

Image Resizing Test (512 vs. 1024)

- **Motivation:** Originally, images were resized to 512×512 to reduce token usage, but the text in tables became blurry.
- **What Changed:** The experiment tested resizing images to 1024×1024 for better clarity.
- **Method:** Used the summary approach (one summary prompt per image) and asked three detailed questions about each image.
- **Findings:**
 - 1024×1024 produced noticeably sharper images and better table extraction.
 - Consistent improvement in similarity scores compared to 512×512.
 - This was tested on two separate images and showed a similar positive trend.

Hypothetical Questions Method vs. Summary Method

- **Quick Definition:** The hypothetical questions approach generates multiple question-based prompts per image (e.g., “What is the base salary for the CEO?”) instead of a single summary prompt.
- **Test Details:**
 - Two images from different documents were used, generating both a single summary (at 1024×1024) and multiple hypothetical question embeddings for each.
 - Three test questions relevant to the table content were asked.
- **Results:**
 - A DataFrame comparison showed that the hypothetical questions approach consistently yielded higher similarity scores than the summary approach for detailed queries.
 - The improvement was especially dramatic when the question asked about specific details within a single image.

Multi-Page Retrieval and Partitioning Strategy

- **Setup:** A scenario was tested where the answer spans multiple pages in a 63-page document.
 - Initially, all images—including figures (title pages, irrelevant graphics)—were included in the similarity search.
 - It was discovered that some irrelevant figures (partitioned as “image” by unstructured.io) appeared at the top of similarity rankings, which was incorrect.
- **Refinement:** Removing these “figures” eliminated the irrelevant hits, leaving only tables and text.
 - With both summary and hypothetical question embeddings in the mix, the highest similarity scores were from summary embeddings—but hypothetical questions were also near the top.
 - The pages that ultimately contained job titles and salaries ranked highly, confirming that the system surfaced the correct pages.

Notable Observations

- **512 to 1024 Image Resizing**
 - Switching from 512×512 to 1024×1024 radically improved image clarity and extraction quality.
- **Impact on Detailed Questions**
 - 1024 consistently outperformed 512 when questions required specific numbers or textual details from the table.
- **Hypothetical Question Embeddings**
 - Dramatically higher similarity scores on single-page, detail-oriented queries compared to summaries.
- **Multi-Page Queries and “Figures”**
 - When the question spanned multiple pages, “figures” sometimes dominated the top similarity scores but were irrelevant.
 - Removing these “figure” images allowed the system to surface relevant tables.
- **Top Matches in Multi-Page Context**
 - Among actual table pages, the best matches were typically summary embeddings —but hypothetical questions also ranked near the top.
 - This indicates that both approaches can help retrieve the right data across many pages, though the summary approach sometimes edged out the hypothetical approach in top rank **for high-level general questions only**.