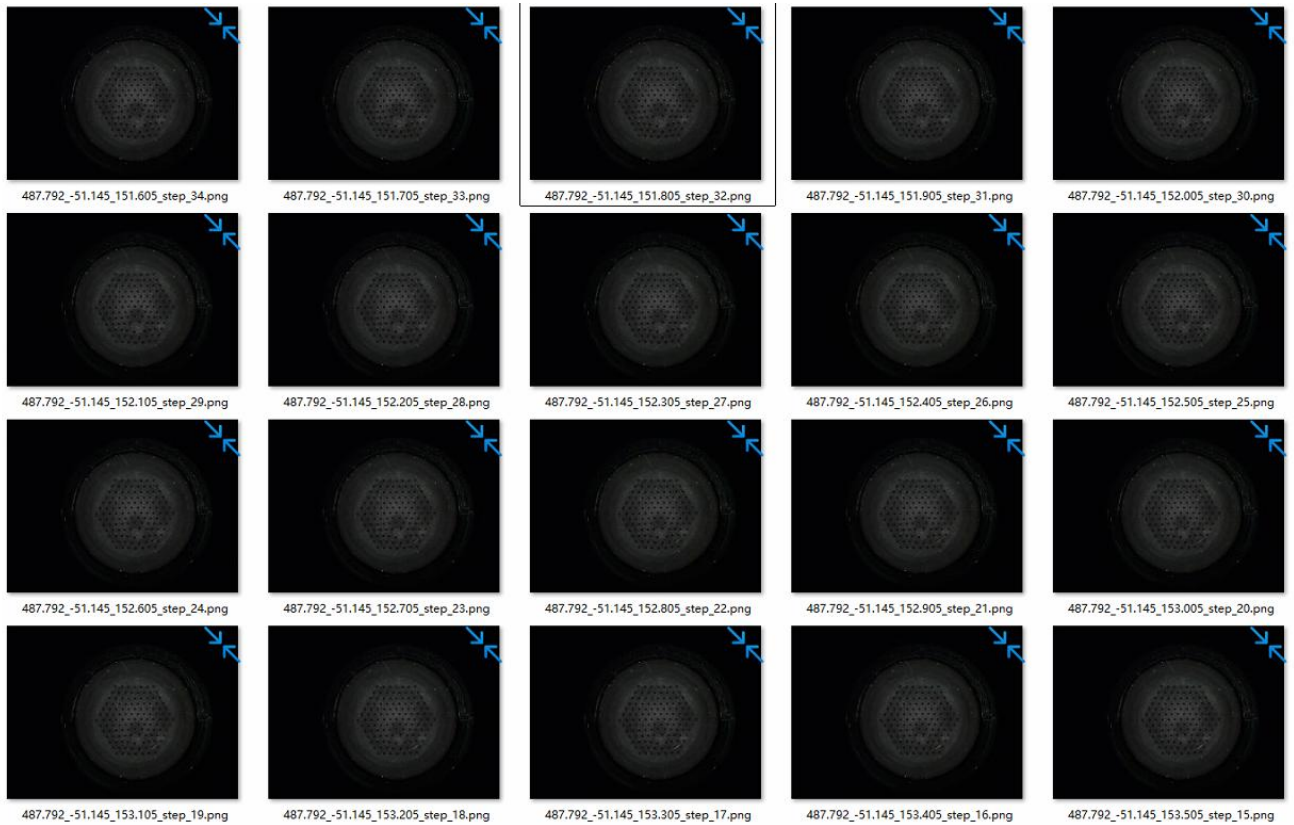

Error of fx vs True fx
Mean Error ± Std Dev
True fx Value

Hysteresis Experiment

1. The robot arm would first press the VBTS to the indenter for a limited depth, image is taken at every step
2. Then the arm would lift at its maximum speed
3. During the lifting process, all images would be saved

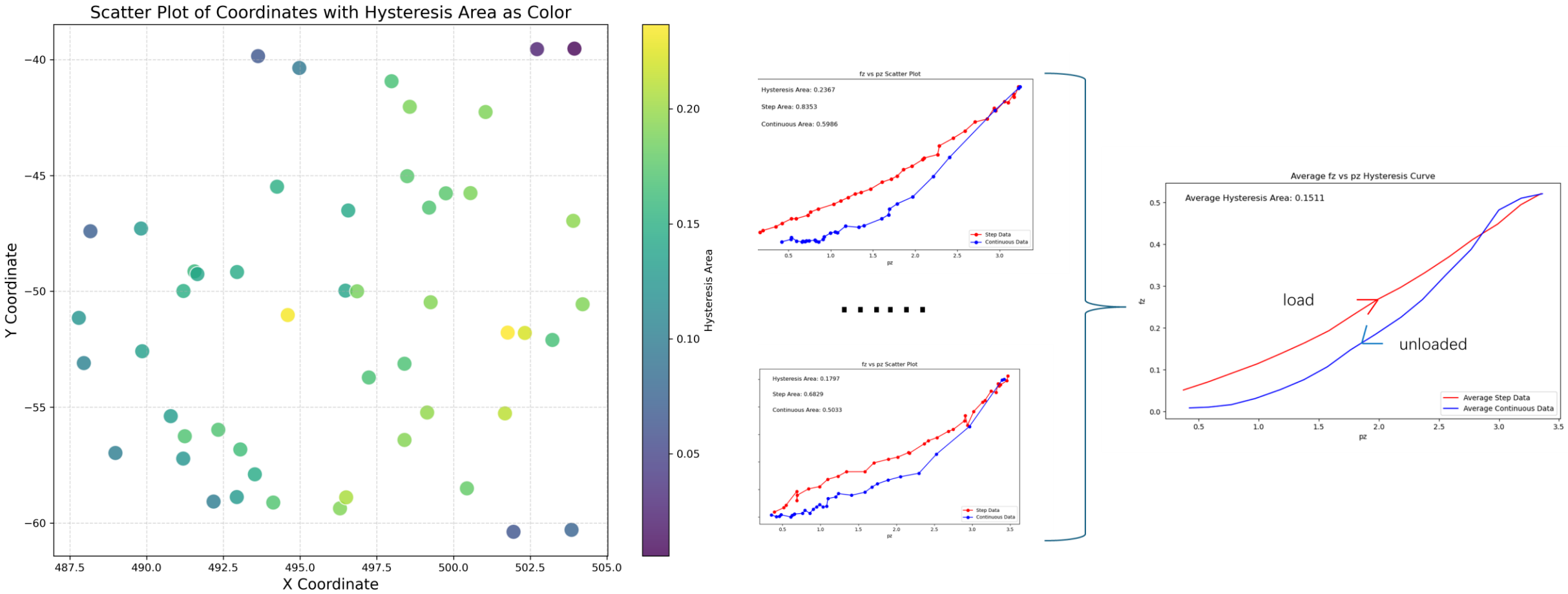




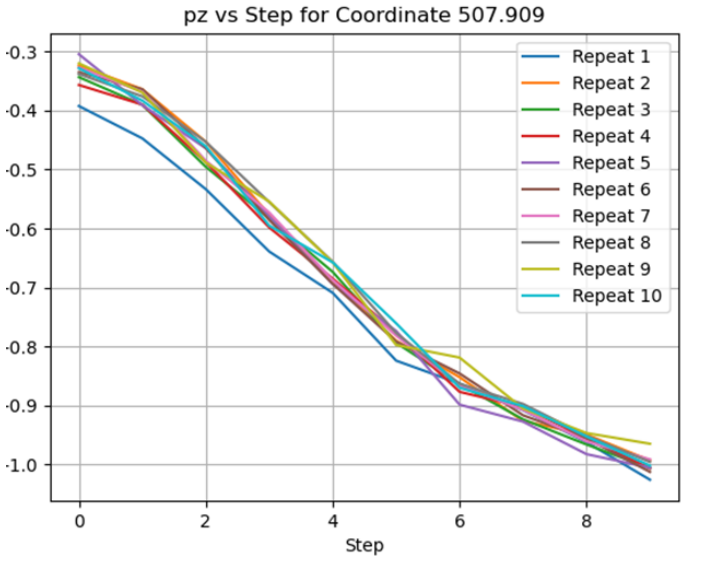


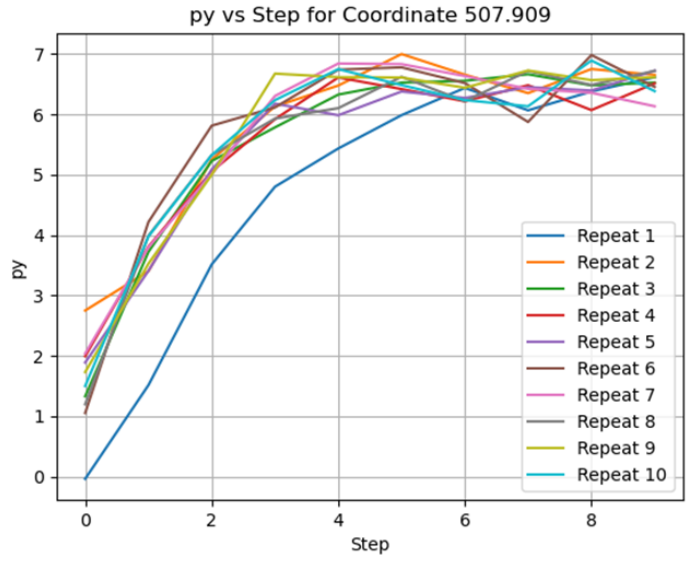
Hysteresis

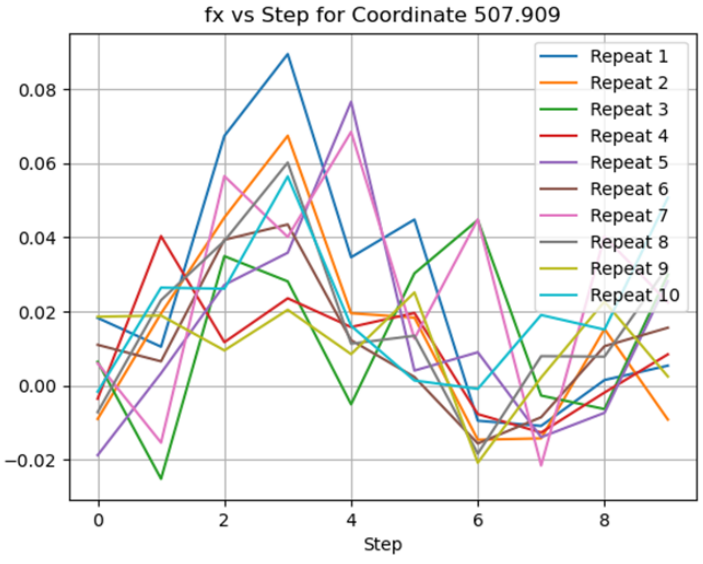
1. 100 points are taken to analysis the hysteresis
2. So, an overall Hysteresis map is concluded as well as the average hysteresis of the sensor

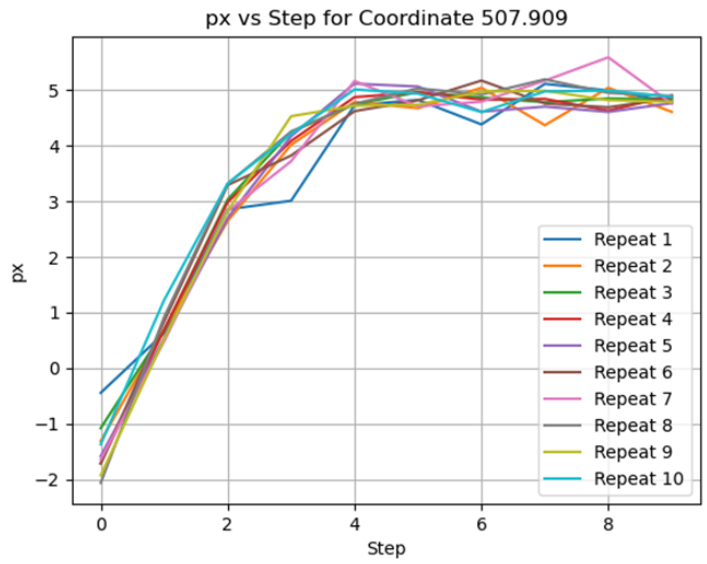


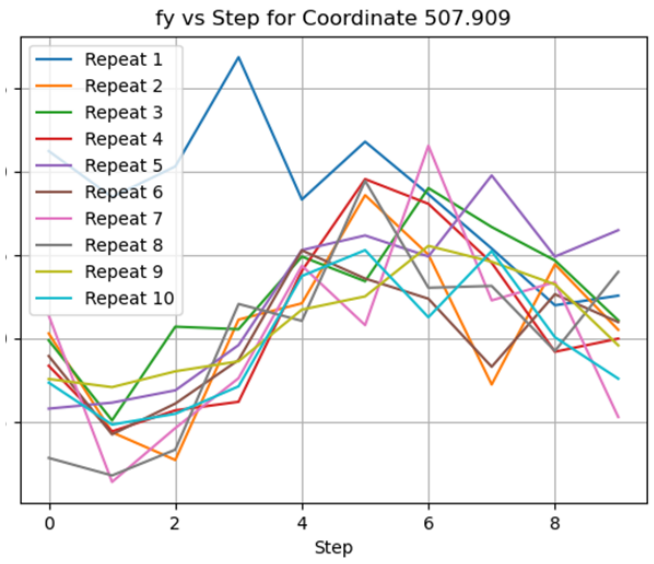
Repeatability

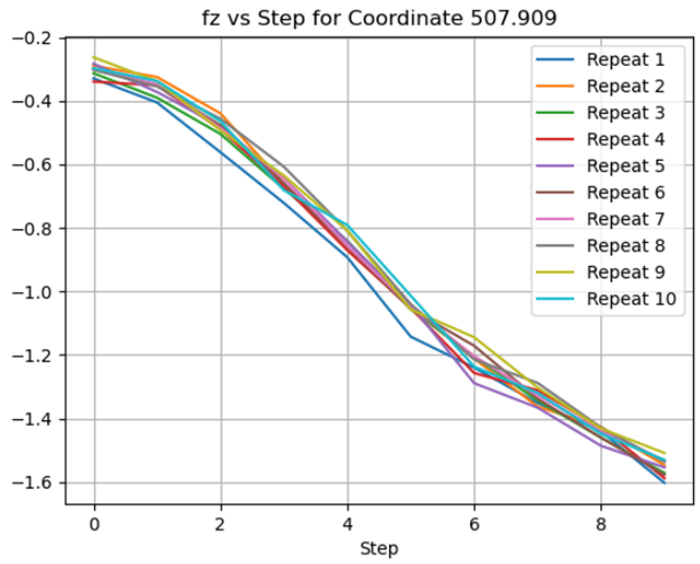


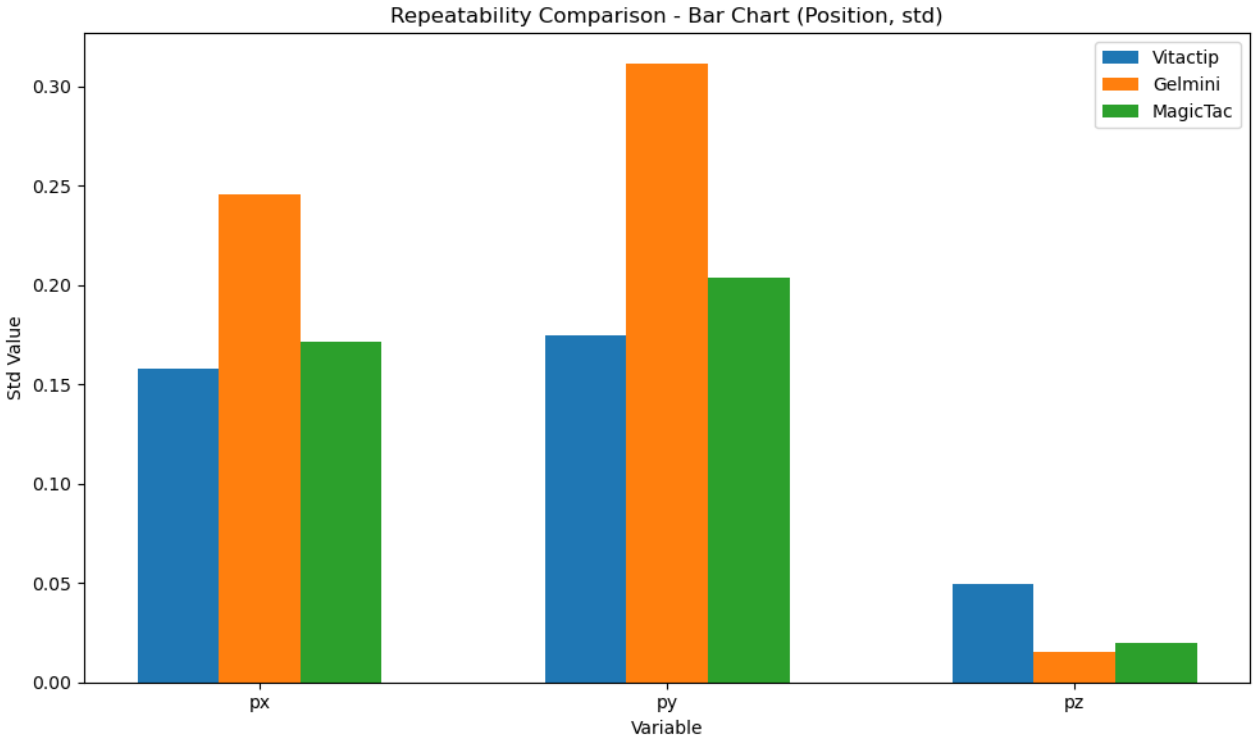


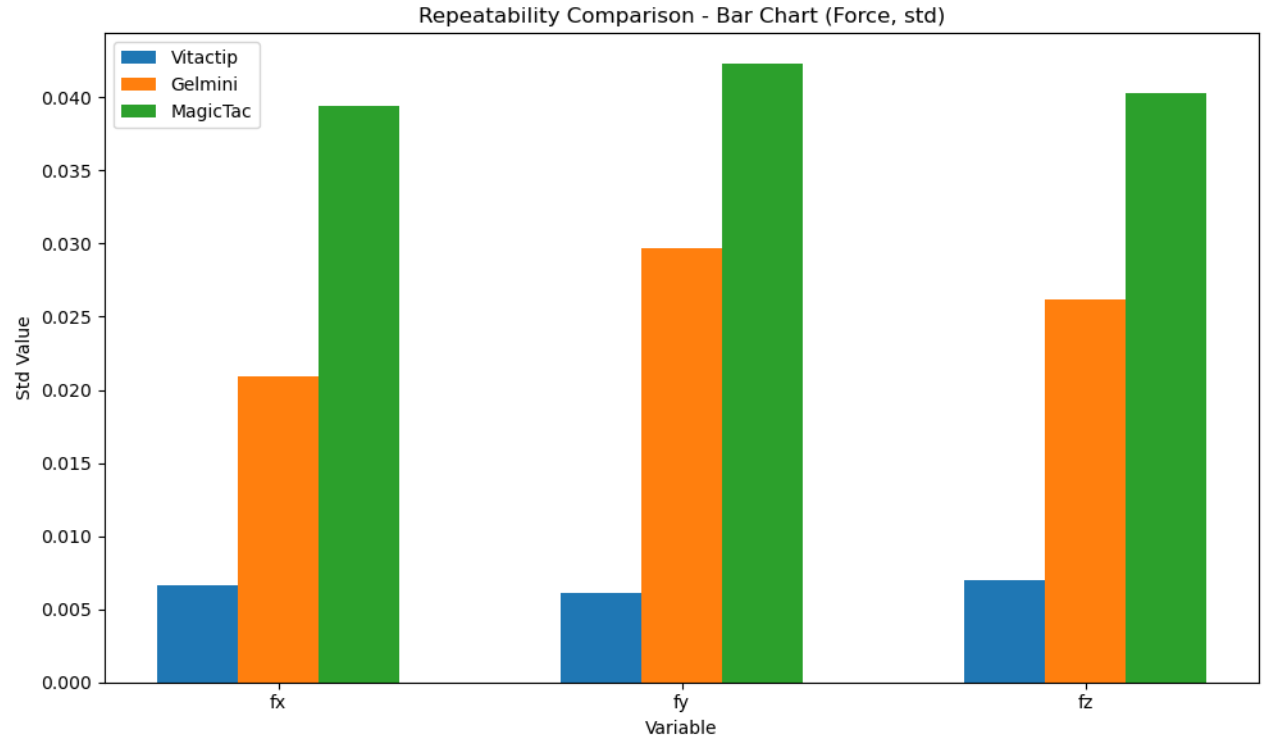




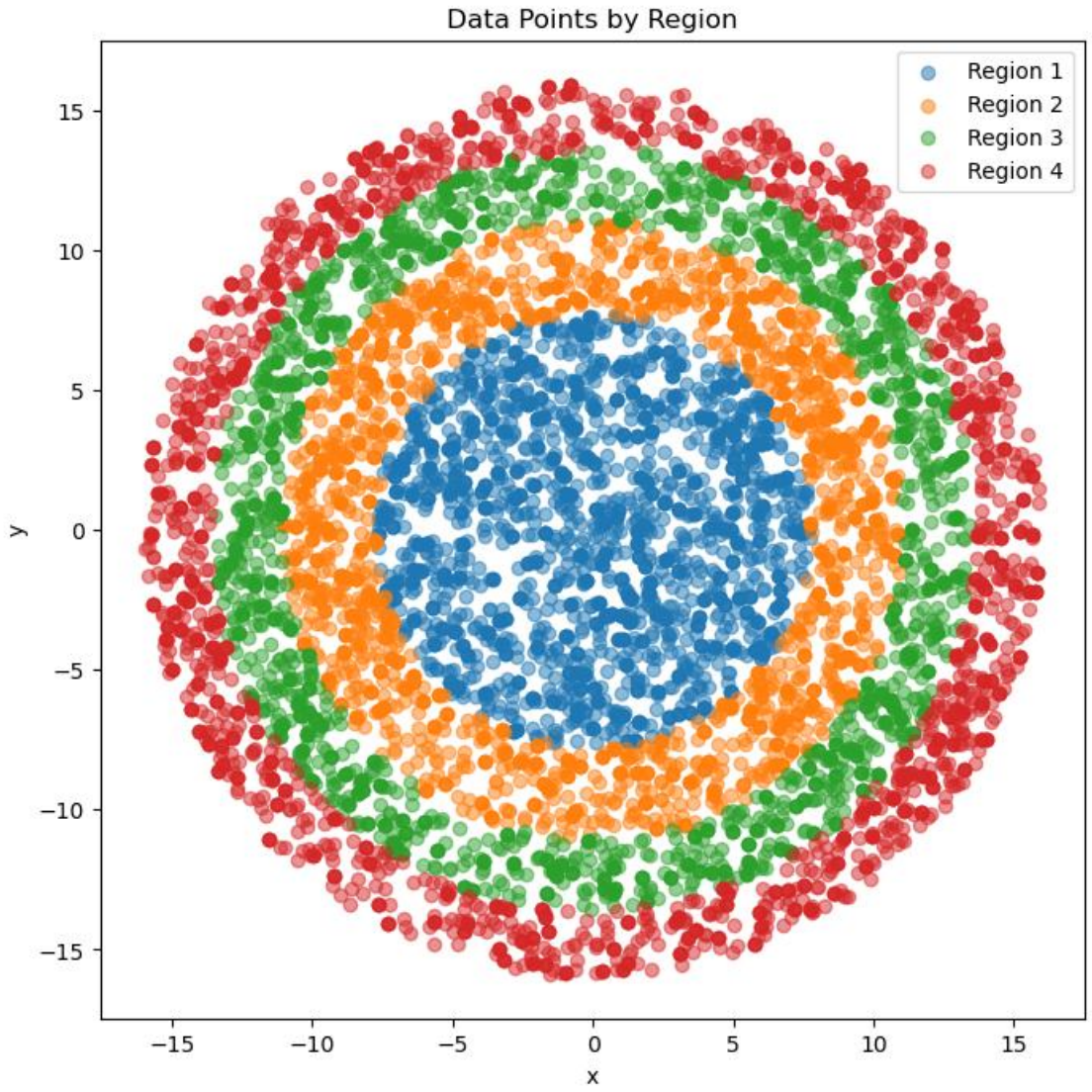


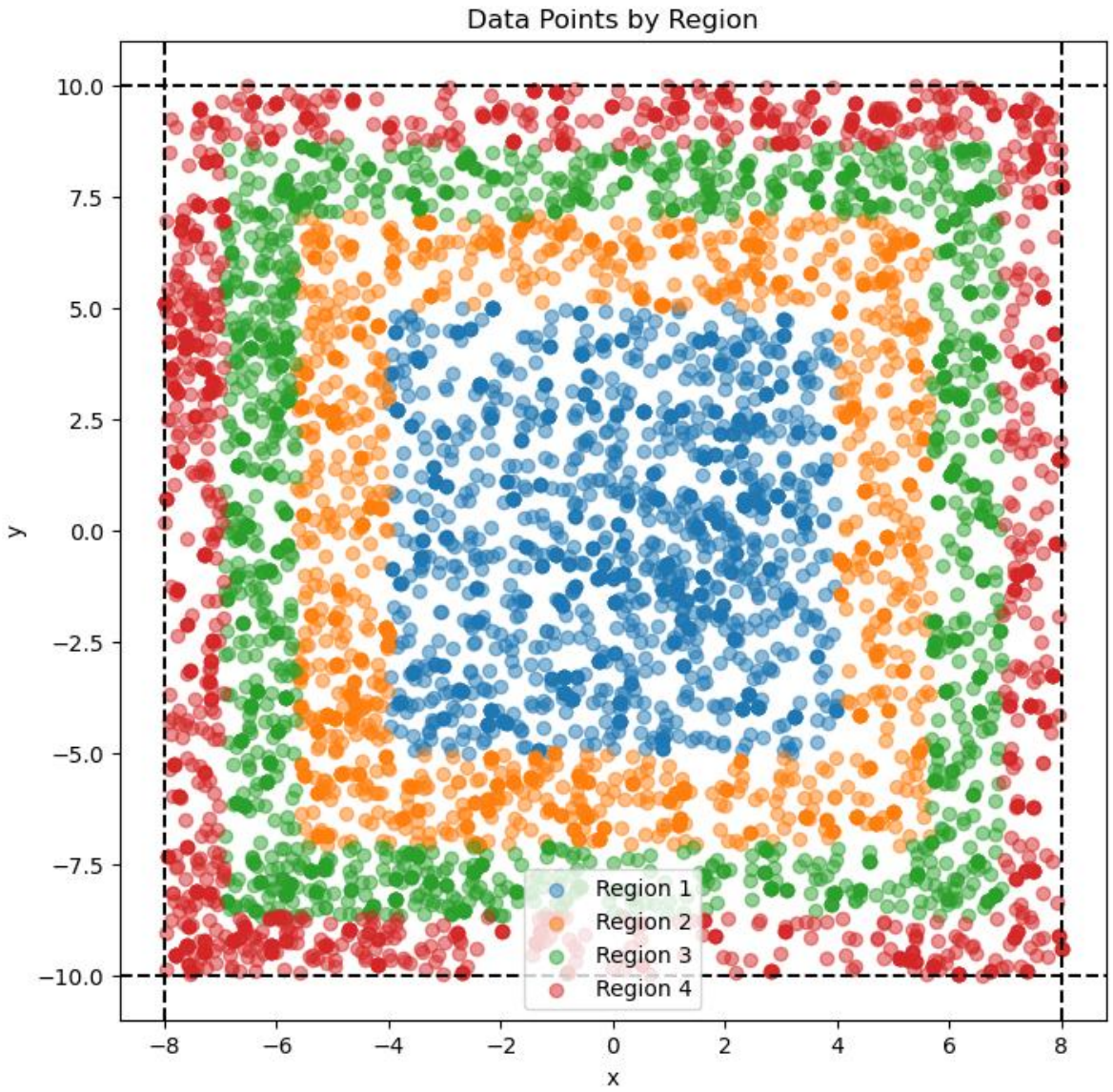


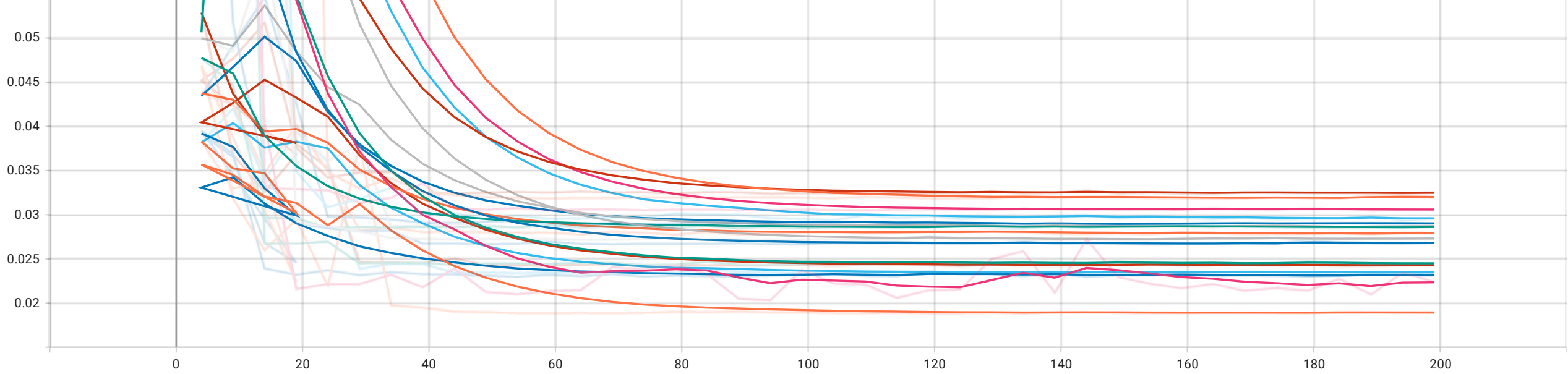

Repeatability Comparison - Bar Chart (Position, std)
Vitactip
Gelmini
MagicTac
Std Value
0.00
0.05
0.10
0.15
0.20
0.25
0.30
px
py
pz
Variable


Repeatability Comparison - Bar Chart (Force, std)
Vitactip
Gelmini
MagicTac
Std Value
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
fx
fy
fz
Variable

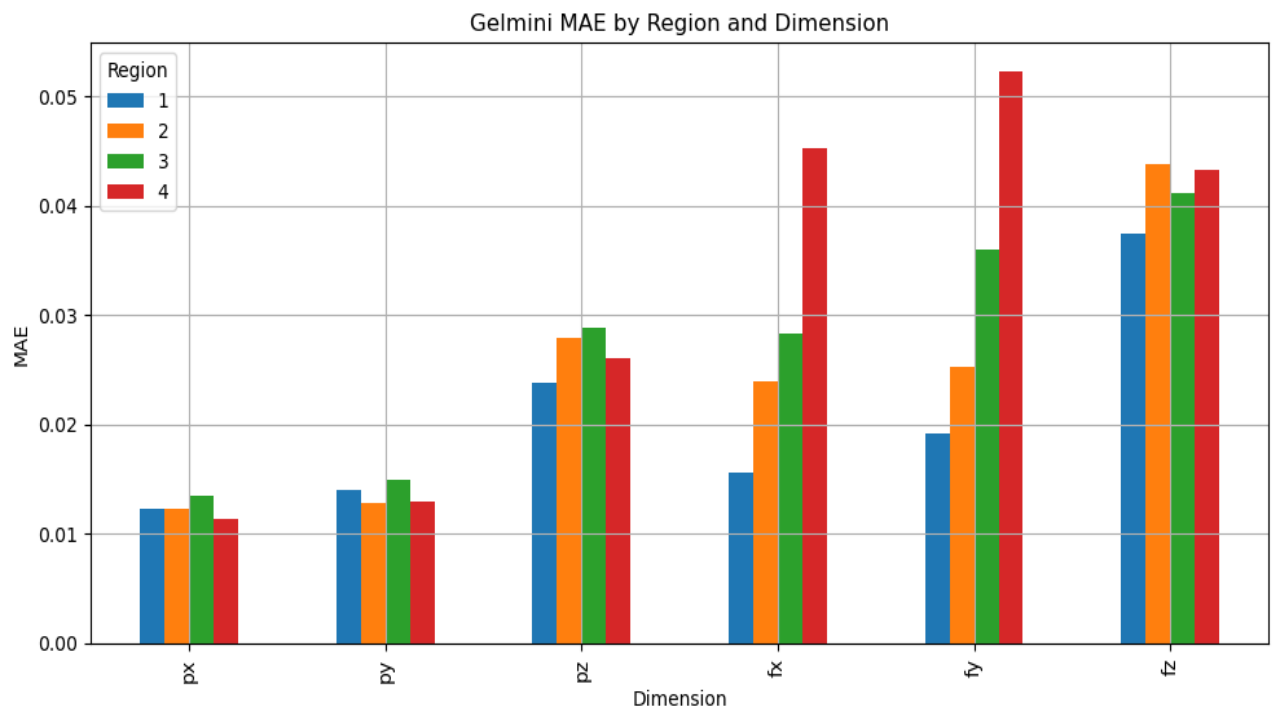
Spatial Robustness

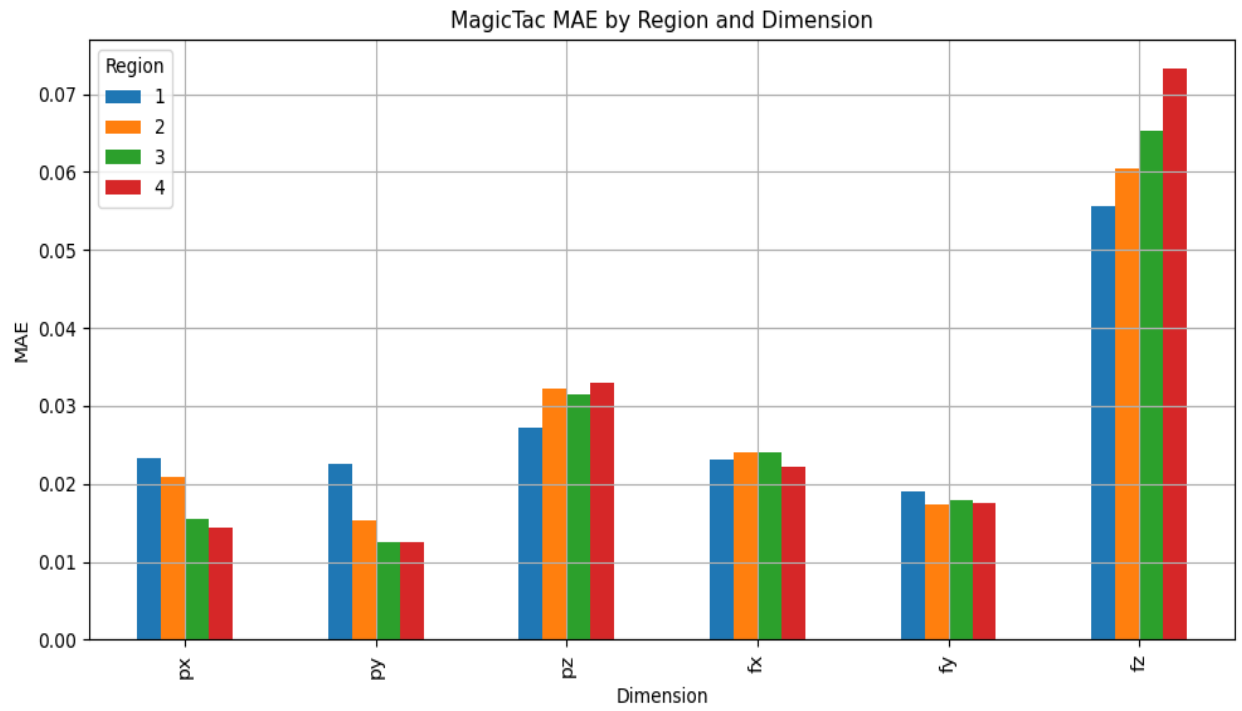


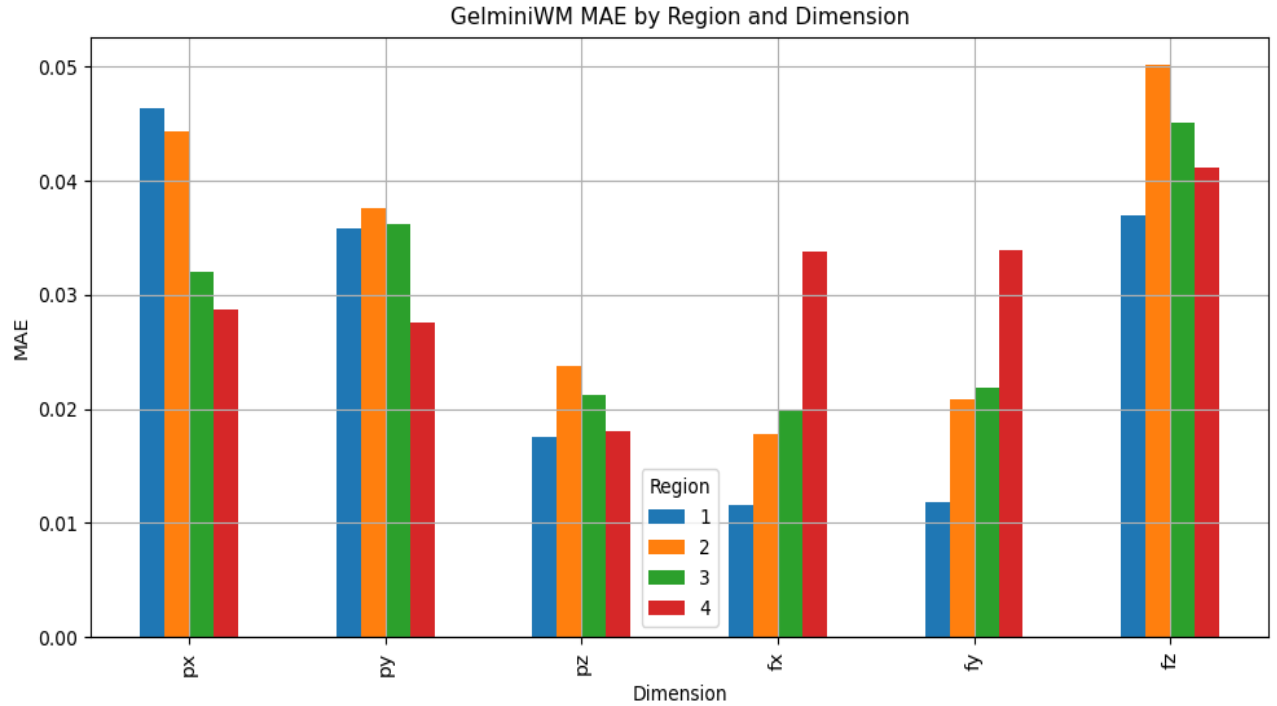


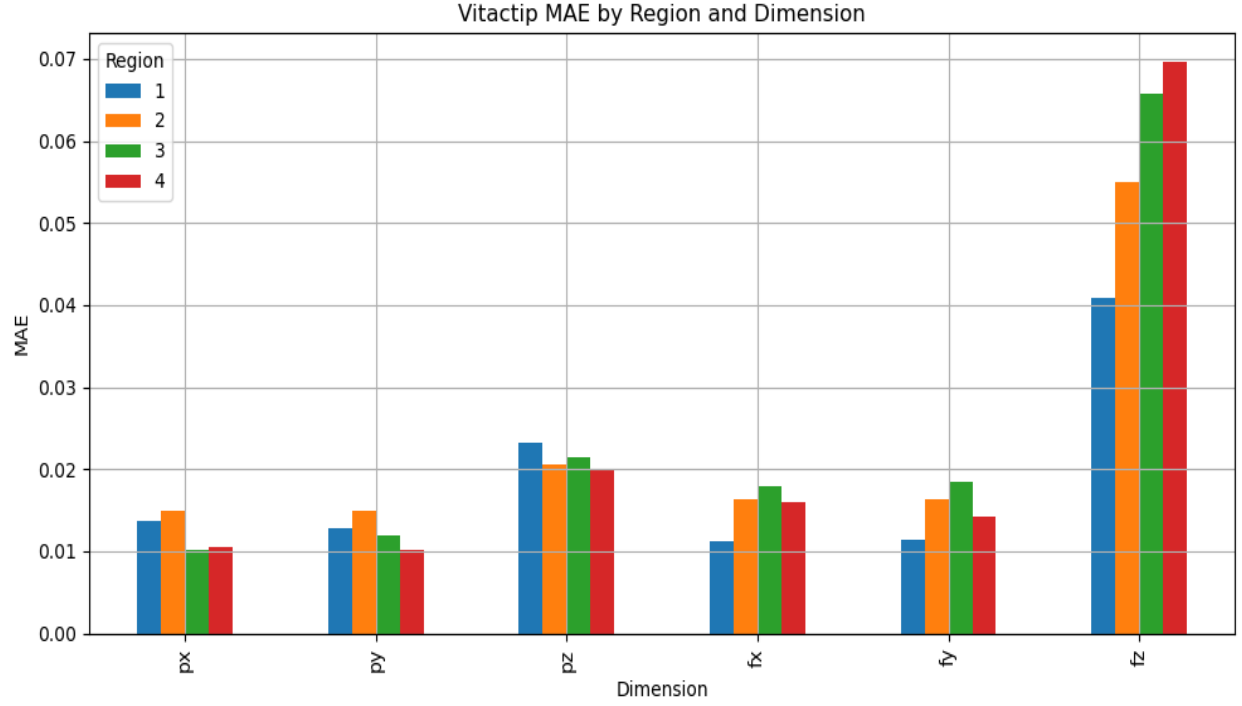


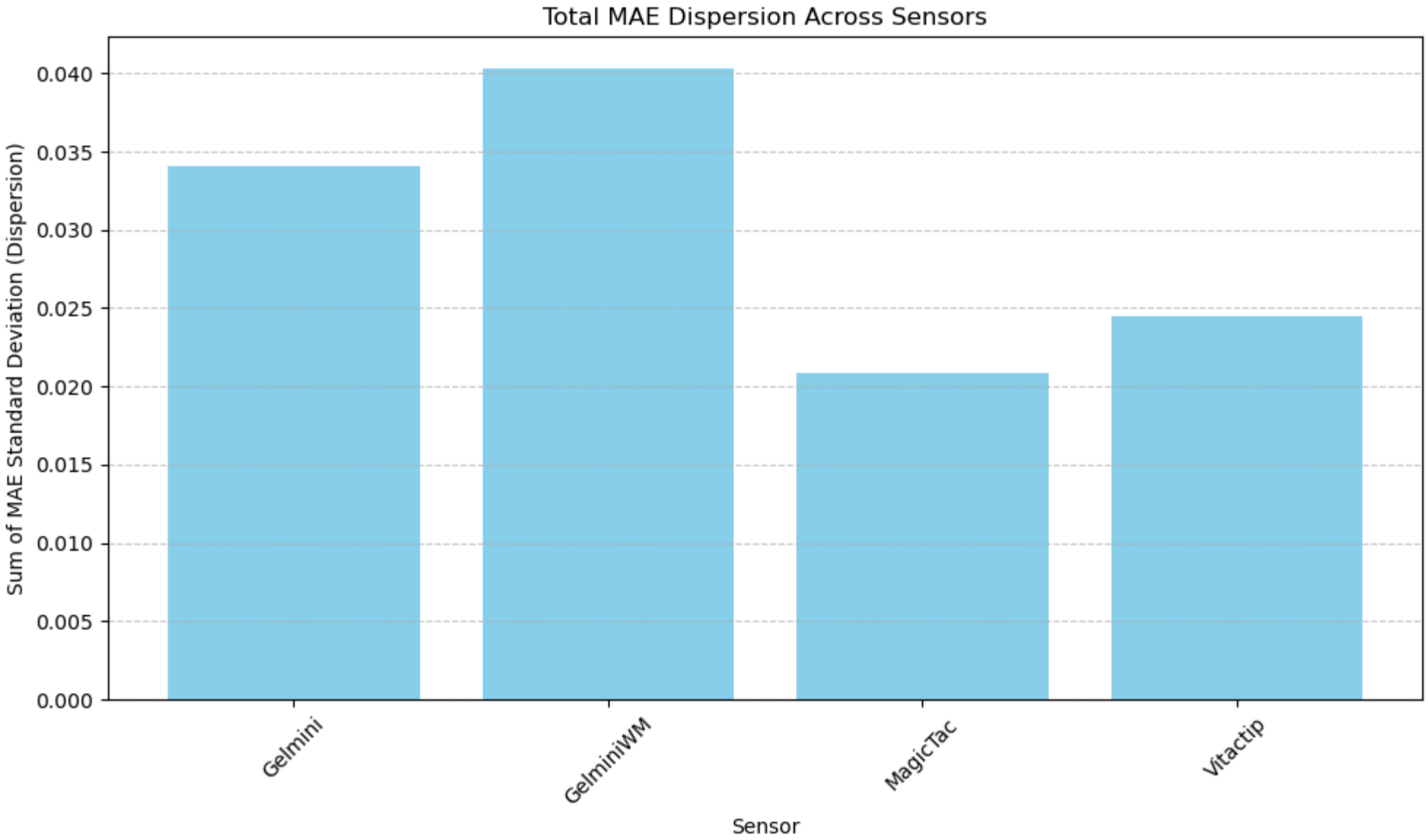
Run ↑	Min	Max	Start Value	End Value	△Value	△%	Start Step	End Step
Gelmini_result\train_region_1\logs	0.0215	0.0994	0.0847	0.0224	↓0.0622	↓-74%	4	199
Gelmini_result\train_region_2\logs	0.0244	0.1099	0.0506	0.0245	↓0.0262	↓-52%	4	199
Gelmini_result\train_region_3\logs	0.0272	0.1084	0.075	0.0273	↓0.0477	↓-64%	4	199
Gelmini_result\train_region_4\logs	0.0319	0.1881	0.1881	0.032	↓0.1561	↓-83%	4	199
Gelminiwm_result\train_region_1\logs	0.0267	0.1146	0.1146	0.0268	↓0.0878	↓-77%	4	199
Gelminiwm_result\train_region_2\logs	0.0324	0.1434	0.1434	0.0325	↓0.1109	↓-77%	4	199
Gelminiwm_result\train_region_3\logs	0.0296	0.132	0.098	0.0296	↓0.0684	↓-70%	4	199
Gelminiwm_result\train_region_4\logs	0.0306	0.2354	0.2155	0.0306	↓0.185	↓-86%	4	199
MagicTac\train_region_1\logs	0.0286	0.0478	0.0478	0.0286	↓0.0191	↓-40%	4	199
MagicTac\train_region_2\logs	0.0286	0.0545	0.05	0.0287	↓0.0212	↓-43%	4	199
MagicTac\train_region_3\logs	0.0279	0.0437	0.0437	0.0279	↓0.0158	↓-36%	4	199
MagicTac\train_region_4\logs	0.0289	0.0509	0.0434	0.029	↓0.0144	↓-33%	4	199
Vitactip_result\train_region_1\logs	0.0189	0.0383	0.0383	0.0189	↓0.0194	↓-51%	4	199
Vitactip_result\train_region_2\logs	0.0231	0.0392	0.0392	0.0232	↓0.016	↓-41%	4	199
Vitactip_result\train_region_3\logs	0.0243	0.0529	0.0529	0.0243	↓0.0286	↓-54%	4	199
Vitactip_result\train_region_4\logs	0.0235	0.0406	0.0382	0.0235	↓0.0147	↓-39%	4	199


Gelmini MAE by Region and Dimension
Region
1
2
3
4
MAE
0.00
0.01
0.02
0.03
0.04
0.05
px
py
pz
fx
fy
fz
Dimension

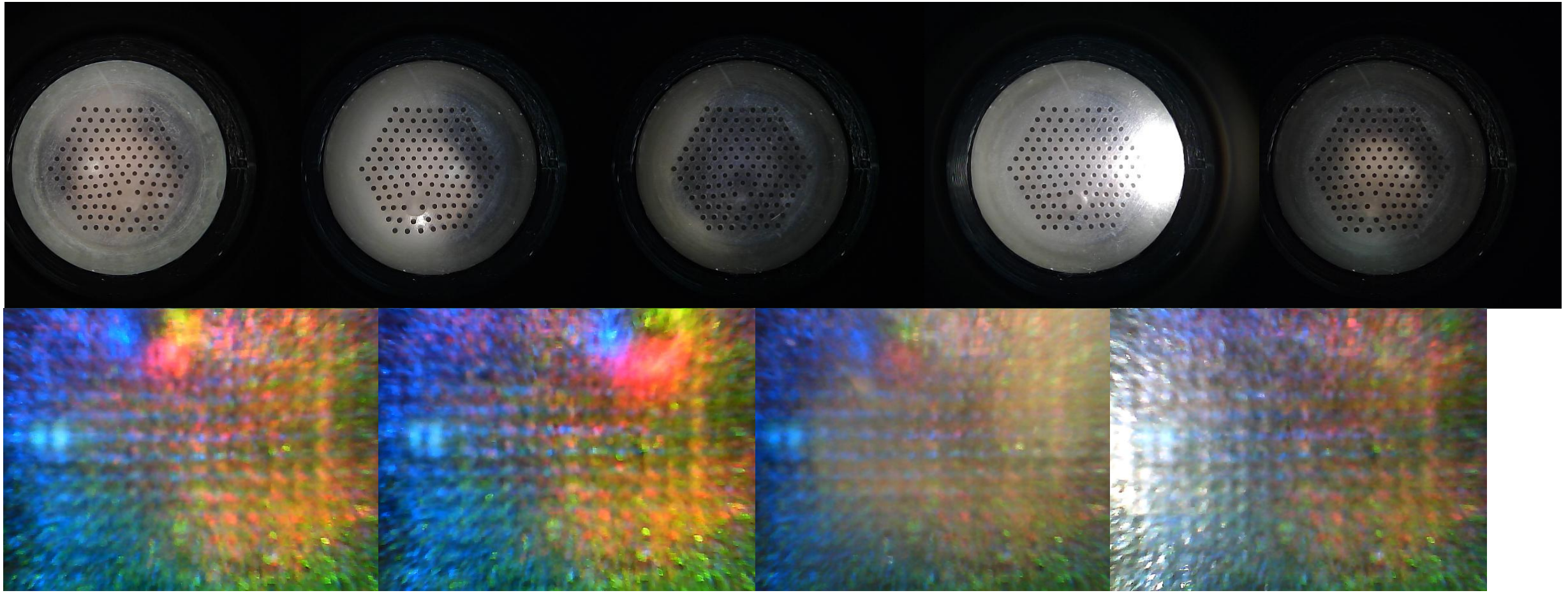

MagicTac MAE by Region and Dimension
Region
1
2
3
4
MAE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
px
py
pz
fx
fy
fz
Dimension

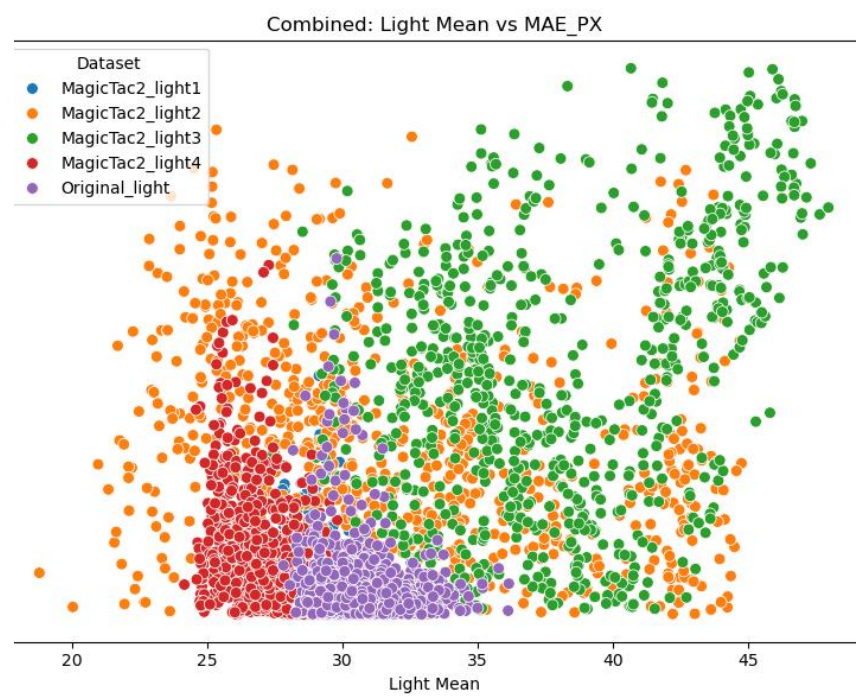

GelminiWM MAE by Region and Dimension
Region
1
2
3
4
MAE
0.00
0.01
0.02
0.03
0.04
0.05
px
py
pz
fx
fy
fz
Dimension

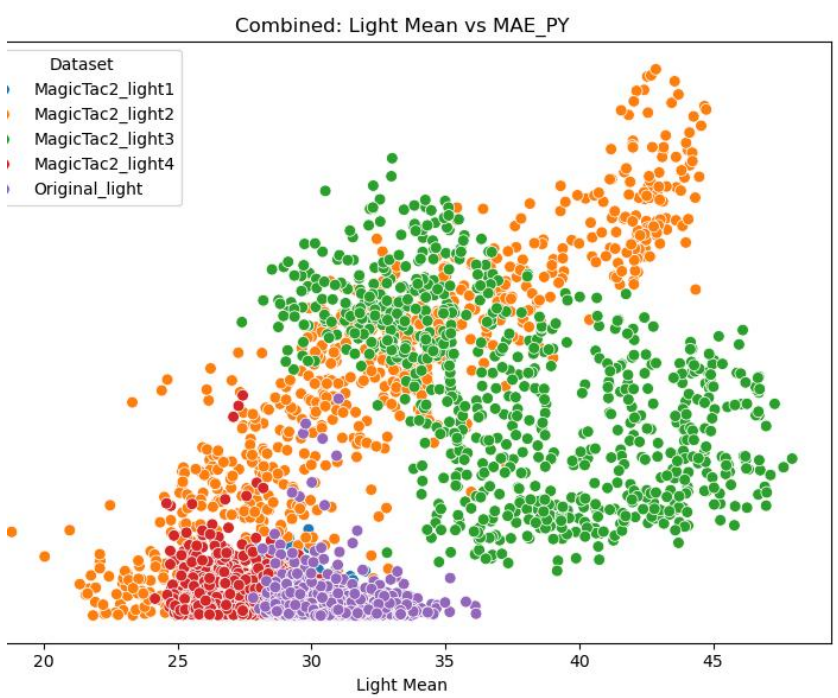

Vitactip MAE by Region and Dimension
Region
1
2
3
4
MAE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
px
py
pz
fx
fy
fz
Dimension

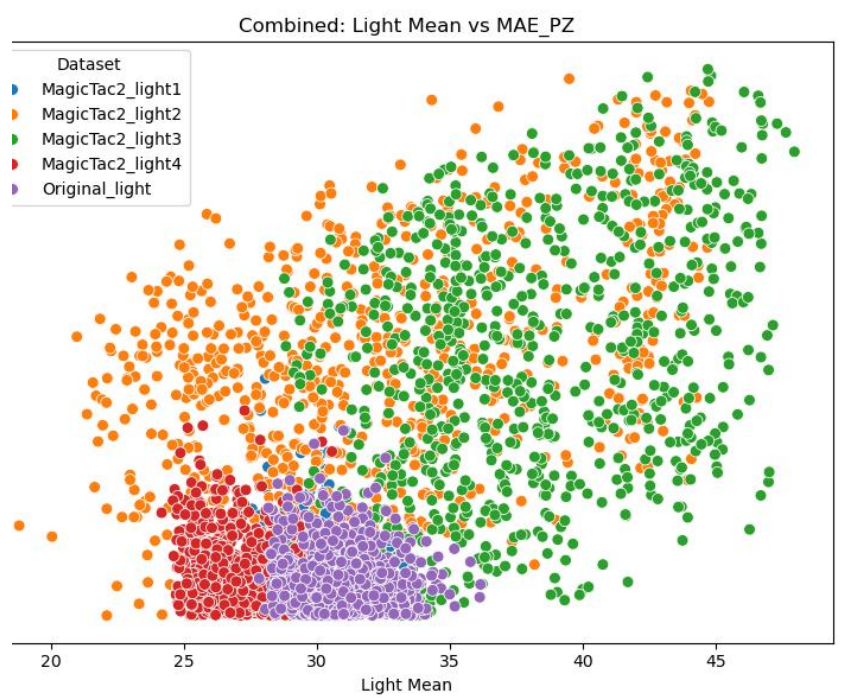

Total MAE Dispersion Across Sensors
Sum of MAE Standard Deviation (Dispersion)
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Gelmini
GelminiWM
MagicTac
Vitactip
Sensor

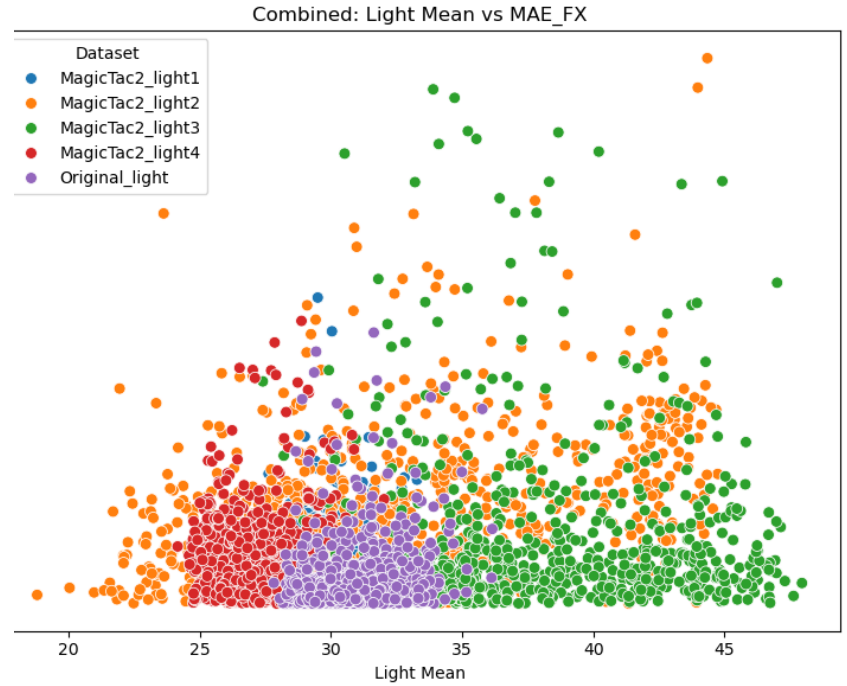
Robustness for external light

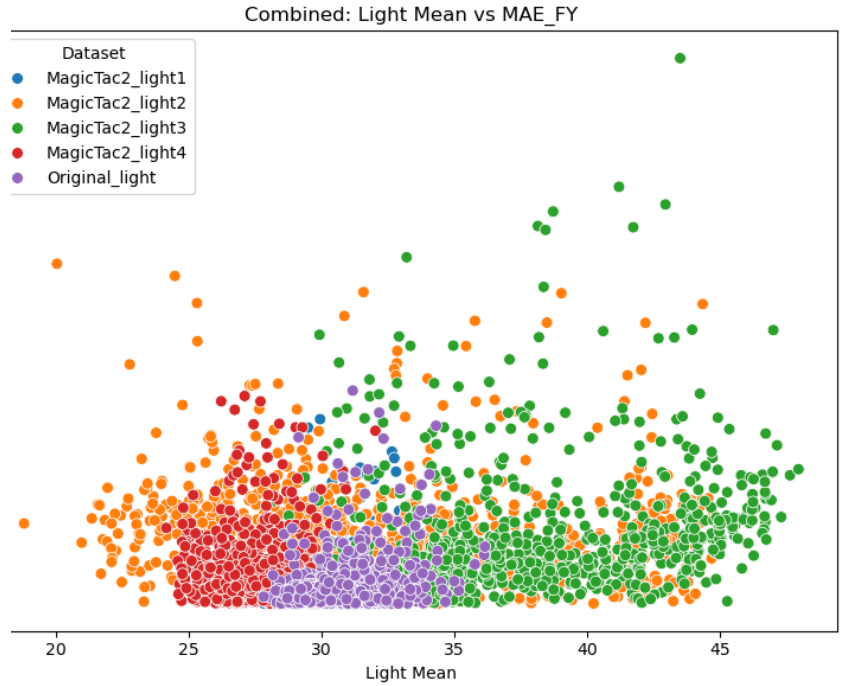


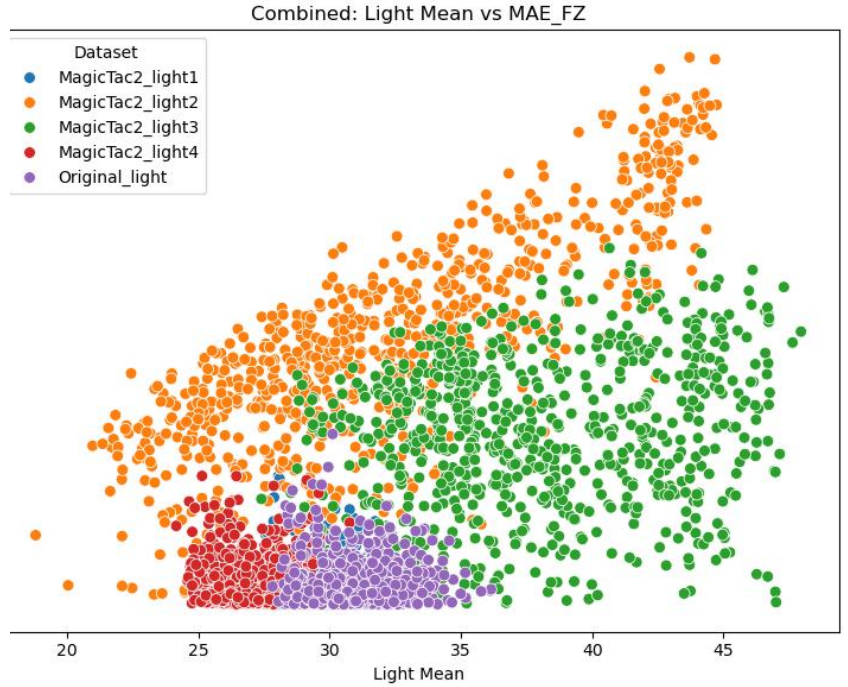

Combined: Light Mean vs MAE_PX
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

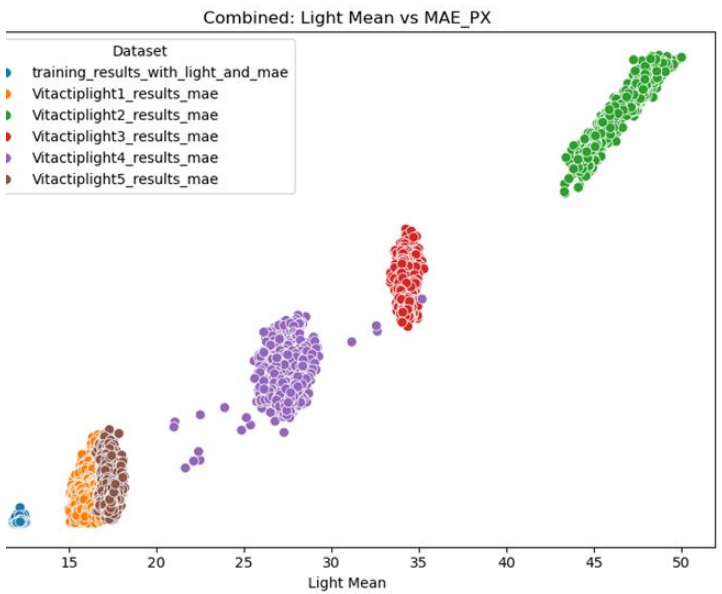

Combined: Light Mean vs MAE_PY
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

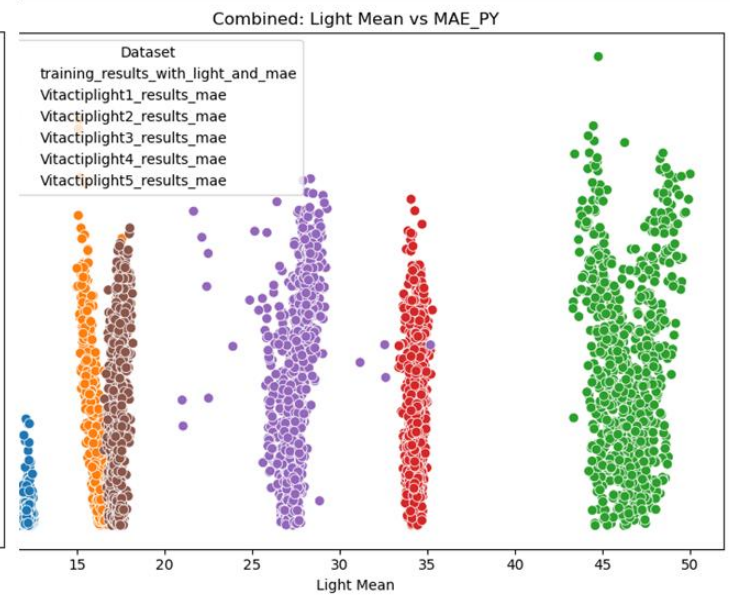

Combined: Light Mean vs MAE_PZ
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

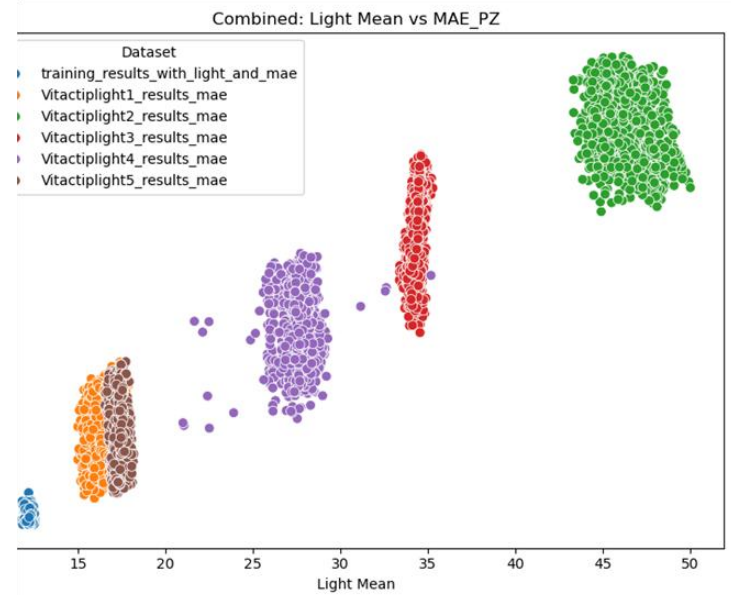

Combined: Light Mean vs MAE_FX
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

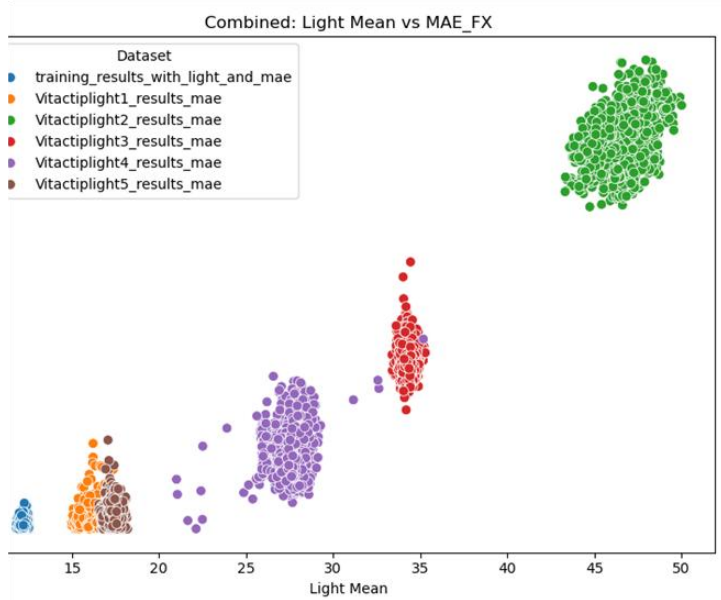

Combined: Light Mean vs MAE_FY
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

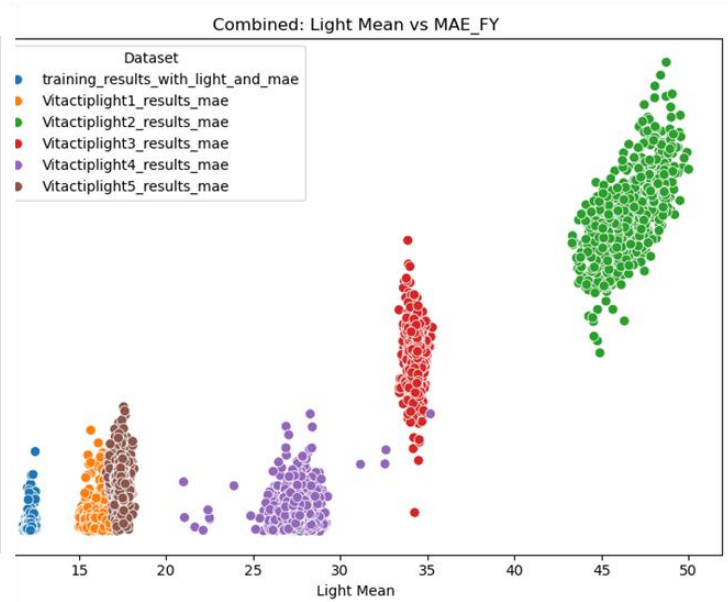

Combined: Light Mean vs MAE_FZ
Dataset
MagicTac2_light1
MagicTac2_light2
MagicTac2_light3
MagicTac2_light4
Original_light
20
25
30
35
40
45
Light Mean

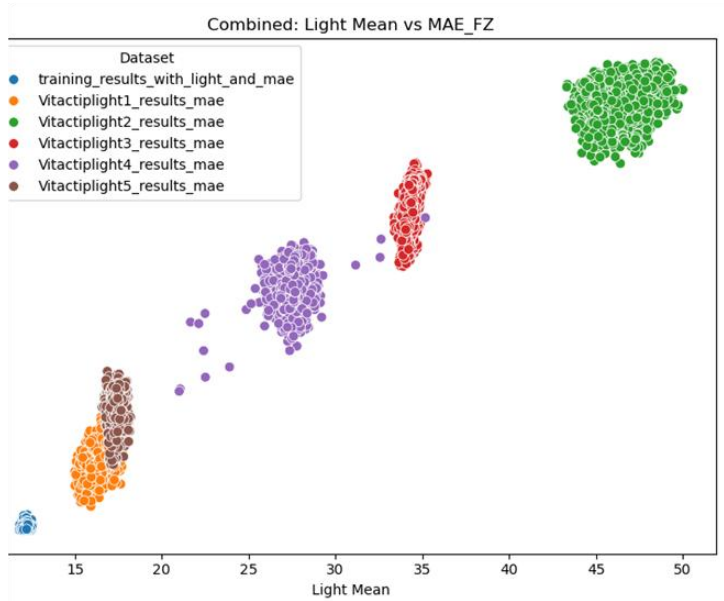

Combined: Light Mean vs MAE_PX
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean


Combined: Light Mean vs MAE_PY
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean


Combined: Light Mean vs MAE_PZ
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean


Combined: Light Mean vs MAE_FX
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean


Combined: Light Mean vs MAE_FY
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean


Combined: Light Mean vs MAE_FZ
Dataset
training_results_with_light_and_mae
Vitactiplight1_results_mae
Vitactiplight2_results_mae
Vitactiplight3_results_mae
Vitactiplight4_results_mae
Vitactiplight5_results_mae
15
20
25
30
35
40
45
50
Light Mean

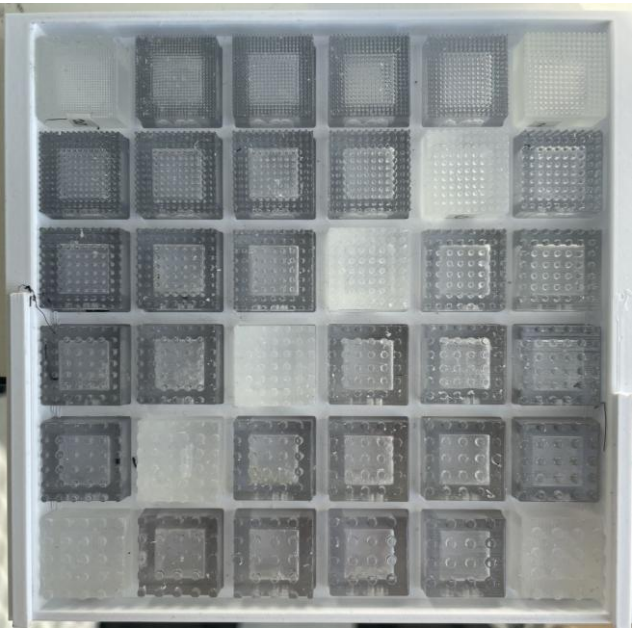
Spatial Resolution

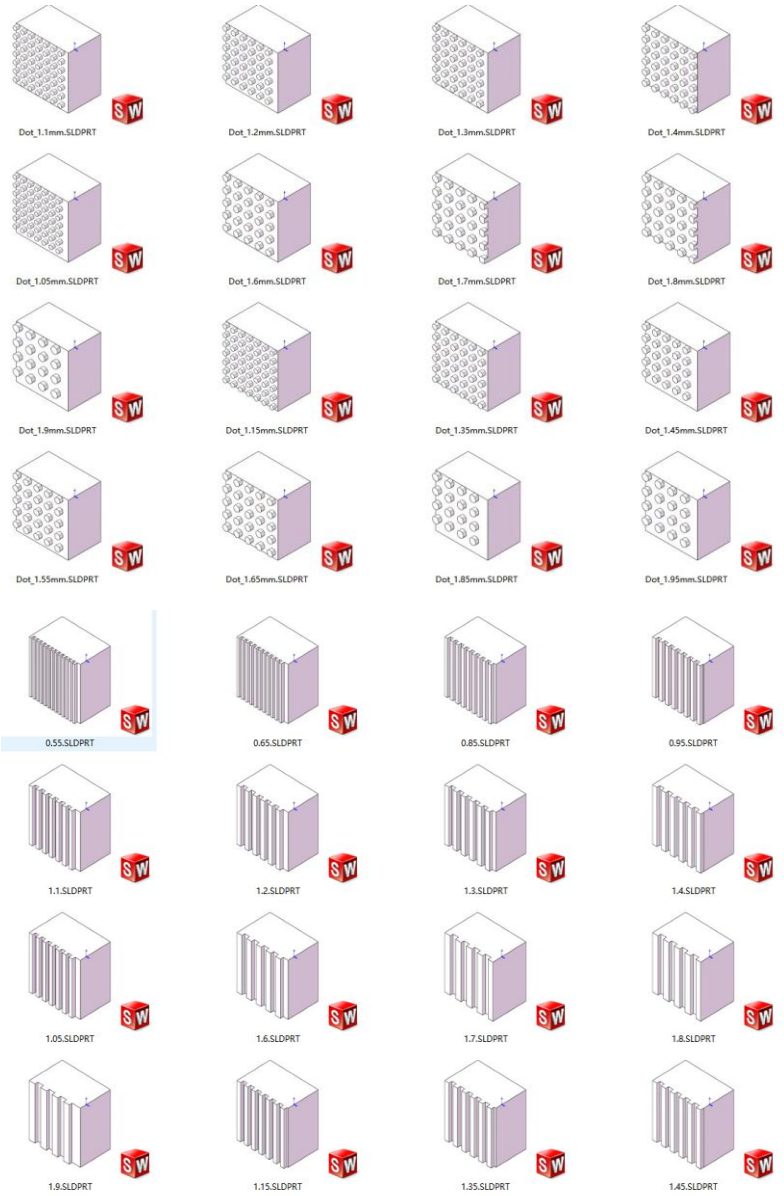
36 classes

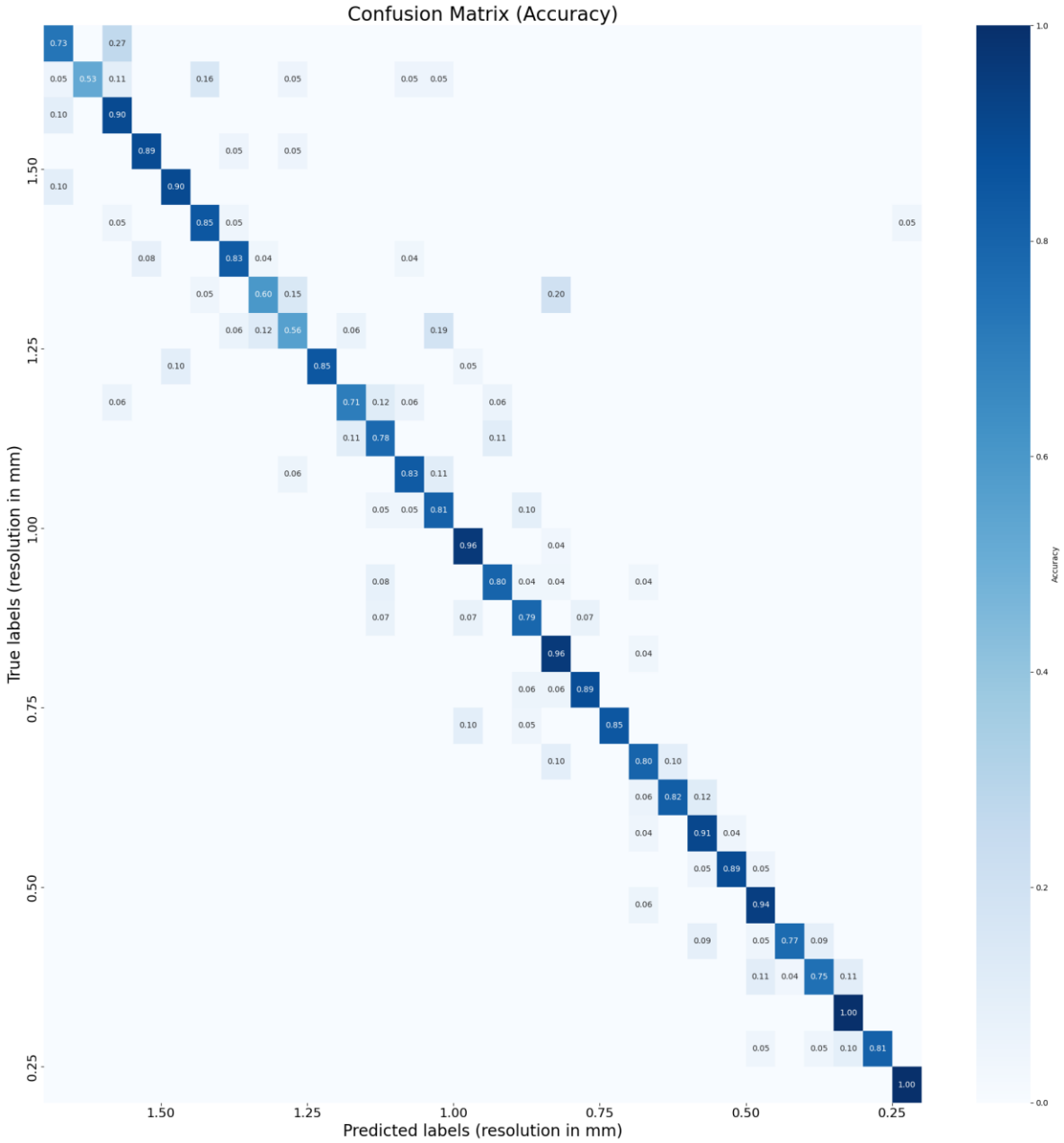
Resolution: **2 - 0.25 mm**
gap = 0.05mm

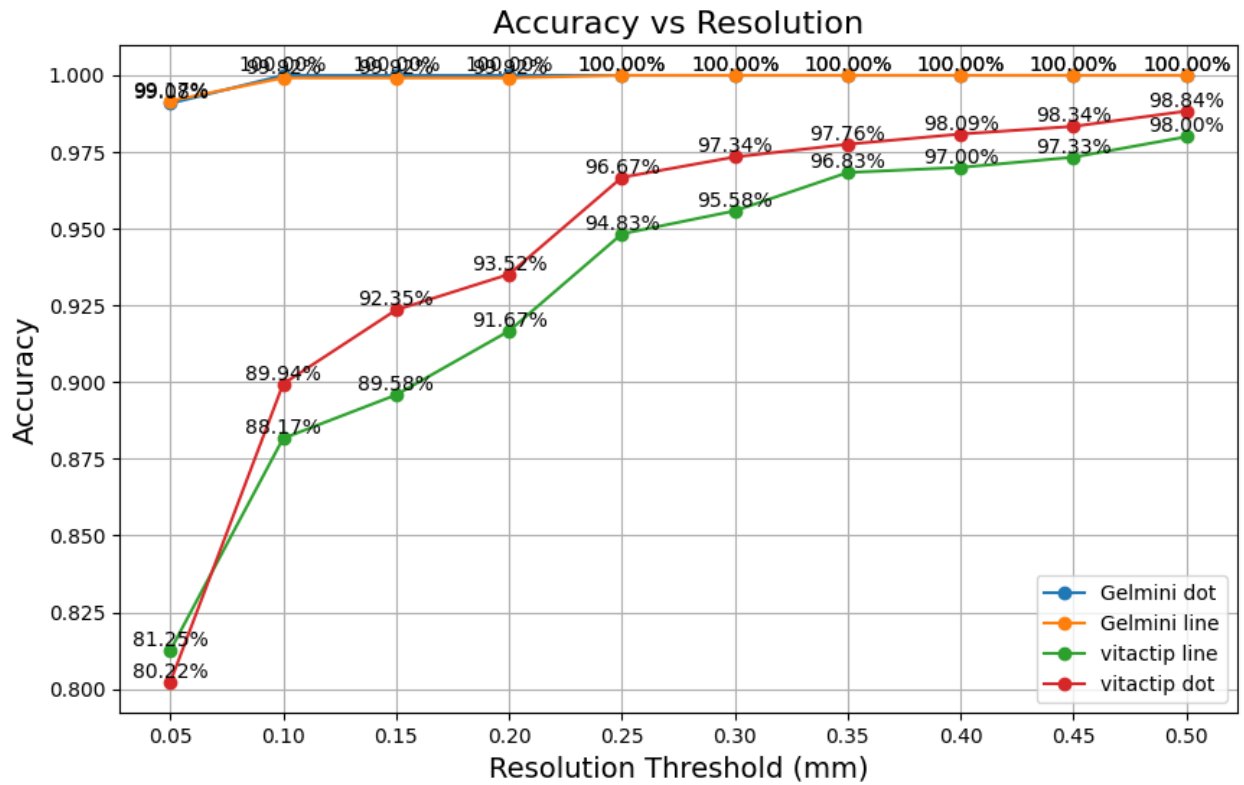
Shape: Dot, Line

Color: Black, Transparent






Confusion Matrix (Accuracy)
True labels (resolution in mm)
Predicted labels (resolution in mm)
1.50
1.25
1.00
0.75
0.50
0.25
Accuracy


Accuracy vs Resolution
Accuracy
Resolution Threshold (mm)
Gelmini dot
Gelmini line
vitactip line
vitactip dot
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
100.00%
99.08%
89.94%
88.17%
92.35%
89.58%
93.52%
91.67%
96.67%
94.83%
97.34%
95.58%
97.76%
96.83%
98.09%
97.00%
98.34%
97.33%
98.84%
98.00%
81.25%
80.22%