Vector Databases in Computer Vision

How vector databases are changing the field

CONTENTS

zilliz

Speaker

Filip Haltmayer

Software Engineer

filip@zilliz.com
linkedin.com/in/filiphaltmayer

zilliz

Zilliz at a Glance

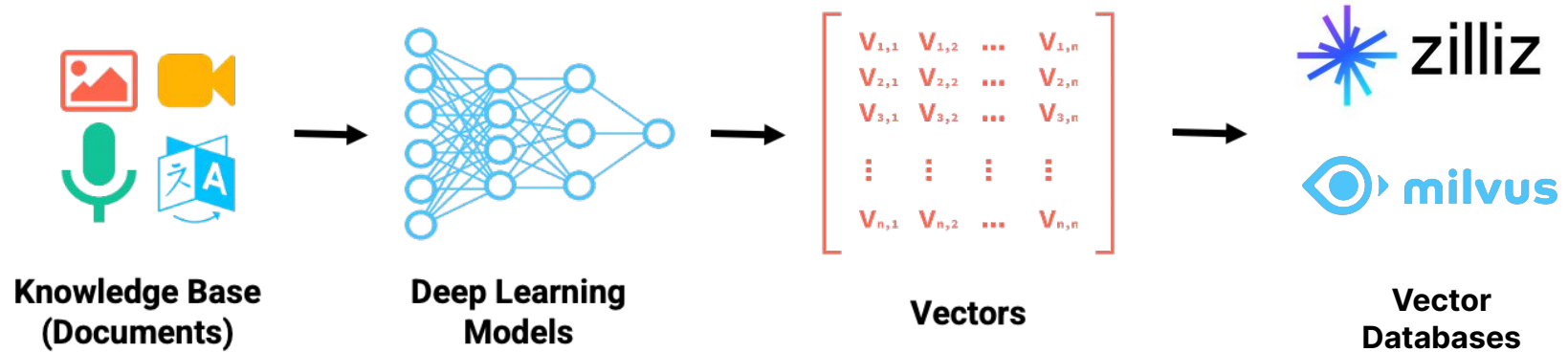|	
Founded	2017
Headquarters	Redwood Shores, CA
Focus	Vector database company for enterprise-grade AI built on Milvus, the popular open-source vector database that helps organizations quickly create AI applications.

Key maintainer of the following Open-Source projects

milvus   Attu   towhee   GPT-Cache   VDB Benchmark   Knowhere

zilliz

01
What is a Vector Database?

zilliz

Working with Unstructured Data

Vector databases are purpose-built to store, index and query vector embeddings from unstructured data.

Knowledge Base (Documents) → **Deep Learning Models** →

$$\begin{bmatrix} V_{1,1} & V_{1,2} & \dots & V_{1,n} \\ V_{2,1} & V_{2,2} & \dots & V_{2,n} \\ V_{3,1} & V_{3,2} & \dots & V_{3,n} \\ \vdots & \vdots & \vdots & \vdots \\ V_{n,1} & V_{n,2} & \dots & V_{n,n} \end{bmatrix}$$

Vectors → **zilliz** / **milvus** **Vector Databases**

zilliz

Purpose-built vector databases

Capabilities

- Advanced filtering (filtered vector search, chained filters)
- Durability (any write in a db is durable, a library typically only supports snapshotting)
- Replication / High Availability
- Sharding
- Aggregations or faceted search
- Backups
- Lifecycle management (CRUD, Batch delete, dropping whole indexes, reindexing)
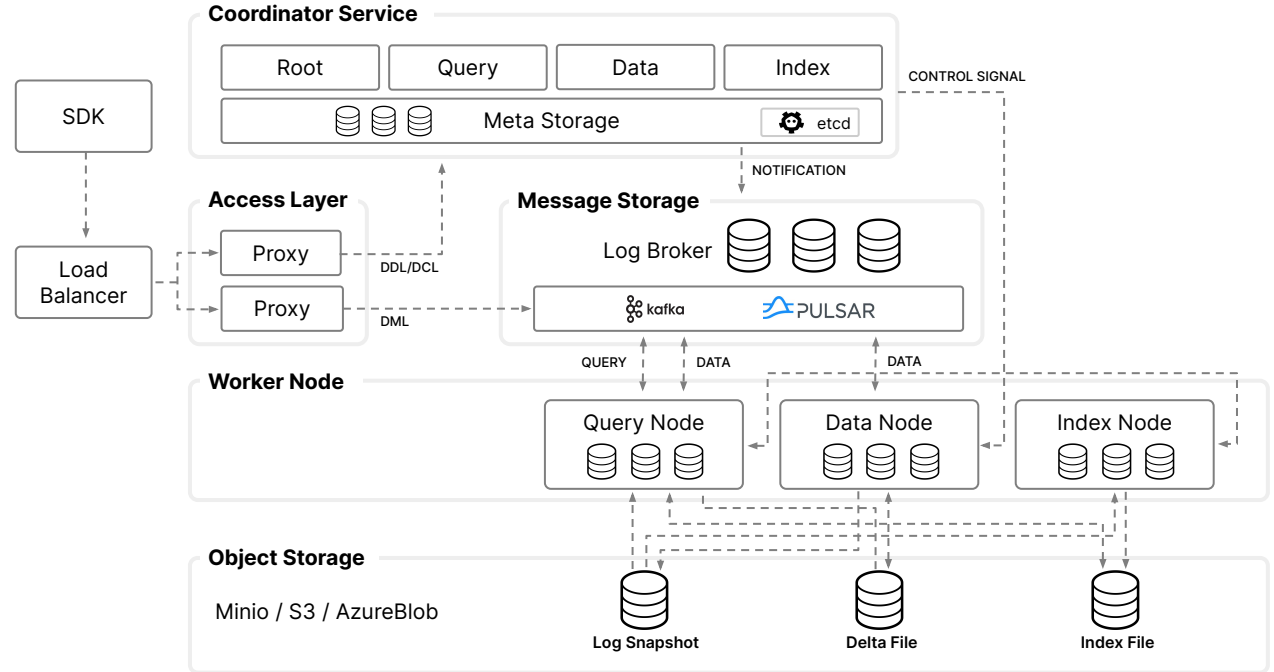- Multi-tenancy

zilliz

Why Milvus?
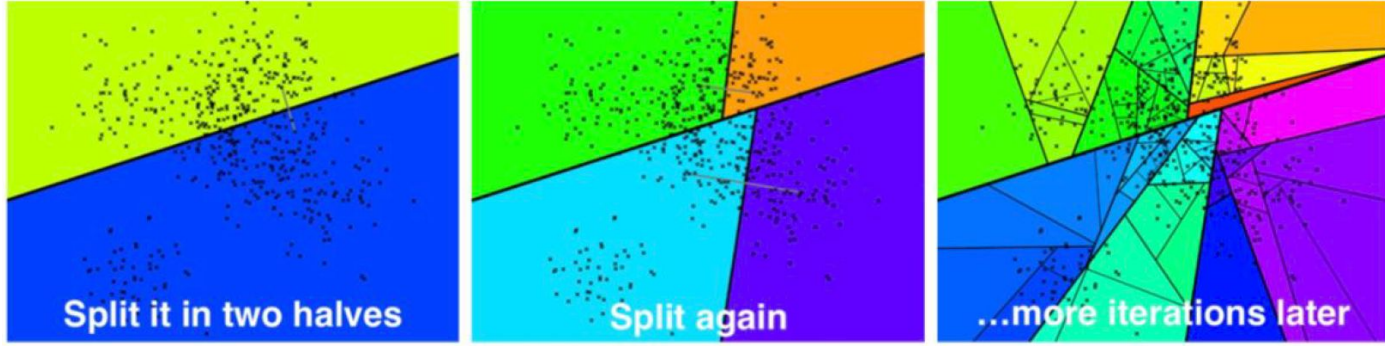
Adaptable to different use cases

- High query load vs. High insertion/deletion
- Full precision/recall
- Accelerator support (GPU)
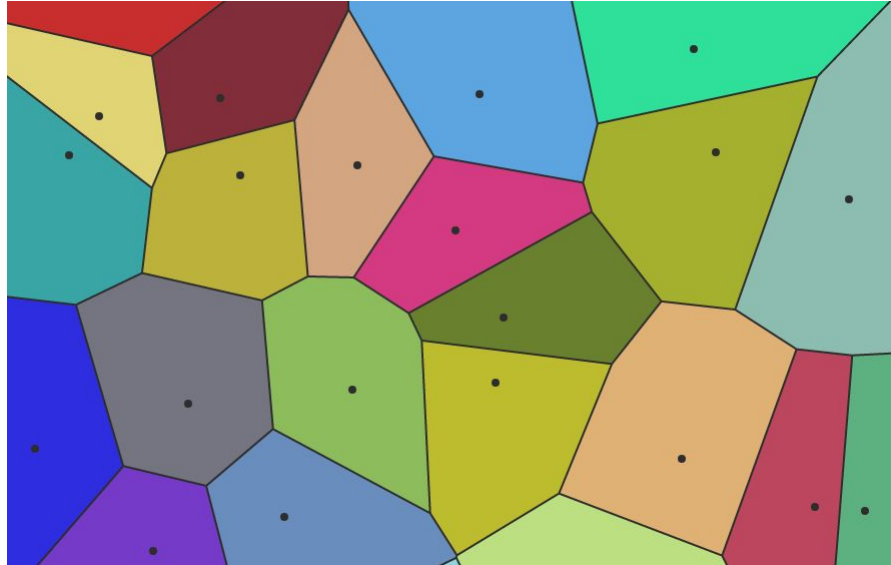- Billion-scale storage

Vector search library

- High-performance vector search
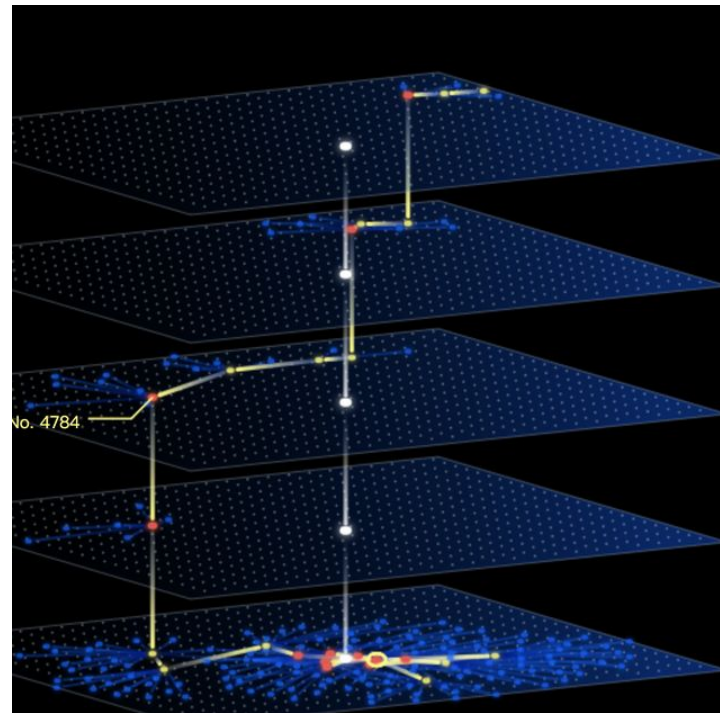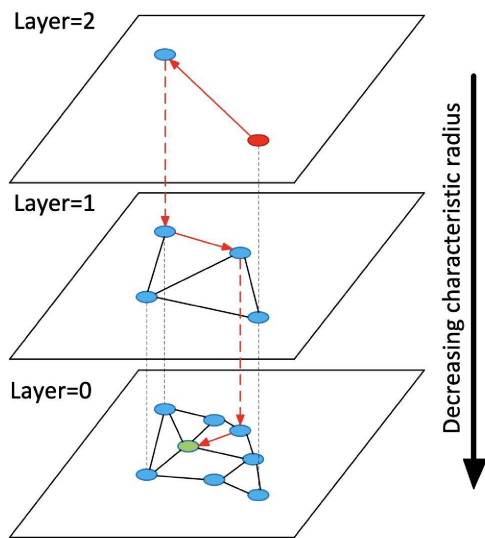- Open sourced and modular

Approximate Nearest Neighbors Oh Yeah

Split it in two halves

Split again

...more iterations later

|

zilliz

Inverted File Index

Hierarchical Navigable Small Worlds (HNSW)

02

Reverse Image Search

zilliz

How Does It Work?

Segmentation (Optional)
- Extract areas of focus using segmentation model

Embedding
- Convert segments into comparable format (embedding)

Search
- Compare embedding across all stored embeddings to find closest matches

zilliz

Segmentation

None
- Good for when matching full images
- Image might carry too much data making it difficult to find good matches

Generic
- Over extracts data
- How can you know which is relevant?

Domain Specific
- Only extracts details that are necessary for search
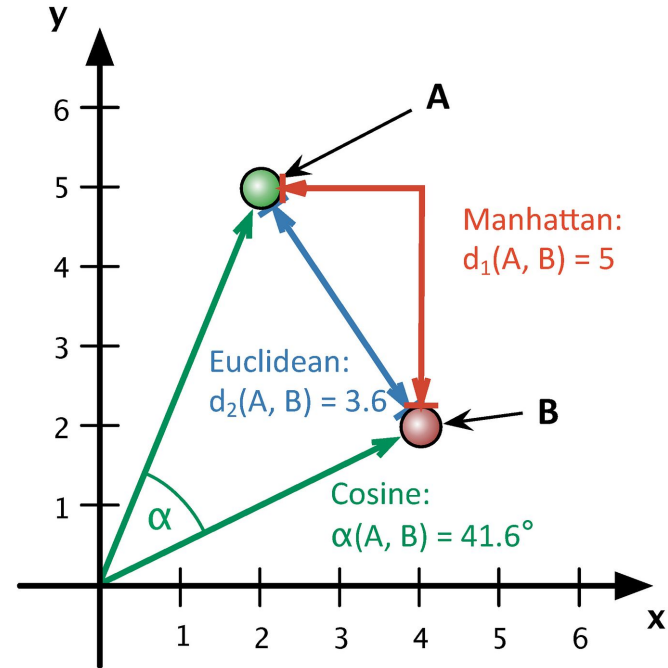- Requires training custom model

Embedding

- **Embedding falls under the same trap as segmentation**

- **Overall image vs certain details?**
  - Clothing search vs scene search
  - Generic model might embed all brown shoes and brown hats as "brown blobs"
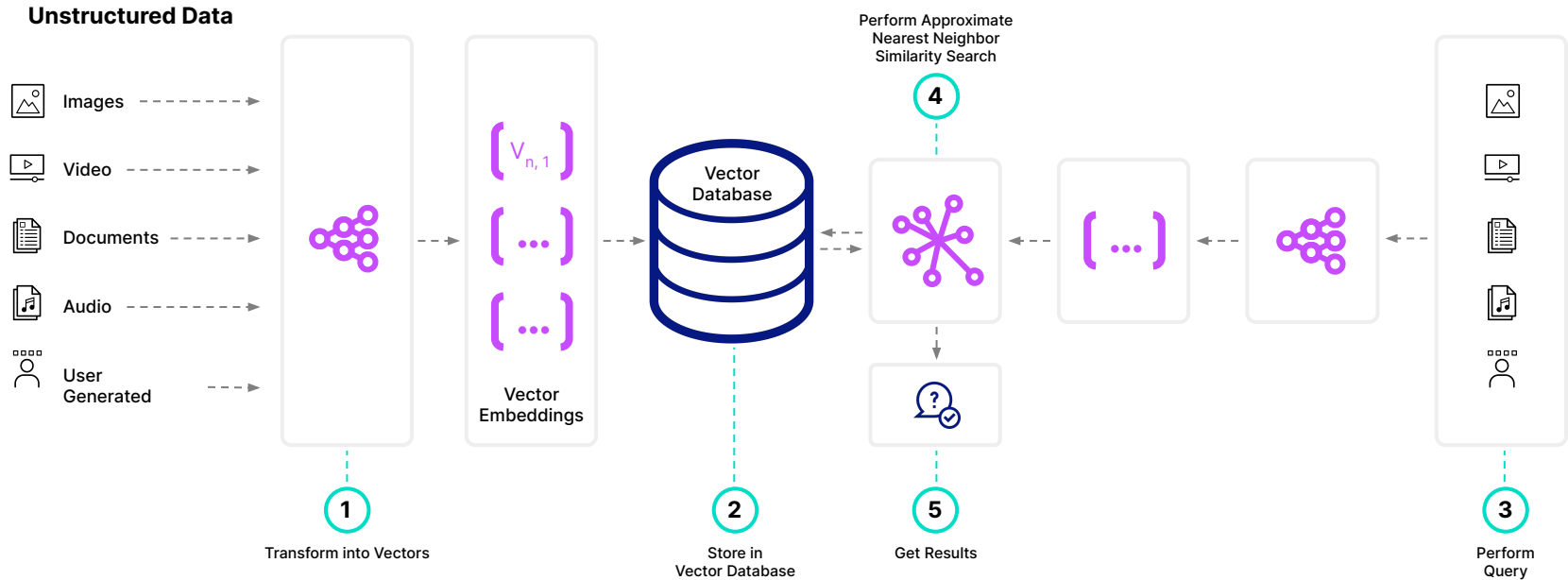

Generic

Domain Specific
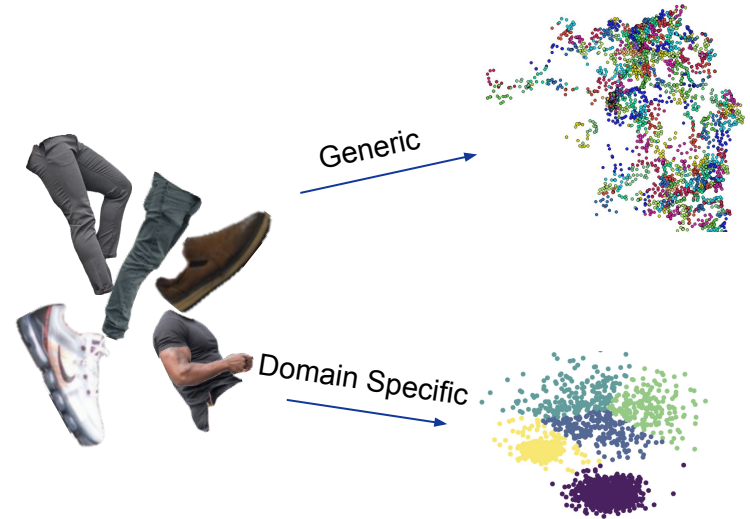
Search

- Vector search is data agnostic

- Performance vs Recall vs Price
  - Pick 2

- Different distance metrics for different results
  - IP (inner product)
  - L2 (euclidean)
  - Cosine (normalized IP)

Manhattan:
$d_1(A, B) = 5$

Euclidean:
$d_2(A, B) = 3.6$

Cosine:
$\alpha(A, B) = 41.6°$

How It Looks

Unstructured Data

Images

Video

Documents

Audio

User
Generated

$$\left[V_{n,\,1} \right]$$

$$\left[\ldots \right]$$

$$\left[\ldots \right]$$

Vector
Embeddings

Vector
Database

Perform Approximate
Nearest Neighbor
Similarity Search

4

$$\left[\ldots \right]$$

1

Transform into Vectors

2

Store in
Vector Database

5

Get Results

3

Perform
Query

zilliz

Computer Vision In Production

Ecommerce
- Product search
  (text→image and image→image)

Law and Finance
- Copyright search

Autonomous driving
- Decision making

zilliz

03
Reverse Video Search

Why is it hard?

Images are a "solved" vision problem

Increased processing power requirements
- Extra temporal dimension

Long term dependencies
- No mechanism for introducing "memory"

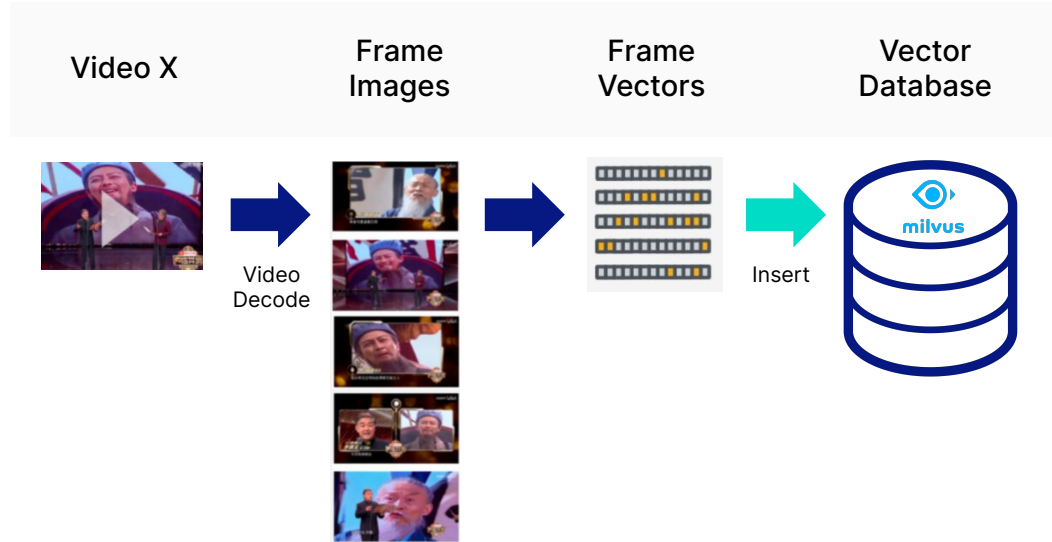Lack of meaningfully labelled data
- LLM's strength

zilliz

General Process

Similar to reverse image search

- Break down to frames
- Group frames
- Embed using video model

Popular video models

- Clip4Clip
- X3d
- VideoMAE

In Production

- Large scale video search

- Video recommendation

- Scene de-duplication

- Targeted advertising

zilliz

04

Demo

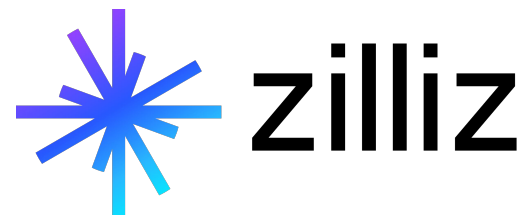# 05
Getting Started

Getting Started with Milvus

- What is Milvus https://zilliz.com/what-is-milvus

- VectorDBBench https://github.com/zilliztech/VectorDBBench

- Open Source Projects https://zilliz.com/product/open-source-vector-database

- FiftyOne Integration: https://docs.voxel51.com/integrations/milvus.html

zilliz

Start building with Milvus today!

zilliz.com

Thank You