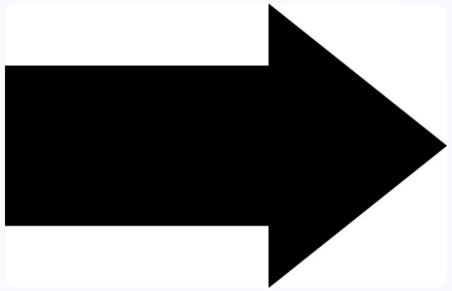
RooMatch

Represented by: Arkadi Doktorovich, Elia Meerson

Room Image Retrieval

Matching images of the same room is challenging due to visual variations. We address this using embedding based image representations and similarity in a learned feature space.

Real Image:



Generative Target Synthetic Image:

Project Goal

The goal is to learn image embeddings that enable reliable retrieval of images of the same physical room. Images are ranked by embedding similarity.

Novelty

We created a synthetic room image dataset with multiple variants per room, changing lighting, decor, and movable objects while preserving room identity. This enables systematic evaluation of exact room retrieval, a task for which no public dataset currently exists.

Recent Advances in Visual Retrieval

Recent work studies embedding-based retrieval for identifying the same physical space. Our project evaluates room-level embeddings using synthetic data and Recall@K metrics.

work/year	Task	Method	Dataset	Results	Relation to our project
Breaking the Frame: Visual Place Recognition by Overlap Prediction 2025	Visual Place Recognition (retrieval under partial overlap)	Predicts visual overlap using global embeddings extracted with ViT (DINOv2) and overlap-based scoring	Large-scale outdoor VPR benchmarks	Improves retrieval accuracy when images share only partial visual content	Uses global embedding-based retrieval as a baseline and analyzes its limitations.
AirRoom: Objects Matter in Room Re-Identification 2025	Room Re-Identification (instance-level retrieval)	global embeddings followed by object-level matching and aggregation	MPReID, HMReID, ReplicaReID	Achieves state-of-the-art room re-identification performance	focuses on retrieving images of the same physical room using embeddings
Patch-NetVLAD: Multi-Scale Fusion of Locally-Global Descriptors 2020	Visual Place Recognition / Image Retrieval	Embedding-based retrieval combining global image descriptors with local patch aggregation	Pitts250k, Tokyo24/7	Outperforms global descriptors by incorporating local information	Supports embedding-based retrieval for identifying the same place, aligning with our room matching approach

Dataset Example:

Generated Synthetic Images:

Real Image:

Room Matching Dataset

Dataset Description

- The dataset contains **50** real room images collected from various sources.
- For each real room image **10** synthetic variants were generated with **Stable Diffusion V1** to increase diversity while preserving the same room layout.
- Each image is assigned a room ID, which acts as the ground truth for evaluation.

Data Labeling

- Room attributes include lighting, rugs, curtains, wall decor, plants, and clutter.
- **BASE_PROMPT** - ensures the same room geometry is preserved while decor, lighting, and movable objects are varied.
- **NEGATIVE_PROMPT** - prevents generating entirely different rooms or altering the room layout.

Dataset Preprocessing

- Images are processed through a PyTorch Dataset class.
- Each image is resized and center cropped to 224×224 pixels.
- Images are normalized for **DINOv2** embeddings.
- The dataset provides image tensor, room ID, and index for embedding computation and evaluation.

DINOv2 Embeddings and Retrieval Evaluation

Baseline Model

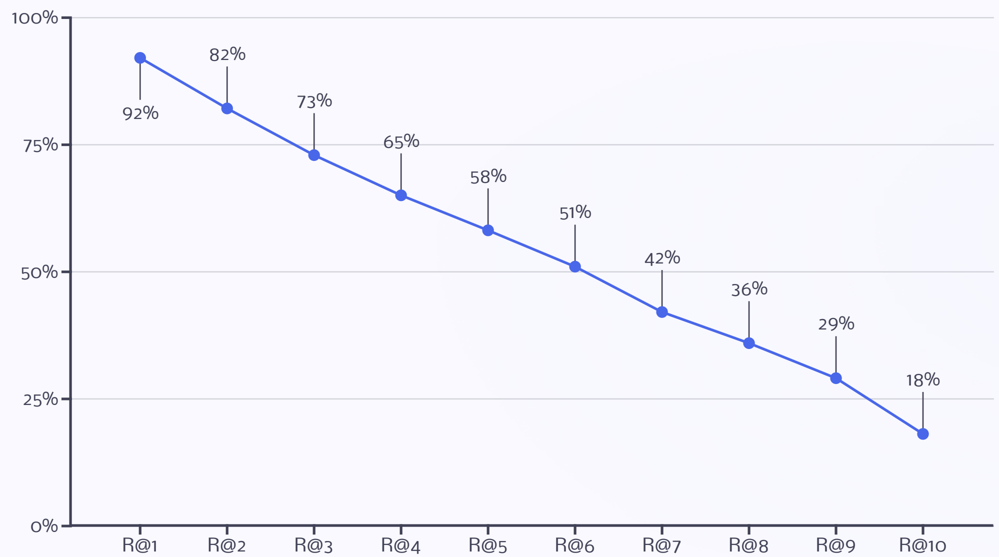
- Pretrained DINOv2 Vision Transformer used as a fixed embedding extractor
- No fine-tuning or task specific training (zero shot)
- Each image is mapped to a 768-dimensional embedding
- Embeddings are L2-normalized (equally long)
- Cosine similarity + k-nearest-neighbors used for retrieval

Error Analysis

- Prompts that are too weak result in nearly identical images.
- Prompts that are too strong may unintentionally alter the room identity.
- Even a single poorly generated synthetic image can cause Recall@K to fail.

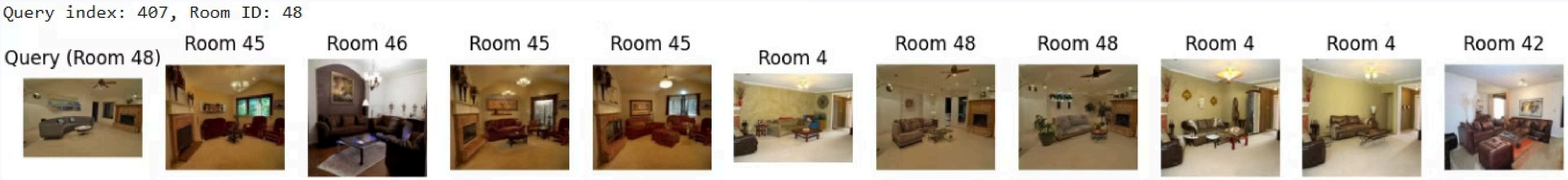
Evaluation Protocol

- **Task:** Given a query image, retrieve the most similar images
- **Ground truth:** images with the same room ID
- A query is counted as correct only if all top-K retrieved images belong to the same room

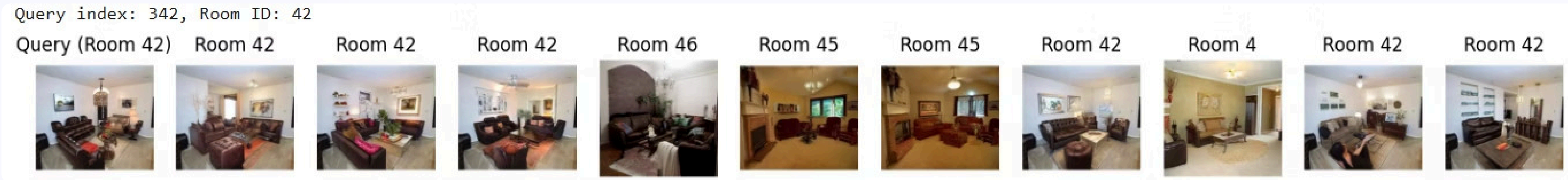


Result Examples:

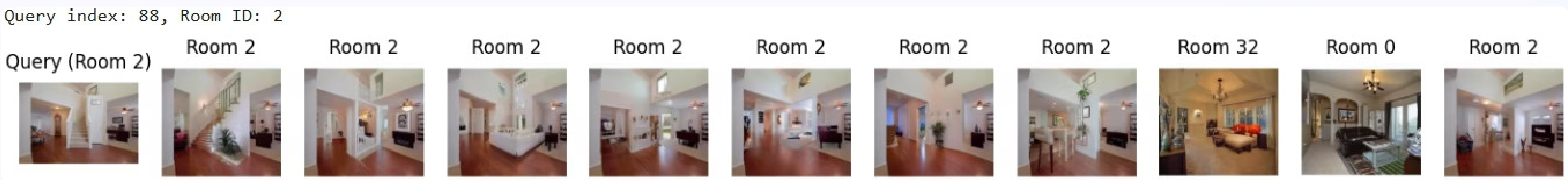
Recall@0 (failure):



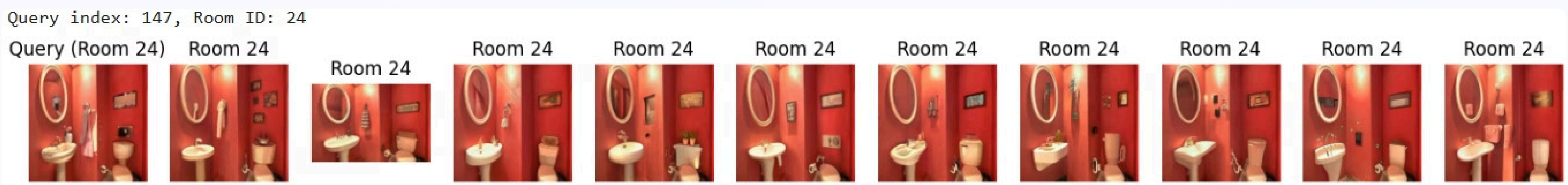
Recall@3:



Recall@7:



Recall@10:



Project Plans

Step	Description	Deadline	Expected Outcome
Prompt Optimization & Testing	Design and Test different prompt strengths and attribute combinations to balance visible changes and room identity preservation.	week 10	Stable prompts that consistently generate diverse yet identity preserving room variants.
Data Creation	Generate synthetic room images using diffusion models and assign room IDs inherited from the original images.	week 10	A balanced dataset of ~11,000 images with multiple visual variations per room.
Model Fine-Tuning	Train with room ID based positive and negative pairs so that images of the same room have closer embeddings, while different rooms are pushed apart.	week 11	Improved embedding space where images of the same room cluster tightly, leading to higher Recall@K on unseen images.
Evaluation & Analysis	Evaluate the model by measuring how well its embeddings retrieve images of the same room using Recall@K.	week 11	Gain insights into errors caused by visually similar rooms, generation artifacts, or outlier images.
Prepare Final Presentation	Summarize methodology, results, and findings. Visualize retrieval examples and evaluation metrics.	week 12	Clear and well structured presentation communicating the project's results.