

# Improve Computer Vision Use Cases With Standardized Implementation Patterns

Published 18 February 2022 - ID G00749354 - 30 min read

By Analyst(s): Daniel Cota

Initiatives: [Analytics and Artificial Intelligence for Technical Professionals](#); [Evolve Technology and Process Capabilities to Support D&A](#)

Computer vision is an evolving AI technology, driving vendors to rapidly develop use-case solutions. Data and analytics technical professionals must understand implementation patterns of computer vision to strategize their choice of vendor versus autogenous implementation.

## Overview

### Key Findings

- Industry use cases for computer vision (CV) introduce integration challenges with business applications, analytical systems and endpoints such as edge devices.
- Efficient and high-accuracy CV solutions require massive amounts of trained and annotated unstructured data, expanding the need for robust computer processing infrastructures.
- Deep learning techniques are the underpinning of CV. Without a thorough understanding of CV technique application and domain knowledge, inaccurate results may occur.
- Image and video collection for facial recognition has heightened concern for private and public entities. Data extraction challenges of facial recognition's unstructured data have only exacerbated the issue for technical professionals.

### Recommendations

Data and analytics technical professionals responsible for implementing computer vision solutions must:

- Improve preprocessing steps and the overall quality of incoming images or videos by formatting, updating and annotating public or domain-specific datasets.

- Strengthen use-case inferencing results and increase operational efficiencies by selecting the appropriate CV technique for the deployment strategy.
- Define the long-term strategy and identify needs for computer vision to support complementary technical use cases such as digital twin, simulation or metaverses.
- Avoid lawsuits or unintended public scrutiny by familiarizing yourself with local and global privacy laws for collecting personally identifiable images or videos.

## Strategic Planning Assumption

By 2023, more than 80% of organizations will use some form of computer vision to analyze images and videos.

## Analysis

### [Download All Graphics in This Material](#)

Computer vision is a technique that involves capturing, processing and analyzing real-world images and videos to allow machines to extract meaningful and contextual information from the physical world. This technique attempts to replicate cognitive abilities that humans use through sight and mental processing of physical objects. Human brains process enormous amounts of visual information throughout their life span, decoding salient details and persisting them into memory. Similarly, the digital world has seen an extraordinary amount of images and videos being stored as unstructured data in various data management ecosystems.

Organizations advancing digital transformation initiatives must understand the importance of leveraging unstructured data to introduce or increase CV capabilities. Failure to do so will impact competitive advantage and result in missed opportunities to improve overall business functions.

To accomplish successful CV implementations, data and analytics technical professionals need to interpret their CV use cases to determine the appropriate implementation strategy. Although the CV market is saturated with robust solutions, they are mostly proprietary and industry- or use-case-specific. Data and analytics technical professionals will have to decide whether an off-the-shelf solution will satisfy requirements for future use cases and integrate within existing infrastructures, or if an end-to-end solution will need to be developed from scratch.

This document provides an initial set of guidance to assist with CV initiatives:

- [Understanding Core Processes](#)
- [Evaluating Use-Case Examples](#)
- [Exploring the Future of Computer Vision](#)

## Understanding Core Processes

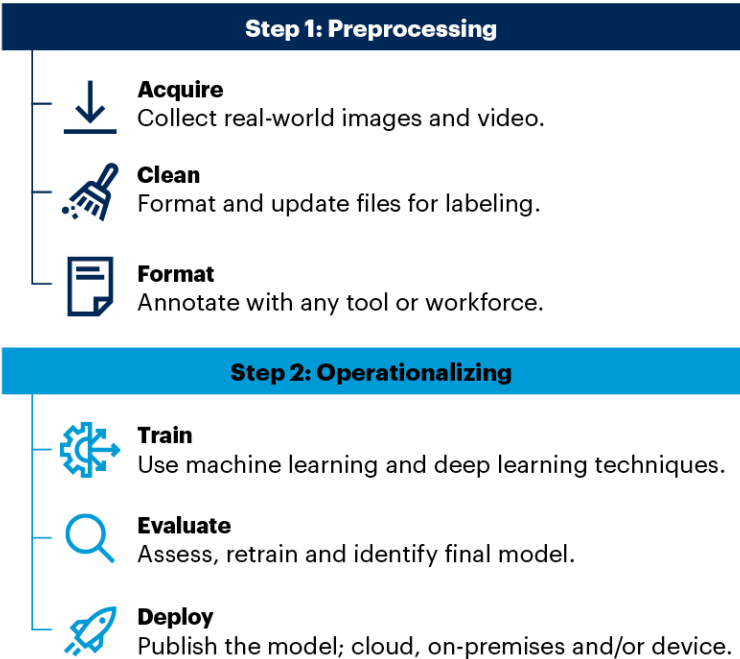
Human vision ability has the advantage of decoding information into “bites” that are later used for decision making — learning from previous experiences and applying them in a new context or problem. To fully prepare computers with the same ability, they need to be fed enough labeled data and real-world scenarios to learn. This requires many forms of unstructured data, such as 2D and 3D images and video, sensor information, and sometimes synthetic data. Additionally, adequate compute processing infrastructure is needed to process this data in a timely manner — real-time facial recognition, for example. Nevertheless, large high-quality unstructured data and high-performing compute engines are the underlying stipulations for effective and complex CV implementations.

Identifying the data and having sufficient infrastructure is only the beginning. Preprocessing steps incorporate cleaning and labeling mechanisms of the data deriving relevant information for contextualization. Fully prepared data is then processed through additional steps to operationalize the final output.

Use cases for CV fall into two common techniques: image analysis and video analytics. Although slightly different, both follow similar core processes when it comes to preprocessing and operationalizing. Figure 1 illustrates the steps within each category.

Figure 1. Computer Vision Core Processes

Computer Vision Core Processes



Source: Gartner  
749354\_C

Gartner

Acquire: Collect Real-World Images and Video

Before embarking on the collection of data, a fundamental understanding of the use case and desired outcome is critical. For instance, object detection, and in some cases 3D tracking, can leverage models pretrained on public datasets. These datasets can be acquired for your use case if the solution calls for deciphering *common* images or objects in videos. A list of several popular public datasets is provided in Table 1.

**Table 1: Public Datasets**

(Enlarged table in Appendix)

Dataset	Number of Files	Description
<a href="#">V7Labs COVID-19 X-Ray Dataset on GitHub</a>	6,500 images	Anterior to posterior (AP) and posterior to anterior (PA) chest X-rays, with pixel-level polygonal lung segmentations.
<a href="#">KITTI-360</a>	320,000 images 100,000 laser scans	Large-scale dataset containing rich sensory information and full annotations. Recorded in several suburbs of Karlsruhe, Germany. Semantic and instance annotation for both 3D point clouds and 2D images.
<a href="#">Common Objects in Context (COCO)</a>	330K images >200K labeled images	Large-scale object detection, segmentation and captioning dataset.
<a href="#">ImageNet</a>	14 million images 21,000 synsets indexed	Image database organized according to the WordNet hierarchy (nouns only), in which each node of the hierarchy is depicted by hundreds and thousands of images.
This is a representative, not exhaustive, list of public datasets.		

Source: Gartner (February 2022)

As you can see, these datasets represent a set of information for particular use cases. Although they can be used to train your CV models, an extensive amount of *domain-specific* unstructured data may be needed.

Consider a scenario where CV is being used to detect anomalies or defects of a certain product you manufacture. Automating this task using cameras on the manufacturing line or within a controlled inspection booth would identify defects before product completion. This would increase overall production and reduce costs.

However, to complete this task effectively, hundreds or thousands of product images would need to be fed into the CV solution for consistency and accuracy. If your product images aren't available in any of the public datasets above, or anywhere else, they will have to be created. The images will also have to account for the many variations of the product and potential anomalies. While not covered in this research, a newer alternative approach from vendors is using augmented variational autoencoders (VAEs) for very precise representation of objects being inspected.

As you can imagine, creating your own domain-specific data can be expensive and time-consuming, eating away at both the potential efficiency and cost gains you would benefit from using CV. As products change over time, so will the need for updated images. A proper balance between CV implementation costs and long-term cost savings must be thoroughly evaluated.

Examine the options above before collecting data. Determine whether a pretrained model is feasible, or domain-specific data must be used or if there is an opportunity to combine both.

## **Clean: Format and Update Files for Labeling**

Accuracy for CV solutions depends on the quality of files coming in. Yet files being used for CV are *not* created equal. On one hand, high-powered equipment capturing high-quality images and videos can be expensive, and files they produce are quite large. On the other hand, less-expensive equipment produces lower-quality images, requiring additional cleaning and updating steps.

When dealing with lower-quality files, the ultimate goal is to remove any substantial noise that would interfere with labeling (the final preprocessing step) and training the CV models. Image enhancement is the process to reconstruct lower-quality files into higher-quality files for further analysis. It consists of techniques like deblurring, denoising, upscaling, enhancing, rotating and compressing. Keep in mind that these same processes will need to be applied for inference so the trained model can recognize and predict correctly.

Popular image enhancement software can be used, but it entails manual efforts by a product-qualified and skilled expert. So, why not automate the process with artificial intelligence (AI)? Deep learning techniques provide the ability to solve the problem through different types of neural networks. Both open-source frameworks and commercial solutions utilize underlying deep neural networks to clean files. Below are a few examples of each.

## **Open-Source Examples**

Photo enhancers are available through several of the open-source frameworks, with Tensorflow being the most widely used. Two common types of neural networks for photo enhancers are:

- **Convolutional neural networks (CNNs)** extract features from images with the help of its many layers. Designed to parse through pixels of images through a series of layers, CNNs assist with image classification techniques. Additional steps can be taken to improve the overall pixelation to generate higher-quality images.
- **Generative adversarial networks (GANs)** use an unsupervised learning approach composed of two models: generator and discriminator. The generator learns to “generate” fake images that look realistic. These images, combined with real examples, are then fed into the discriminator model to determine which images are fake and which are real. The models iterate through this process until realistic high-quality images are created. GANs are known to be used for image generation and manipulation.

Figure 2 provides output examples from the 2018 Conference on Computer Vision and Pattern Recognition (CVPR) spotlight paper, “Deep Photo Enhancer: Unpaired Learning for Image Enhancement From Photographs With GANs” <sup>1</sup> and “IBM Developer Model Asset Exchange: Image Resolution Enhancer.” <sup>2</sup> Both solutions used TensorFlow and a set of various GANs.

Figure 2. Output Examples Using GANs

**Output Examples Using GANs**

Source: Gartner  
749354\_C

**Gartner****Commercial Solutions**

Commercial solutions offer an automated approach for enhancing images or videos, which can be used when internal staffing skills in deep learning are scarce. Although proprietary in nature, most solutions include API integration or the ability to upload the enhanced versions for consumption. Using these solutions, you'll be able to integrate final output for additional analysis or continue preprocessing for CV operationalization. Two example vendors are:

- **Deep-Image.** Upscales a photo, image, frame or picture (from low resolution up to 4K); removes JPG artifacts, removes unwanted noise and sharpens blurry photos.
- **Topaz Video Enhance AI.** Improves the quality of videos by using information from multiple frames to achieve high-end results for video upscaling, denoising, deinterlacing and restoration.







Higher quality equals higher accuracy and minimized processing. Implementing neural networks for cleaning and enhancing images and videos requires comprehensive skills in deep learning, regardless of whether you are creating your own or incorporating a commercial solution. This step institutes a basis to improve labeling and prepare data for model training and inferencing.

## **Label: Annotate With Any Tool or Workforce**

CV will not properly work by collecting and cleaning images alone. Once you have acquired the data, labeling — often referred to as annotating — is a crucial and final step during preprocessing. This step adds the necessary metadata or tags for identifying features that the CV model needs in order to learn and recognize new incoming images or frames of videos.

Advances in machine learning (ML) can assist with automating annotation tasks, but in most cases, it is still completed manually. Human-in-the-loop annotation options include outsourcing, crowdsourcing or in-house, whereas ML annotation automation can be done programmatically (see Table 2).

Table 2: Annotation Options and Risks

	Human in the Loop (HITL)			ML Automation
Option	 Outsource	 Crowdsourcing	 In-House	 Programmatic
Risks	<ul style="list-style-type: none"> <li>■ High cost</li> <li>■ Time-consuming</li> <li>■ Secure data sharing required</li> </ul>	<ul style="list-style-type: none"> <li>■ Medium cost</li> <li>■ Difficult to manage</li> <li>■ Multiple external partner coordination</li> </ul>	<ul style="list-style-type: none"> <li>■ High cost</li> <li>■ Time commitments</li> <li>■ Internal skills required</li> </ul>	<ul style="list-style-type: none"> <li>■ Medium cost</li> <li>■ ML skills required</li> <li>■ Application integration</li> </ul>

Source: Gartner (February 2022)

Similar to the acquire step, a decision will have to be made on which annotation option will have the most impact without introducing significant costs or risks. A thorough understanding of the use case is important in this step as well. Not only will you have to decide which annotation option to select, but you will also have to determine which annotation *technique* to use.

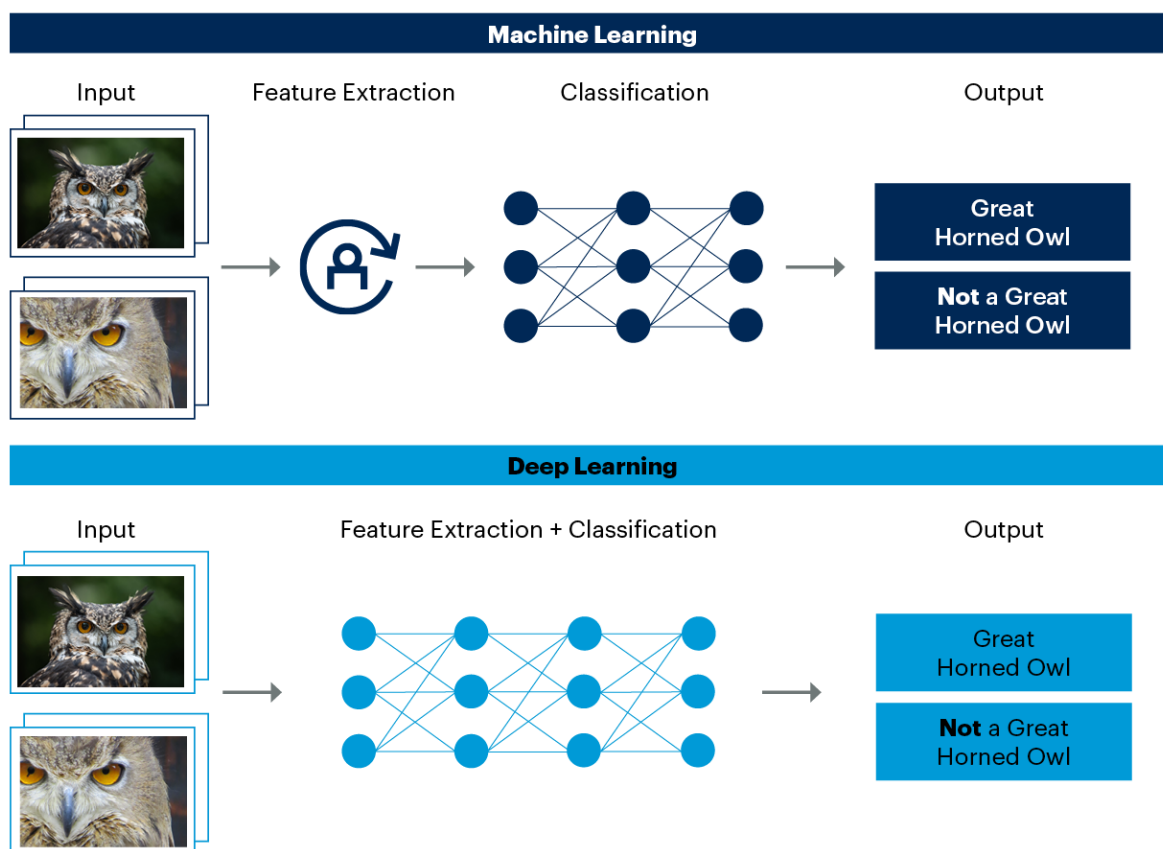
### Train: Use Machine Learning and Deep Learning Techniques

Conducting the steps during preprocessing creates a dataset ripe for feature extraction to contextualize the information. In simple terms, features are merely the characteristics of the data in the use case. Features, often referred to as independent variables, are used to train CV models so they can learn and produce the desired outcome.

CV has been around for decades, but applying machine learning algorithms has increased the capabilities of feature extraction. Deep learning, considered a subset of ML, has been at the center stage of CV innovation because of its ability to extract and learn features from high-quality data, mirroring the human learning process. Deep learning and machine learning are used to train models on the features to generate the final model that will be evaluated and deployed. See Figure 3 for an illustration.

**Figure 3. Machine Learning vs. Deep Learning for Image Classification**

### Machine Learning vs. Deep Learning for Image Classification



Source: Gartner  
749354\_C

Deciding whether to use ML or DL for feature extraction and training will be further understood through a simple CV image classification use case. Imagine you work for an animal conservation facility and are required to classify great horned owls from thousands of images sent to you from across the country. You have been presented with the option to implement a standard ML solution or to enhance the CV use case with a more automated DL approach. The two options seem viable, but an understanding of each will help determine which to use:

- **ML approach.** Requires a human in the loop to clearly tell the algorithm what to look for during feature extraction. Explicitly annotates visual features of the images so the classification model can derive the information and determine whether each image is a great horned owl or not.
- **DL approach.** Uses a deep learning algorithm like a CNN or GAN to iterate through the images while automatically processing and learning the visual features. Removes the need for a human in the loop, but may require additional images that do not contain great horned owls so the classification model can learn to decipher the difference.

In either approach, the ultimate goal of this step is to have the first iteration of the model that will then be processed into the next operationalizing step: evaluate.

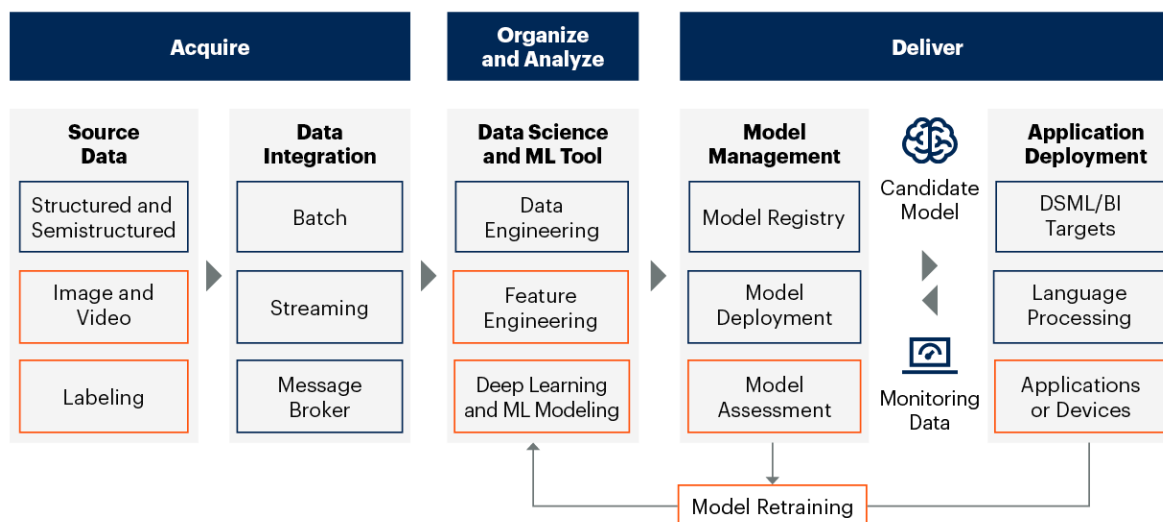
### **Evaluate: Assess, Retrain and Identify Final Model**

Operationalizing CV models follow a similar process as ML or data science solutions found in MLOps. The models flow through standard DevOps cycles with an understanding that they adhere to continuous integration/continuous delivery (CI/CD) guidelines and incorporate governing policies. However, there are several differentiating factors to consider when introducing CV models to the workflow. See Figure 4, which highlights CV-specific considerations.

Figure 4. Computer Vision Considerations for MLOps

## Computer Vision Considerations for MLOps

Computer Vision (CV)

Source: Gartner  
749354\_C

Gartner

**Assess.** There are three key areas where formal assessments must be made for CV use cases:

1. Once the files are acquired, an initial check to ensure they came from trusted sources is critical. In some countries, legal policies and privacy laws prohibit the use of images or videos containing confidential or private information. Assessing sources and conducting validation checks will prevent misuse of the files to avoid lawsuits.
2. AI governance and responsible AI are at the forefront of delivering solutions. A second assessment check must be conducted after labeling is complete to ensure bias, or any other nefarious content, wasn't included in any of the annotations. Avoid AI governance issues by including additional human-in-the-loop peer review checks in the process.
3. Constant assessments of deep learning models are similar to model management activities found in ML and data science solutions. But dealing with unstructured data for CV is different. If the model is not performing to accuracy standards, additional high-quality labeled images or videos may need to be added to the initial dataset.

**Retrain.** Most CV models fail at the beginning because of the lack of enough annotated data (as described in the Label: Annotate With Any Tool or Workforce section). Despite the increased technical capabilities of computer processing, deep learning models still demand a significant amount of resources. When retraining the model with additional images and videos joined to the initial dataset, consider the infrastructure impact it may cause.

**Identify the final model.** If all assessment checks pass and retraining adheres to accuracy standards, a candidate model is identified. Prevent an unnecessary deployment by understanding the delivery requirements of the candidate model before preparing it as the final model. For instance, a CV video analytics solution requiring the final model output be deployed to an edge device such as a video surveillance camera has drastically different infrastructure and processing requirements than a cloud or on-premises deployment.

### **Deploy: Publish the Model: Cloud, On-Premises and/or Device**

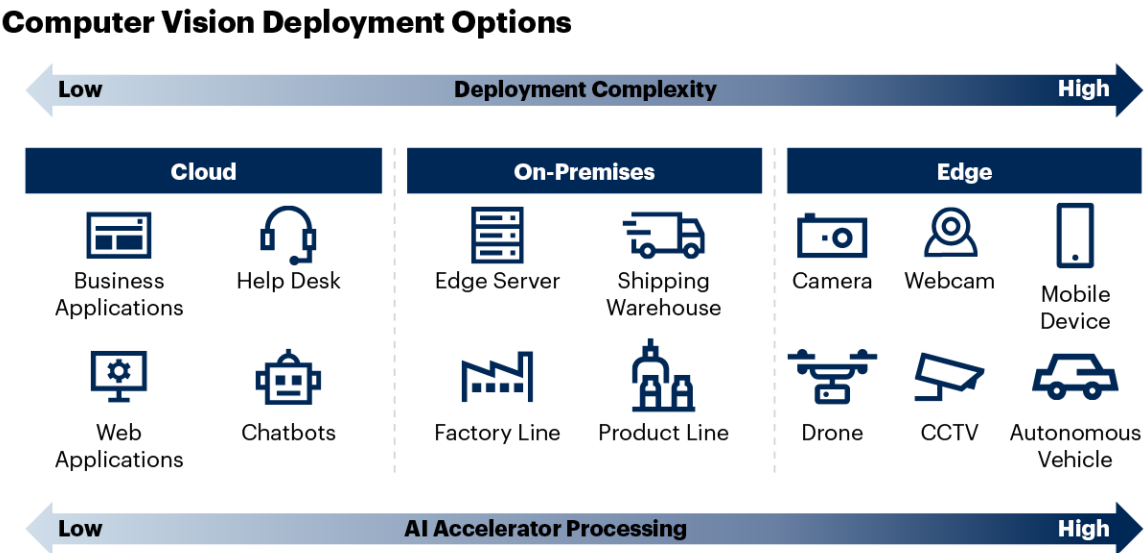
CV capabilities have exploded in the market, creating a dynamic environment where a majority of AI solutions incorporate CV in one way or another. This has introduced challenging infrastructure and processing demands, since CV solutions can be deployed almost anywhere.

Final models can be delivered in many different ways. Instead of deploying all the preprocessed data and models, the more common approach is deploying the set of output or inferencing — a *vision*, in this case. Figure 5 illustrates the different levels of deployment complexity and AI accelerator processing requirements across each of the following:

- **Cloud.** Provides options to build end-to-end solutions or utilize prebuilt CV applications that simplify overall deployment and uses scalable AI accelerators, such as GPUs, for model training or inferencing. Low-level deployment complexity may be introduced because of integration challenges such as sending output to a business application or chatbot.
- **On-premises.** A “build your own” approach that increases deployment complexity, since you will have to deploy to isolated environments such as a set of devices in a shipping warehouse. There are many highly flexible AI accelerator options to choose from. However, you will have to create and manage your own infrastructure.

- **Edge.** Requires lightweight output and low-power AI accelerators to be effective. High deployment complexity is an issue, since most solutions require near- or real-time capabilities. The AI accelerator must be able to quickly make adjustments as new output is fed into the solution. Autonomous vehicles are a perfect example of edge deployment.

Figure 5. Computer Vision Deployment Options



Evaluating Use-Case Examples

Have you ever wondered how your mind recalls movies or short video clips that you have already seen? Does your mind *replay* the entire video or does it visually recollect static *images* of scenes? Or both? Consider the definition of “video.”

**video** — a digital recording of an image or set of images (such as a movie or animation)

— Source: Merriam-Webster

CV essentially follows a similar approach, where videos are simply a set of images broken into frames that can be contextualized. This indicates that image analysis and video analytics share common underlying process techniques to perform CV activities. It also indicates, however, that video analytics require more resources to store, process and perform complex deep learning algorithms, since it has to decipher each image, or *frame*, in the video.

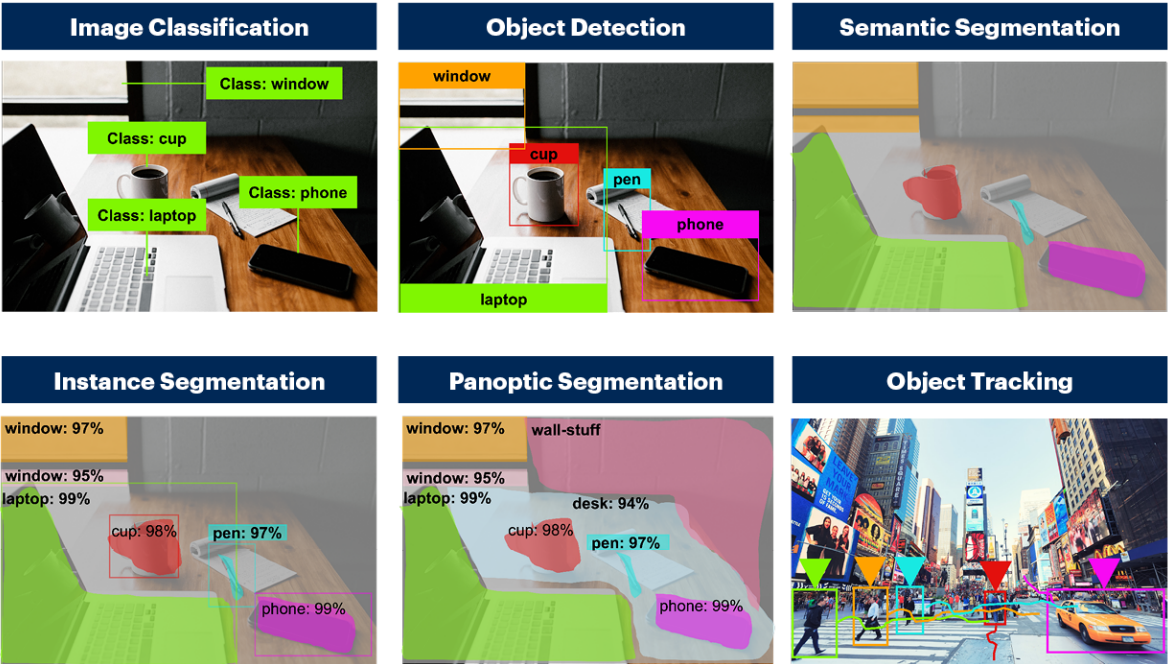
Figure 6 represents six techniques used for image analysis and video analytics: image classification, object detection, semantic segmentation, instance segmentation, panoptic segmentation and object tracking:

- **Image classification** scans through a set of input images and classifies an object or set of objects within the image. It requires labeled data to process new images to place them into specific categories. *Analysis:* Identified the various objects on the desk.
- **Object detection** identifies and defines objects and labels them with bounding boxes. It also applies classification and localization to the objects. *Analysis:* Bounded each of the classified objects.
- **Semantic segmentation** divides the entire image into groups of pixels. You can then classify every pixel to identify what objects it contains. *Analysis:* Grouped pixels into the classified objects.
- **Instance segmentation** advances semantic segmentation by not only classifying objects in the image at a pixel level, but also differentiating those similar objects. *Analysis:* Grouped pixels into the classified object and also recognized two windows and separated them.
- **Panoptic segmentation** combines both semantic segmentation to assign a class label to each pixel and instance segmentation to detect and segment each object instance. *Analysis:* Combined segmentation and instance capabilities to classify at the pixel level and identify the multiple instances of the window. “Panoptic” is defined as considering all parts or elements, so the wall and desk were added to account for every pixel in the image.
- **Object tracking** is the process of following a particular object or multiple objects in a video. It uses a generative method to describe characteristics of the objects and a discriminative method to separate the objects from the background. *Analysis:* Detected four people and one vehicle, and tracked the objects throughout a set of frames in a video.



Figure 6. Computer Vision Techniques Used in Image Analysis and Video Analytics

Techniques Used in Image Analysis and Video Analytics



Source: Gartner  
749354\_C






Gartner

The techniques described above introduce a multitude of useful applications for use cases across many industries. To provide further context, the Image Analysis and Video Analytics sections below supply examples of where the techniques can be applied and a use-case sample process.

Image Analysis

The previous section established a base understanding of the techniques used for CV. The next step is to evaluate the options, align to the appropriate technique and verify whether the approach can be used for your use case(s). To assist with this endeavor, Table 3 provides a list of options for image analysis along with a description and examples of use cases.

Table 3: Image Analysis Use Cases

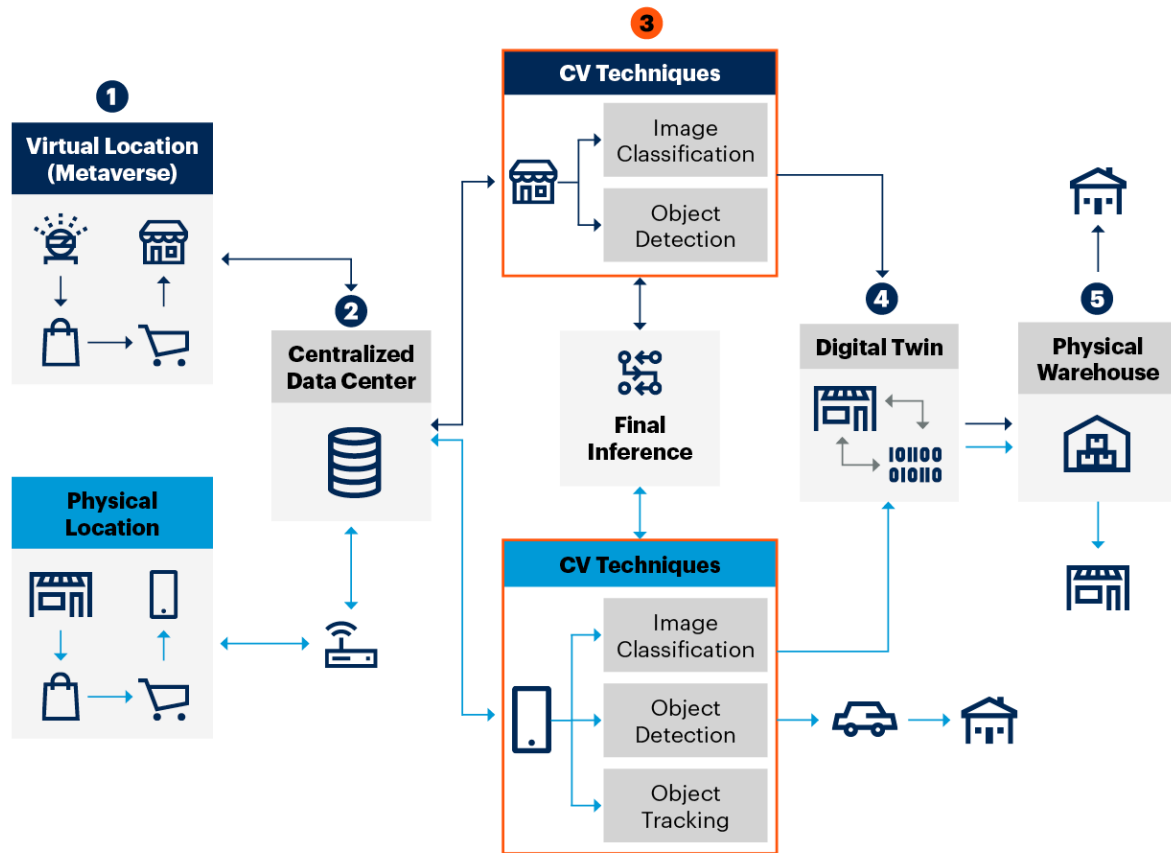
 Image Analysis	 Tagging	 Detecting	 Categorizing	 Describing
Description	Identify and tag visual features in an image.	Detect objects or faces in an image.	Categorize tags or pixels of an image into classes.	Describe an image into a readable format.
Example Use Cases	<ul style="list-style-type: none"><li>■ Object tracking</li><li>■ Facial expression</li><li>■ Advertisement analysis</li><li>■ SEO product tagging</li></ul>	<ul style="list-style-type: none"><li>■ Facial recognition</li><li>■ Object identification</li><li>■ Visual search</li><li>■ Biometric pattern recognition</li></ul>	<ul style="list-style-type: none"><li>■ Product hierarchy detection</li><li>■ No-scan check-out</li><li>■ Object and people grouping</li></ul>	<ul style="list-style-type: none"><li>■ Help desk support</li><li>■ Chatbot interaction</li><li>■ Language transcription</li><li>■ Damage evaluation</li></ul>

Source: Gartner (February 2022)

When was the last time you visited a grocery store and utilized the *quick* self-check-out? OK, maybe it wasn't as quick as you thought it would be. Visiting next time, envision scanning your phone at a turnstile, placing all of your items in bags as you shop, loading your cart with your bags of items and walking out through the same turnstile. Better yet, visualize the ability to don your favorite virtual reality (VR) headset, walk through a virtual representation (digital twin) of the grocery store, select your items, finalize your order and wait until your items are delivered. Architecting and applying CV techniques for these two scenarios are illustrated in Figure 7.

Figure 7. Virtual and Physical No-Scan Sample Process

## Virtual and Physical No-Scan Sample Process



Source: Gartner  
749354\_C

Gartner

In the image analysis process:






1. Customers don a VR headset or walk into a physical grocery store. The VR customer uses their credentials to log into the virtual environment, which is a digital twin virtual representation of the physical location (see Exploring the Future of Computer Vision section below). The physical customer scans their app at the turnstile. Both customers peruse through their respective environments, collecting items and placing them into their carts.
2. The entire experience for both customers is captured using high-quality image and video collection. This data is fed directly into a centralized data center for the virtual customer. Data for the physical customer is collected from their app, sent to an edge server on-site and then processed to the same centralized data center.

3. CV techniques are used differently for each experience:
  - **VR customer.** *Image classification* is used to categorize grocery items and translate them into a virtual representation. As the customer reviews and selects the items, *object detection* is used to detect the items the customer has put into their cart, and an itemized list is generated. Final inference is completed and looped back through the data center and then back to the customer for check-out.
  - **Physical customer.** *Object tracking* is used to track the customer throughout the store. *Object detection* identifies the items placed into the customer's cart, and *image classification* categorizes the items to generate the final itemized list. Final inference is passed back through the data center and to the customer's app for checking out.
  - When shopping is complete, the VR customer finalizes the order through a virtual kiosk, and the physical customer scans their app at the turnstile and exits the store.
4. A digital twin processes information from both experiences to ensure physical inventories have sufficient supply. For the VR customer, the digital twin verifies that the selected items are available and prepares a fulfillment order. Behind the scenes of the physical customer, the digital twin receives inventory information from the store, validates inventory levels and prepares a fulfillment order by using object detection.
5. Fulfillment orders for the VR customer are processed at the physical warehouse, and items are shipped directly to the customer's address on file. Items that require replenishing at the physical store are also processed and fulfilled.

## Video Analytics

Video analytics utilizes many of the same CV techniques as image analysis (see the earlier Understanding Core Processes section). To properly analyze and apply CV techniques, the videos must be extracted into frames. Once this is complete, various CV techniques can be applied. Table 4 represents four areas in which video analytics can be deployed, with several example use cases for an additional reference point.

Table 4: Video Analytics Use Cases

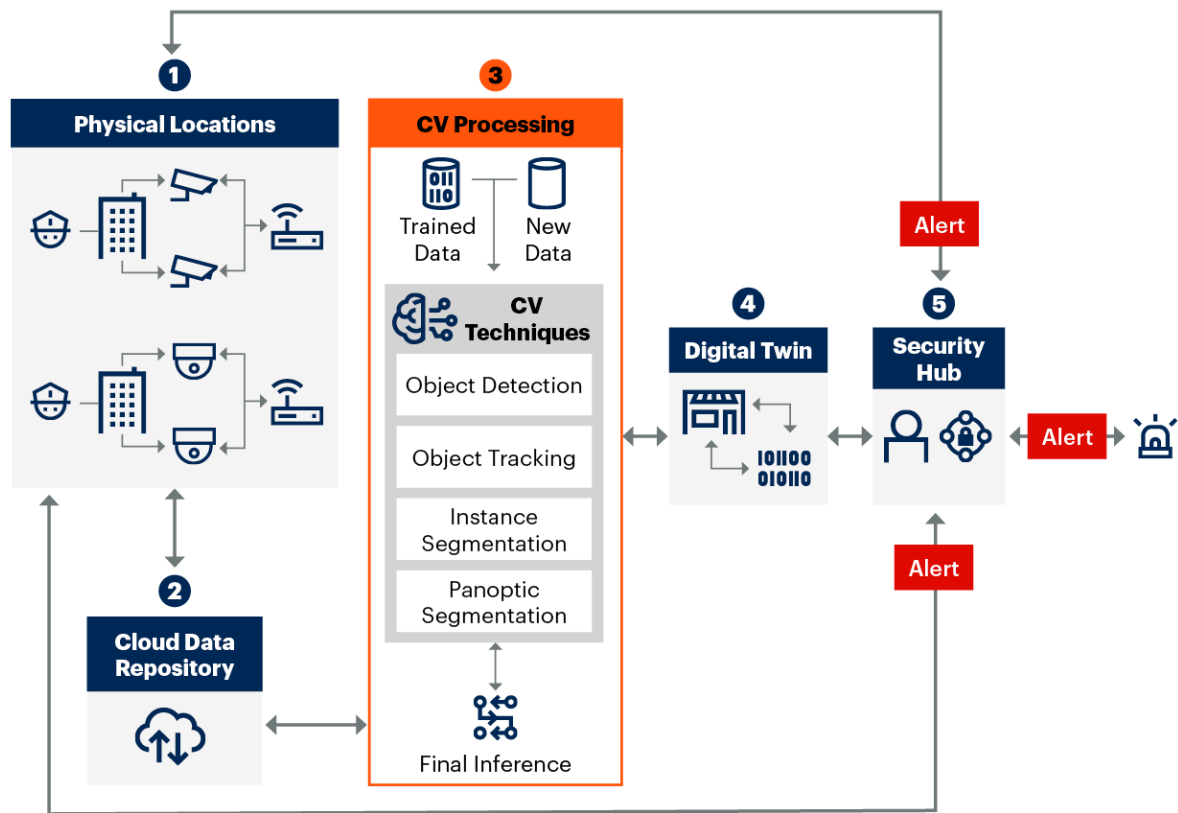
	 Video Analytics	 Security	 Safety	 Process	 Assets
Description	Monitor environments and issue alerts.	Identify safety or noncompliance concerns.	Evaluate objects for defects or anomalies.	Scan physical objects for inventory control.	
Example Use Cases	<ul style="list-style-type: none"><li>■ Threat alerting</li><li>■ Crowd control monitoring</li><li>■ Hazardous driving detection</li></ul>	<ul style="list-style-type: none"><li>■ Hygiene testing</li><li>■ Incident exposure</li><li>■ Remote site access</li><li>■ Unsafe machinery usage</li></ul>	<ul style="list-style-type: none"><li>■ Defect detection</li><li>■ Production line inspection</li><li>■ Product scanning and weighing</li></ul>	<ul style="list-style-type: none"><li>■ Digital asset management</li><li>■ Digital twin development</li><li>■ Product inventory</li></ul>	

Source: Gartner (February 2022)

Video surveillance has been around for quite some time, providing security measures for various purposes. Incorporating CV techniques has automated several components of the process and incorporates mechanisms to improve security protocols. In a scenario where building security is critical, such as secure government facilities or confidential and proprietary engineering rooms, CV techniques can play a vital role. Figure 8 highlights five areas where CV can automate tasks and improve overall security for an intelligent video surveillance use case.

Figure 8. Intelligent Video Surveillance Sample Process

## Intelligent Video Surveillance Sample Process



Source: Gartner  
749354\_C

Gartner

In the intelligent video surveillance process:

1. Video from surveillance cameras of secure areas within physical locations is fed into edge servers at each location. Partial preprocessing techniques of cleaning and labeling are conducted before sending to a cloud data repository.
2. Remaining cleaning and labeling are completed in the cloud data repository. Video feeds are also parsed into frames and formatted in preparation for training the CV models.

3. Two sets of data are used for CV processing. Pretrained data that includes video frames of people's actions in a building setting and the new preprocessed data. *Object detection* identifies each object in the frames and classifies them — adding a class for humans. *Object tracking* uses the human class, tracks the objects, and identifies the difference between the background and human objects. Human objects are differentiated with their own unique identifier using *instance segmentation*. *Panoptic segmentation* is then used to identify every pixel in each frame to safeguard from any false alarms. The final inference is fed into the digital twin and, in some cases, retrained or reprocessed.
4. A digital twin representation of each building provides digital boundary lines where *human* objects are not allowed to breach. If the *object tracking* inference recognizes a breach of a boundary line, a security analyst is automatically notified.
5. The security analyst, the human in the loop, quickly validates whether the breach is legitimate. In the event the breach is real, an alert is sent to physical security at the building where the breach occurred. If deemed necessary, the security analyst can also send an alert to local law enforcement.

## Exploring the Future of Computer Vision

The world of CV is exciting, and technical capabilities are advancing. CV can be used in almost every corner of the digital business, and unstructured data continues to be tapped into for deep analysis. A few key areas will see an increase in CV capabilities and integration over the next several years and beyond.

**Digital twin.** Several goals of a digital twin are the ability to feed it information, make adjustments, determine thresholds and apply modifications in a digital environment before making massive changes to physical objects, personas or processes. CV object detection and object tracking can supply a significant set of real-world information to a digital twin for increased precision. Take airport operations as an example. Airplanes, people and support vehicles can be tracked using images, videos and sensors. Historical data collected for each of these objects can be fed into a digital twin. Operational managers can use the digital twin, along with the historical CV data, to identify inefficiencies within the operational process. They can also replay and simulate certain scenarios, such as weather conditions, to determine potential impact to operations.

**Simulation.** Successful and accurate CV solutions thrive on high-quality, pixel-perfect data. Images or videos coming from low-tech equipment hardly produce the quality files needed for increased accuracy. Using CV image reconstruction techniques can improve image quality by cleaning imperfections in each and every pixel — panoptic segmentation, as described previously. Combining this with simulation data from 3D rendering software advances panoptic segmentation even further. The outcome of these approaches is considered synthetic data (see [Maverick\\* Research: Forget About Your Real Data — Synthetic Data Is the Future of AI](#)). Cleaner and high-quality synthetic data will increase accuracy in your CV solutions.

**Deep fakes.** CV re-creates and replaces the original video with precise visual facial features from another video or set of images. The video is then reconstructed frame by frame with the new set of images. Audio files are re-created with a similar approach and also added to the final video. By using pixel analysis, as seen with panoptic segmentation, every pixel of an image or video can be updated to whatever the developer wishes. A popular example of a deep fake is presented in a TODAY segment, [‘Deep Fakes’ Are Becoming More Realistic Thanks to New Technology](#).<sup>3</sup> Although a comical illustration, deep fakes are raising many ethical questions, arousing substantial attention from government agencies across the world because of its potential misuse.

**Metaverse.** There has been an overwhelming hype around the metaverse (see [Quick Answer: What Is a Metaverse?](#) for more details). Essentially, a metaverse will combine capabilities from many avenues to create an interactive virtual environment. CV techniques combined with some form of digital twins will facilitate real-time virtual interaction, as with the earlier example of the no-scan shopping experience, where a customer utilizes VR to shop in a digital twin representation of a physical grocery store. Several retailers have stepped into the metaverse space with their version of a virtual commerce experience. View GMA's segment, [Walmart Steps Into Metaverse to Sell Virtual Goods](#), for a glimpse of how retailers may use the metaverse.<sup>4</sup>

## Recommendations

Digitalizing the replication of human cognitive abilities is not easy. Data and analytics technical professionals will be challenged with developing infrastructure to adequately process large amounts of unstructured data — sometimes in near-real or real time. Solutions will have to incorporate complex deep learning techniques to ensure effective results and then operationalize them into end-to-end pipelines.

Use the following recommendations to navigate through the noise and adequately plan a CV implementation:



- **Evaluate annotation requirements before kick-starting a CV project.** To be effective, CV requires massive amounts of labeled data. Most unstructured data is retrieved from devices that produce low-quality images or videos that are not properly labeled. Understanding these requirements upfront will assist with making a decision to move forward with the project or not.
- **Integrate your CV solution with a hardware-intensive data management and analytics infrastructure.** Not only will you have to collect, label and update unstructured data, but you will also have to run complex deep learning models. Using a robust, scalable data management and analytics environment will increase efficiencies in preprocessing and operationalizing CV initiatives. It will also consolidate data that can be used by other business applications or analytical solutions like data science or business intelligence.
- **Include a comprehensive CV deployment strategy with requirements and design.** Final CV inferencing will be delivered to one or multiple endpoints. A miscalculation of the final inference deployment can be catastrophic; for example, delivering incorrect information to a customer's mobile device or alerting security unintentionally. Additional time incorporating detailed deployment information to requirements will reduce implementation risks.
- **Familiarize yourself with local and global legal and policy laws while evaluating CV use cases.** Capturing images or videos of people may interfere with certain privacy laws, which could result in expensive lawsuits or unwanted publicity. Facial recognition software has been challenged in many countries with massive amounts of social backlash. Additionally, General Data Protection Regulation (GDPR) and other regulations have been enacted, giving customers more power to request their personal information scrubbed from data environments. Imagine searching for one individual in thousands, or maybe millions, of images used for training a CV algorithm. Know the law and potential impact before initiating the use case.
- **Assess capabilities and integration among vendor use-case-specific solutions.** If a vendor solution is the option selected, ensure it has full CV capabilities and integrates with other applications and the final deployment endpoints. Ask the vendor for these details, and request a full demo of the product before making financial decisions. To get started, a randomly selected list of five vendors is represented in Table 5; see [Emerging Technologies: Tech Innovators for Computer Vision](#) for more.

Table 5: Computer Vision Vendors

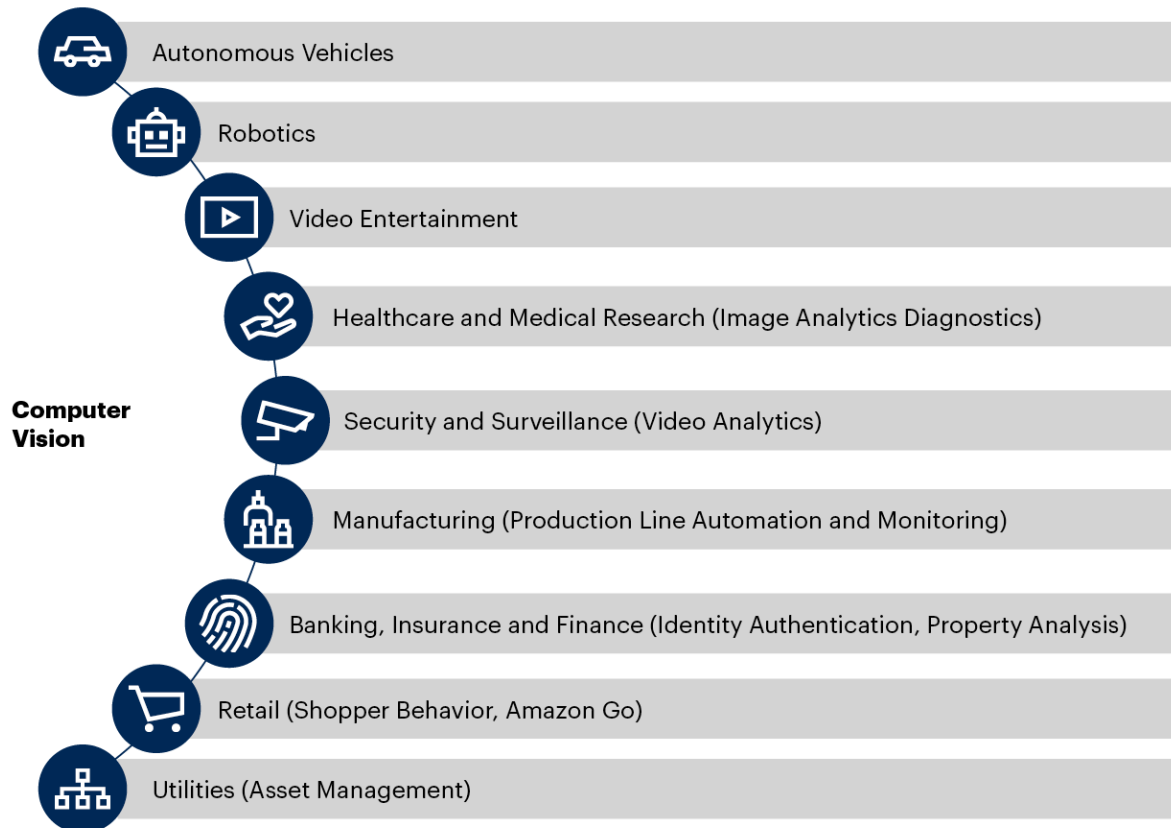
Vendor	Innovation
<a href="#">Arcarithm</a>	Pioneering application of augmented and synthetic data generation to train and test drone and weapon detection systems at military grade, cost-effectively.
<a href="#">Binah.ai</a>	Breakthrough remote detection and analysis of a human's vital signs via a smartphone camera.
<a href="#">Deep Recognition</a>	Flexible platform architectures and services that enable users to massively scale video analytics to deliver complex object and activity recognition.
<a href="#">RE'FLEKT</a>	Revolutionizes augmented reality experiences for frontline workers in an industrial environment.
<a href="#">Tencent</a>	AI and autofusion technology is used to generate simulated data for training diagnostic models for motor function disease detection.

Source: Gartner (February 2022)

## Conclusion

Preprocessing and operationalizing steps illustrate a high-level process. Individual use cases deviate from this process slightly, but enough to force vendors to create use-case-specific solutions and force customers to decide which solution to implement. For instance, object detection for X-ray analysis to identify cancer will be drastically different from object detection for mobile device facial recognition. Both use cases require different steps in annotation, training and more than likely, infrastructure. Figure 9 represents a list of sample use cases for CV.

Figure 9. Sample Use Cases for Computer Vision

**Sample Use Cases for CV**

Source: Gartner  
735994\_C

Gartner

Examine the potential steps required for each of these samples, and determine how they may relate to your use case. Follow the preprocessing and operationalizing steps outlined in this research to conclude whether a prebuilt use-case solution is feasible or if a fully developed in-house option satisfies the requirements.

**Evidence**

<sup>1</sup> [Deep Photo Enhancer: Unpaired Learning for Image Enhancement From Photographs With GANs](#), GitHub.

<sup>2</sup> [IBM Developer Model Asset Exchange: Image Resolution Enhancer](#), GitHub.

<sup>3</sup> [‘Deep Fakes’ Are Becoming More Realistic Thanks to New Technology](#), TODAY, 28 December 2021, YouTube.

<sup>4</sup> [Walmart Steps Into Metaverse to Sell Virtual Goods](#), GMA, 18 January 2022, YouTube.

---

## Recommended by the Author

Some documents may not be available as part of your current Gartner subscription.

[Emerging Technologies: Tech Innovators for Computer Vision](#)

[Emerging Technologies: Top Use Cases for Edge AI Computer Vision](#)

[Emerging Technologies: Top Edge AI Use Cases in Manufacturing Industries](#)

[Emerging Technologies: Top Use Cases in Machine Vision](#)

[Hype Cycle for Artificial Intelligence, 2021](#)

[Emerging Technologies: Top Advanced Computer Vision Use Cases for Retail](#)

[Emerging Technologies Research Roundup: Machine Vision Drives Business Value](#)

---





© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)."

Table 1: Public Datasets

Dataset	Number of Files	Description
<a href="#">V7Labs COVID-19 X-Ray Dataset on GitHub</a>	6,500 images	Anterior to posterior (AP) and posterior to anterior (PA) chest X-rays, with pixel-level polygonal lung segmentations.
<a href="#">KITTI-360</a>	320,000 images 100,000 laser scans	Large-scale dataset containing rich sensory information and full annotations. Recorded in several suburbs of Karlsruhe, Germany. Semantic and instance annotation for both 3D point clouds and 2D images.
<a href="#">Common Objects in Context (COCO)</a>	330K images >200K labeled images	Large-scale object detection, segmentation and captioning dataset.
<a href="#">ImageNet</a>	14 million images 21,000 synsets indexed	Image database organized according to the WordNet hierarchy (nouns only), in which each node of the hierarchy is depicted by hundreds and thousands of images.
This is a representative, not exhaustive, list of public datasets.		






Source: Gartner (February 2022)

Table 2: Annotation Options and Risks

	Human in the Loop (HITL)			ML Automation
Option	 Outsource	 Crowdsource	 In-House	 Programmatic
Risks	<ul style="list-style-type: none"> <li>■ High cost</li> <li>■ Time-consuming</li> <li>■ Secure data sharing required</li> </ul>	<ul style="list-style-type: none"> <li>■ Medium cost</li> <li>■ Difficult to manage</li> <li>■ Multiple external partner coordination</li> </ul>	<ul style="list-style-type: none"> <li>■ High cost</li> <li>■ Time commitments</li> <li>■ Internal skills required</li> </ul>	<ul style="list-style-type: none"> <li>■ Medium cost</li> <li>■ ML skills required</li> <li>■ Application integration issues</li> </ul>






Source: Gartner (February 2022)

**Table 3: Image Analysis Use Cases**

 Image Analysis	 Tagging	 Detecting	 Categorizing	 Describing
Description	Identify and tag visual features in an image.	Detect objects or faces in an image.	Categorize tags or pixels of an image into classes.	Describe an image into a readable format.
Example Use Cases	<ul style="list-style-type: none"><li>■ Object tracking</li><li>■ Facial expression</li><li>■ Advertisement analysis</li><li>■ SEO product tagging</li></ul>	<ul style="list-style-type: none"><li>■ Facial recognition</li><li>■ Object identification</li><li>■ Visual search</li><li>■ Biometric pattern recognition</li></ul>	<ul style="list-style-type: none"><li>■ Product hierarchy detection</li><li>■ No-scan check-out</li><li>■ Object and people grouping</li></ul>	<ul style="list-style-type: none"><li>■ Help desk support</li><li>■ Chatbot interaction</li><li>■ Language transcription</li><li>■ Damage evaluation</li></ul>

Source: Gartner (February 2022)

Table 4: Video Analytics Use Cases

 Video Analytics	 Security	 Safety	 Process	 Assets
Description	Monitor environments and issue alerts.	Identify safety or noncompliance concerns.	Evaluate objects for defects or anomalies.	Scan physical objects for inventory control.
Example Use Cases	<ul style="list-style-type: none"><li>■ Threat alerting</li><li>■ Crowd control monitoring</li><li>■ Hazardous driving detection</li></ul>	<ul style="list-style-type: none"><li>■ Hygiene testing</li><li>■ Incident exposure</li><li>■ Remote site access</li><li>■ Unsafe machinery usage</li></ul>	<ul style="list-style-type: none"><li>■ Defect detection</li><li>■ Production line inspection</li><li>■ Product scanning and weighing</li></ul>	<ul style="list-style-type: none"><li>■ Digital asset management</li><li>■ Digital twin development</li><li>■ Product inventory</li></ul>

Source: Gartner (February 2022)



Table 5: Computer Vision Vendors

Vendor	Innovation
<a href="#">Arcarithm</a>	Pioneering application of augmented and synthetic data generation to train and test drone and weapon detection systems at military grade, cost-effectively.
<a href="#">Binah.ai</a>	Breakthrough remote detection and analysis of a human's vital signs via a smartphone camera.
<a href="#">Deep Recognition</a>	Flexible platform architectures and services that enable users to massively scale video analytics to deliver complex object and activity recognition.
<a href="#">RE'FLEKT</a>	Revolutionizes augmented reality experiences for frontline workers in an industrial environment.
<a href="#">Tencent</a>	AI and autofusion technology is used to generate simulated data for training diagnostic models for motor function disease detection.

Source: Gartner (February 2022)