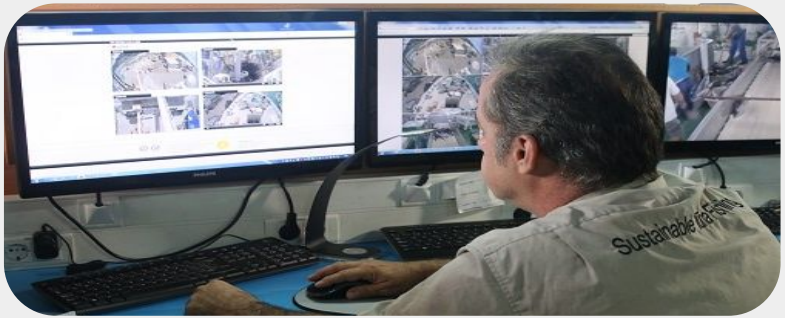
Open-domain Vision with RAG for Ocean Conservation

Sepand Dyanatkar, Angran Li, Alexander Dungate

Tackling Climate Change with AI Workshop 2024

3 billion people depend on healthy oceans

For Example



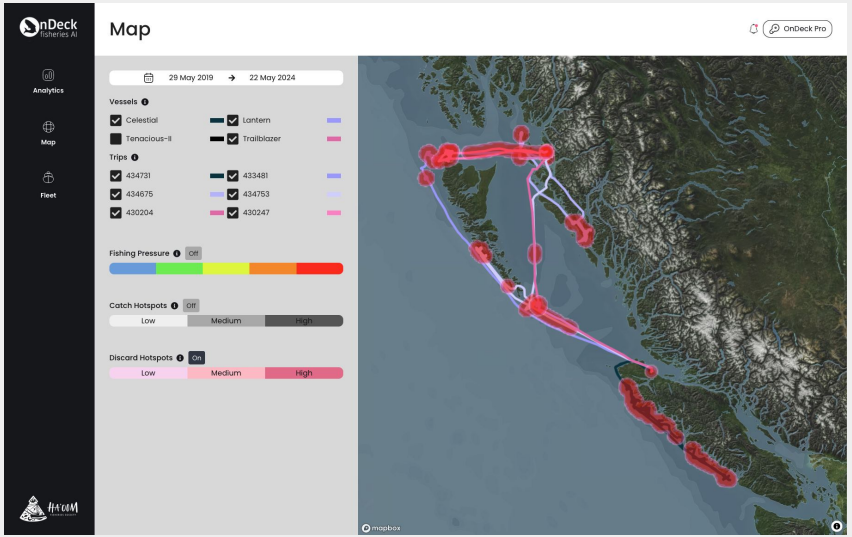
Review In Fisheries

- Crushingly Expensive
- Too slow for timely fishing data
- Cannot scale

© OnDeck Fisheries AI Inc. 2024

Solved!

[1]



[1] J. Kay and M. Merrifield, "**The Fishnet Open Images Database**" Jun. 16, 2021, doi: 10.48550/arXiv.2106.09178.

Well... Not Quite

Limitations of Top-down Computer Vision

Generalization

- Traditional models struggle with dynamic and diverse environments, even within the same domain.

Long-Tailed Distributions

- Difficulty in identifying rare or unseen species due to imbalanced data. Or entirely missing data.

Limitations of Top-down Computer Vision

Domain Transfer

- Models require retraining or tuning for new environments or conditions.

Decision Provenance

- Explainability is possible only with reverse engineering and not a first-class feature.

Limitations of Top-down Computer Vision

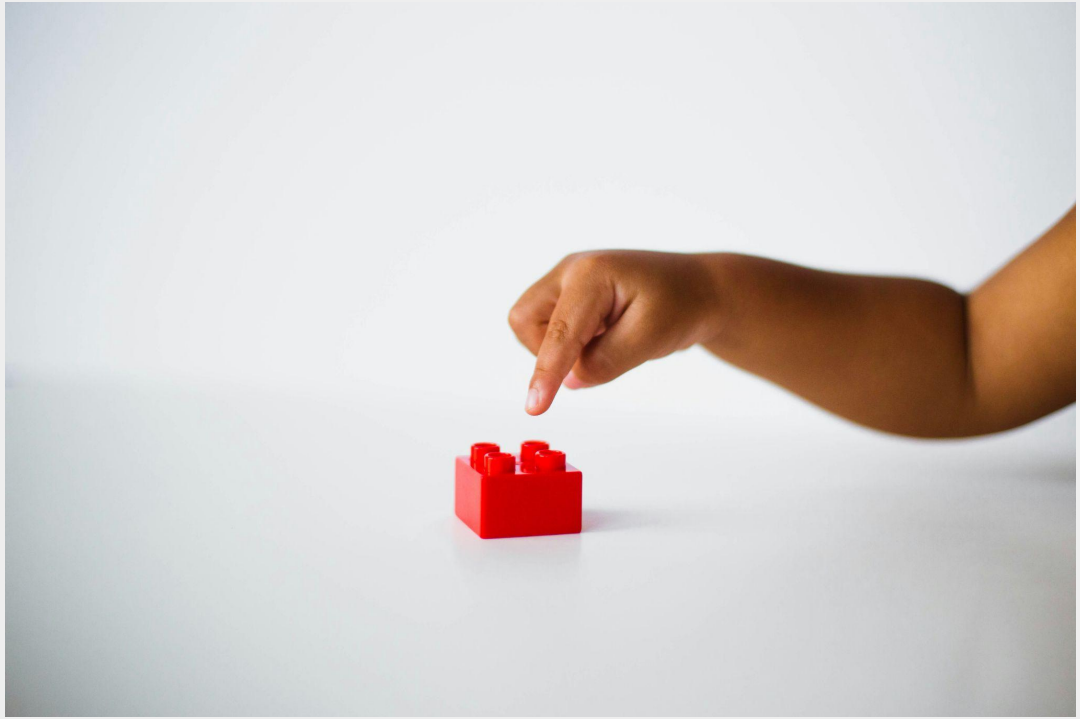
Generalization

Long-Tailed Distributions

Domain Transfer

Decision Provenance

Bottom-Up Learning



Vision Language Models

Self-supervised Pre-training

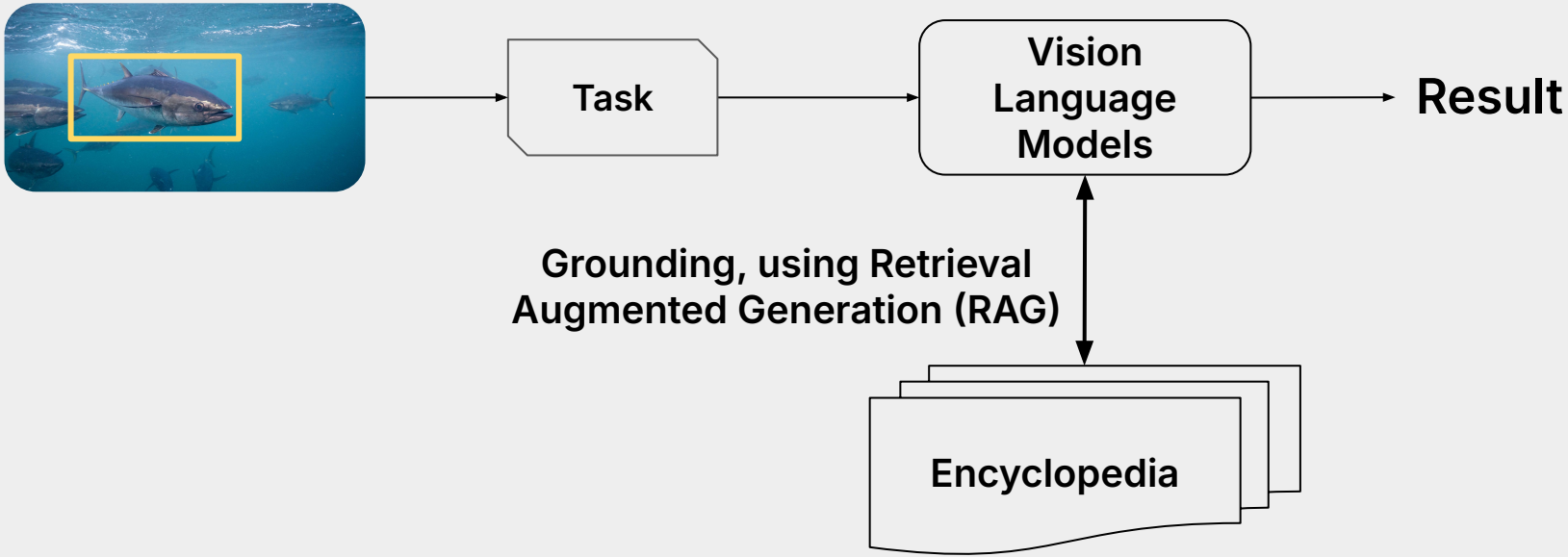
+

Contrastive Learning

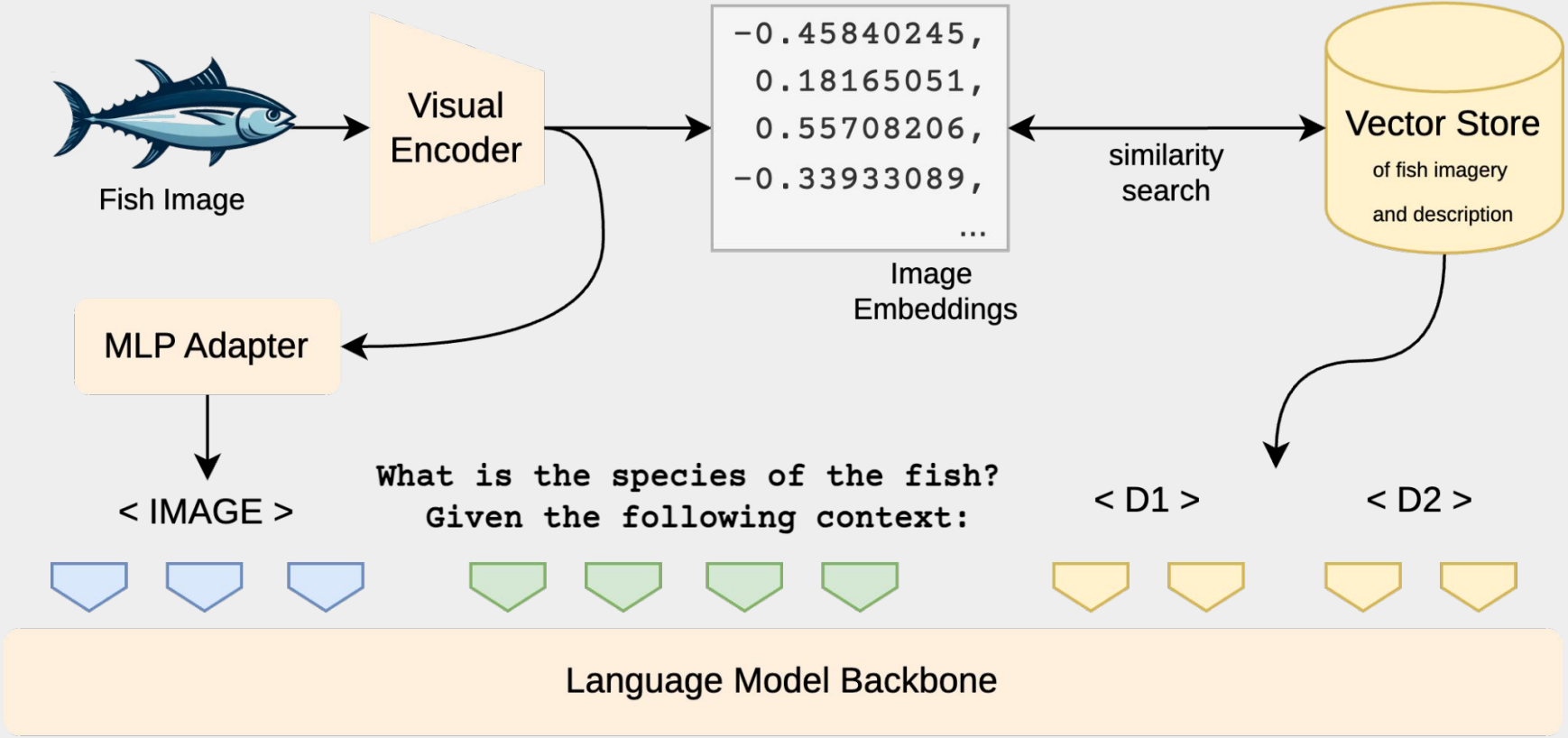
+

Diverse and Varied Training Data

Method


Task
Vision Language Models
Result
Grounding, using Retrieval Augmented Generation (RAG)
Encyclopedia

Architecture

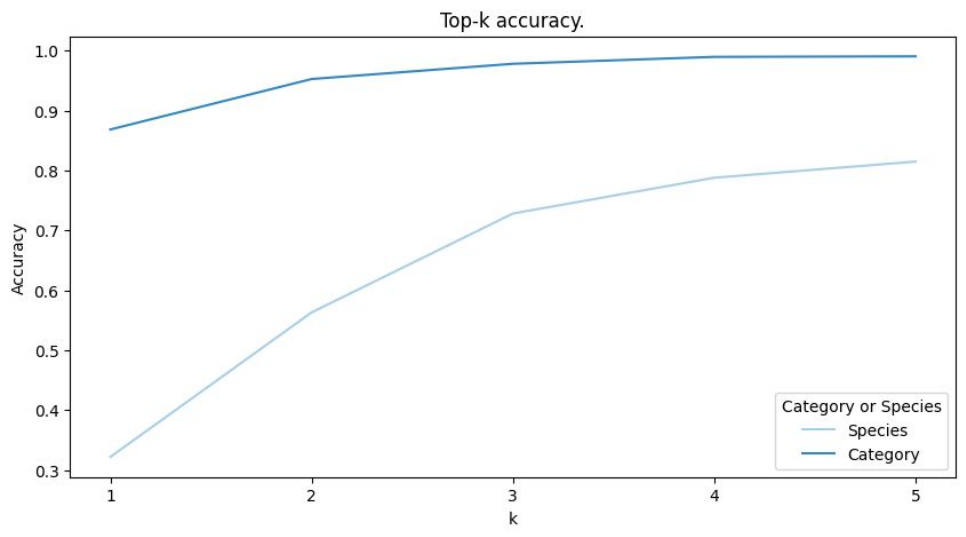


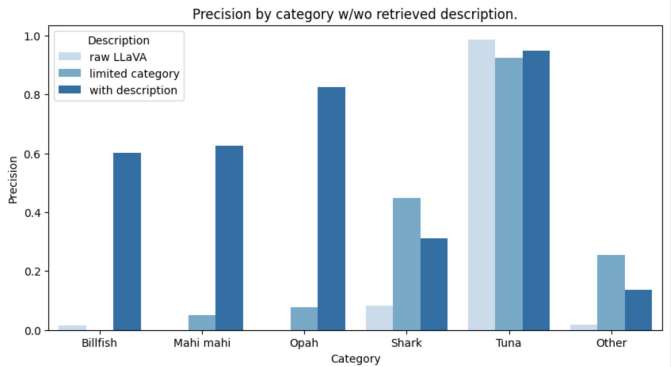
Preliminary Results

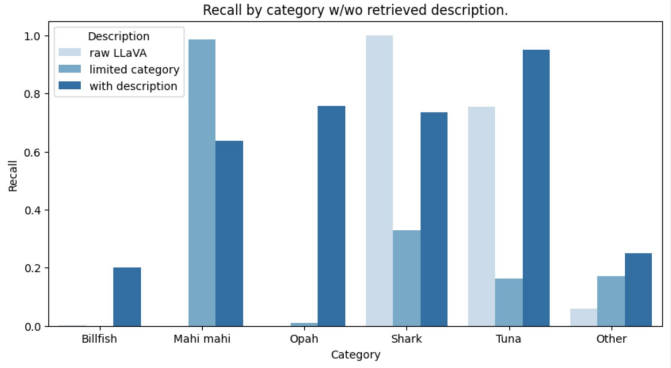
Table 1: Classification accuracy of baseline vs. our VLM-RAG approach on 5 categories. We measure both performance of final prediction (single answer response) and intermediate RAG retrieval.

	Accuracy		
Method	Top-1	Top-2	Top-3
InceptionV3 (Baseline)	0.7501	0.8312	0.9408
VLM-RAG (Ours, Final Prediction)	0.8403	N/A (single answer)	N/A (single answer)
VLM-RAG (Ours, RAG Retrieval)	**0.8684**	**0.9527**	**0.9781**

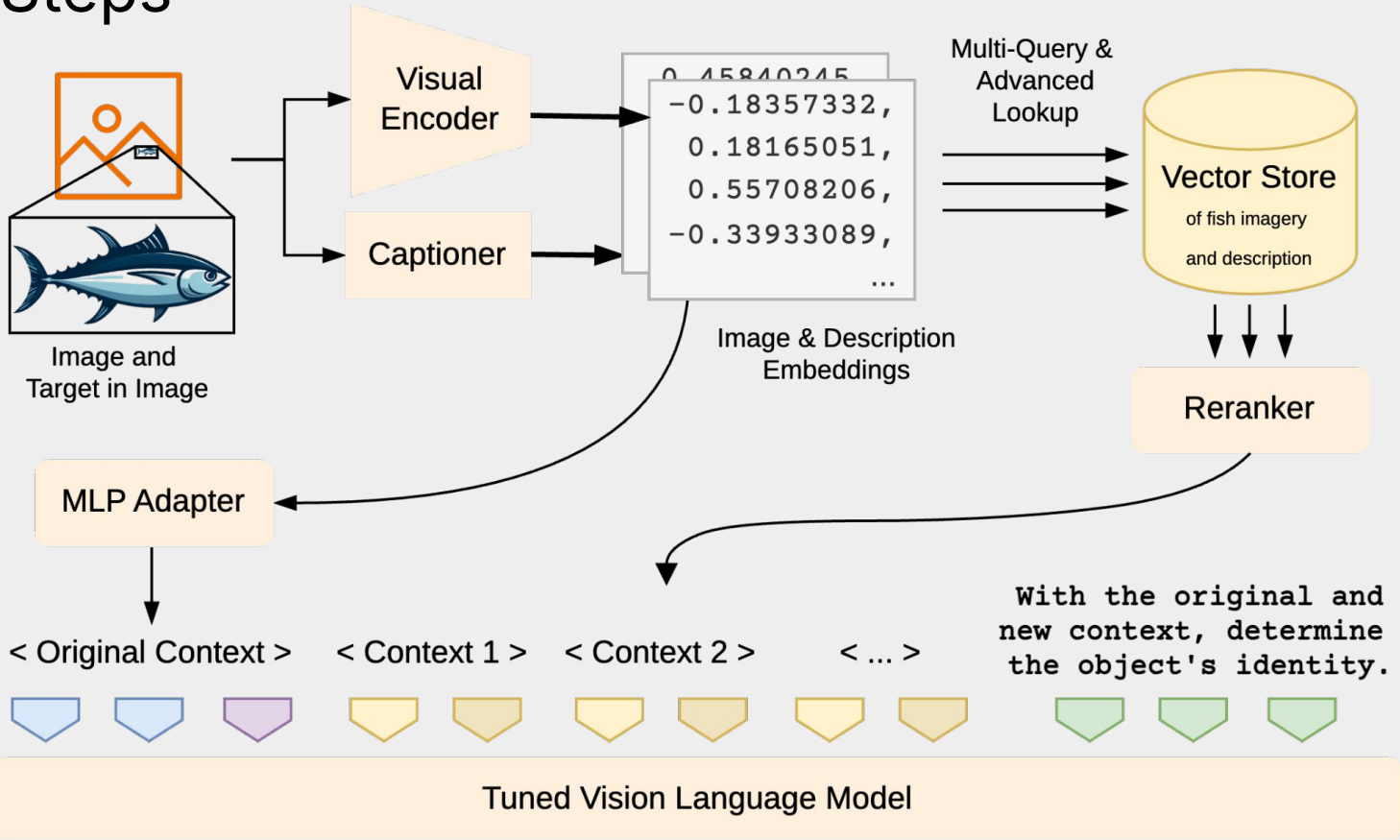
Preliminary Results







Next Steps


Visual Encoder
Captioner
-0.18357332,
0.18165051,
0.55708206,
-0.33933089,
...
Multi-Query & Advanced Lookup
Vector Store
of fish imagery
and description
Image and Target in Image
Image & Description Embeddings
Reranker
MLP Adapter
< Original Context >
< Context 1 >
< Context 2 >
< ... >
With the original and new context, determine the object's identity.
Tuned Vision Language Model

Impact and Conclusion

Accurate, faster and more accessible deployments of marine life monitoring.

=

Significantly more informed and effective responses to changing climates.

© OnDeck Fisheries AI Inc. 2024

OnDeck fisheries AI

Paper

Our Work

CANADA'S OCEAN SUPERCLUSTER

NATIONAL GEOGRAPHIC

Environmental Defense Fund

OCEAN IMPACT ORGANISATION

SOA SUSTAINABLE OCEAN ALLIANCE

CREATIVE DESTRUCTION LAB

Institute for the Oceans and Fisheries

OCEAN WISE