Research Article

Prototype-based Few-Shot Learning for Industrial Object Detection

Aalto University
School of Engineering

Hari Prasanth S.M., Nilusha Jayawickrama, Risto Ojala
Autonomy & Mobility Lab, Mechatronics Research Group, Aalto University

TwinFlow

Introduction

Industrial object detection is a challenging task due to its reliance on large annotated datasets and frequent retraining as the object inventory changes. In such cases, new objects must be detected with few labeled samples. To address this, we propose **Decoupled Prototype Matching with Vision Foundation Models (DPM-VFM)**, a training-free framework that decouples object localization and identification. By utilizing the segmentation and representation capabilities of the vision foundation models, DPM-VFM facilitates scalable and efficient few-shot object detection for real-world industrial applications.

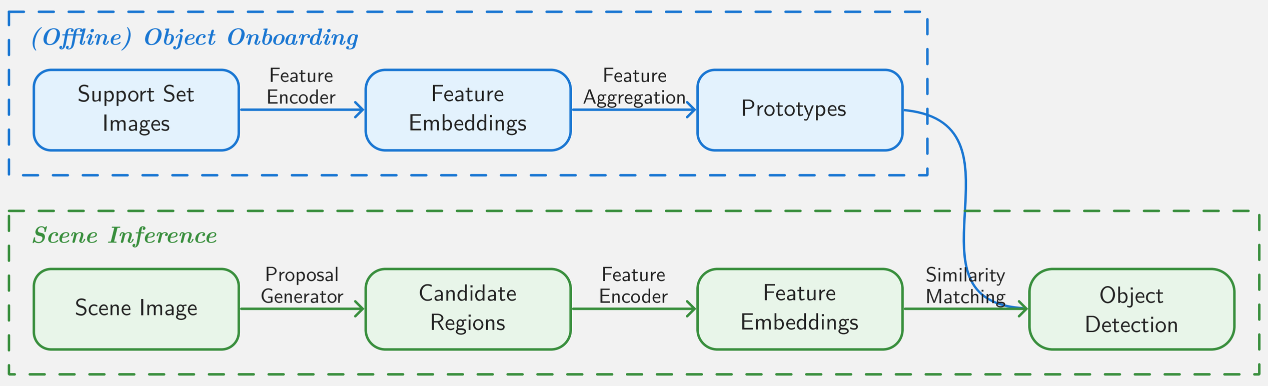

Figure 1. Overview of the proposed DPM-VFM approach

Methods

The proposed **DPM-VFM** is presented in the following figure:

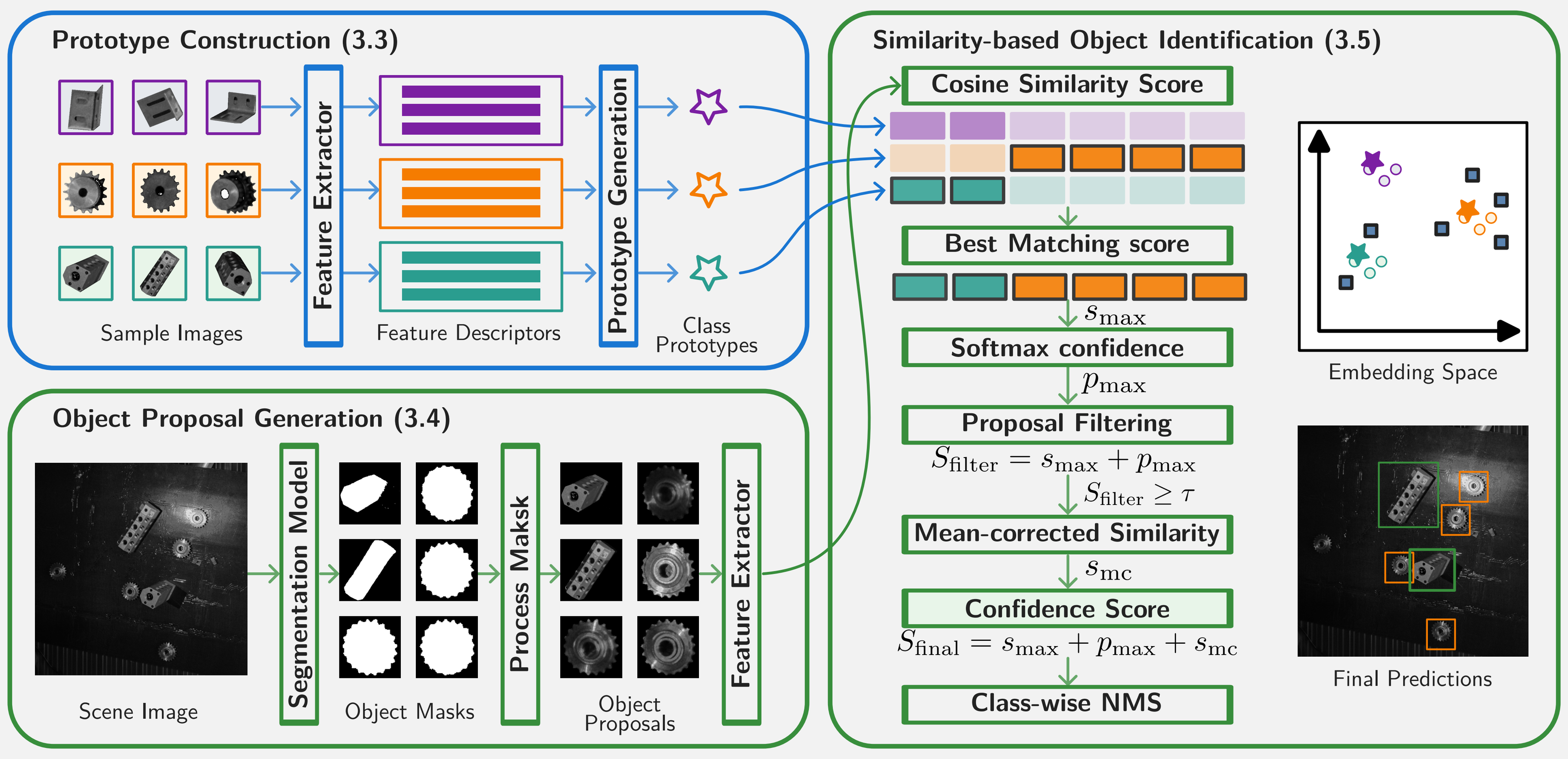

Figure 2. The architecture of the proposed DPM-VFM Pipeline

Decoupled Detection:
Object detection is divided into class-agnostic proposal generation and similarity-based prototype-matching identification.
Proposal Generation:
A segmentation foundation model (e.g., SAM) generates object proposals and masks from the input scene image.
Prototype Matching:
Each proposal is represented using feature embeddings and matched to class prototypes using similarity-based matching.
Embedding Space:
As shown in Fig. 3, the IPD object class samples form distinct clusters in the embedding space, enabling reliable matching with few samples.
Datasets:
The proposed approach is evaluated on BOP industrial datasets with challenges such as clutter, occlusion, and low-texture objects (**Table 1**).

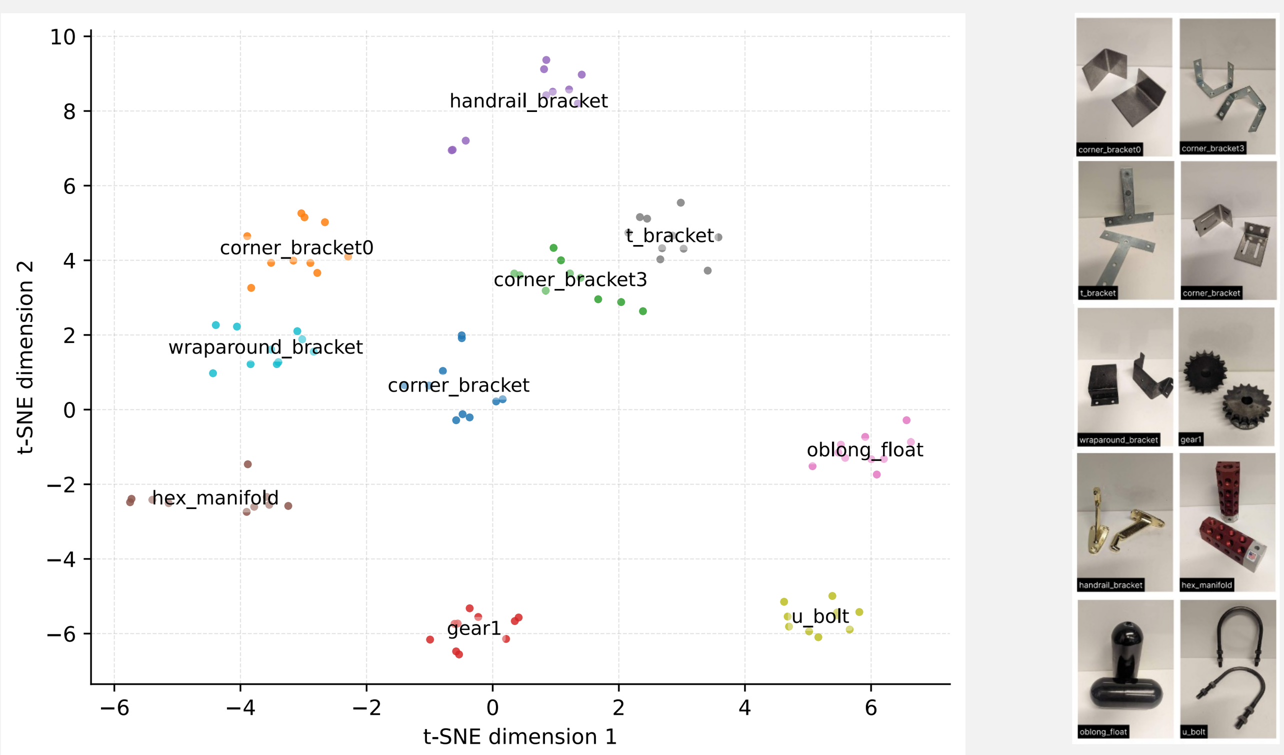

Figure 3. The tSNE visualization of DINOv3 CLS embeddings of IPD object classes

Table 1: BOP-Industrial datasets used in evaluation.

Dataset	# Objects	# Scenes	# Images	Key challenges
ITODD	28	1	721	Realistic bin-picking scenarios
IPD	10	15	1232	Varying lighting conditions
XYZ-IBD	15	60	297	Textureless objects with heavy occlusions

Results

The proposed DPM-VFM achieves strong performance across all BOP industrial datasets, outperforming existing training-free methods such as CNOS and SAM-6D. The AP results are reported in the following table.

Table 2: **Performance comparison on BOP Industrial datasets.** The AP metric (higher is better) reported here. Each SOTA method is evaluated with two segmentation frameworks.

Seg. Model	Method	BOP Industrial			Mean
		ITODD	IPD	XYZ-IBD	
SAM v1 + DINOv2	CNOS[1]	44.3	23.8	22.4	30.2
	SAM6D[2]	**48.1**	33.8	31.3	37.7
	DPM-VFM	45.2	**49.5**	**37.1**	**43.9**
FastSAM + DINOv2	CNOS[1]	48.3	31.5	27.9	35.9
	SAM6D[2]	**49.9**	31.9	26.7	36.2
	DPM-VFM	45.6	**48.8**	**34.4**	**42.9**
SAM v1 + DINOv3	**DPM-VFM (best)**	**46.0**	**51.5**	**36.4**	**44.6**

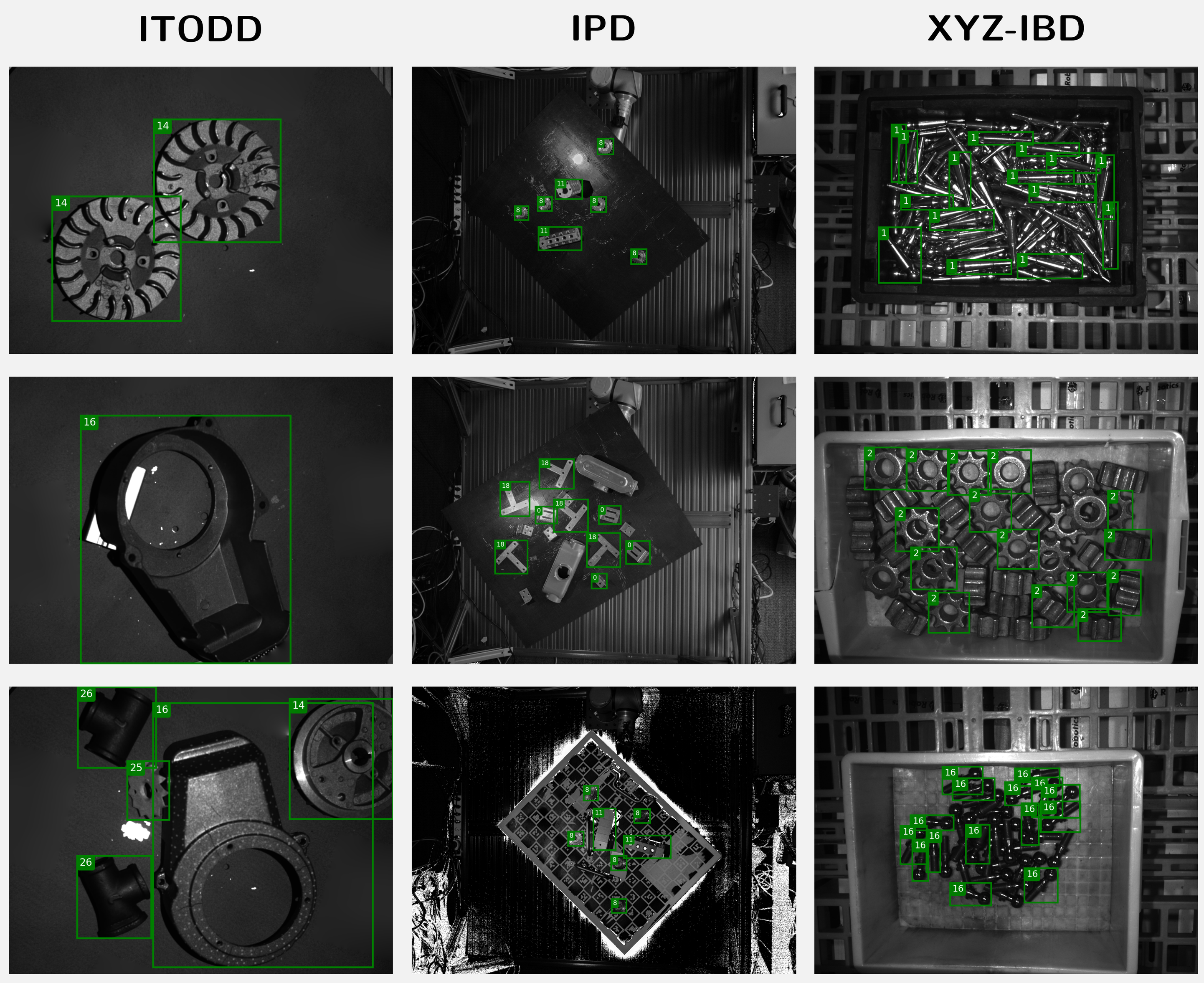

Figure 3. Qualitative results of successful detections produced by DPM-VFM approach

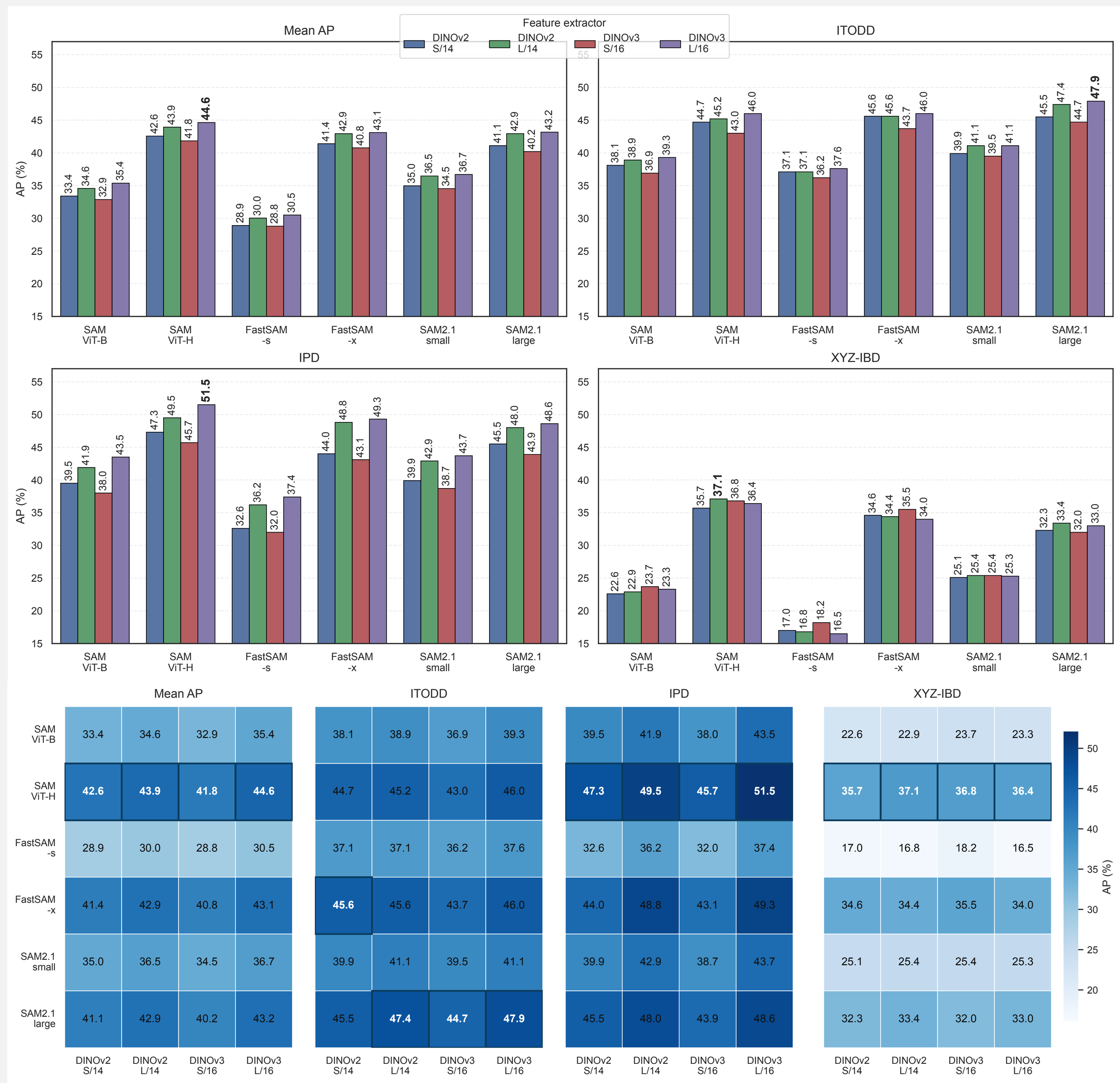

Figure 4. Accuracy comparison of segmentation models and feature extractors for the proposed DPM-VFM pipeline across three BOP datasets.

Summary

1) The best configuration (SAM v1 + DINOv3) achieves a mean AP of **44.6**, with significant high AP score on IPD dataset (**51.5**).
2) The method detects more true positives, including small and cluttered objects, as shown in the examples.
3) DPM-VFM performs well in challenging scenarios with occlusion, low texture, and dense object arrangements.
4) Failure cases mainly arise from inaccurate proposals (e.g., merged or partial objects) and visually similar distractors.