

# VisionChain: Taming Foundational Vision Models for use in the Real World

George H. W. Pearce<sup>1,2</sup>

## Abstract

Since LLMs came into the spotlight via ChatGPT, there has been an explosion of tooling which make it possible to deploy specialist NLP applications without the need to train a model for a specific problem. This has greatly increased the accessibility of developing advanced Deep Learning powered applications. Despite the tools and technology being available, the same has not occurred for Computer Vision. VisionChain seeks to close this gap.

## Keywords

Vector Database, Computer Vision, Object Detection, Grounding Dino, Zero-Shot, Heuristics

## 1. Problem Statement

Training Deep Learning models for custom problems is resource and time intensive. The typical process involves many sequential steps of labelling and training. The developers of Foundational, Zero-Shot Models such as Grounding Dino have sought to solve this problem via generalisation, however, as of 2023, without significant adaptation, most are not sufficiently accurate to solve real business problems out of the box.

## 2. Proposed Solution

A simple, user-friendly API to make developing and evaluating Computer Vision systems on top of Foundational Computer Vision Models extremely simple and fast. This can be achieved via a combination of retrieval-augmented detection and post-processing heuristics.

## 3. Demo

The demo will involve the real-time improvement of a foundational Object Detection model on the realistic industrial dataset zerowaste (<https://ai.bu.edu/zerowaste/>). The demo will begin with the raw output of GroundingDino. Performance will then be improved by the retrieval-augmented detection approach, and finally by the addition of a few post-processing heuristics.

---

HITLAML '23: *Human-in-the-Loop Applied Machine Learning*, September 04–06, 2023, Belval, Luxembourg


✉ [georgehwp26@gmail.com](mailto:georgehwp26@gmail.com) (G. H. W. . Pearce)

🌐 <https://www.linkedin.com/in/georgepearce/> (G. H. W. . Pearce)

🆔 0000-0002-0877-7063 (G. H. W. . Pearce)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)