

Google Cloud Vision API Technology Review

The Vision API by Google Cloud allows users to draw insights from images stored in the cloud. The API offers facial detection, landmark detection, logo, label, text, document, handwriting and questionable content recognition. Some notable Vision API features leveraged across various industries include image detection and classification, emotion detection and insights on document security and sentiment. Pre-trained machine learning models can be accessed using a REST API.

Relevant to the CS410 coursework, the Vision API leverages optical character recognition (OCR) to detect text within images across numerous languages and also leverages the Google Document AI to ensure documents are accurate, compliant and can offer insights to improving documentation.

Among the many practical applications of the Vision API is with historic photography collections. Archival of photos via traditional card catalogue can log an overview of the photograph contents and details. To preserve this physical and visual history, where the photo imagery as well as content written on the back of the photo for example, can seem daunting to digitize. Using The Cloud Vision API, with no additional preprocessing, users can obtain a digital text transcription for additional context written on the back of the photo. Paired with image classification and detection, the API affords a robust archival of photo collections.

The Vision API supports the following image types:

- JPEG
- PNG8/24
- GIF/Animated GIF
- BMP
- WEBP
- RAW
- ICO
- PDF
- TIFF

Limitations and alternatives

While the Google Cloud Vision API gives you access to pretrained models making it simple to add machine learning capabilities to programs, there are limitations to its applications. Alternative frameworks should be considered when a specialized or customized training dataset would be needed. For example, medical classifications or specialized research topics relying on custom training data for a highly specialized model may benefit from a framework such as TensorFlow.

Additional limitations of the Vision API include requiring larger file sizes or high quality for image formats for accuracy. Minimum image size should be roughly 300 pixels. It is possible the API may not be able to perform proper detection with lower quality photos. Alternatively, this also means there is a tradeoff with larger file sizes and processing time and bandwidth consumption. Image files leveraging the Vision API should not exceed 20MB.

Google Cloud's Vision API appears to offer the most robust features for a paid service although there is a free tier available. Similar offerings in the paid industry include the Amazon Web Services Rekognition API and the Microsoft Vision API. The AWS Rekognition API offers robust video detection features and its facial detection reviews set it apart from competitors. For beginners, the Microsoft Vision API has been rated easier to get started.

Conclusively, Google is an industry leading in computer vision artificial intelligence. The Google Cloud Vision API affords a robust feature set for many industry applications with primary limitations where specialized custom training data is required.

References

Vision AI | Derive Image Insights via ML | Cloud Vision API. Retrieved October 26, 2020, from <https://cloud.google.com/vision>

Guide: TensorFlow Core. Retrieved October 26,2020, from <https://www.tensorflow.org/guide>

Machine learning in the AWS Cloud: add intelligence to applications with Amazon SageMaker and Amazon Rekognition. Retrieved October 26, 2020, from <https://aws.amazon.com/rekognition/>

Microsoft Azure Computer Vision. Retrieved October 26, 2020 from <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>