Moritz Ottenweller, Michael Kalb & Steffen Rüger

Kaskadierte Anwendung von Foundation Models als Verfahren zur Beschreibung von Leichtverpackungsstoffströmen im Recycling

Fraunhofer EZRT

Entwicklungszentrum Röntgentechnik des Fraunhofer-Instituts für Integrierte Schaltungen IIS

K3I CYCLING

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

FONA Forschung für Nachhaltigkeit

Agenda

Kaskadierte Anwendung von Foundation Models als Verfahren zur Beschreibung von Leichtverpackungsstoffströmen im Recycling

1. Motivation
2. Objekt-Lokalisierung:
   Segmentieren von LVP mit SAM
3. Objekt-Klassifikation:
   „transparenter“, „nicht-transparenter“ LVP und dessen Farbkategorien
4. Zusammenfassung & Ausblick

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Motivation

Ziel und Nutzen für Forschung und Industrie

Hauptpunkt 1: Verbesserung der Sortierung fürs Recycling

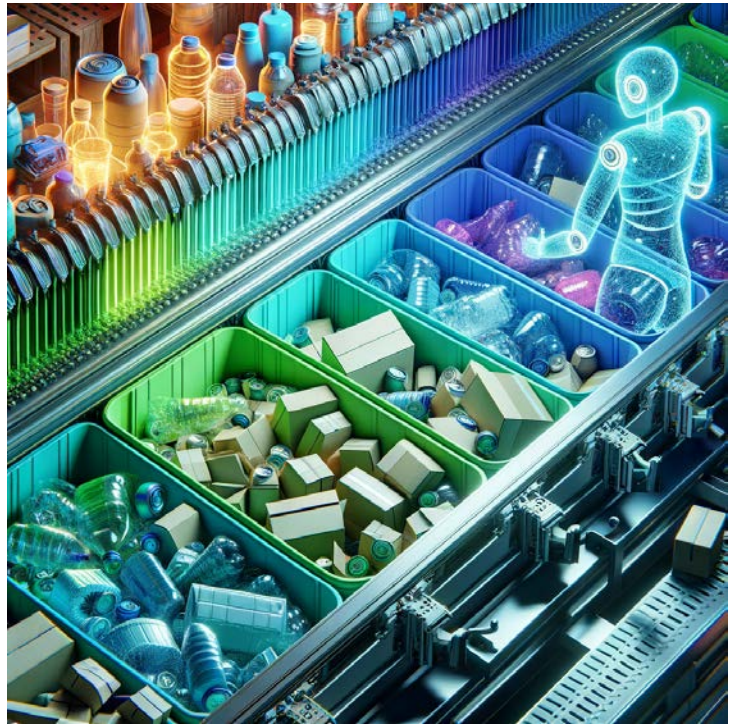
Hauptpunkt 2: Entwicklung, Einsatz und Datenanforderung von KI-Methoden

Herangehensweise

- Verwendung von RGB-Bildern aus Wertstoffaufbereitungsanlagen
- Aufgaben der Lokalisierung und Klassifikation von LVP-Objekten
- Detektion mit adaptiertem SAM und Klassifikation mit DINOv2 Foundation Model

Ziel und Nutzen unserer Arbeit

- Automatisierte Annotation zur Verbesserung der Sortiertiefe
- Trainingsdatensatz für KI-Modellen zur zukünftigen Optimierung der Wertstoffsortierung



DALL-E Prompt:

Verbesserung der Sortiertiefe von Wertstoffen im Leichtverpackungsabfall durch KI zur Gewinnung von Sekundärrohstoffen.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

SAM ist ein Foundation Modell, welches mittels Eingabe-Prompts Instanz Segmentierung ermöglicht

Foundation Modelle

- vortrainierte KI-Modelle
- vielseitig einsetzbar
- durch Feinabstimmung an spezifische Aufgaben anpassbar

Segment Anything

1 Billion (SA-1B) Dataset [7]

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

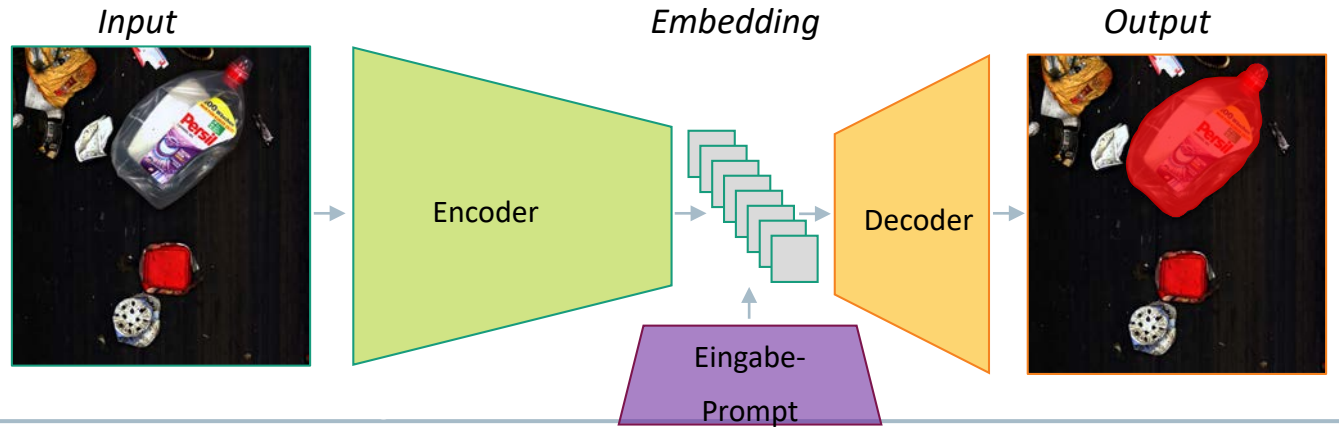
Objekt-Lokalisierung: Segmentieren von LVP mit SAM

SAM ist ein Foundation Modell, welches mittels Eingabe-Prompts Instanz Segmentierung ermöglicht

Aufbau und Funktionsweise

Segment Anything Model (SAM) [7] bestehend aus drei Komponenten:

1. Encoder: Verarbeitung des Eingabebildes Image Embeddings
2. Eingabe-Prompts: zur gezielten Ansteuerung der Segmentierungs-ergebnisse
3. Decoder: liefert Segmentierungsergebnis mit der höchsten Wahrscheinlichkeit



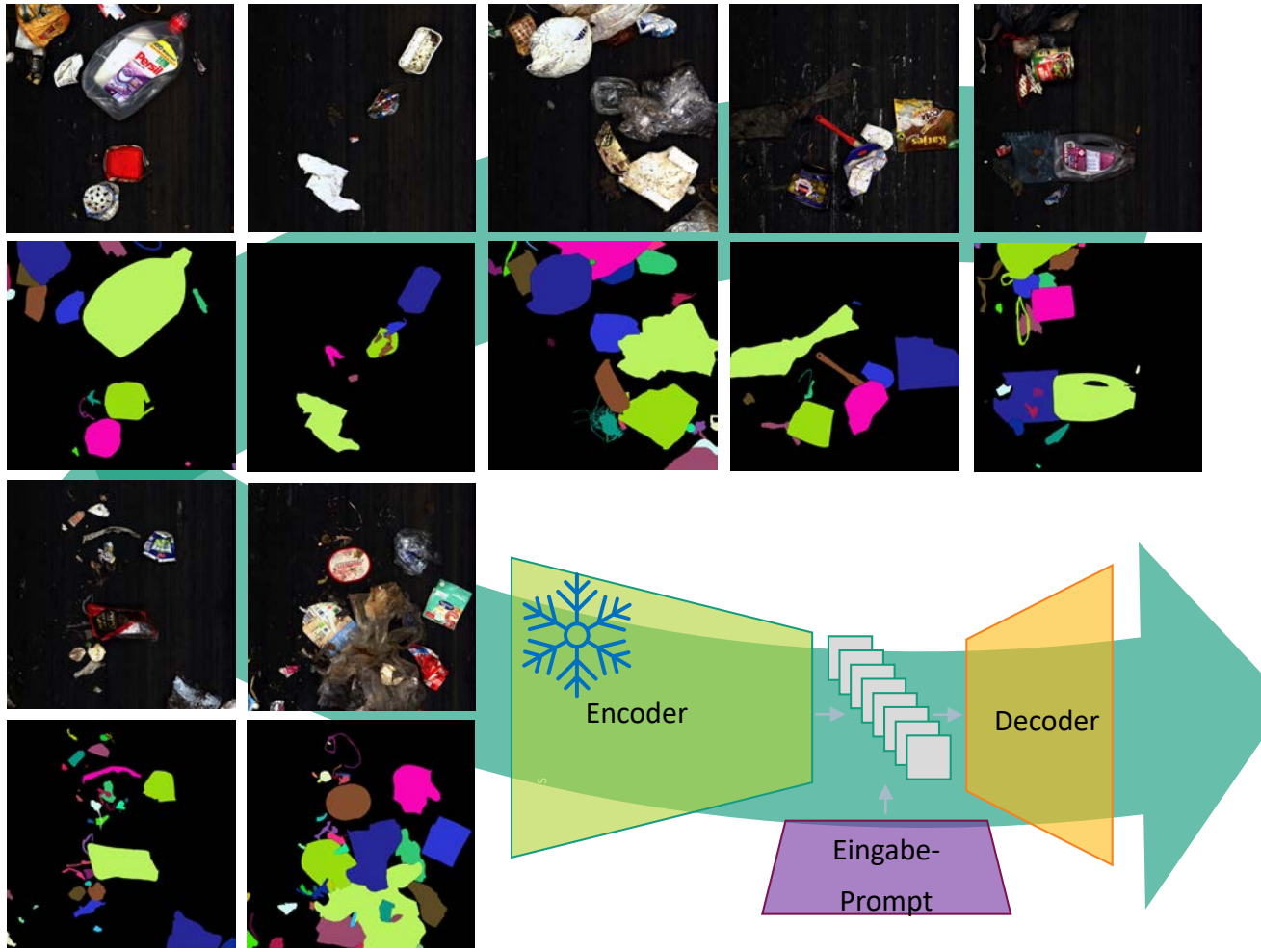
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

Fine-Tuning ermöglicht die Spezialisierung des Modells auf eine dedizierte Aufgabe

Fine-Tuning: Unsere Herangehensweise und Umsetzung

- Feinabstimmung mit ersten annotierten Daten der Zieldomäne hier LVP-Abfall + Instanz Segmenten/Masken
- Diese wurden mittels semi-automatisierter Annotation erhoben
  - Händisches Ansteuern der Eingabe-Prompt von SAM in einem Annotation-Tool
  - Ggf. Pixelgenaue Korrektur der Ergebnismasken
- Fine-Tuning Setup:
  - Datenbasis: 155 Bilder mit 2058 Instanzen
  - Frozen Encoder -> Remaining Parameters Trainable
  - Point Prompt
  - Epochen: 300



Ziel: Adaption auf Regionen von Interesse

- Ganzheitliche Erkennung der Objekte
- Unterdrückung der Erkennung von Unrelevantem (z.B. Förderband, Verschmutzungen)

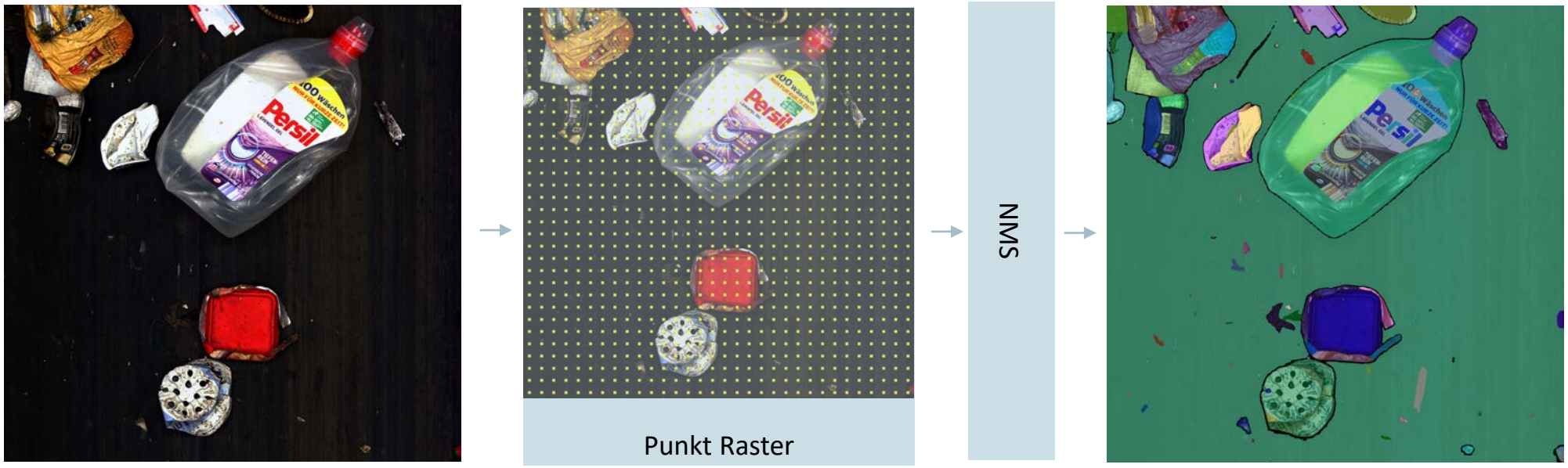
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

Gezielte Ansteuerung der Eingabe-Prompts – Prompt Engineering

Methode 1: AutoMaskGenerator (7)

- Point Prompts: Bestehend aus äquidistanten Rastern aus N x N Vorhersagepunkten
- Filterung von Segmenten mit niedriger Qualität und Duplikaten



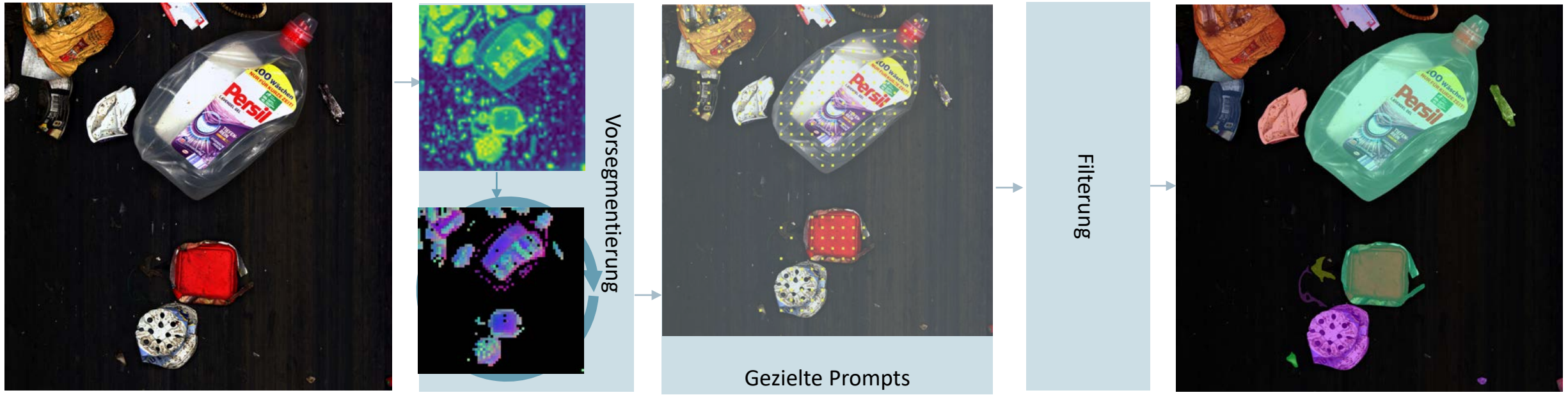
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

Gezielte Ansteuerung der Eingabe-Prompts – Prompt Engineering

Methode 2: PCAMaskGenerator

- Vorsegmentierung: Basierend auf Image Embeddings mittels PCA [9] und Schwellenwertberechnung
- Gezielte Ansteuerung der Point Prompts
- Filterung
  - CCA + Fragment Anpassung [1]
  - Duplikat Erkennung
  - Maskengröße



Alle Parameter der Methoden wurden mittels Rastersuche optimiert

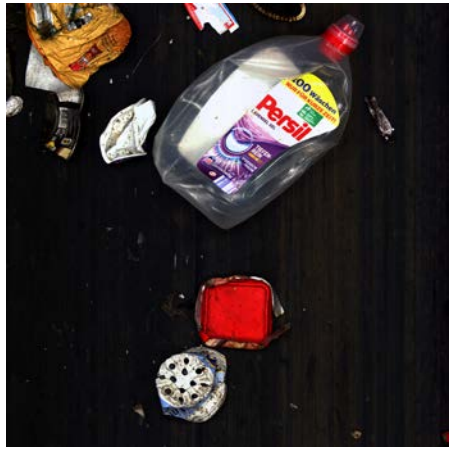
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

Fine-Tuning und Prompt Engineering erzielen eine deutlich gesteigerte Ergebnisqualität

	Modell	Prompt Engineering	Test-Ergebnisse			
			TP	FN	FP	F1-Score (2*TP/(2*TP+FP+FN))
A	pretrained	AutoMaskGenerator	413	61	1.972	0,29
B	pretrained	PCAMaskGenerator	358	116	307	0,63
C	Fine-Tuning	AutoMaskGenerator	357	117	650	0,48
D	Fine-Tuning	PCAMaskGenerator	358	116	98	**0,77**

Eingabebild



A

Maskenvorhersage: 125

B

Maskenvorhersage: 49

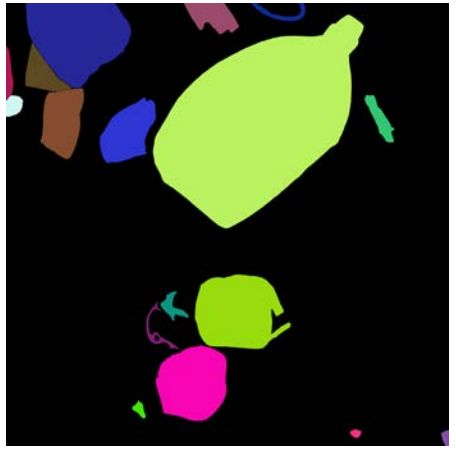
C

Maskenvorhersage: 56

D

Maskenvorhersage: 30

Grundwahrheit



Masken: 17

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Lokalisierung: Segmentieren von LVP mit SAM

Fine-Tuning und Prompt Engineering erzielen eine deutlich gesteigerte Ergebnisqualität

Eingabebild

D

Maskenvorhersage: 30

Grundwahrheit

Masken: 17

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation mit DINOv2

Das Foundation Modell DINOv2 (9)

Informationsverdichtung zu 1D-Vektoren

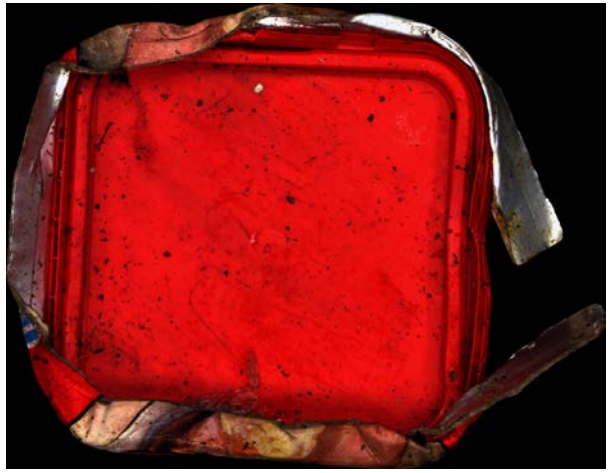
- Semantische Bildbetrachtung
- Clusterbildung zueinander ähnlicher 1D-Vektoren
- Training mit geringen Datenmengen ermöglicht

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation

Datensatz





GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation

Datensatz

	Training-Satz	Validierung-Satz	Test-Satz
RGB-Bilder	155	13	28
Pixel-Masken	2058	68	474
transparent	298	37	37
nicht-transparent	1.376	171	171

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation

Ergebnisse Transparent

Modell	Eingangsformat	Test-Ergebnisse [%]						
		TP	FN	TN	FP	Transparent	Nicht-transparent	Genauigkeit
ResNet18	RGB	25	12	163	8	**67,57**	**95,32**	**90,38**
MLP	DINOv2	34	3	160	11	**91,89**	**93,57**	**93,27**
MLP	SAM	7	30	170	1	18,92	99,42	85,10
SVM	DINOv2	23	14	168	3	62,16	98,25	91,83
SVM	SAM	0	37	171	0	0,00	100	82,21

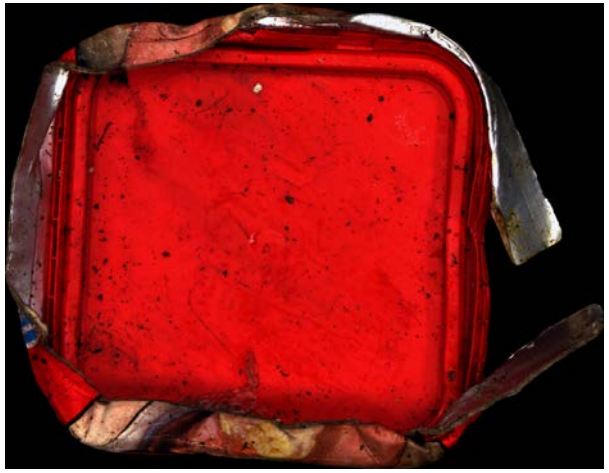
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation

Farberkennung

Klassischer Algorithmus: Schwellwertbasierte Entscheidung über Farbzugehörigkeit

Farbklasse	Genauigkeit [%]
Schwarz	?(96-100)
Weiß	
Rot	
Grün	
Blau	
bunt	



GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Objekt-Klassifikation

Farberkennung

Grenzfälle und Subjektivität:

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Zusammenfassung & Ausblick

Iterative Kaskade aus:

- LVP-Segmentierung mit adaptiertem SAM
- LVP-Klassifikation mit DINOv2
- Initial Aufwand für menschl. Annotator
- Welcher je Iteration geringer wird

In Zukunft:

- Ausweitung der Datenmenge
- Umfangreichere Objektbeschreibung (Sortiertiefe)
- Initial geringer Aufwand für menschl. Annotator

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT

Finanzierung

Die Finanzierung erfolgte durch das Bundesministerium für Bildung und Forschung (BMBF) mit der Fördermaßnahme „KI-Anwendungshub Kunststoffverpackungen – nachhaltige Kreislaufwirtschaft durch Künstliche Intelligenz“ unter dem Förderkennzeichen 033KI201.

Danksagung

Die Autoren bedanken sich für die Möglichkeit einer Messdurchführung bei dem Unternehmen Lobbe Holding GmbH & Co KG in der Wertstoffaufbereitungsanlage Iserlohn, im Speziellen für die Unterstützung von Simon Sadowski und Arne Prior. Ein weiterer Dank geht an Lukas Roming vom Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB für die wissenschaftlich Messdurchführung und das Zurverfügungstellen der Messdaten.

Kontakt:

Moritz Ottenweller

Tel. +49 911 58061 7656

moritz.ottenweller@iis.fraunhofer.de

Fraunhofer EZRT

Entwicklungszentrum Röntgentechnik des Fraunhofer-Instituts für Integrierte Schaltungen IIS

K3I CYCLING

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

FONA Forschung für Nachhaltigkeit

Literaturverzeichnis

(1) Bolelli, F., Cancilla, M., & Grana, C. (2017). Two More Strategies to Speed Up Connected Components Labeling Algorithms. In S. Battiato, G. Gallo, R. Schettini, & F. Stanco (Ed.), *Image Analysis and Processing - ICIAP 2017* (pp. 48-58). Cham: Springer International Publishing. doi:10.1007/978-3-319-68548-9_5

(2) Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging Properties in Self-Supervised Vision Transformers. *arXiv*. doi:https://doi.org/10.48550/arXiv.2104.14294

(3) Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*(20), 273-297. doi:https://doi.org/10.1007/BF00994018

(4) Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*. doi:https://doi.org/10.48550/arXiv.2010.11929

(5) He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked Autoencoders Are Scalable Vision Learners. *arXiv*. doi:https://doi.org/10.48550/arXiv.2111.06377

(6) Kingma, D., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv*. doi:https://doi.org/10.48550/arXiv.1412.6980

(7) Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., . . . Girshick, R. (2023). Segment Anything. *arXiv*. doi:https://doi.org/10.48550/arXiv.2304.02643

(8) Klotz, M., Haupt, M., & Hellweg, S. (2022). Limited utilization options for secondary plastics may restrict their circularity. *Waste Management, 141*, 251-270. doi:https://doi.org/10.1016/j.wasman.2022.01.002

(9) Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., . . . Bojanowski, P. (2024). DINOv2: Learning Robust Visual Features without Supervision. *arXiv*. doi:https://doi.org/10.48550/arXiv.2304.07193

(10) Sudre, C., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. *arXiv*. doi:https://doi.org/10.48550/arXiv.1707.03237

(11) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., . . . Polosukhin, I. (2023). Attention Is All You Need. *arXiv*. doi:https://doi.org/10.48550/arXiv.1706.03762

(12) Volk, R., Stallkamp, C., Steins, J., Yogish, S., Müller, R., Stapf, D., & Schultmann, F. (2021). Techno-economic assessment and comparison of different plastic recycling pathways: A German case study. *Journal of Industrial Ecology, 25*(5), 1318-1337. doi:https://doi.org/10.1111/jiec.13145

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
FONA Forschung für Nachhaltigkeit
Fraunhofer EZRT