

# 1. PROJECT REVIEW & DEFINITION

## The "Intent" Gap

Bridging the gap between POS sales data and actual customer browsing behavior on the shop floor.

## The CCTV Challenge

Identifying vinyl albums from low-quality, grayscale, and noisy CCTV "crops".

## Objective

Developing an Image Retrieval system to match degraded query images to a clean reference catalog.



## MODELS, DATA & METRICS

⌚ Core Architecture: ResNet50 Siamese Network.

⚡ Training Strategy: Optimized using Triplet Margin Loss for robust embeddings.

\_DISK\_ Synthetic Dataset: 5,000 images generated using GLIGEN AI-inpainting.

↖ Evaluation Metrics: Measured by Precision1 (P1) and Precision5 (P5) accuracy.

## 2. ACHIEVEMENTS & NOVELTY

### Key Achievements

- **High-Accuracy Retrieval:** Achieved 89% Top-5 and 77% Top-1 accuracy on entirely unseen album covers.
- **Robust Domain Adaptation:** Successfully mapped heavily degraded, grayscale CCTV crops to sharp, high-resolution catalog embeddings.
- **Performance vs. Complexity:** Demonstrated that ResNet50 outperforms shallow architectures in extracting features from noisy inputs.

### Core Novelty

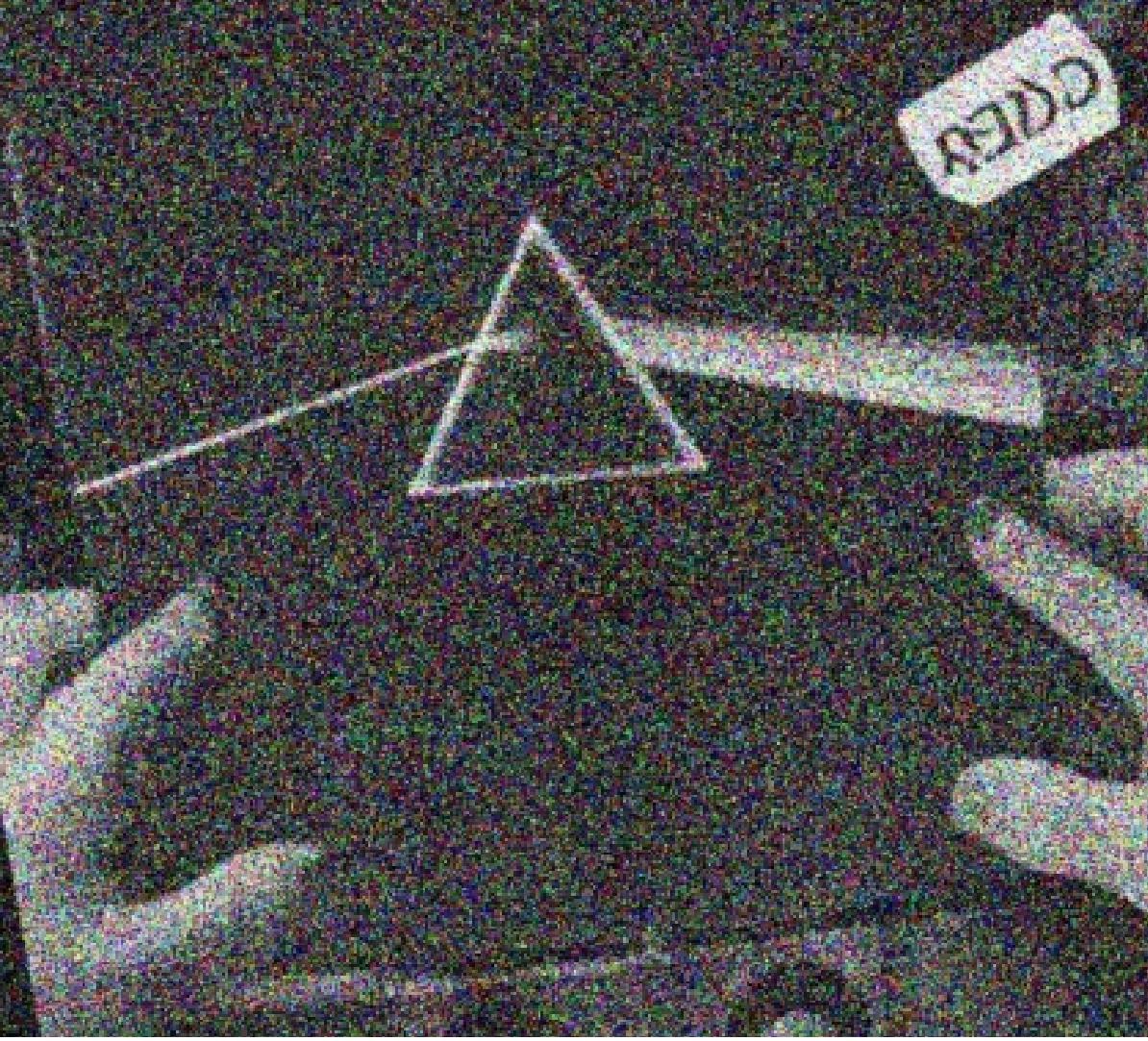
- **AI-Powered Occlusion Synthesis:** Innovative use of GLIGEN to generate realistic human-hand occlusions and price stickers, surpassing traditional methods.
- **Overcoming Data Scarcity:** Created a specialized synthetic-to-real pipeline to train models where real samples are missing.

### 3. METHODOLOGY & PIPELINE

1. Clean Catalog Image



2. Generated CCTV Query



#### The Pipeline Workflow

**Step 1:** Data Gen. 5,000 samples with GLIGEN hands + CCTV filters.

**Step 2:** ResNet50 backbone creating 128-dim embeddings.

**Step 3:** Triplet Margin Loss with hard negative mining.

**Step 4:** Generalization test on 100 images not seen in training.

**Technique:** We used **spatially-precise** inpainting to ensure the added hands realistically overlap the record geometry.

## 4. EXPERIMENTAL RESULTS

| Model / Setup       | P1 Accuracy | P5 Accuracy |
|---------------------|-------------|-------------|
| Random Guess        | 2.0%        | 10.0%       |
| ResNet18 (Baseline) | 67.0%       | 78.0%       |
| ResNet50 (Final)    | 77.0%       | 89.0%       |

### Key Insights:

- **Generalization:** 89% Top-5 accuracy on 100 unseen images.
- **Superior Extraction:** ResNet50 outperformed baseline by 11%.
- **Reliability:** Robust even under extreme noise/occlusions.



## 5. FINAL CONCLUSION

### Goals Achieved

Successfully built a system that identifies vinyl covers from messy footage with 89% top-5 accuracy, solving the intent gap.

### Key Lessons

AI-driven data generation (GLIGEN) is superior to standard augmentations for training on real-world occlusions.

### Future Work

Scaling to 10,000+ items and testing on real-time video streams for live store analytics.

