

Product recognition AI

Mariam Sayed Ahmad
Mahmoud Nasser

Problem statement

Manual Tagging

- Manual Product tagging is a time consuming task, it involves:
 - Searching for each product label
 - Searching for similar products
 -

Automatic Tagging

- Streamline business needs
- More efficient

🏠 Home



ALL PRODUCTS



Search by name or Product ID...

Summary

Sales

Orders

Content



Posts



All posts

Galleries



Curate



Facebook

Instagram

acumencollection

acumencollection

acumencollection

acumencollection

Upload



Products



All Products

Product Feeds



Reviews



Proposed Solution

Object Detection

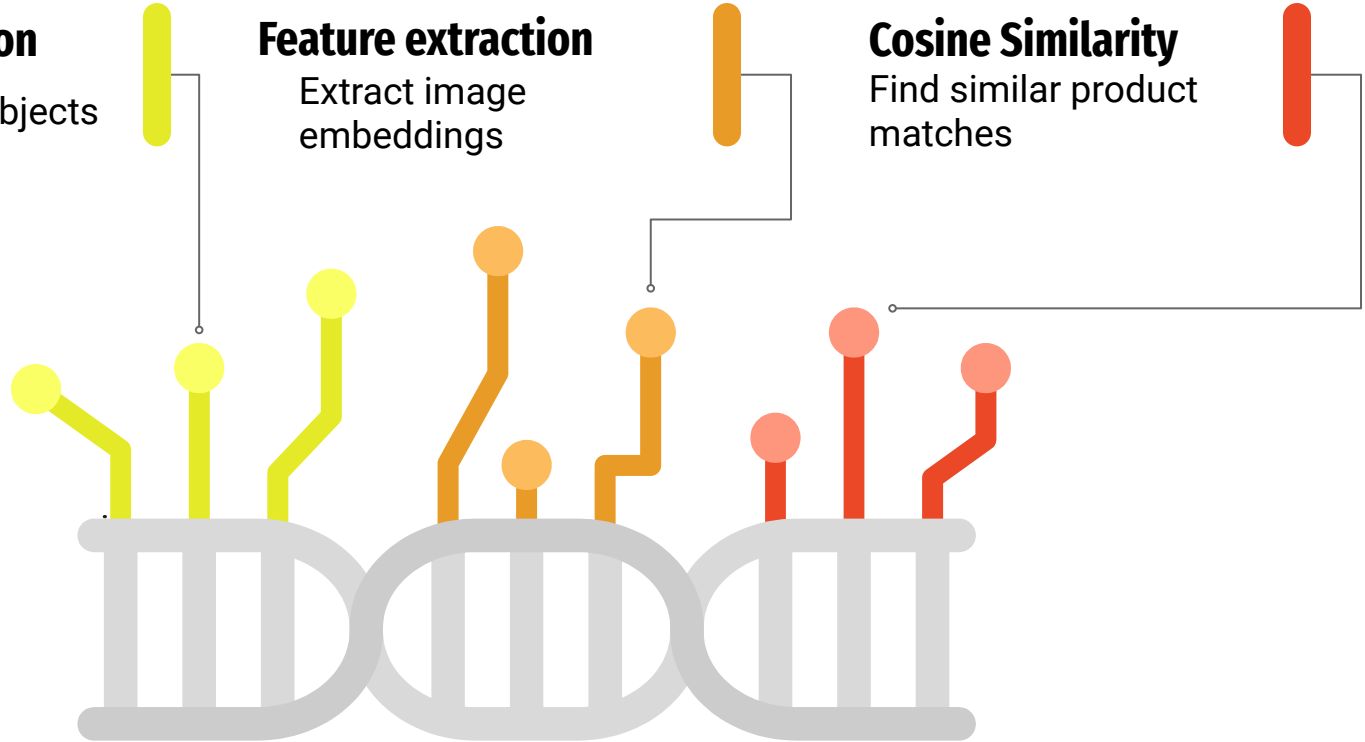
Detect desired objects

Feature extraction

Extract image embeddings

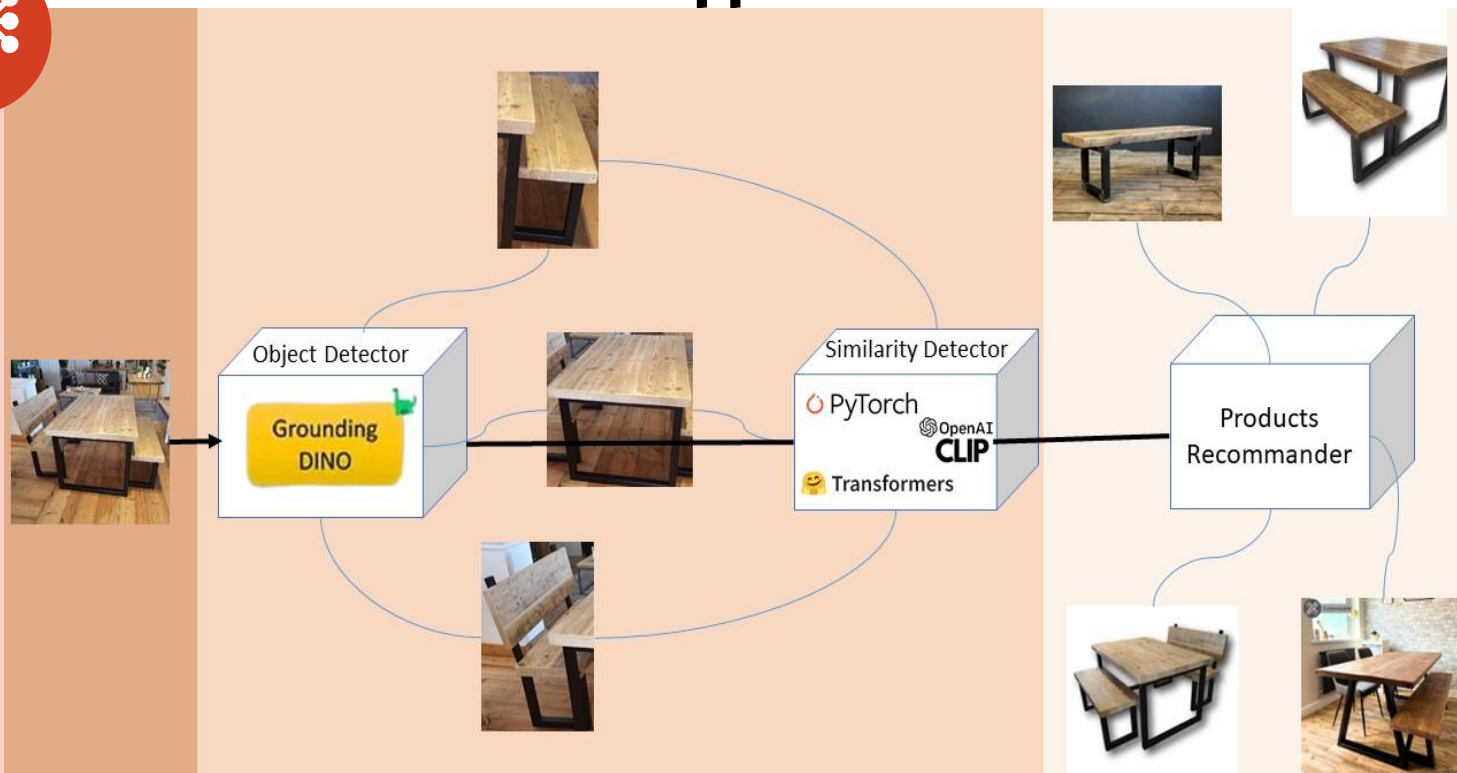
Cosine Similarity

Find similar product matches

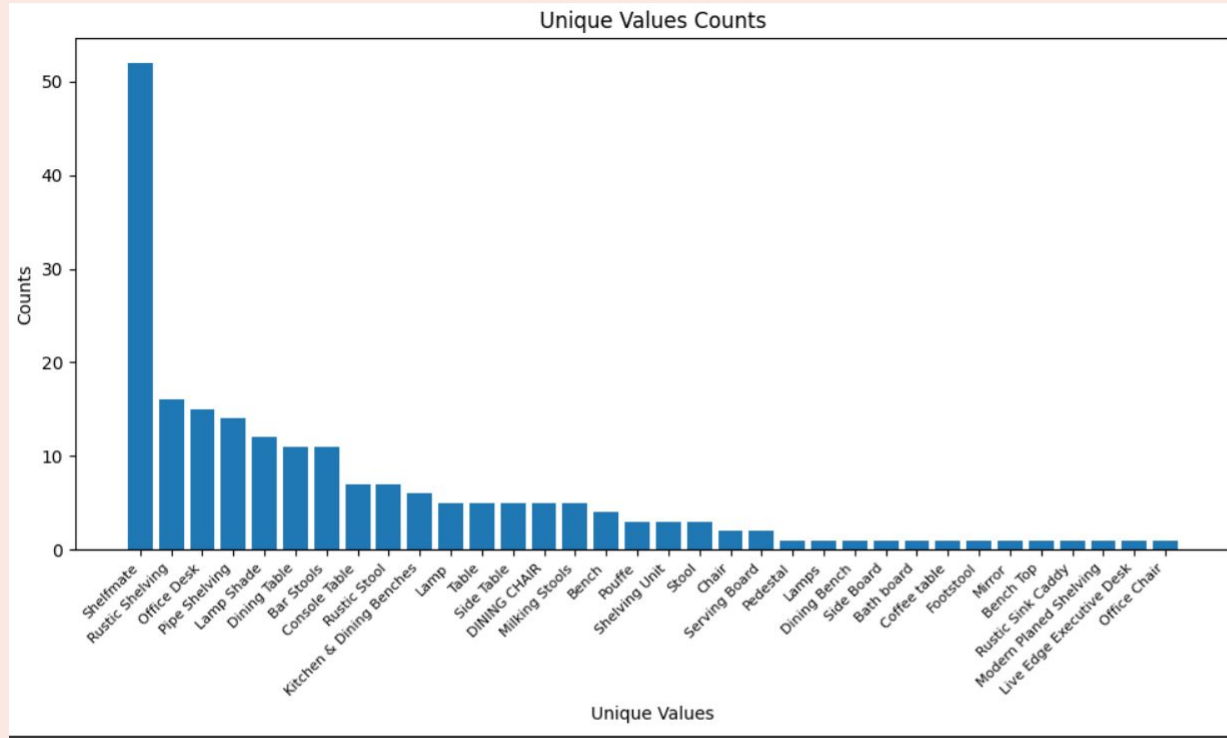




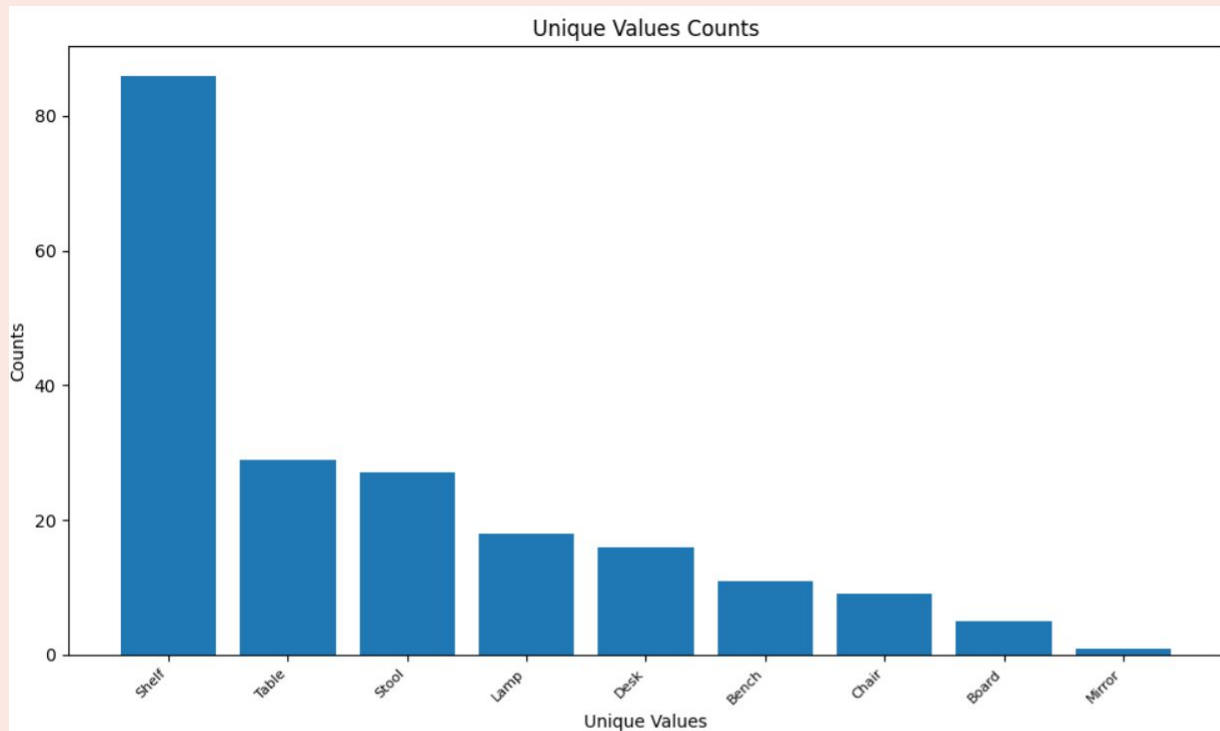
Our Approach



Data Before Cleaning



Data After Cleaning





1

Object Detection

What's grounding DINO

Grounding Dino



Zero-Shot Object Detection

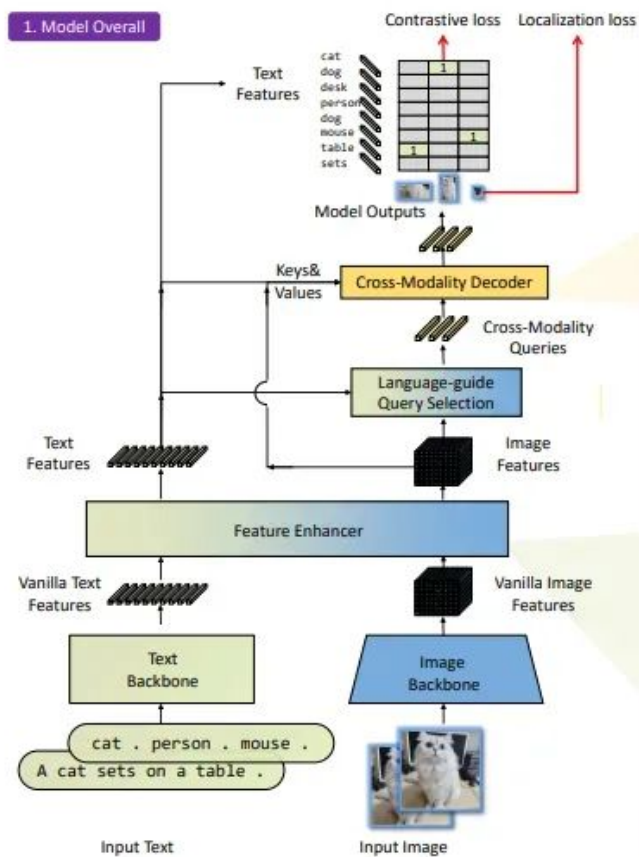
detecting objects even when they are not part of the predefined set of classes in the training data

Referring Expression Comprehension (REC)

Identifying and localizing a specific object or region within an image is based on a given textual description

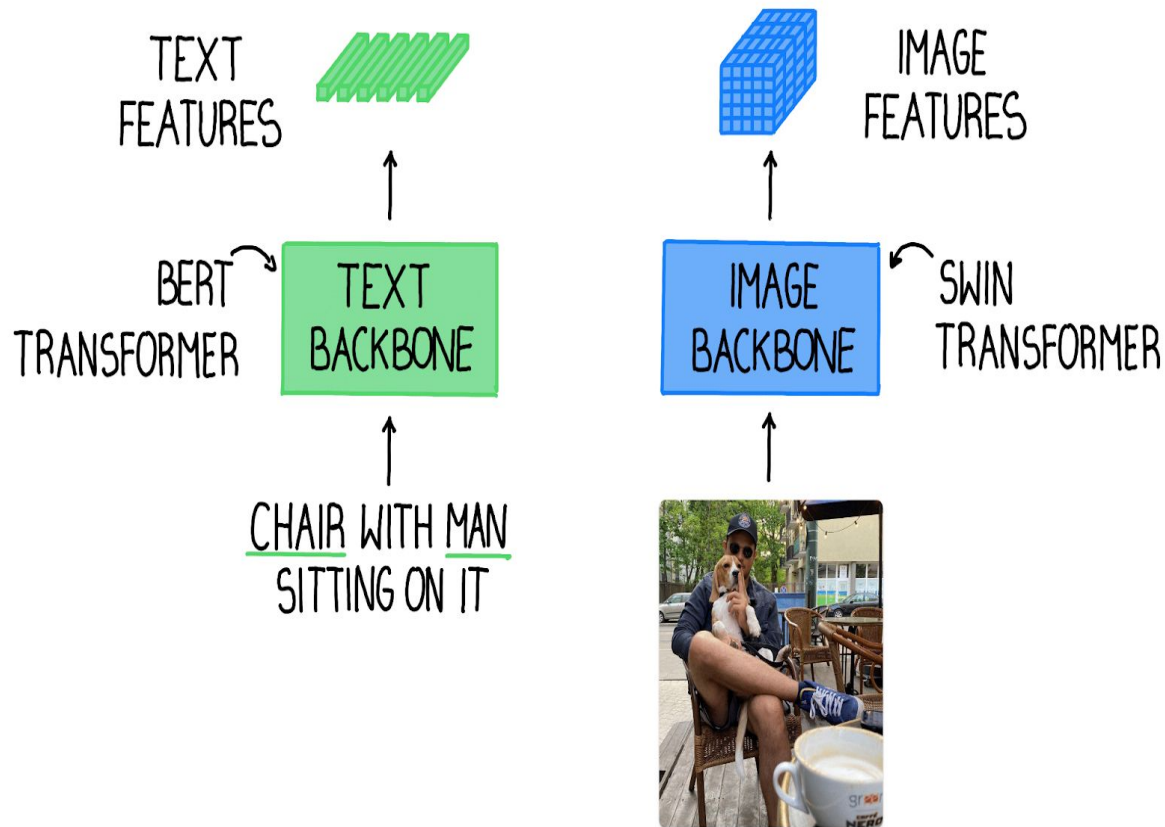
Elimination of Hand-Designed Components

simplifies the object detection pipeline by removing the need for hand-designed components

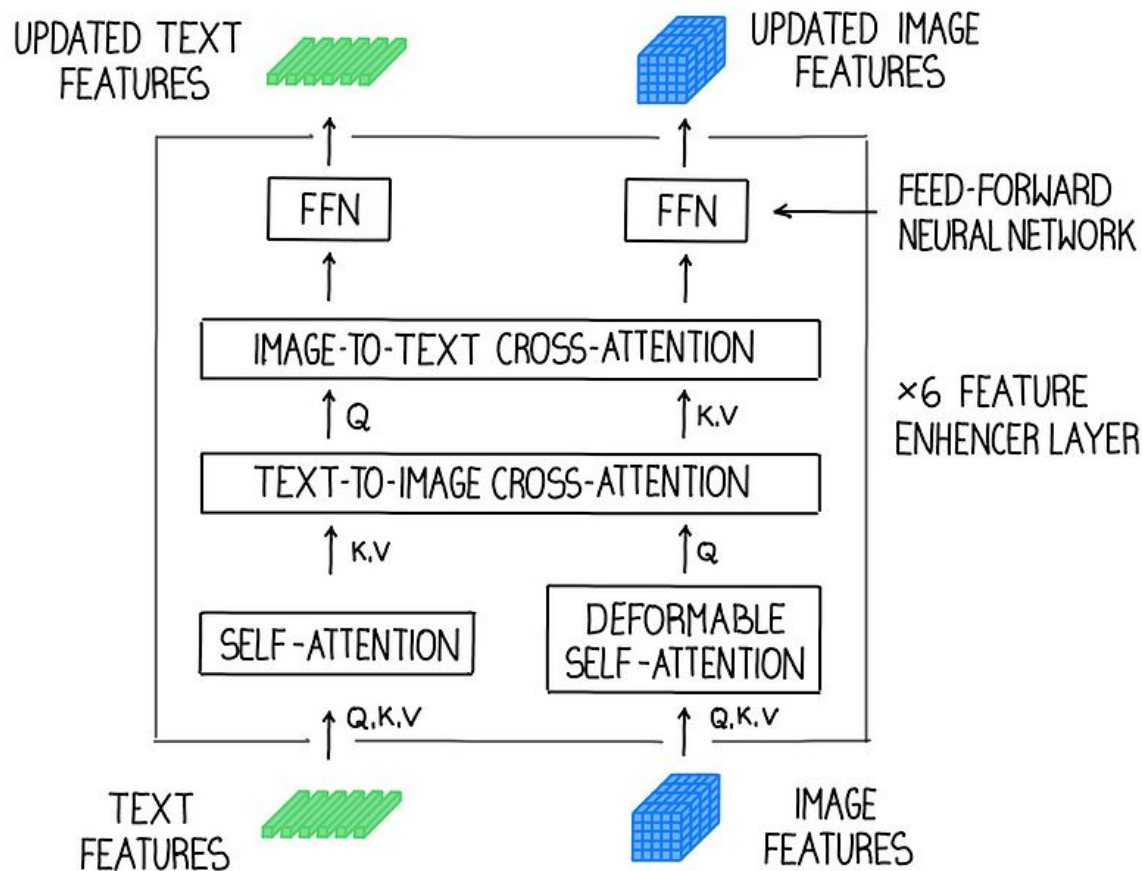




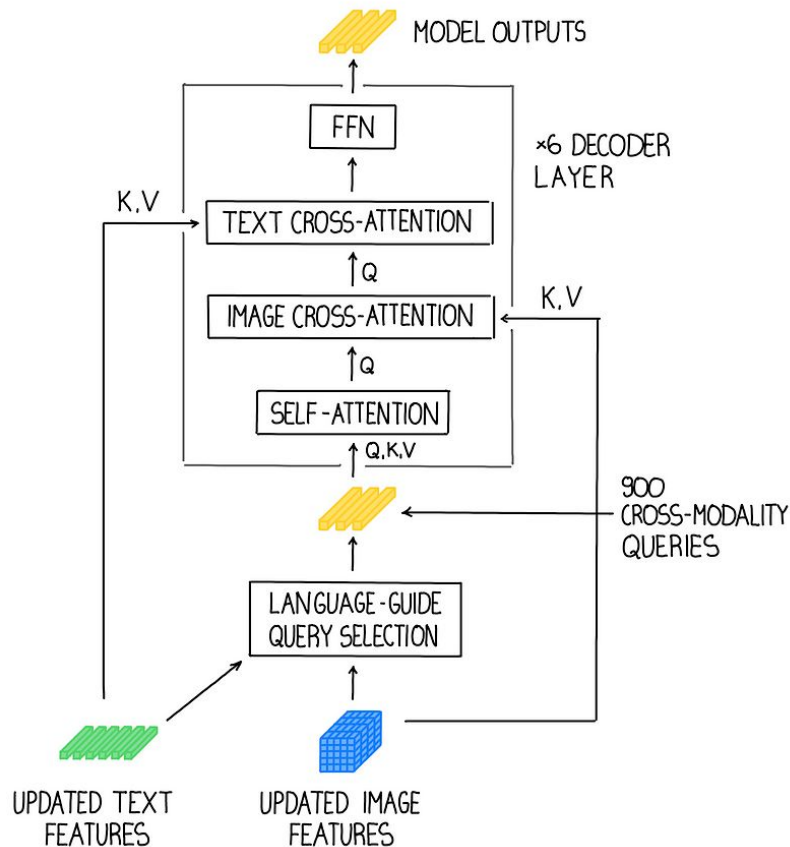
Grounding dino Text and Image Backbone



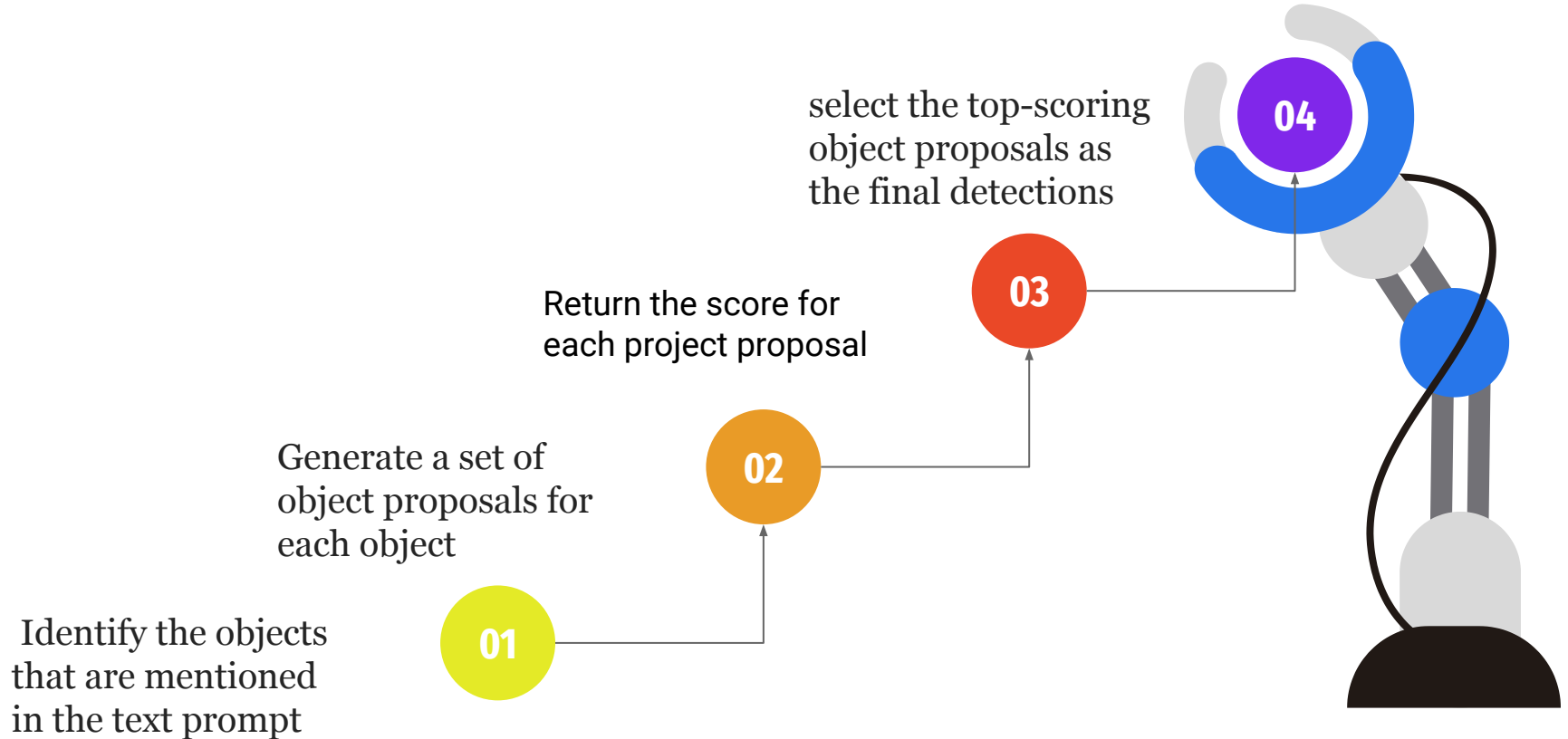
Grounding dino Feature enhancer



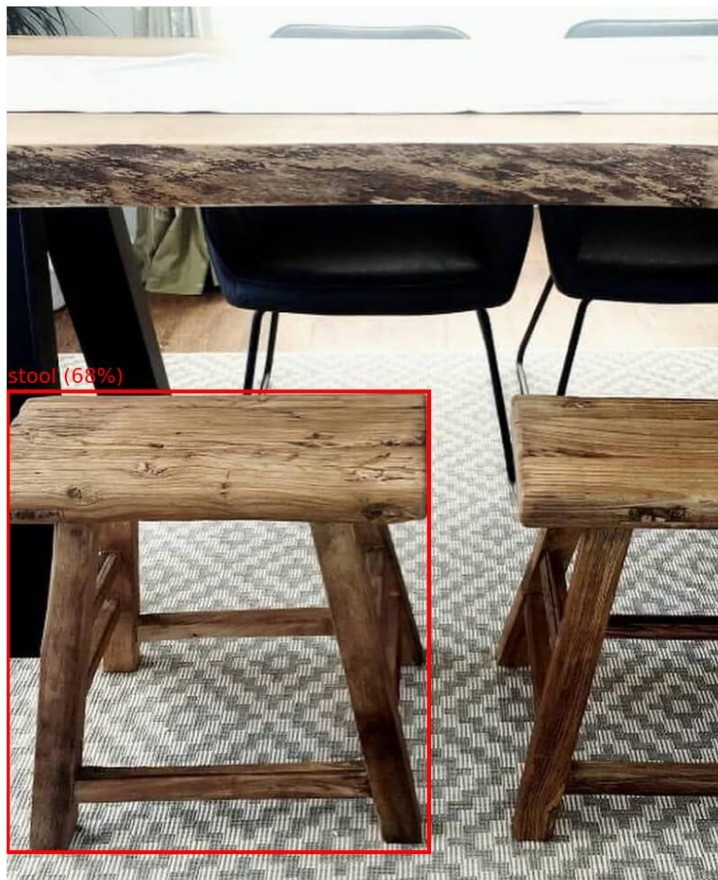
Grounding dino language-guided query selection and cross-modality decoder



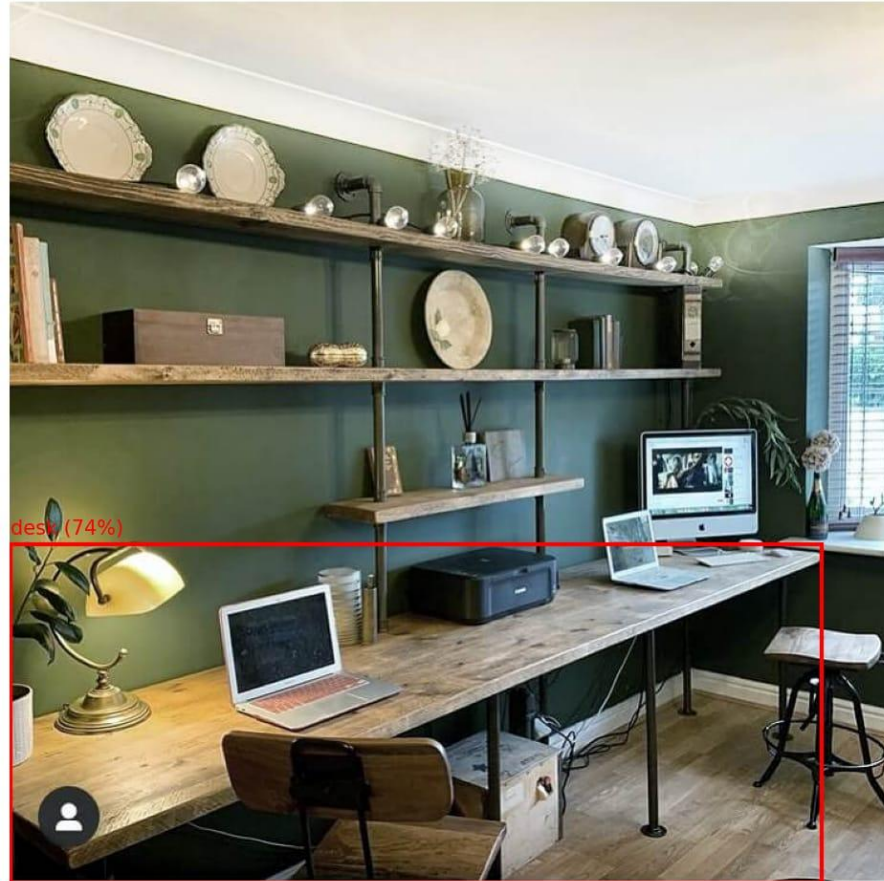
How it works



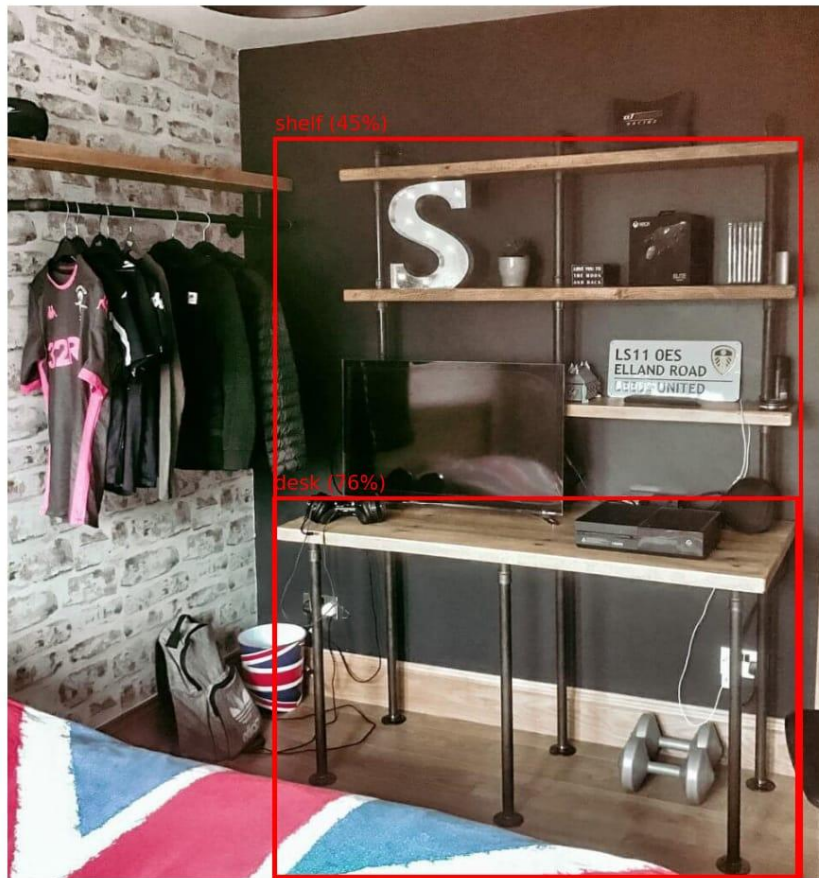
Example 1



Example 2



Example 3





2

Feature extraction

What's Clip(Contrastive Language–Image Pre-training)

01

Clip is a neural network

built on hundreds of millions of images and captions,

02

can return the best caption given an image

03

CLIP is a bridge between CV and NLP

Clip image embeddings

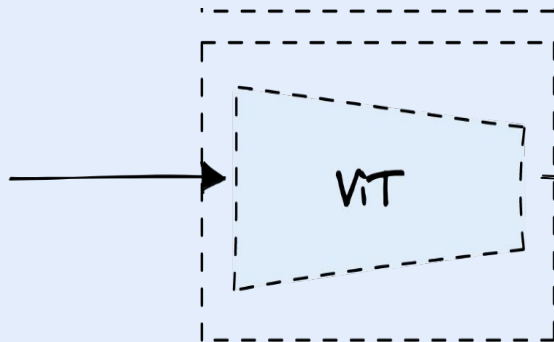
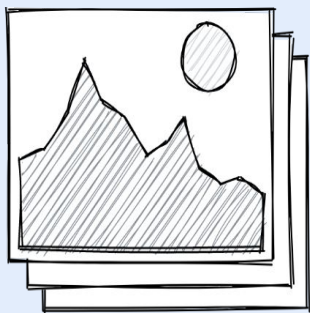
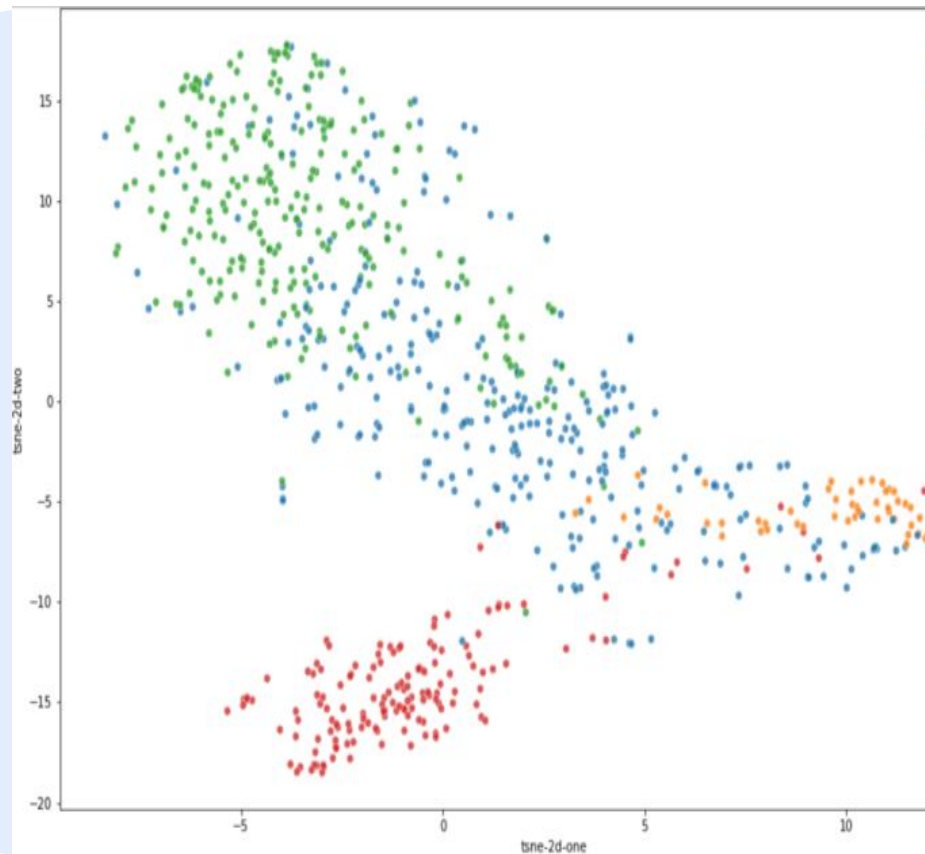


Image encoder

[1.21, 0.81, ...],
[-.31, 2.75, ...],
[2.04, -.11, ...]

Image Embeddings

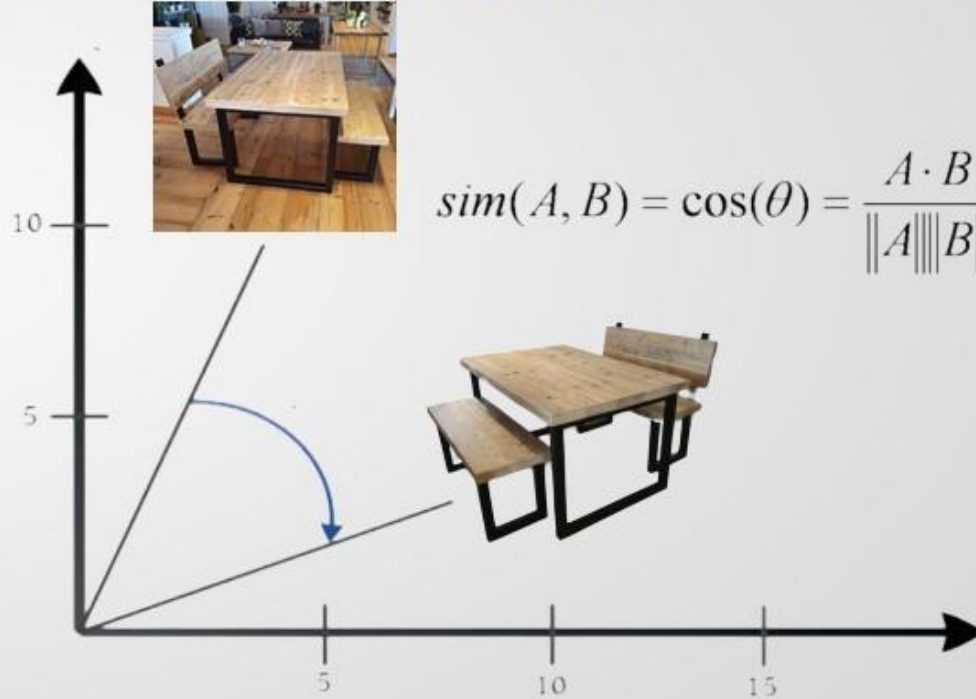




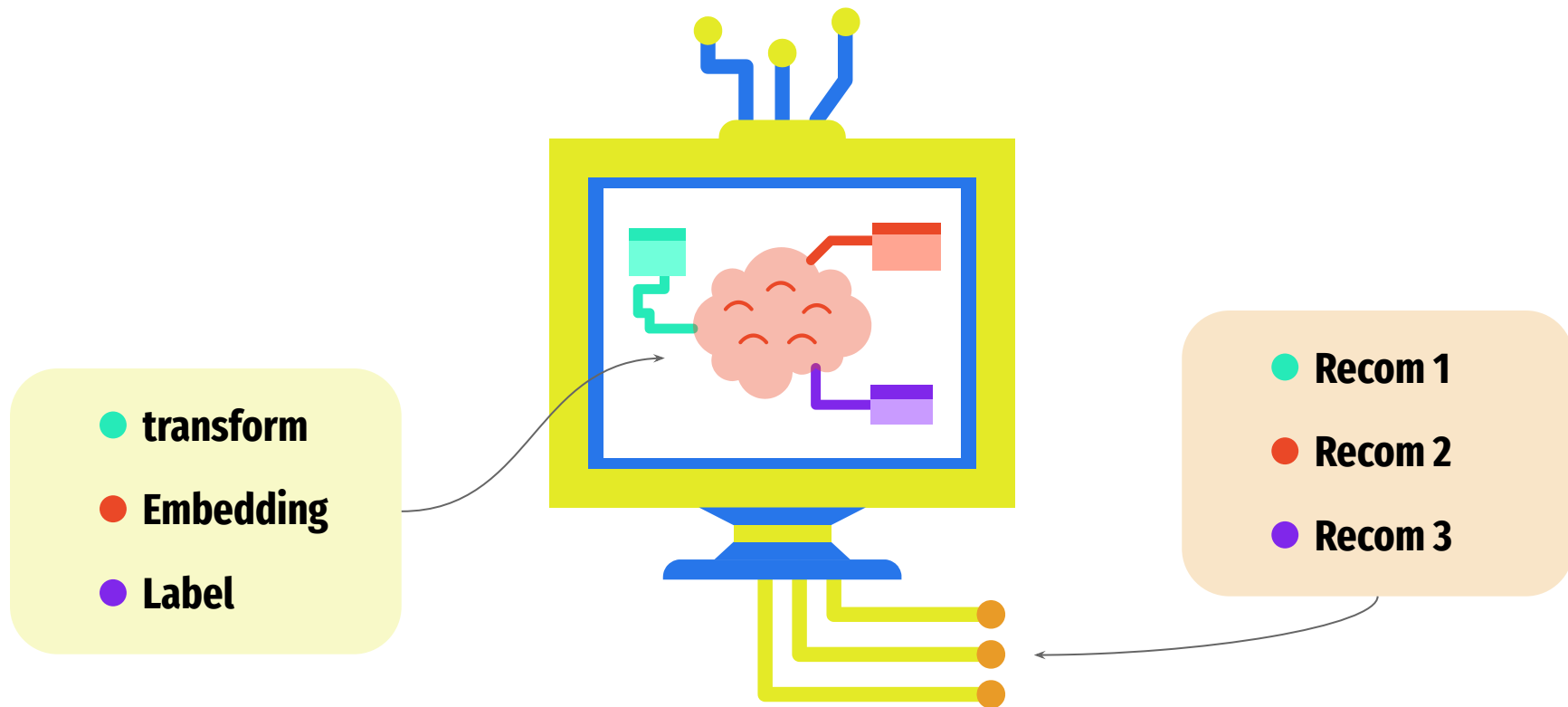
3

Cosine Similarity

What's cosine similarity



Similarity pipeline



Example



Similarity: 79%
ID: 4632199233591



Similarity: 77%
ID: 6622128472119



Similarity: 75%
ID: 4290573697079



Similarity: 75%
ID: 4465363648567



Example



Similarity: 79%
ID: 6624599736375



Similarity: 79%
ID: 6871472537655



Similarity: 77%
ID: 6820334108727



Similarity: 77%
ID: 3528439038016



Example



Input Image table (table)

Similarity: 87%
ID: 6624603111479



Similarity: 84%
ID: 4632992907319



Similarity: 84%
ID: 4453882724407



Similarity: 83%
ID: 4467025641527



Example



Input Image table (table)

Similarity: 87%
ID: 6624603111479



Similarity: 84%
ID: 4632992907319



Similarity: 84%
ID: 4453882724407



Similarity: 83%
ID: 4467025641527



Deployment



Future Work



Our roadmap includes:

1. Expanding the model to diverse product categories like clothing.
2. fine-tuning the grounding DINO approach through semi-supervised learning, effectively bridging labeled and unlabeled data.
3. Enhance evaluation metrics by introducing a new dimension for measuring similarity, ensuring a more comprehensive assessment.

**THANK
YOU!**



Q&A