# Image and Video Processing

This section explores how generative AI is transforming image and video processing, covering computer vision applications, image enhancement, and object detection, with examples from research, startups, and S&P 500 companies.

Christophe Pennetier, André-Benoît De Jaegere

Generative Engineering · Volume I · written November 2024 to February 2025, two years after ChatGPT; its text is kept as published · https://generative-engineering.aquia.studio/fields/eitt/image-video/

CONTENTS

Image and Video Processing 1
Computer Vision Applications: Object Detection and Scene Understanding 2
Image Enhancement and Super-Resolution 2
Object Detection and Recognition 3
Conclusion 3
How are S&P 500 companies integrating generative AI in object recognition? 4
What specific object recognition technologies are Adobe and NVIDIA developing 5
How does Microsoft's Custom Vision service enhance object detection? 6
How does Amazon Rekognition integrate with AWS technologies for object recognition? 8
What are the key features of Google's Vision AI and Cloud Vision API? 9
Citations for Section on Image and Video Processing 10
General References for Section on Image and Video Processing 11

## Image and Video Processing

Generative AI has significantly impacted the fields of image and video processing, particularly in areas such as:

- computer vision applications
- image enhancement
- object detection

Below is an exploration of these categories with relevant examples from academic research, startups, and major companies.

# Computer Vision Applications: Object Detection and Scene Understanding

**Academic Papers**: Generative AI plays a crucial role in advancing object detection and scene understanding. Research papers often focus on how generative models can synthesize realistic images that aid in training more robust object detection algorithms. These models help improve the accuracy of identifying objects within complex scenes by providing diverse training datasets.

**Startups and SMEs**: Companies like Megvii [4] and viso.ai are at the forefront of using AI for computer vision applications. Megvii, known for its Face++ platform, has developed deep learning frameworks that enhance object detection capabilities [1]. Viso.ai offers an end-to-end computer vision platform that supports various applications, including object detection and scene analysis [1].

**S&P 500 Companies**: Nvidia is a key player in this space, providing the necessary hardware for running advanced AI applications. Its GPUs are widely used in training and deploying generative models for computer vision tasks [2] [3]. Nvidia's advancements in AI have made it a leader in supporting complex image processing tasks.

# Image Enhancement and Super-Resolution

**Academic Papers**: Research in this area often explores how generative adversarial networks (GANs) can be used to enhance image quality. These models are capable of producing high-resolution images from low-resolution inputs, which is particularly useful in fields like medical imaging and satellite imagery.

**Startups and SMEs**: Startups such as AnyClip use AI to process video data, extracting valuable insights and enhancing video quality for business applications [1]. These enhancements aid in better content delivery and user experience.

**S&P 500 Companies**: Adobe has integrated generative AI into its suite of creative tools, such as Photoshop, to improve image editing capabilities through features like super-resolution [3]. This allows users to upscale

images while maintaining quality, demonstrating AI's potential in creative industries.

## Object Detection and Recognition

**Academic Papers**: Generative models contribute to object detection by creating synthetic datasets that improve model robustness against variations in object appearance. This research helps in developing systems that can accurately identify and classify objects across different environments.

**Startups and SMEs**: Tractable uses AI for visual inspection tasks, leveraging generative models to enhance object recognition processes [1]. This is particularly beneficial in industries like automotive insurance, where accurate damage assessment is crucial.

**S&P 500 Companies**: Microsoft has incorporated AI into its Azure cloud services to offer scalable solutions for object detection and recognition. By integrating OpenAI's technologies, Microsoft enhances its ability to provide intelligent cloud services that support various AI-driven applications [2] [3].

## Conclusion

In conclusion, generative AI is transforming image and video processing across multiple sectors by enhancing capabilities in object detection, scene understanding, image enhancement, and super-resolution. This transformation is driven by both innovative startups and established tech giants leveraging advanced AI technologies to push the boundaries of what's possible in digital media processing.

# How are S&P 500 companies integrating generative AI in object recognition?

| COMPANY NAME | INTEGRATION STRATEGY | KEY PRODUCTS | PARTNERSHIPS |
|---|---|---|---|
| **Adobe Inc.** | Adobe integrates generative AI into its workflows primarily through Firefly, a suite of AI models that assist in creating, editing, and extending images and multimedia. Notable features include Generative Fill in Photoshop, which allows users to manipulate images via text prompts, enhancing creative processes and streamlining production. The AI capabilities are also embedded in Adobe Experience Manager and other products, assisting marketers in real-time content generation [5]. | Key products related to generative AI include Adobe Photoshop (with features like Generative Fill), Adobe Premiere Pro, Adobe Express, and Adobe Experience Manager, all utilizing Firefly for various content generation tasks [3]. | Adobe has formed notable partnerships with Microsoft to integrate generative AI into Microsoft 365 applications, enhancing marketing workflows. Additionally, Adobe collaborates with NVIDIA to develop advanced generative AI models for its products, and has partnered with Google to integrate its Firefly into Google's technologies for ethical image generation improvements [7]. |
| **Nvidia Corp.** | Nvidia integrates generative AI into object recognition by using NIM microservices and deploying them across popular frameworks like LangChain and Haystack. This includes leveraging GPU acceleration for real-time inference in applications such as autonomous vehicles and robotics [8]. | NVIDIA NeMo and Picasso image, video and 3D service for generative AI applications, as well as their Jetson platform for edge AI object recognition [9]. | NVIDIA has formed partnerships with SAS for machine learning and computer vision, Google Cloud for AI infrastructure and software, VMware for private AI solutions, and ServiceNow for enterprise workflow applications [10]. |
| **Microsoft Corp.** | Microsoft integrates generative AI into object recognition through its Azure AI services, particularly Azure AI Vision, which provides capabilities like object detection and optical character recognition (OCR). Additionally, the Custom Vision service allows users to create and train their own | Key products include Microsoft Azure AI Vision, which offers image analysis and object detection services, and Microsoft Copilot, which integrates AI capabilities across Microsoft applications to enhance productivity [12]. | Microsoft has formed significant partnerships with OpenAI and NVIDIA to enhance its AI technology development [13]. |

| COMPANY NAME | INTEGRATION STRATEGY | KEY PRODUCTS | PARTNERSHIPS |
|---|---|---|---|
| | object detection models, enhancing customization and accuracy. Their integration strategies also involve developing extensive APIs and utilizing cloud infrastructure to deliver AI capabilities seamlessly within existing applications [11]. | | |
| **Apple Inc.** | Apple is integrating generative AI through its 'Apple Intelligence' system, which includes multiple generative models fine-tuned for various daily tasks, including image recognition and context-aware actions, across its devices [14]. | The main offerings include the Apple Intelligence platform, enhanced Siri features, and on-device object recognition capabilities in iOS, iPadOS, and macOS [6.1.2-33]. | Apple has established partnerships with OpenAI and acquired Datakalab as part of its AI and computer vision strategy [16]. |
| **Google LLC** | Google integrates generative AI into object recognition through its Vision AI and Cloud Vision API, leveraging multimodal models and machine learning via Google Cloud's services like Vertex AI and the Gemini model [17]. | Key products include Google Vision AI and Cloud Vision API [17]. | Google has formed partnerships with Hugging Face and Plainsight, among others [18]. |
| **Amazon.com Inc.** | Amazon integrates generative AI into its object recognition processes through Amazon Rekognition for image and video analytics, supported by AWS technologies such as SageMaker and Bedrock [19]. | Key products include Amazon Rekognition, Amazon Textract, and Project P.I. [20]. | Amazon partners with institutions like Johns Hopkins University and companies such as TechSee and Scale [21]. |

# What specific object recognition technologies are Adobe and NVIDIA developing

Adobe and NVIDIA are developing advanced object recognition technologies with distinct approaches and applications:

### Adobe

**Adobe Sensei**: Adobe uses its AI and machine learning platform, Adobe Sensei, to power object-specific image searches in Adobe Stock. This technology allows users to find similar images by specifying the object and its position within an image, enhancing search precision for content creators [22].

**Project Stardust**: This is an AI-powered object-based image editor that detects objects similarly to how humans do. It enables users to edit images by selecting, moving, scaling, or deleting objects with ease. The tool automates background removal or replacement using Adobe Firefly, facilitating seamless editing [26].

**Adobe Camera Raw**: The Detect Objects feature allows users to quickly remove objects from photos by circling or scribbling over them. This is part of a broader suite of tools that include Generative Expand, which extends the canvas size while maintaining the original subject [25].

### NVIDIA

**Object Detection Models**: NVIDIA provides state-of-the-art object detection models like Single Shot Detector (SSD) and YOLO (You Only Look Once), which are optimized for real-time performance using NVIDIA GPUs. These models are used in various applications, including autonomous driving and intelligent video analytics[23] [24].

**TensorRT Optimization**: NVIDIA employs TensorRT to optimize deep learning models for inference on GPUs. This enhances the speed and efficiency of object detection pipelines, making them suitable for real-time applications [24].

These technologies reflect Adobe's focus on enhancing creative workflows and NVIDIA's emphasis on high-performance computing for complex object detection tasks.

## How does Microsoft's Custom Vision service enhance object detection?

Building a detection system from scratch can be daunting, but Microsoft's Custom Vision aims to simplify the journey—from labeling your images to refining model performance:

## Custom Model Training

The Custom Vision service allows users to build custom AI models tailored to specific object detection needs. Users can upload their own images, label them with custom tags, and train models to detect these objects. This process involves using a machine learning algorithm that analyzes images for custom features, allowing for the creation of highly specialized object detection models [27].

## Flexibility and Ease of Use

Custom Vision provides flexibility in how it can be accessed and used. It supports a web-based interface as well as SDKs for various programming languages, enabling both code-based and no-code approaches to model training and deployment. This makes it accessible to users with varying levels of technical expertise [29].

## Performance Metrics and Optimization

The service provides tools to evaluate model performance through metrics such as precision, recall, and mean average precision (mAP). Users can adjust the probability threshold to balance between precision and recall according to their project's needs. This iterative process helps in refining models for better accuracy and reliability [28].

## Quick Prototyping

Custom Vision is optimized for quick prototyping, allowing users to start building models with a relatively small dataset. This is particularly useful for projects where rapid development and testing are required. The service also offers domain-specific optimizations, which can enhance model performance for particular types of images, such as those featuring retail items or landmarks [27].

Overall, Microsoft's Custom Vision service enhances object detection by providing customizable, easy-to-use tools that enable users to create tailored AI models with robust performance evaluation features.

# How does Amazon Rekognition integrate with AWS technologies for object recognition?

Amazon Rekognition integrates with AWS technologies to enhance object recognition through several key methods:

## Integration with AWS Services

**Amazon S3**: Rekognition seamlessly integrates with Amazon S3, allowing users to analyze images and videos stored in S3 buckets without needing to move data. This integration facilitates scalable image analysis and storage management [30].

**AWS Lambda**: By integrating with AWS Lambda, Rekognition can automatically trigger image processing workflows. This enables real-time analysis and automation, such as processing images as they are uploaded to S3 [30].

**AWS Amplify**: Developers can use AWS Amplify to connect Rekognition to their applications, leveraging its APIs for image and video analysis. This integration simplifies the process of adding powerful object recognition capabilities to web and mobile applications [31].

## Object Recognition Capabilities

**Deep Learning Models**: Rekognition uses advanced deep learning models to detect objects, scenes, and faces in images and videos. These models are trained on large datasets, enabling accurate identification and categorization of visual content [32] [34].

**Real-Time Video Analysis**: With Amazon Rekognition Video, users can perform real-time object detection on streaming video inputs via Amazon Kinesis Video Streams. This service provides low-latency alerts for detected objects, enhancing applications like security monitoring and smart home automation [33].

## Additional Features

**Custom Labels**: Users can train custom models using their own datasets to detect specific objects that are unique to their business needs. This feature allows for tailored object recognition solutions without requiring extensive machine learning expertise [32].

**Scalability and Cost Efficiency**: Rekognition offers a pay-as-you-go pricing model and scales automatically with demand, making it cost-efficient for analyzing large volumes of images and videos [30].

These integrations and features make Amazon Rekognition a versatile tool for enhancing object recognition capabilities across various applications within the AWS ecosystem.

# What are the key features of Google's Vision AI and Cloud Vision API?

Google's Vision AI and Cloud Vision API offer a comprehensive suite of features that enable developers and businesses to perform advanced image analysis and object recognition. Here are the key features of these services:

## Google Cloud Vision API

1. **Label Detection**: This feature allows the API to detect and classify objects within an image, providing descriptive labels that help in understanding the content of the image [35] [38].
2. **Optical Character Recognition (OCR)**: The API can extract text from images, including printed and handwritten text, making it useful for document scanning and text analysis [35] [38].
3. **Facial Detection and Analysis**: It can identify faces within an image, analyze facial expressions, and detect facial landmarks and attributes such as age and gender [35] [38].
4. **Landmark Detection**: The API recognizes famous landmarks in images, providing information about the landmark and its geographic location [35] [39].
5. **Logo Detection**: This feature identifies brand logos within images, which can be used for brand monitoring and marketing analysis [38] [39].
6. **SafeSearch Detection**: It automatically detects explicit or inappropriate content within images, allowing for content moderation and filtering [35] [39].
7. **Image Properties Analysis**: Provides additional information about an image, such as dominant colors and other attributes that can be used for further analysis [35] [38].

8. **Web Entity Detection**: Identifies web entities related to the content of an image, linking images to relevant web pages or online resources [39].
9. **Object Localization**: Detects multiple objects within an image and provides bounding boxes around them, which is useful for applications like product recognition in e-commerce [38] [39].

### Google Vision AI

- **Pre-trained Machine Learning Models**: Vision AI uses sophisticated models trained on extensive datasets to provide highly accurate recognition capabilities across a wide range of categories [36].
- **Scalability**: Leveraging Google Cloud's infrastructure, Vision AI can process from a few images to millions, making it suitable for both small-scale applications and large enterprises [36].
- **Ease of Use**: The service is accessible via a simple REST API and user-friendly interfaces, making it approachable for developers with varying levels of expertise [36].
- **Versatility**: Suitable for diverse industries such as retail, media, healthcare, and more, Vision AI supports various use cases including image editing, research, and AI detection tasks [36].
- **Continuous Improvement**: Benefits from Google's ongoing investment in AI technologies, ensuring that the service remains cutting-edge with regular updates and improvements [36].

These features make Google's Vision AI and Cloud Vision API powerful tools for businesses looking to integrate advanced image recognition capabilities into their applications. They offer robust solutions for tasks ranging from basic image labeling to complex facial analysis and object detection.

## Citations for Section on Image and Video Processing

[1] Viso.ai – Most Popular Computer Vision Companies and Startups
[2] Fool – AI Stocks
[3] US News Money – Best AI Companies
[4] Megvii
[5] Adobe Blog – Bringing Gen AI to Video Editing Workflows in Adobe Premiere Pro
[7] WorkflowOTG – Adobe and NVIDIA Generative AI
[8] NVIDIA Developer – Deploying Generative AI with NVIDIA NIM
[9] Aethir – AI Applications Using GPUs
[10] NVIDIA News – Generative AI for Enterprises

[11] Microsoft Azure Blog – Next Generation AI-Powered Applications
[12] Microsoft Azure AI Vision
[13] Microsoft Blog – Microsoft and OpenAI Extend Partnership
[14] Forbes – For Apple, AI is Personal
[15] Apple Newsroom – Introducing Apple Intelligence
[16] PYMNTS – Apple Unveils Apple Intelligence Suite
[17] Google Cloud Vision
[18] InfoQ – Hugging Face and GCP AI
[19] About Amazon – How Amazon Uses Generative AI
[20] Eden AI – Amazon Web Services
[21] Johns Hopkins Ventures – Amazon Collaboration on AI
[22] Adobe Developer Blog – AI-Powered Object-Specific Search in Adobe Stock
[23] NVIDIA NGC – Object Detection Collection
[24] NVIDIA Developer – Object Detection on GPUs in 10 Minutes
[25] Adobe Camera Raw – What's New 2025
[26] Adobe Labs – Project Stardust
[27] Microsoft Learn – Custom Vision Service Overview
[28] Microsoft Custom Vision – Characteristics and Limitations
[29] Microsoft Custom Vision – Object Detection Quickstart
[30] AWS Rekognition – What Is Rekognition
[31] AWS Amplify – Connect Amazon Rekognition
[32] CloudVisor – Amazon Rekognition Guide
[33] AWS Rekognition Video Features
[34] AWS Rekognition Image Features
[35] Educative – What is Cloud Vision API?
[36] Futurepedia – Google Cloud Vision AI
[37] TLV Tech – Understanding Google Vision API Easily
[38] Google Cloud Vision – Features List
[39] YouTube – Google Vision API Overview

# General References for Section on Image and Video Processing

[G1] Nasdaq – New SP 500 AI Insights
[G2] Datamation – AI Companies
[G3] Technology Magazine – Magnificent Seven Tech Companies Driving Forward with AI
[G4] SenseTime
[G5] Nauto
[G6] Verkada
[G7] Tractable
[G8] AI Robotics Drones
[G9] Hawkeye Innovations
[G10] Trigo
[G11] Intel – Movidius VPU
[G12] Standard AI
[G13] Orbital Insight
[G14] Regna AI
[G15] AnyClip
[G16] Bossanova
[G17] Descartes Labs

[G18] Zebra Medical
[G19] Neuromation
[G20] Shield AI
[G21] Eyesight Tech
[G22] Onfido
[G23] Nuro
[G24] DeepMind
[G25] UiPath
[G26] Cerence
[G27] Cyngn
[G28] Xnor.ai
[G29] Built In – Generative AI Companies
[G30] RBC GAM – Generative AI: The 5th Revolution
[G31] Fool – New AI Stock in SP 500
[G32] Arize – Rise of Gen AI in SEC Filings
[G33] Fool – Best AI Stock in SP 500: Nvidia
[G34] US News Money – Best AI Companies
[G39] Elastic – Image Recognition at Adobe with Elasticsearch and Sensei
[G40] Adobe Research – Computer Vision, Imaging & Video
[G41] Baydachnyy – Detecting Objects Using Custom Vision Service