QUALITEST™

Leading Multinational Technology Organization Improves Project Overhead by Streamlining Their Data Annotation Strategies

When Our Client required tens of thousands of images to be collected and classified, we helped them reduce their costs were reduced by 40 - 45%.

Challenges

The large and increasing dataset took the Client months to complete and there was an urgent need to reduce timelines

Highly skilled data analysts on the onsite team increased costs for the project

Solutions

Ramped up a highly skilled offshore data analysts' team in a very short span of time for data collection and tagging

Proposed an offshore – onshore hybrid model with offshore resources based in Madagascar for the data collection & tagging and a minimal onsite team operated as Quality Control

Results

The Client was able to process a thousands of images in just one month.

The cost reduced by 40% and our Client has the flexibility to ramp up and ramp down team size based on work-load increase

Client overview

Our Client is a leading multinational technology company that is primarily focused on online advertising, search engine technology, artificial intelligence, quantum computing, e-commerce and consumer electronics.

They have significant technological advantages in the field of artificial intelligence, search and online ads and is one of the world's most valuable brands. They strive to organise the world's information and make it universally accessible and useful for all.

Say cheese! The challenges of working with image recognition systems

Our Client required tens of thousands of images to be collected continuously and perform a complex image tagging & classification process to identify objects and properties of the image without localizing them in the image. It was taking them months to complete this process and there was an urgent need to reduce timelines.

Their image recognition platform leveraged deep learning, which is a machine learning framework that mimics a human brain which is achieved by training the machine to recognize visual objects on the image. Machines use a large repository of tagged images to train them, and then based on an image feed, can derive trends and patterns. One of the areas of application of this technique is product discoverability, whereby users search for products using a reference image taken from their camera or downloaded from an online source.

They also had a large number of data analysts attached to the onsite team, which significantly increased the costs for the project.

The following were the constraints with the current process:

1   Automated tagging tools available were limited and unstable.

2.   Image processing tools are expensive and cannot scale to process large datasets.

3.   Difficulty in assembling a large team of data technicians onsite, due to availability of skilled data analysts and costs.

These constraints posed a huge challenge for the program that resulted in:

1.   Cost escalations and increasing the overall project budgets due to:
    a. People costs
    b. Image processing tooling licensing costs for a large team

2.   Increased pressure on the talent acquisition teams to find the right talent within budgets.

3.   The backlog of data to be processed increased, with minimal progress to reverse the trend

Studying machine learning technologies to generate accurate results from large data sets

One of the key aspects for machine learning technologies to be successful is to learn from a large data set to produce accurate results.

For visual recognition the platform is trained on large volumes of tagged images and considering the volume of data to be processed, automating the tagging process is beneficial, but with the challenges and constraints of automation, the process needed to be done manually with a large team of data technicians.

When algorithms get camera shy: diving into image recognition testing for greater success

The project consisted of the following phases:

The pilot phase

Analysis & planning was based on initial requirements, design workflows and project delivery mechanism. The delivery team researched optimal technology and tooling based on project objectives. The next logical step was a dry run of the solution on a sample data set, to optimize and make changes to the technology/tools. Once the proof of concept was established, the solution was showcased to the client for approval.

One-time effort

This phase was an extension of the pilot phase, with the goal being to scale the proof of concept to ramp up the project delivery capabilities in terms of processes, tools and people, address the thousands of backlog pending images and reduce it to a manageable level for the core team. There was a spike in resources to achieve the goal in a fixed time period, after which there was a ramp down of the team after achieving a steady state and then activities like process optimization, delivery performance metrics & measurement were implemented.

Ongoing effort

Based on the learning from the pilot and one-time phases, this phase established a long-term delivery strategy to address current & future in-flight projects. The team was ramped down to a "core" group for the long term that optimized the delivery model which resulted in overall cost savings.

Laughing in the face of image recognition: a journey in ground truth data collection

The service delivery solution that was deployed was designed to ensure that all the heavy lifting throughout the project lifecycle was done by the offshore team at the offsore Q Studios, with all the customer-facing activities being addressed by the onsite team.

Some of the project deliverables included:

- The streamlining of project related updates and communications since the onsite team and our Client were in the same time zone.

- The onsite team was responsible only for the review and 2nd level quality control, which ensured faster turnaround since the obvious issues were audited and fixed by our offshore Q Studios team.

- The core delivery and repetitive tasks were owned by the offshore Q Studios team, which reduced costs and flexibility to ramp up and down at short notice.

A large percentage of the project delivery workload was handled by the offshore team and the onsite site team bridged the gap between the Client and offshore teams by being the liaison layer to mitigate any potential risk of miscommunication or delayed responses due time zone difference.

Key benefits

- Costs were reduced by 40 - 45% for our client, since the major percentage of their workload was delivered by our QStudios which was in a more cost-effective location.

- There was a 35 - 40% productivity improvement as a direct impact of the onsite-offshore model that increased output.

- The data analyst's expertise on the opensource/free GIMP tool was opted for the project, saved licensing costs.

- A 95% first-time client acceptance rate was achieved through the creation of an internal QC process.

- 132953 images have been processed to date, which is an increase from 30000 at the start of the project, with the project still ongoing.

- Our Client benefitted from a highly scalable model with access to a large talent pool of highly skilled data analysts.

- The offshore delivery team was ramped up to full capacity within 1 - 2 weeks and was productive from week 3 onwards.