

Cross-Domain Applications of Data Science in Autonomous Robotics: A Study of Methods and Insights

M. Carson^{1,2}

Dr. T. Tiedemann¹

¹Department of Computer Science, University of Applied Sciences Hamburg (HAW), Hamburg, Germany

²Department of Computer Science, California State University Long Beach (CSULB), Long Beach, CA, USA

ICSOFT 2023
July 11th, 2023



- 1 Motivation/Objectives
- 2 Methodology
- 3 Results
- 4 Conclusion and Outlook



Motivation

- In 3 separate autonomous robotics projects, progress with the current data was stagnating.
- Making significant contributions without domain knowledge is difficult.
- By doing a broad analysis between projects, I could potentially draw new insight.

This work was partly funded by the Federal Republic of Germany, Ministry for the Environment, Nature Conservation and Nuclear Safety, grant no. 67KI1013A/B/C.

Motivation

- In 3 separate autonomous robotics projects, progress with the current data was stagnating.
- Making significant contributions without domain knowledge is difficult.
- By doing a broad analysis between projects, I could potentially draw new insight.

This work was partly funded by the Federal Republic of Germany, Ministry for the Environment, Nature Conservation and Nuclear Safety, grant no. 67KI1013A/B/C.



Motivation

- In 3 separate autonomous robotics projects, progress with the current data was stagnating.
- Making significant contributions without domain knowledge is difficult.
- By doing a broad analysis between projects, I could potentially draw new insight.

This work was partly funded by the Federal Republic of Germany, Ministry for the Environment, Nature Conservation and Nuclear Safety, grant no. 67KI1013A/B/C.

Motivation

- In 3 separate autonomous robotics projects, progress with the current data was stagnating.
- Making significant contributions without domain knowledge is difficult.
- By doing a broad analysis between projects, I could potentially draw new insight.

This work was partly funded by the Federal Republic of Germany, Ministry for the Environment, Nature Conservation and Nuclear Safety, grant no. 67KI1013A/B/C.

Objectives

- Determine what analysis what can be done with the current data and with limited domain knowledge.
- Perform a ground-up analysis of each project in parallel.
- Highlight potential improvements and considerations before collecting more data.

Objectives

- Determine what analysis what can be done with the current data and with limited domain knowledge.
- Perform a ground-up analysis of each project in parallel.
- Highlight potential improvements and considerations before collecting more data.

Objectives

- Determine what analysis what can be done with the current data and with limited domain knowledge.
- **Perform a ground-up analysis of each project in parallel.**
- Highlight potential improvements and considerations before collecting more data.

Objectives

- Determine what analysis what can be done with the current data and with limited domain knowledge.
- Perform a ground-up analysis of each project in parallel.
- **Highlight potential improvements and considerations before collecting more data.**

Projects

3 **data driven** autonomous robotics projects:

- SmartRecycling-UP
 - Autonomous sorting of coarse waste for recycling.
- RoLand
 - Semi-autonomous strawberry harvesting robot.
- AuTag BeoFisch
 - Autonomous deep-sea robot for fish monitoring and tracking.



Projects

3 **data driven** autonomous robotics projects:

- SmartRecycling-UP
 - Autonomous sorting of coarse waste for recycling.
- RoLand
 - Semi-autonomous strawberry harvesting robot.
- AuTag BeoFisch
 - Autonomous deep-sea robot for fish monitoring and tracking.



Projects

3 **data driven** autonomous robotics projects:

- SmartRecycling-UP
 - Autonomous sorting of coarse waste for recycling.
- RoLand
 - Semi-autonomous strawberry harvesting robot.
- AuTag BeoFisch
 - Autonomous deep-sea robot for fish monitoring and tracking.



Projects

3 **data driven** autonomous robotics projects:

- SmartRecycling-UP
 - Autonomous sorting of coarse waste for recycling.
- RoLand
 - Semi-autonomous strawberry harvesting robot.
- AuTag BeoFisch
 - Autonomous deep-sea robot for fish monitoring and tracking.



SmartRecycling-UP

Labelled multi-spectral images of unsorted construction waste
(UV, VIS, NIR, SWIR)

- Complex and unpredictable environment
 - 15 dimensional data from images
 - Labelled between 7 materials
-
- Goal: Material classification



Scene in a recycling facility (Lange, 2023).



SmartRecycling-UP

Labelled multi-spectral images of unsorted construction waste
(UV, VIS, NIR, SWIR)

- Complex and unpredictable environment
- 15 dimensional data from images
- Labelled between 7 materials
- Goal: Material classification



Scene in a recycling facility (Lange, 2023).



SmartRecycling-UP

Labelled multi-spectral images of unsorted construction waste
(UV, VIS, NIR, SWIR)

- Complex and unpredictable environment
 - **15 dimensional data from images**
 - Labelled between 7 materials
-
- Goal: Material classification



Scene in a recycling facility (Lange, 2023).



SmartRecycling-UP

Labelled multi-spectral images of unsorted construction waste
(UV, VIS, NIR, SWIR)

- Complex and unpredictable environment
- 15 dimensional data from images
- Labelled between 7 materials

-
- Goal: Material classification



Scene in a recycling facility (Lange, 2023).



SmartRecycling-UP

Labelled multi-spectral images of unsorted construction waste
(UV, VIS, NIR, SWIR)

- Complex and unpredictable environment
 - 15 dimensional data from images
 - Labelled between 7 materials
-
- Goal: Material classification



Scene in a recycling facility (Lange, 2023).



RoLand

Multi-spectral images of strawberry plants (UV, VIS, NIR, SWIR)

- Challenging natural environments
- 15 dimensional data from images
- Labelled between 4 classes
 - 4 subclasses for strawberry ripeness
- Goal: Detection of ripe and unripe/defective strawberries.



Setup used for a first in-field data collection (Tiedemann, 2022).



RoLand

Multi-spectral images of strawberry plants
(UV, VIS, NIR, SWIR)

- Challenging natural environments
- 15 dimensional data from images
- Labelled between 4 classes
 - 4 subclasses for strawberry ripeness
- Goal: Detection of ripe and unripe/defective strawberries.



Setup used for a first in-field data collection (Tiedemann, 2022).



RoLand

Multi-spectral images of strawberry plants
(UV, VIS, NIR, SWIR)

- Challenging natural environments
- **15 dimensional data from images**
- Labelled between 4 classes
 - 4 subclasses for strawberry ripeness
- Goal: Detection of ripe and unripe/defective strawberries.



Setup used for a first in-field data collection (Tiedemann, 2022).



RoLand

Multi-spectral images of strawberry plants
(UV, VIS, NIR, SWIR)

- Challenging natural environments
- 15 dimensional data from images
- Labelled between 4 classes
 - 4 subclasses for strawberry ripeness

-
- Goal: Detection of ripe and unripe/defective strawberries.



Setup used for a first in-field data collection (Tiedemann, 2022).



RoLand

Multi-spectral images of strawberry plants
(UV, VIS, NIR, SWIR)

- Challenging natural environments
- 15 dimensional data from images
- Labelled between 4 classes
 - 4 subclasses for strawberry ripeness

-
- Goal: Detection of ripe and unripe/defective strawberries.



Setup used for a first in-field data collection (Tiedemann, 2022).



AuTag BeoFisch

Underwater video and sonar data from the Baltic Sea.

- Turbid and challenging underwater environments
 - Unlabelled
 - Needs to be temporally aligned
-
- Goal: Detection of fish and other aquatic animals



Video frame of a moon
jelly (Zach, 2021)



AuTag BeoFisch

Underwater video and sonar data from the Baltic Sea.

- Turbid and challenging underwater environments
 - Unlabelled
 - Needs to be temporally aligned
-
- Goal: Detection of fish and other aquatic animals



Video frame of a moon jelly (Zach, 2021)



AuTag BeoFisch

Underwater video and sonar data from the Baltic Sea.

- Turbid and challenging underwater environments
 - **Unlabelled**
 - Needs to be temporally aligned
-
- Goal: Detection of fish and other aquatic animals



Video frame of a moon jelly (Zach, 2021)



AuTag BeoFisch

Underwater video and sonar data from the Baltic Sea.

- Turbid and challenging underwater environments
- Unlabelled
- Needs to be temporally aligned

-
- Goal: Detection of fish and other aquatic animals



Video frame of a moon jelly (Zach, 2021)



AuTag BeoFisch

Underwater video and sonar data from the Baltic Sea.

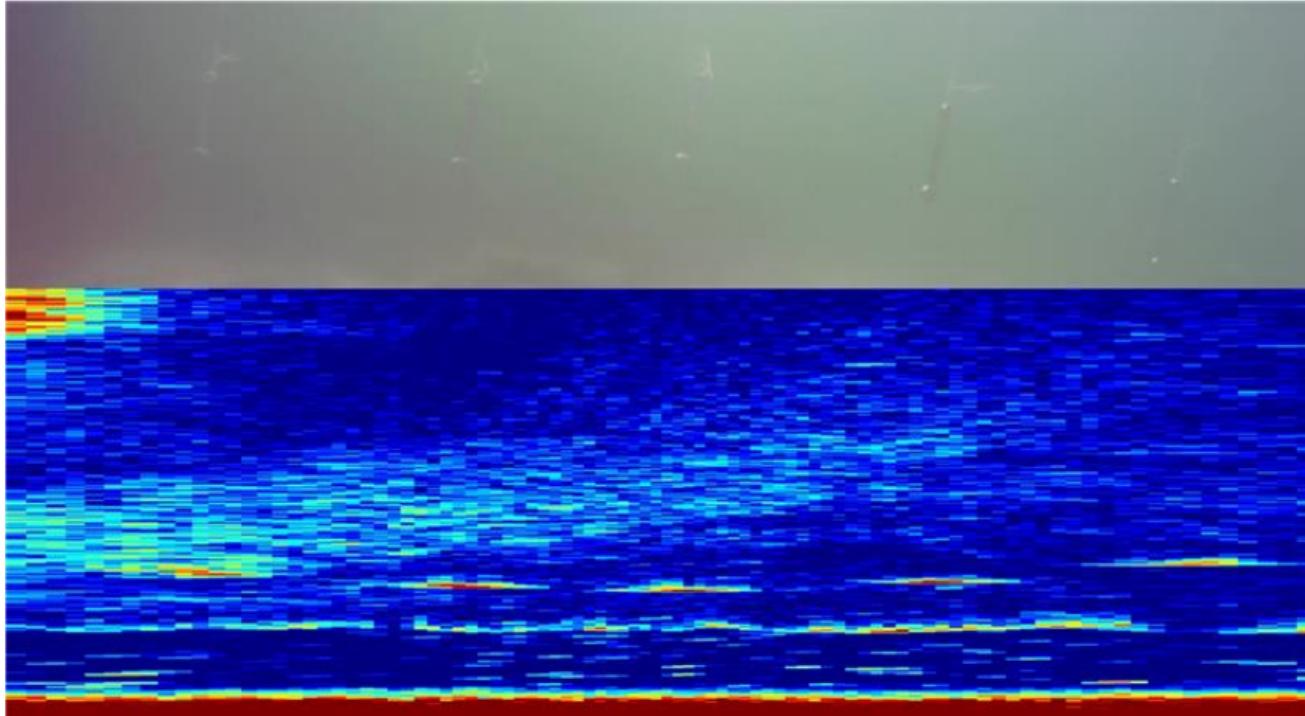
- Turbid and challenging underwater environments
 - Unlabelled
 - Needs to be temporally aligned
-
- Goal: Detection of fish and other aquatic animals



Video frame of a moon jelly (Zach, 2021)



AuTag BeoFisch



Data Summary

- Characteristically challenging environments.
- Two datasets of high dimension labelled images.
- A strongly dissimilar dataset of especially difficult data.

Data Summary

- Characteristically challenging environments.
 - Two datasets of high dimension labelled images.
 - A strongly dissimilar dataset of especially difficult data.

Data Summary

- Characteristically challenging environments.
- Two datasets of high dimension labelled images.
- A strongly dissimilar dataset of especially difficult data.

Data Summary

- Characteristically challenging environments.
- Two datasets of high dimension labelled images.
- A strongly dissimilar dataset of especially difficult data.

Methodology

Ground-up data analysis

- ① Pre-processing
- ② Histograms and Correlation Maps
- ③ Outlier Detection and Box Plots

- ④ Dimensionality reduction and visualization
- ⑤ Clustering



Methodology

Ground-up data analysis

- ① Pre-processing
- ② Histograms and Correlation Maps
- ③ Outlier Detection and Box Plots

- ④ Dimensionality reduction and visualization
- ⑤ Clustering

Methodology

Ground-up data analysis

- ① Pre-processing
 - ② Histograms and Correlation Maps
 - ③ Outlier Detection and Box Plots
-
- ④ Dimensionality reduction and visualization
 - ⑤ Clustering

Methodology

Ground-up data analysis

- ① Pre-processing
- ② Histograms and Correlation Maps
- ③ Outlier Detection and Box Plots

- ④ Dimensionality reduction and visualization
- ⑤ Clustering

Methodology

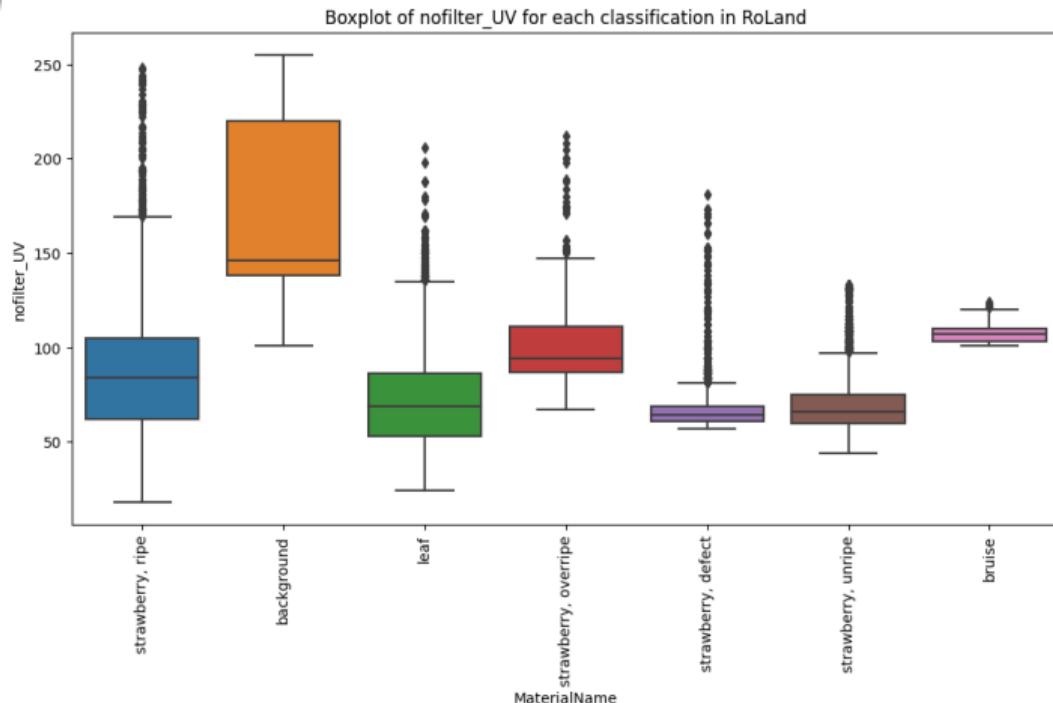
Ground-up data analysis

- ① Pre-processing
- ② Histograms and Correlation Maps
- ③ Outlier Detection and Box Plots

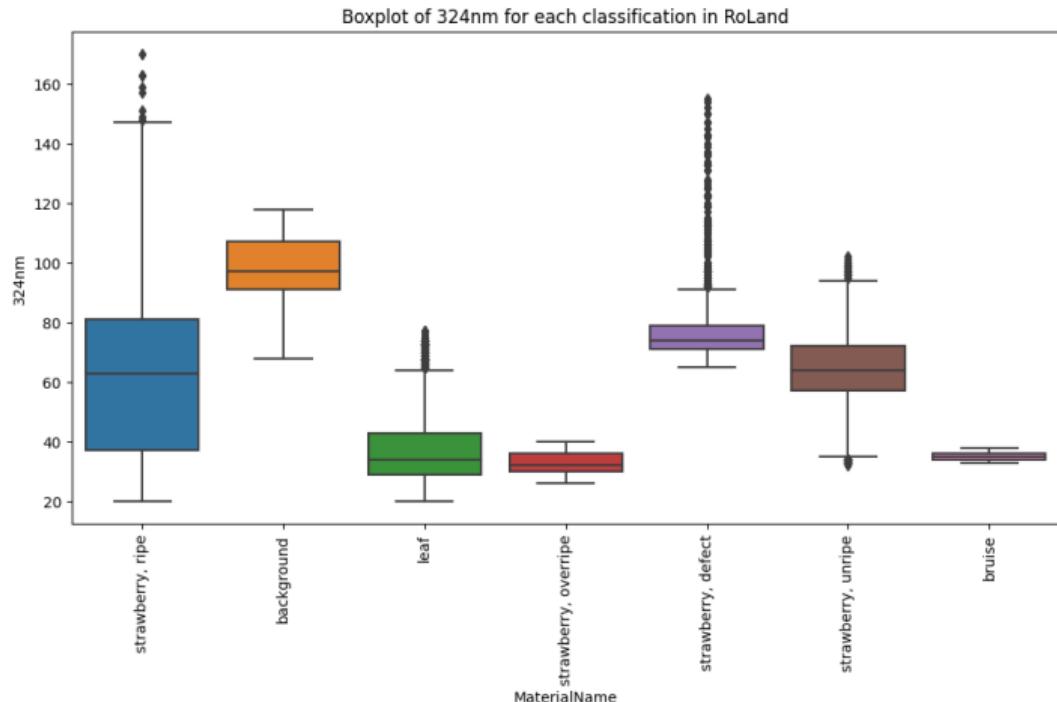
- ④ Dimensionality reduction and visualization
- ⑤ Clustering



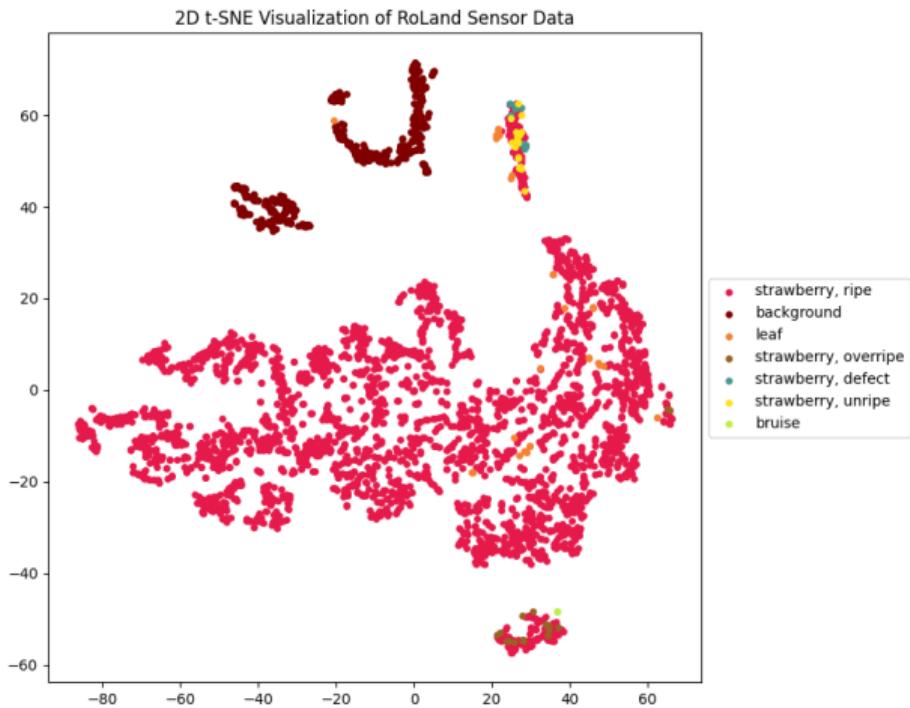
RoLand



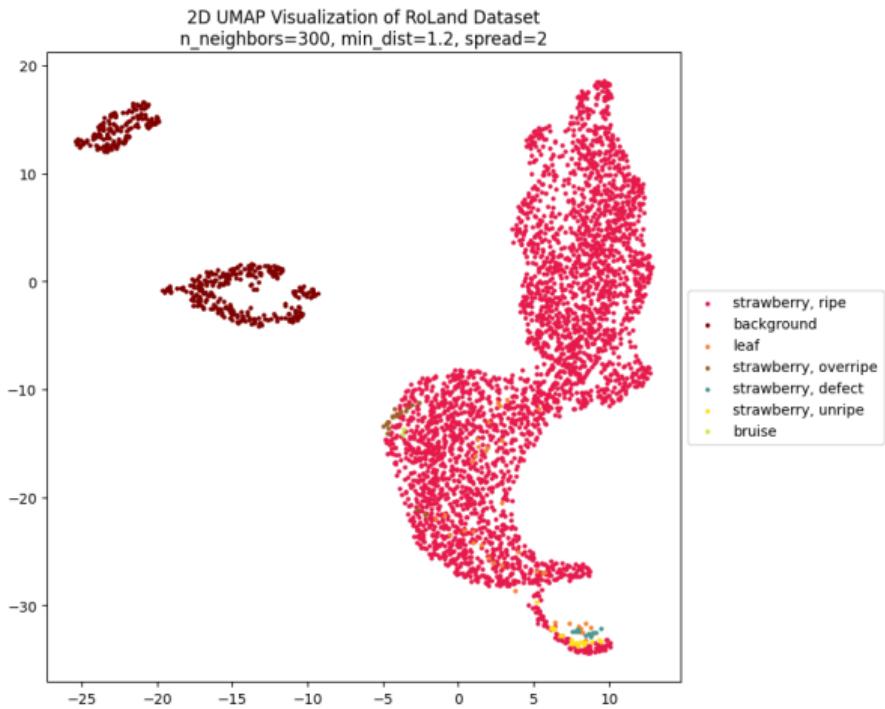
RoLand



RoLand

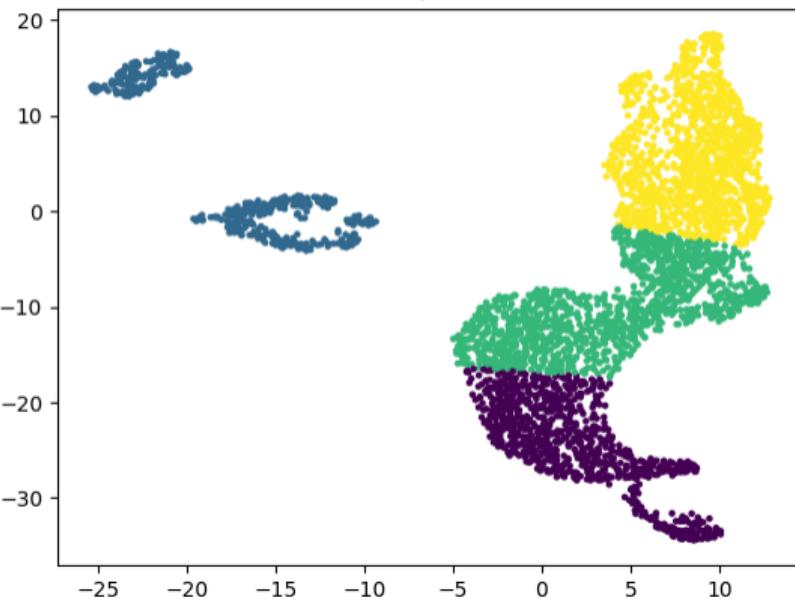


RoLand



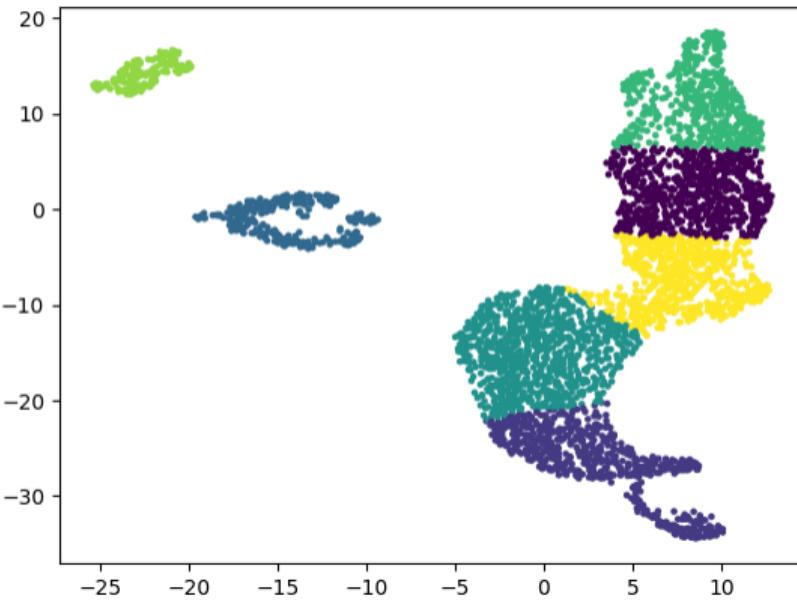
RoLand

K-means Clustering of RoLand Dataset
clusters=4
ARI=0.18, NMI=0.42



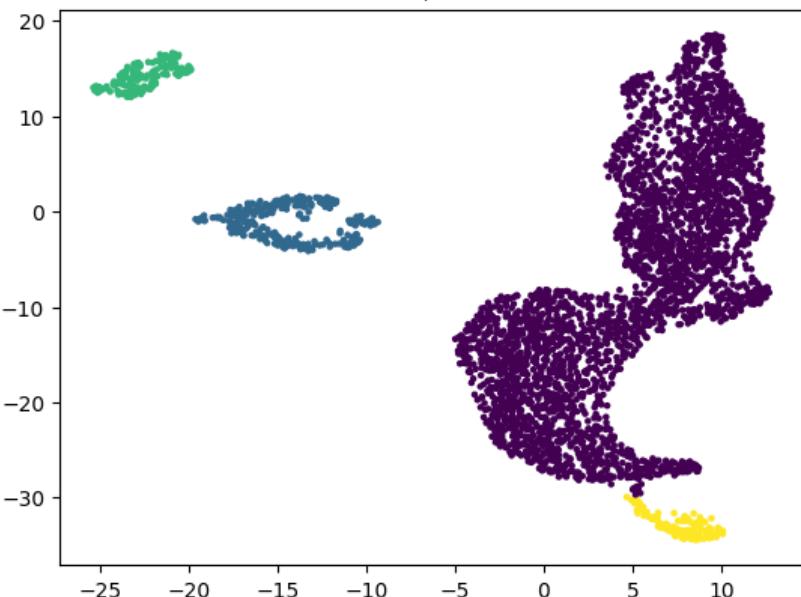
RoLand

K-means Clustering of RoLand Dataset
clusters=7
ARI=0.10, NMI=0.33



RoLand

DBSCAN Clustering of RoLand Dataset
 $\text{eps}=1.2$, $\text{min_samples}=18$
ARI Score: 0.83, AMI Score: 0.72



RoLand

- Strong under-representation of non-ripe strawberry data.
- Background and non-ripe clustering is relatively easy.
 - Clustering by specific ripeness is much more difficult.
- UV and 324nm wavelengths had greatest differences in ripeness subclasses.

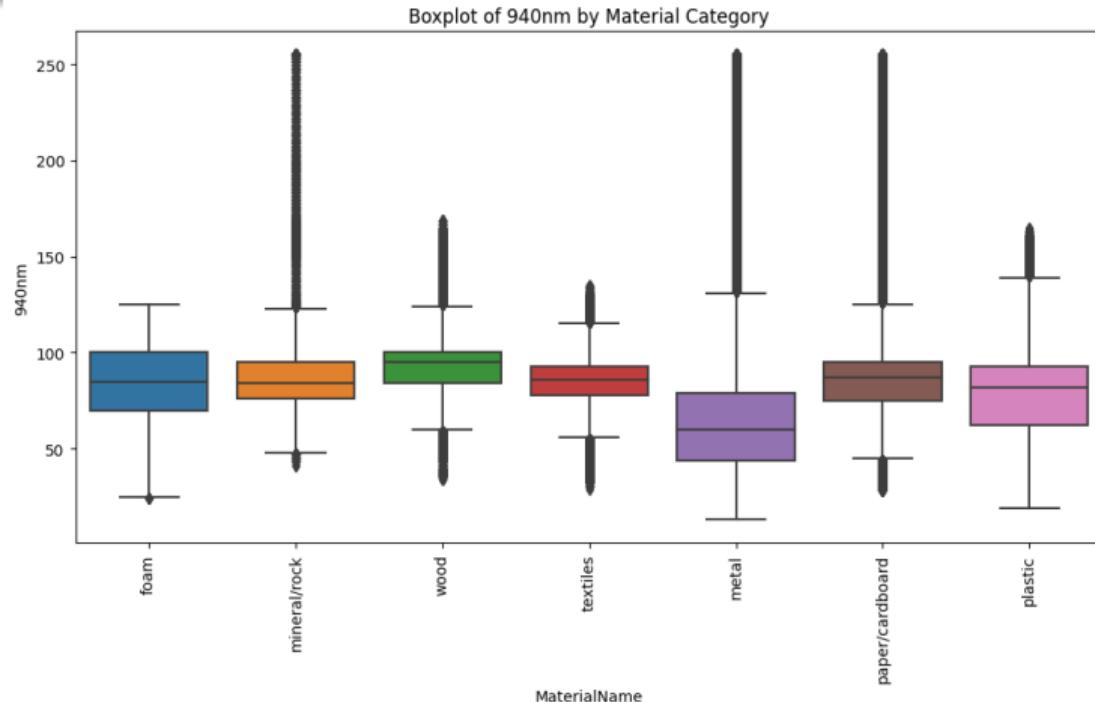
RoLand

- Strong under-representation of non-ripe strawberry data.
- Background and non-ripe clustering is relatively easy.
 - Clustering by specific ripeness is much more difficult.
- UV and 324nm wavelengths had greatest differences in ripeness subclasses.

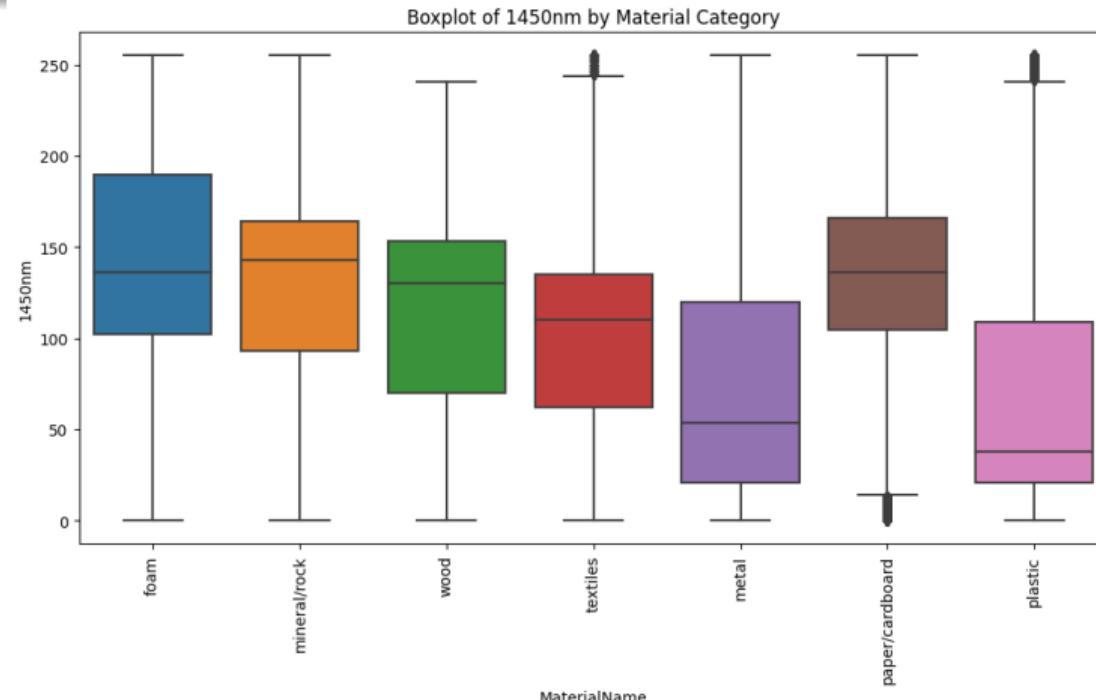
RoLand

- Strong under-representation of non-ripe strawberry data.
- Background and non-ripe clustering is relatively easy.
 - Clustering by specific ripeness is much more difficult.
- UV and 324nm wavelengths had greatest differences in ripeness subclasses.

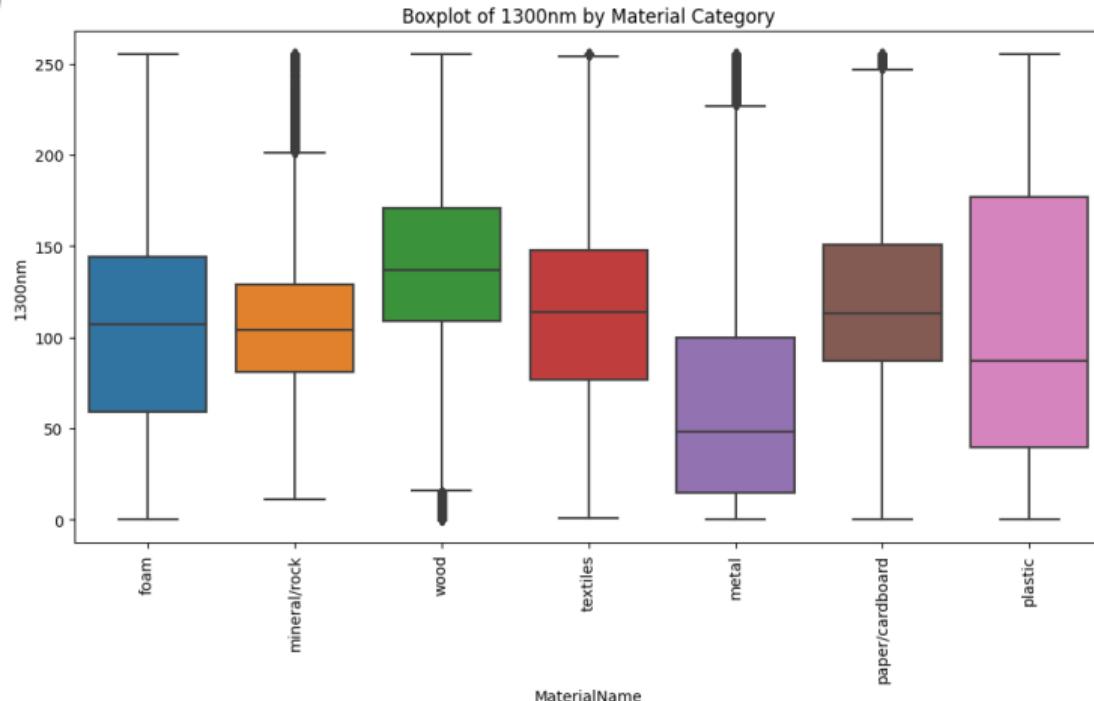
SmartRecycling-UP



SmartRecycling-UP

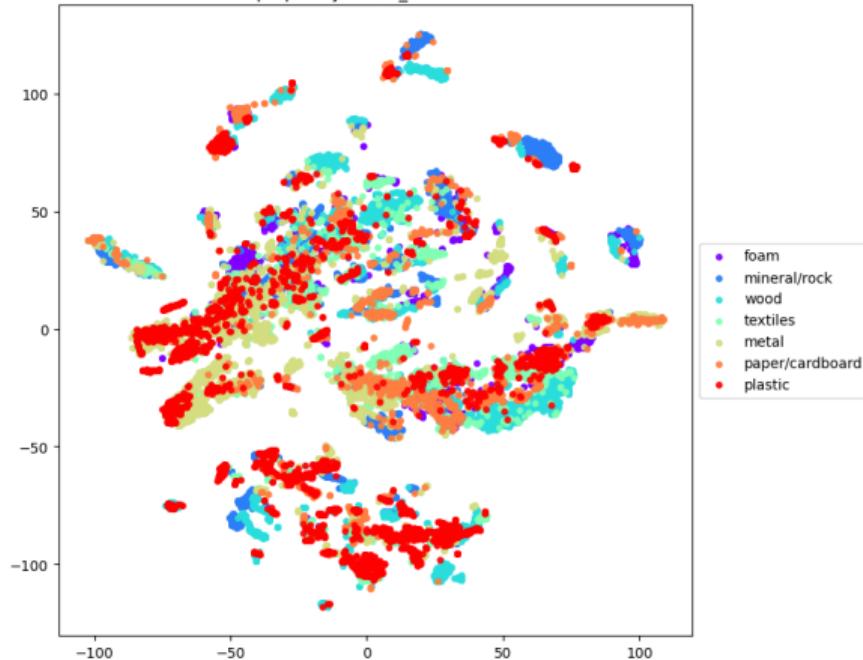


SmartRecycling-UP



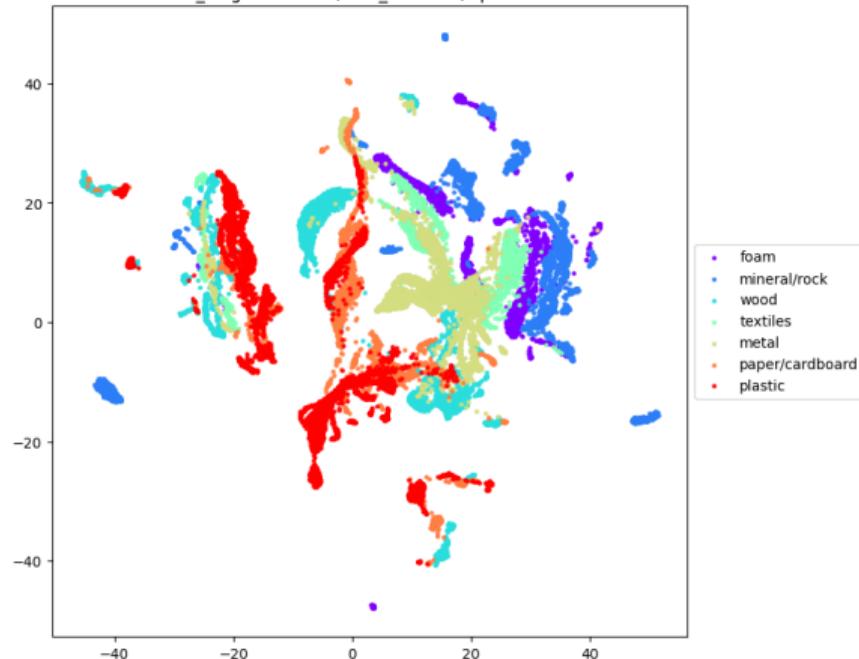
SmartRecycling-UP

2D t-SNE Visualization of SmartRecycling-UP Dataset
(perplexity=75, n_iter=2000)



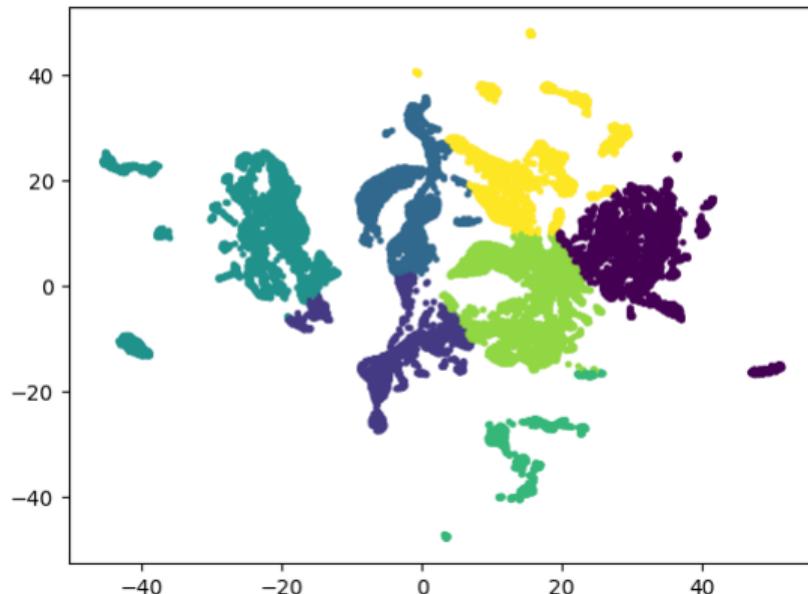
SmartRecycling-UP

2D UMAP Visualization of SmartRecycling-UP Dataset
n_neighbors=200, min_dist=1.2, spread=3



SmartRecycling-UP

K-means Clustering of SmartRecycling-UP Dataset
clusters=7
ARI=0.22, NMI=0.34

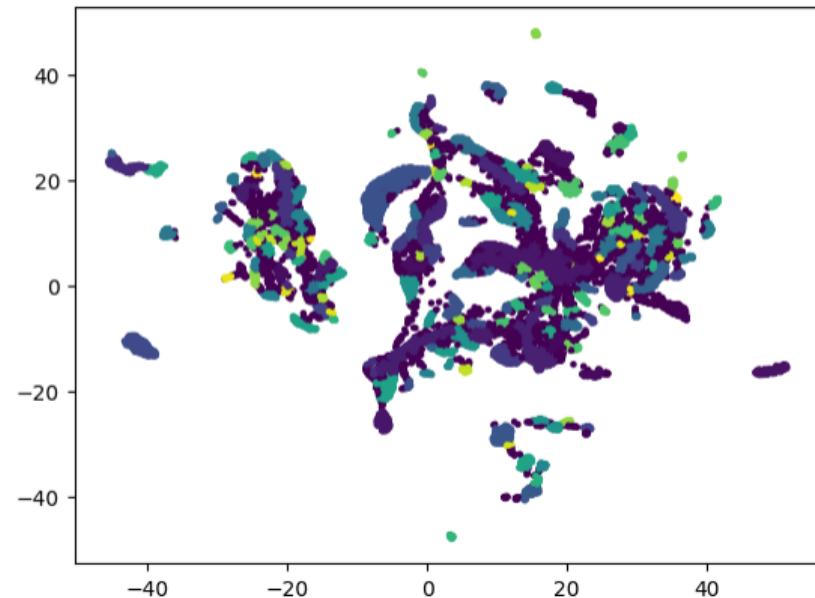


SmartRecycling-UP

DBSCAN Clustering of SmartRecycling-UP Dataset

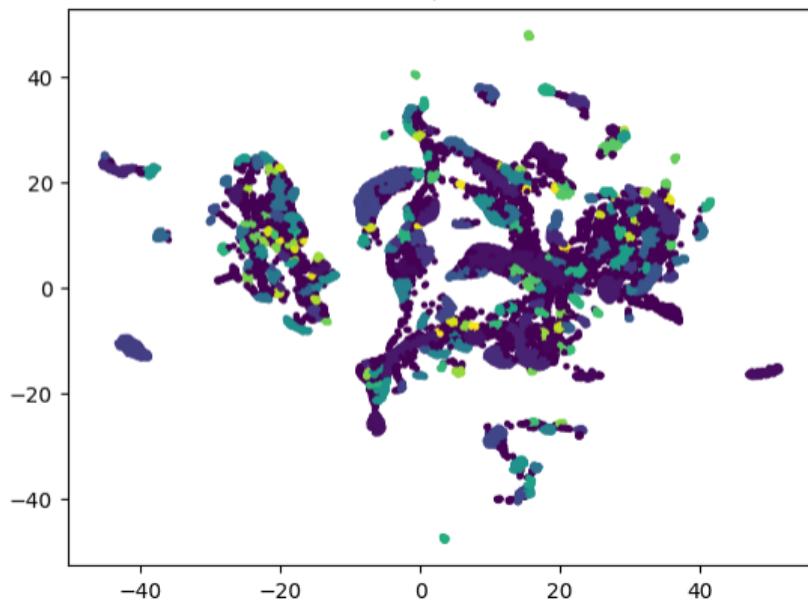
$\text{eps}=0.53$, $\text{min_samples}=17$

ARI Score: 0.10, AMI Score: 0.45



SmartRecycling-UP

HDBSCAN Clustering of SmartRecycling-UP Dataset
cluster_selection_epsilon=0.51, min_samples=17
ARI Score: 0.15, AMI Score: 0.44



SmartRecycling-UP

- Clustering with the current data looks promising, but further distinction is needed for better results.
- Frequencies weaker than 1450nm may have stronger distinctions.
- Clustering outcomes typically improve with HDBSCAN.

SmartRecycling-UP

- Clustering with the current data looks promising, but further distinction is needed for better results.
- Frequencies weaker than 1450nm may have stronger distinctions.
- Clustering outcomes typically improve with HDBSCAN.



SmartRecycling-UP

- Clustering with the current data looks promising, but further distinction is needed for better results.
- Frequencies weaker than 1450nm may have stronger distinctions.
- Clustering outcomes typically improve with HDBSCAN.



AuTag BeoFisch

What happened with the 3rd project?

- Not much data (1000 usable frames).
- Not much **valuable** data.
- Too dissimilar to draw any comparative insight.

AuTag BeoFisch

What happened with the 3rd project?

- Not much data (1000 usable frames).
- Not much **valuable** data.
- Too dissimilar to draw any comparative insight.

AuTag BeoFisch

What happened with the 3rd project?

- Not much data (1000 usable frames).
- Not much **valuable** data.
- Too dissimilar to draw any comparative insight.

Common techniques and outcomes

- Data visualization using UMAP outperformed t-SNE in both RoLand and SmartRecycling-UP
- DBSCAN looks to be ideal for the current level of fitting offered from UMAP.
- Cross-domain hyperparameter adjustments

Common techniques and outcomes

- Data visualization using UMAP outperformed t-SNE in both RoLand and SmartRecycling-UP
- DBSCAN looks to be ideal for the current level of fitting offered from UMAP.
- Cross-domain hyperparameter adjustments

Common techniques and outcomes

- Data visualization using UMAP outperformed t-SNE in both RoLand and SmartRecycling-UP
- DBSCAN looks to be ideal for the current level of fitting offered from UMAP.
- Cross-domain hyperparameter adjustments

Common difficulties

- Highly complex and closely related data
- Dirty dataset

Common difficulties

- Highly complex and closely related data
- Dirty dataset

Conclusion

- Benefits of performing data analysis on projects with **similar data structure** in parallel
 - Greatly improves:
 - Analysis speed
 - Algorithm implementation
 - Understanding of used techniques
 - Easier hyper-parameter adjustment

Conclusion

- Benefits of performing data analysis on projects with **similar data structure** in parallel
 - Greatly improves:
 - Analysis speed
 - Algorithm implementation
 - Understanding of used techniques
 - Easier hyper-parameter adjustment

Conclusion

- Benefits of performing data analysis on projects with **similar data structure** in parallel
 - Greatly improves:
 - Analysis speed
 - Algorithm implementation
 - Understanding of used techniques
 - Easier hyper-parameter adjustment

Future Work

- Working with the project teams with new insights.
- Attempt a similar analysis approach with data more similar to AuTag BeoFischs'.
- Continue this analysis deeper into the training of classification models.

Future Work

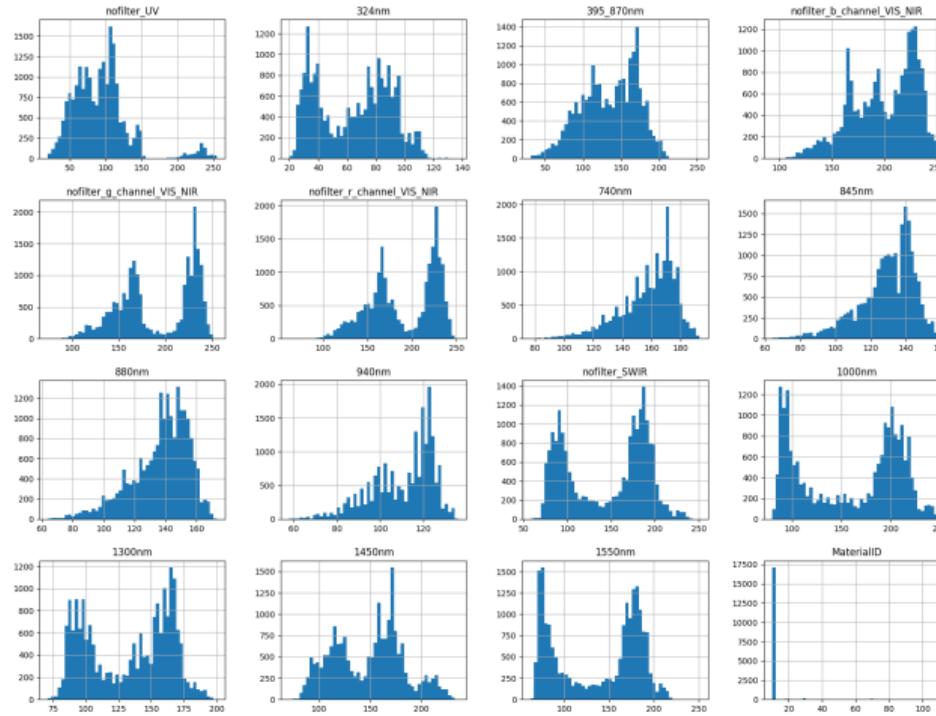
- Working with the project teams with new insights.
- Attempt a similar analysis approach with data more similar to AuTag BeoFischs'.
- Continue this analysis deeper into the training of classification models.

Future Work

- Working with the project teams with new insights.
- Attempt a similar analysis approach with data more similar to AuTag BeoFischs'.
- Continue this analysis deeper into the training of classification models.

Thank you for your attention!
Questions?

Appendix



Appendix

3D t-SNE visualization of dataset

3D t-SNE visualization of dataset

- foam
- mineral
- wood
- textile
- metal
- paper
- plastic

Appendix



Appendix



References I

-  Lange, T., Babu, A., Meyer, P., Keppner, M., Tiedemann, T., Wittmaier, M., Wolff, S., and Vögele, T.
First lessons learned of an artificial intelligence robotic system for autonomous coarse waste recycling using multispectral imaging-based methods.
In *19th International Symposium on Waste Management and Sustainable Landfilling*, pages 0–0, 0, 2023.
-  Tiedemann, T., Cordes, F., Keppner, M., and Peters, H.
Challenges of autonomous in-field fruit harvesting and concept of a robotic solution.
In *Proceedings of the 19th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2022)*, volume 1, pages 508–515, 2022.



References II

-  Tiedemann, T., Keppner, M., Runge, T., Vögele, T., Wittmaier, M., and Wolff, S. Concept of a robotic system for autonomous coarse waste recycling.
In *Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021)*, pages 493–500, 2021.
ISBN 978-989-758-522-7.
-  Zach, J., Busse, C., Funk, S., Möllmann, C., Renner, B. C., and Tiedemann, T. Towards non-invasive fish monitoring in hard-to-access habitats using autonomous underwater vehicles and machine learning.
In *OCEANS 2021: San Diego – Porto*, pages 1–8, San Diego, CA, USA, 2021.

