Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-365-1_4

Chapter 4: Exploring the intersection of computer vision, artificial intelligence at the edge, and IoT

Priyambada Swain

Infosys Limited

1 Introduction: Introduction to Computer Vision and AI at the Edge

Recent advances in computer vision are paving the way for smarter artificial intelligence at the edge. These advances are rapidly enhancing the capacity of IoT devices to analyze images beyond basic recognition tasks. Object detection techniques enable cameras to identify not only the presence of objects but also their precise locations, while image segmentation can differentiate between the object and its background [1-3]. Furthermore, generative models possess the ability to create images based on learned examples. Object detection involves locating objects within an image and classifying them. Originally implemented through traditional image processing techniques, state-of-the-art methods now rely on deep-learning-based techniques [2,4,5]. Image segmentation divides an image into superpixels—meaningful regions or objects. By focusing on the primary objects, it simplifies the background. Semantic segmentation classifies each pixel into a category, while instance segmentation classifies pixels belonging to different instances of the same class. Generative models are capable of creating images that resemble those from the training dataset. To fully grasp the importance of these techniques, it is essential to understand the challenges of AI at the edge and the principles behind IoT.

2. Fundamentals of Internet of Things (IoT)

The Internet of Things (IoT) is an interconnected network of physical objects embedded with sensors, software, and hardware infrastructure. These connected things not only sense and collect data but also communicate it for monitoring, control, and analysis. IoT devices come in a variety of forms within varying industries and domains, including smart homes, connected cars, industrial applications, and many more. The IoT value chain connects multiple industry segments and areas of expertise for effective functioning and uptake [6-8]. IoT deployments comprise the things that generate, act upon and process data, the networking infrastructure, and the IoT hub and platform that correlate and augment the data. A common pattern emerges in most of the industrial and consumer IoT applications: the connected things collect telemetry data through sensors and cameras and transmit the data for further processing and analysis. The analysis often involves combining the data with other sources of information, including historical data and data coming from other related connected things.

The major problem with performing data transmission and processing in centralized locations such as the cloud or traditional data centers is the increasing latency and bandwidth usage generated by the volume of data sent from connected things. Traditional data centers are not ideally suited for real-time control, as the request from control equipment to the data center and the data center's response both have different hops to handle through the private or public networks [9,10]. The long distance between the compute power and the control equipment will generate additional latency and affecting the real-time control and reaction of the system. One effort to overcome this challenge is performing data processing closer to the sensors, instead of in centralized locations. This physical proximity of processing resources to the connected things generating or acting upon the data has led to the terms "mobile cloud computing" and "cloudlet". Bringing the AI models—specifically the AI inference—to the edge where the data is captured is "AI at the edge". Typically, the IoT is associated with definitions for an ecosystem of connected physical things that can collect, communicate, analyze and act on data [11-13]. IoT uses data created by connected sensors and IoT devices, along with AI models, to add business insight; AI-at-the-edge is the delivery of such AI models—specifically for inference—to the edge of an enterprise or service provider network or into an end-user device.

3. Advancements in Object Detection Techniques

Object Detection — which entails localization-plus-identification — has always enticed researchers and developers. Traditional detection algorithms use a sliding-window method for generating bounding box proposals. These proposals are classified using extracted features such as Haar, HOG, CIFAR. Deep Learning also enables object detection by combining the detection and classification steps, producing rich and robust feature maps for class prediction [2,14-17]. Although it achieved higher accuracy and precision, its computations for training and inference make the end system slower. Still, resource-hungry and latency-sensitive applications require the feasibility of Video Analytics with inference-on-the-edge, and this conflict leaves space for innovation. Researchers have developed custom object-detection CNN architectures and other tricks to enhance performance on edge devices.

Today's world runs on connectedness, and Computer Vision forms an integral part of this equation — cameras in diverse locations generate a vast amount of real-time intelligence, yet real-time decisions are lacking due to computational cost [9,18-21]. The growing Internet of Things (IoT) ecosystem has advanced Smart Surveillance with automation requirements, but Smart Cities are equally eager for Object Detection to supply smart data for traffic lights, accident alerts, and warning indicators. Object Detection's high accuracy and precision also make it suitable for imaging in neurology, gastroenterology, and medical robotics. A related fundamental task, Image Segmentation, generates predictions pixel-by-pixel on the image and finds application in disease diagnosis and treatment, environmental health assessment, and prosthetics design.

3.1. Deep Learning Approaches

Object detection is the task of locating notable instances of predefined classes within an input image. Much research in object detection has focused on bounding box detection, which involves detecting objects within the surrounding rectangle, rather than specifically outlining the boundaries to encompass its entirety [22,23]. This is in contrast to object segmentation, which involves outlining the actual boundaries of the detected objects.

Advancements in object detection have relied heavily on deep learning models, notably DCNNs such as AlexNet, VGG, GoogleNet, and the Residual Net. Faster R-CNN pioneered a two-stage approach with dedicated proposal and detection subnetworks. Single-shot-detection (SSD) and You Only Look Once (YOLO) employ a single-stage detection model for real-time constraints. More recently, RetinaNet incorporated the focal loss function to mitigate the class imbalance encountered during single-stage object detection. These developments facilitated accurate, real-time object detection in IoT environments.

3.2. Traditional Methods

Handcrafted algorithms played an important role in object detection before deep learning became popular. The idea of handcrafted features is mimicking the human visual cortex, where lower-level neurons act as edge and bar detectors. Filter banks are applied against an image to extract low-level features that encode the image in different ways [24-26]. Some common filter banks include: the Gabor filter bank, Gaussian derivatives, and the Leung-Malik filter bank.

One of the most standard filter banks is the Gabor filter bank, which consists of multiple even and odd symmetric Gabor filters at different scales and orientations. This filter bank is biologically inspired and mimics the orientation columns in the primary visual cortex. Other feature extractors, such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG), create histograms based on the output of these filters. HOG is generally used as a feature descriptor for shape, while SIFT is often used to detect interest points in images.

3.3. Real-Time Object Detection

In recent years, the field of object detection has seen significant progress. Research is currently focused on the detection of a large number of classes, the detection of tiny objects, and real-time object detection. Deep learning approaches outperform traditional computer vision methods on every objective. Nevertheless, traditional approaches such as Template Matching, which is used as a baseline, continue to be relevant due to their low inference time and simplicity.

Object detection applied to the Internet of Things must account for the latency implied by inference. Devices based on Raspberry Pi, for example, are limited in power supply, size, and cost, making them attractive candidates for smart devices that rely on computer vision. They are suited to the development of integrated systems capable of controlling object detection with the possibility of managing other sensors.

4. Image Segmentation: Techniques and Applications

Image segmentation involves dividing an image into multiple regions or sections to simplify and facilitate solving various computer vision problems [27,28]. This process can extract distinct objects within images and describe the relationship between them.

Techniques include semantic segmentation, which partitions images into semantically memorable parts, and instance segmentation, which locates and classifies all objects in the image at the pixel level. Current research also explores simultaneous 3D reconstruction, pose estimation, and image labeling for images depicting multiple persons.

Like most computer vision problems, image segmentation has witnessed the introduction of many CNN-based algorithms. In healthcare, for example, image segmentation can assist in the early detection and classification of diseases by segmenting medical images. In computed tomography, CT images of lung cancer patients can be thus segmented to help assess the progress of the disease during treatment. The segmented images can be used to analyze variations in volume, size, and position of the lesions during radiation therapy, which can help predict the patients' response to the therapy, bringing about significant improvements in diagnosis and treatment planning.

4.1. Semantic Segmentation

Semantic segmentation is the process of grouping similar pixels in an image into a set of classes. Unlike object detection, segmentation indicates precisely the location of the object or person in the image. In segmentation, each pixel in the image is allocated a class label, and hence, each pixel of the image is classified. Taking the previous example of the street scene in object detection, the segmentation output would be the group of pixels belonging to the car, group of pixels belonging to the road, and so on. The main types of segmentation used in computer vision are semantic and instance segmentation.

Semantic segmentation classifies each pixel in the image belonging to a particular class, for example, car, road, building, and so on. Instance segmentation distinguishes different objects in the image by classifying each pixel belonging to a particular object. For example, in a street scene, two cars that differ from each other but belong to the same category will be assigned different labels [19,29-31]. Hence, instance segmentation is also called panoptic segmentation. Semantic segmentation has found widespread use in medical imaging. Broadly segments like the heart, the lungs, the spine, etc. are predicted. Specific organs or cell segmentation like the heart, liver, red blood cells, white blood cells, etc., are also undertaken.

4.2. Instance Segmentation

Instance segmentation methods largely rely on deep learning due to their high demand for diverse training data. Fully convolutional instance-aware semantic segmentation proposes an end-to-end fully convolutional framework for accurate and efficient instance segmentation. In a different approach, fully convolutional neural networks classify each pixel of the feature representation into different instances [32,33]. The Mask-RCNN approach combines a Faster-RCNN architecture and semantic segmentation, tackling both tasks simultaneously for improved performance.

Several studies target real-time instance segmentation, often emphasizing inference speed at the expense of accuracy. For instance, YOLACT employs parallel prototype masks with per-instance mask coefficients. In the same vein, a real-time approach utilizes deconvolution layers and an upsampling mask branch for instance mask prediction. Other methods incorporate a multi-task loss function to balance bounding-box detection, mask prediction, and classification. Within these approaches, the edge detection network operates first; subsequently, an edge refinement network reconstructs boundary details, generating a precise mask for the detected object. Enhanced Residual Networks are also proposed for practical reasons, demonstrating improved real-time performance in specific datasets.

4.3. Applications in Healthcare

The medical field also profits from image segmentation. For instance, in studying tumors, the aim is to identify the tumor and see if it has been affected by cancer, so the disease might or might not spread [34-36]. If primary features in the tumor can be correctly identified, various mathematical methods can be applied on detected features for diagnosis and assessment of patients' healthcare. Semantic-level segmentation of medical images in a real-time manner would help doctors be more effective and efficient while working towards an accurate diagnosis.

Another application area in the medical field is the segmentation of large 3D medical images. In this case, the computer generates hierarchical brain segments by using radiological atlases and assesses pathological tissues and abnormalities in the images. Image segmentation is broadly used in hospital areas, from diagnostics involving neurocysticercosis and capsule endoscopy to community radiological screening tests for population health research in the healthcare field.

5. Generative Models in Computer Vision

Section 5 covers generative models in computer vision, with an overview of generative adversarial networks (GANs) and variational autoencoders (VAEs), two state-of-the-art techniques typically employed by neural networks capable of generating novel images [37-40]. Additional considerations include the underlying mechanisms of these approaches and their applications in image synthesis, thereby completing the landscape addressed by the preceding segmentation section.

Generative models present a novel approach to image synthesis. Typically, in machine learning, a model is trained to classify an input image into one of several possible categories, a challenge known as a discriminative task [41-43]. The alternative, known as a generative task, involves presenting the model with an image that belongs to a particular class and requesting it to generate novel images that belong to the same class. From a technical standpoint, generative models represent the conditional distribution $P(x|y)$ of the input image x given the class y, whereas discriminative models represent the conditional distribution $P(y|x)$ of the class y given the input image x.

5.1. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) comprise a class of neural networks conceived for unsupervised machine learning. Their core architecture features two models—the generator and the discriminator—which compete in a zero-sum game framework. The generator aims to yield realistic, synthetic data samples, while the discriminator endeavors to differentiate between genuine and artificial samples of a particular class. The rivalry between the generator and discriminator motivates the creation of increasingly realistic samples.

The initial development of generative models utilizing adversarial training is often attributed to the work of Schmidhuber. However, the seminal approach known as Generative Adversarial Networks was introduced by Goodfellow et al. in 2014. The impact of this generative framework has been substantial; numerous variants and enhancements have emerged, enabling the creation of photorealistic images of human faces, facial transformation, the generation of 3D objects, and artistic style generation. In essence, the goal of a GAN model is to generate samples that appear convincingly real to a human observer.

5.2. Variational Autoencoders (VAEs)

Deep learning has also significantly impacted generative models, as previously demonstrated in Sect. 3.4 that introduced Generative Adversarial Networks (GANs). In contrast to the discriminator-generator adversarial setup of GANs, Variational Autoencoders (VAEs) approach their objective probabilistically. Adopting the encoder-decoder framework of traditional Autoencoders, VAEs model the encoder output distribution at the latent space, rather than individual values. This probabilistic modeling facilitates the generation of new images by decoding Gaussian samples, yielding high-quality results.

Given these merits, VAEs have occasionally been employed to support image synthesis. The VAE loss comprises a supervised reconstruction component and an additional Gaussian prior. Unlike GANs—renowned as the fastest and highest-quality method for raw image synthesis—VAEs often rely on optical flow or probabilistic maps. Optical flow charts the spatial displacement of each image pixel, clarifying movement directions from one frame to the next. Probabilistic maps depict a distribution of possible next pixel positions, each assigned a likelihood, thus aiding in planning for varying scenarios. Examples of VAE application include the VMeta model for forecasting future viewpoints.

5.3. Applications in Image Synthesis

Image generation primarily employs two generative model families: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). A GAN comprises a generative model mapping encoded latent vectors to images, and an adversarial discriminator tasked with discerning real images from generated counterparts [28,44-47]. In contrast, a VAE consists of an encoder that transposes images into a latent space meant to emulate a standard normal distribution, and a decoder that reconstructs images from sampled latent vectors. The operational dynamics of VAEs and GANs differ markedly. VAEs tend to yield images exhibiting a degree of blur, attributable to the training process focused on minimizing pixel-wise deviations between reconstructions and original inputs. Conversely, GAN-generated images are rendered with considerable sharpness, a quality stemming from the discriminator's penalty on generated images it identifies as artificial.

Models in image synthesis frequently utilize a conditional mechanism, wherein the latent vector or generation process incorporates specific informational cues—such as textual descriptions or class labels—to guide outcomes [48,49]. Application domains for image synthesis encompass attribute-to-face translation, semantic image synthesis, text-to-image generation, and harmonic image completion, among others. Each area harnesses generative models tailored to fulfill particular creative or restorative objectives.

6. AI at the Edge: Challenges and Solutions

Artificial intelligence has found its way into many uses in society, including financial, health-care and others. One of the most explored areas of AI is computer vision, mainly the understanding of images. Along with this, the recent improvement of Internet of Things (IoT) brings higher connectivity and quantity of low-cost sensors [3,50-52]. These sensors, when combined, are able to identify information in the environment, connect one to another and send it to the cloud for further processing. The development of Internet Protocol version 6 (IPv6) brought the possibility of providing a unique IP addresses to countless newly created devices and the expansion of the Internet.

The AI at the edge topic indexes recent researches relating to the execution of Artificial Intelligence on edge devices, where the problems are how to fit the inference in resource-constrained devices and how to handle the latency when connecting to the cloud. The advent of Deep Learning last decade changed the scenario for Computer Vision, so many problems in the area have been tackled, especially object detection, image segmentation and generation of images generated by Generative Adversarial Networks (GANs).

6.1. Resource Constraints

Recent years have witnessed tremendous advances in computer-vision methods, in particular object-detection and image-segmentation algorithms based on deep-learning techniques. The huge demand for new applications requiring the capabilities of powerful object-detection and image-segmentation algorithms and the wide adoption of artificial-intelligence (AI) techniques by the industry have promoted the deployment of neural-network-based applications on the resource-constrained devices currently found in the so-called Internet of Things (IoT) [53-57]. These new AI-at-the-edge deployments are characterized by inference deployment on low-power, low-memory devices, which can also be subject to low-latency constraints.

IoT also demands object detection with very low detection latency that, in some cases, is impossible to achieve because images have to be sent to a remote server first. This leads to efforts devoted to the development of AI at the edge, where AI inference is performed on the resource-constrained device itself, without communicating the information to a powerful but remote server level. This provisioning of AI at the edge addresses power-consumption issues because images do not have to be sent to a remote server all the time. Regarding image segmentation, several algorithms for accurately outlining the regions of the objects of interest have emerged in recent years. Only deep-learning methods have been considered, because traditional methods are lacking in accuracy for a wide range of general-purpose applications. In particular, the most important approaches, based on a convolutional encoder–decoder architecture, are reviewed. New applications requiring the capabilities of powerful object-detection and image-segmentation algorithms, and the wide adoption of AI techniques by industry, have promoted the deployment of neural-network-based applications on the resource-constrained devices currently found in the IoT [58,59].

6.2. Latency Issues

Latency is inevitable when performing AI-related inference. For some applications—such as seated-car security or grocery-store basket calculations—latency is not a significant concern. When the results are needed in real time, such as with autonomous driving or patient diagnosis, latency is critical. Latency at the edge can be broken into two types. First, the object's distance from the edge device impacts latency due to the proximity of the edge device. Second, the capability of the edge device also has an impact on latency because each device has various specifications (e.g., memory size and processing capability);

during inference, the larger the model, the more computing power is required for the edge device to finish making its prediction in a reasonable amount of time. When an edge device has insufficient resources to perform inference, the recommendation is to use cloud services. Cloud services have the resources and capabilities to run larger deep-learning models; however, severe latency increases are experienced due to the physical distance between the cloud service and the client device. Low latency can be achieved with powerful and small models at the edge. However, the downside is that smaller models usually lead to lower accuracy.

7. Integration of Computer Vision with IoT

With the increasing ubiquity of IoT wireless connectivity and the tremendous growth of the Internet of Things, these two technologies converge in practical smart-functionality setups. The main challenges for intelligent computer-vision and IoT integration reside in the differing domains and scales of the two. Connection links and networking processes in the context of IoT need to ensure reliability, low latency, adaptability, and flexibility [2,5]. At the same time, IoT-enabled services frequently require robust decision-making, significant understanding of the surrounding environment and context, as well as supporting highly dynamic real-time interactions. Semantic information derived from video sensors can complement key technologies in other fields and serve as a new kind of IoT data. Vision sensors provide semantically rich contextual information, whereas classic sensors mainly collect physical information. Examples of prototypes successfully integrating computer vision and IoT can be found in smart surveillance systems, industrial Internet, and smart cities.

Computer vision services have many application scenarios in the smart city context. Typical applications include intelligent construction-site surveillance, traffic monitoring and control, public-safety monitoring, and smart home security. These use cases combine sophisticated decision-making with real-time responses and often rely on the integration of external information to provide better explanation and understanding of the scene. Industrial Internet of Things (IIoT) applications share some of these requirements but involve a different operating environment. Typically, responses are shorter for IIoT applications, and the environment and context are also less dynamic in general. Currently, documented case studies tend to concentrate on the integration of IoT data with a layer of domain knowledge to support rapid decision-making. Additionally, the integration of IoT and computer vision technologies enables the implementation of innovative real-world applications, such as flooding disaster prevention.

7.1. Smart Surveillance Systems

With the rapid expansion of the Internet of Things, smart surveillance systems have attracted interest from both academic institutions and industrial firms. These systems are typically comprised of smart IoT cameras that: (1) monitor a wide area, (2) capture images continuously, (3) detect unsafe or suspicious situations, and (4) send alerts immediately. Recent developments in object detection have contributed to the effectiveness and reliability of smart surveillance. However, the continuous monitoring and immediate alert requirements pose major challenges to implementations.

The first challenge derives from the tight latency requirement and limited communication bandwidth of LTE networks. Consider a large area covered by surveillance systems. When the system captures an alarm, it immediately sends the information to the emergency responder. The time for executing the object-detection algorithm, compressing and transmitting the images, and generating the response must be minimized. Typically, the network bandwidth is scarce, and so the image-data transmission time can be unacceptably long. The second challenge stems from the requirement of continuous monitoring and detection. When dangerous events are detected, the captured images or videos must be promptly transmitted to the emergency responder. However, continuous transmission of the surveillance-image/video streams can consume excessive network bandwidth. Consequently, existing surveillance systems generally either perform the analysis on the cloud, or transmit the captured images or videos to the cloud for analysis. In the former case, the system cannot meet the tight latency requirement of emergency-response services; in the latter, it cannot operate continuously. Lastly, execution of state-of-the-art object-detection algorithms on IoT cameras may exceed the computing power of fabricated smart devices.

7.2. Industrial Automation

An essential step towards Industry 4.0 is the digitization of the manufacturing process with the aid of sensors, IoT, and deep learning, among others. These transformations form a highly connected, real-time changing factory that integrates all the pieces of the manufacturing process. In this context, certain visual concepts need to be addressed, such as parcel and product identification, defect detection, person detection, and robotic grasping.

The integration of artificial intelligence systems with human work has given rise to a new concept called Industry 5.0. This approach involves systems working alongside the worker, assisting with repetitive, laborious, and physically demanding tasks, thereby improving the workers' quality of life and safety. AI cameras can also help measure the effectiveness of a production system by monitoring aspects such as product quality, worker performance, machine state, and waste. This monitoring enables predictive maintenance, minimizing the negative impacts of equipment failures. AI cameras can also measure the consumption of materials and fluids in real time, classifying them as necessary or wasteful. Information collected by AI cameras can be combined with other sensor data to respond quickly to unplanned events and ensure industrial safety.

8. Case Studies in Object Detection and Image Segmentation

The concurrent surge of Internet of Things (IoT) components, sensors, and connectivity technologies has ignited the use of computer vision. Smart surveillance, object locating and tracking, activity analysis and recognition, face recognition, autonomous vehicle, face identification, visual assistants, and various healthcare imaging applications are mostly related to object detection, image segmentation, and categorization. In recent years, object detection has become the foundation of image processing. A large number of deep learning methods for the object detection task have emerged. The recent growth in IoT devices has created an ever more connected and heterogeneous environment. The collecting, processing, and exchanging of data have changed the way people communicate and interact with each other. The growing number of detectors, sensors and generating data create massive amounts of data produced at the edge of the network. The expansion of applications for IoT systems, for instance, smart city and smart home, requires the integration of AI into IoT systems.

High demand in AI performance results in additional computing power and higher latency during the inference process to achieve the quality of service from the cloud. Compared to the on-cloud approach, AI at the edge approach decreases service latency and network costs, increases data privacy, and can finish computing even when the connection between edges and the cloud is broken. On the other hand, inference on edge devices tends to be resource-limited due to the lightweight and smaller AI model requirements. Operations on compressed, low-quality images often degrade accuracy in the computer vision task by using models trained with high-quality images. The process of creating synthetic images with a high visual quality is the fundamental problem for various image-processing applications.

8.1. Autonomous Vehicles

Autonomous Vehicles (AVs) represent a successful realization of the Internet of Things (IoT) vision. Connected to a continually growing network of other vehicles and road infrastructure, AVs are able to collect scans of their surroundings, make smart decisions, and plan their path within the environment [1,12]. Their contributions to safety, efficiency, and sustainability in urban and traffic domains are very well established, with reductions of accidents by up to 50%, fuel consumption by 15%, and increased road capacity by 25%. The principal objective of autonomous vehicles is to sense external conditions and make real-time decision-making. Driving involves a stream of decision-making, linking the past with the present to make a plan for the near future. However, in an ever-changing environment, these decisions must be made moment by moment, without a pause, accounting for the current situation and temporal references. Prediction of the future is required for the success of these decisions, whether the future is moments ahead for factoring distance-shift in vehicles, or seconds in the near future for determining where a vehicle should head in a dense highway scenario.

The primary objectives of different Advanced Driver Assistance Systems (ADAS) include maintaining the vehicle in the center of the lane, avoiding collisions with other vehicles and pedestrians, controlling the speed of the vehicle, and accurately predicting the state of the road and the presence of other external objects. However, accurate prediction of the behavior of other agents may seem as though it will only be possible in the hypothesis of a fully autonomous scenario. Yet, prediction of the future is necessary even if the vehicle is not being driven by a fully autonomous system. For example, whenever a vehicle changes a lane in a highway or city traffic scenario, it would be natural and plausible for the driver to look into the mirror to check if changing lanes is possible without causing accidents. This type of evaluation requires a sense of the future, asking the questions: will the other vehicles on the road be in this predicted position? What trajectory or path are they projecting? Why is the other vehicle decelerating in front of me? Do the pedestrians look like they are in a hurry to cross the road? Is the vehicle indicating a turn? Is the bus slowing down for pick-up or for a stop? Many of these questions require future prediction; thus, understanding the current scenario is not enough, and optimal and safe driving requires knowledge of the possible future.

8.2. Smart Cities

Smart City cities and towns use Internet of Things (IoT) sensors to collect their internal data to all share. AI at the edge can analyze the data that flows from the city and proactively take action to improve environmental conditions for its inhabitants. One area that has seen a lot of interest for Smart City applications is Object Segmentation. Semantic or Instance Segmentation of images from autonomous vehicles provides a holistic understanding of the scene. When applied to identify buildings with solar panels, environment agencies can strategize effectively on effective use of available space. Self-driving cars can avoid pedestrians for safety, and avoid potholes and other damaged parts of the road.

Smart traffic management for autonomous vehicles can also rely on instance segmentation results to detect the traffic signs and identify their status. Masks allow isolation of the traffic signs from the background, making it easier to apply pattern recognition techniques to determine the signs. Semantic segmentation can also be applied to medical images. Image segmentation is used to detect infected areas or anatomical features like blood vessels and organs.

9. Ethical Considerations in AI and Computer Vision

Computer vision and AI at the edge raise a number of ethical issues related to intrusive surveillance, racial or gender bias in algorithms, as well as other privacy and social concerns. Modern ethics require that such systems are implemented with proper security mechanisms, assure anonymity of subjects, include bias mitigation stages, and provide clear information to the public and authorities before use.

The Internet of Things and associated Industrial Internet of Things can help democratize a whole sphere of Machine Learning and Object Detection Applications in Computer Vision. These interconnect societies, cultural events, the media, social relations, and even public and industrial services in a way that would not be possible without this technology. However, the deployment of computer vision algorithms in an IoT environment requires special attention to a set of challenges such as resource-constraint devices, low latency, regulatory compliance, issues of connectivity, security, and data privacy.

10. Future Trends in Computer Vision and AI at the Edge

Artificial intelligence (AI), it will inevitably continue its conquest into other specific sub-areas of robotics and automation. A clear example of a trend that is already in motion is the implementation of AI techniques on mobile robots with a high level of self-management, better known as drones. These aerial devices find their natural applications in disaster relief, surveillance, precision agriculture, forestry, and more. Except for multifactorial disasters, the use of AI in many of these tasks is generally to facilitate the intelligent control of these robots and thus reduce the workload of human operators. However, it is essential to highlight applications in smart cities such as maintenance, infrastructure monitoring, and waste collection. Even in situations like forest fires, AI is exploited to fly the drone in a «leader-follower» formation, reaching their destination faster.

However, the AI revolution is not limited to robotic applications or automation. In the domain of smart cities or smart homes, the irresistible proliferation of diverse IoT and other mobile/portable smart devices with powerful sensing, computing, and communication capabilities has generated a new interaction pattern called the Intelligent Mobile Crowd Sensing (IMCS). The essential benefits of this interaction pattern are the low cost of sensing, the greater coverage area, and the diversity of data generated. On the other hand, the data generated by each node varies in nature due to the ad hoc deployment, heterogeneity, and autonomy of the devices. At the same time, there is a growing concern about data privacy at all levels, whether it is the transmitted information (location, bank information, preferences) or who the audience is. During this period, an urgent need for AI was born, not only to classify and analyze the massive information generated but also to address privacy and heterogeneity in the different collaboration scenarios.

10.1. Edge Computing Innovations

The concept of edge computing has existed for many years, but only recently has it gained widespread attention. The rapid growth of connected IoT devices, commonly seen in applications such as smart homes and smart cities, has led to an increase in the quantities of data, enabling new innovations based on pattern recognition. The challenge lies in implementing these tasks near the sensors, where the data is generated. Indeed, the sensors in an AIoT ecosystem provide the input necessary for AI models to learn how to recognize patterns in data, which allows IoT devices to provide decision-making support to end users. Examples include surveillance and monitoring applications in buildings, community areas, airports, coastal and marine areas, roads, and highways. Other areas where the need for high-quality image/video analysis is growing are in robots and autonomous vehicles, such as self-driving cars. In these last examples, decision-making support is crucial for the safe operation of these vehicles in publicly accessible environments.

Common image/video analysis techniques include object detection and classification, image captioning, and image segmentation. These problems have been studied for several years, with initial attempts based on traditional computer vision algorithms that rely on the mathematical and statistical analysis of image and video pixel information to detect and classify objects or areas/regions of interest. With the recent advancements of deep learning and CNNs, the state of the art has shifted toward powerful algorithms that provide much better results, but with a high compute cost. To date, the high demand for high resources means that many of the edge devices are still employed as data-gathering and sensing entities, with processing and analytics services provided close to, but not at the edge, in fog computing nodes or cloud data centers.

10.2. Evolving IoT Applications

The Internet of Things infrastructure aims to connect numerous devices, vehicles, buildings, and other consumer products for data collection and communications. Connected security cameras, VICs, and traditional motion sensors collaborate to keep homes, offices, factories, and cities secure and protected. Other prominent adjacency applications include urban noise monitoring, overpressure detection in autoseismology, and protection of natural resources, such as forests, water resources, and green areas. One primary focus is the quality of collected data.

With 24×7 active monitoring supported by dedicated sensors and data acquisition boards, the challenge is to understand the information needed for a specific security problem or needed to support decision making in complex circumstances. Sensors provide ample information, and cyber-physical systems (e.g., Internet of Robotic Things) act as sensing and performing units to complete different tasks (e.g., patrolling, protecting, helping) in the most challenging environments. Understanding the relationships between events detected by the sensors and the context of decision making to support responsive actions is fundamental when categorizing the data coming from a multitude of sources.

11. Conclusion

This examination emphasized the intersection of computer vision and AI at the edge, key enablers of the paradigm shift brought about by IoT. After introducing computer vision, IoT, and AI at the edge, particular attention was paid to object detection, image segmentation, and generative models. The specific challenges for AI at the edge, including detection and segmentation, were then discussed, as was the use and implementation of AI at the edge in an IoT environment. Real-world examples for object detection and image segmentation were highlighted.

Computer vision endows IoT sensor data with meaning and context, ushering in true artificial intelligence within the IoT paradigm. Object detection pinpoints regions of interest, enabling the system to recognize vehicles, persons, dangerous or non-compliant behavior, fires, and more. Image segmentation introduces pixel-level accuracy, further detailing current conditions. Generative models complement the two by concealing identity in camera images while still providing analytical capabilities. The constraints of the IoT landscape demand inference in near real time very close to the camera, a task fraught with challenges but critical for a broad array of IoT applications. When the sensor is also the computer, AI at the edge makes the AIoT.

References

[1] Jakhar D, Kaur I. Artificial intelligence, machine learning and deep learning: definitions and differences. Clinical and experimental dermatology. 2020 Jan 1;45(1):131-2.

[2] Siau K, Wang W. Building trust in artificial intelligence, machine learning, and robotics. Cutter business technology journal. 2018;31(2):47.

[3] Panda SP. The Evolution and Defense Against Social Engineering and Phishing Attacks. International Journal of Science and Research (IJSR). 2025 Jan 1.

[4] Kühl N, Schemmer M, Goutier M, Satzger G. Artificial intelligence and machine learning. Electronic Markets. 2022 Dec;32(4):2235-44.

[5] Karthikeyan A, Priyakumar UD. Artificial intelligence: machine learning for chemical sciences. Journal of Chemical Sciences. 2022 Mar;134(1):2.

[6] Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, Spitzer AI, Ramkumar PN. Machine learning and artificial intelligence: definitions, applications, and future directions. Current reviews in musculoskeletal medicine. 2020 Feb;13(1):69-76.

[7] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[8] Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, Spitzer AI, Ramkumar PN. Machine learning and artificial intelligence: definitions, applications, and future directions. Current reviews in musculoskeletal medicine. 2020 Feb;13(1):69-76.
[9] Panda SP. Augmented and Virtual Reality in Intelligent Systems. Available at SSRN. 2021 Apr 16.
[10] Bhat M, Rabindranath M, Chara BS, Simonetto DA. Artificial intelligence, machine learning, and deep learning in liver transplantation. Journal of hepatology. 2023 Jun 1;78(6):1216-33.
[11] Tyagi AK, Chahal P. Artificial intelligence and machine learning algorithms. InChallenges and applications for implementing machine learning in computer vision 2020 (pp. 188-219). IGI Global Scientific Publishing.
[12] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[13] Kühl N, Goutier M, Hirt R, Satzger G. Machine learning in artificial intelligence: Towards a common understanding. arXiv preprint arXiv:2004.04686. 2020 Mar 27.
[14] Rane J, Chaudhari RA, Rane NL. Data Analysis and Information Processing Frameworks for Ethical Artificial Intelligence Implementation: Machine-Learning Algorithm Validation in Clinical Research Settings. Ethical Considerations and Bias Detection in Artificial Intelligence/Machine Learning Applications. 2025 Jul 10:192.
[15] Neapolitan RE, Jiang X. Artificial intelligence: With an introduction to machine learning. CRC press; 2018 Mar 12.
[16] Rane NL, Paramesha M, Choudhary SP, Rane J. Artificial intelligence, machine learning, and deep learning for advanced business strategies: a review. Partners Universal International Innovation Journal. 2024 Jun 25;2(3):147-71.
[17] Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. Injury. 2023 May 1;54:S69-73.
[18] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular diversity. 2021 Aug;25(3):1315-60.
[19] Entezari A, Aslani A, Zahedi R, Noorollahi Y. Artificial intelligence and machine learning in energy systems: A bibliographic perspective. Energy Strategy Reviews. 2023 Jan 1;45:101017.
[20] Mohapatra PS. Artificial Intelligence and Machine Learning for Test Engineers: Concepts in Software Quality Assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27:17.
[21] Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?. The Journal of arthroplasty. 2018 Aug 1;33(8):2358-61.

[22] Ongsulee P. Artificial intelligence, machine learning and deep learning. In2017 15th international conference on ICT and knowledge engineering (ICT&KE) 2017 Nov 22 (pp. 1-6). IEEE.
[23] Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. Journal of global health. 2018 Oct 21;8(2):020303.
[24] Das S, Dey A, Pal A, Roy N. Applications of artificial intelligence in machine learning: review and prospect. International Journal of Computer Applications. 2015 Jan 1;115(9).
[25] Balyen L, Peto T. Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. The Asia-Pacific Journal of Ophthalmology. 2019 May 1;8(3):264-72.
[26] Michalski RS, Carbonell JG, Mitchell TM, editors. Machine learning: An artificial intelligence approach. Springer Science & Business Media; 2013 Apr 17.
[27] Soori M, Arezoo B, Dastres R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cognitive Robotics. 2023 Jan 1;3:54-70.
[28] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015 May 28;521(7553):452-9.
[29] Wang W, Siau K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. Journal of Database Management (JDM). 2019 Jan 1;30(1):61-79.
[30] Nichols JA, Herbert Chan HW, Baker MA. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophysical reviews. 2019 Feb 7;11(1):111-8.
[31] Jones LD, Golan D, Hanna SA, Ramachandran M. Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern?. Bone & joint research. 2018 Mar 1;7(3):223-5.
[32] Nuka ST. Next-Frontier Medical Devices and Embedded Systems: Harnessing Biomedical Engineering, Artificial Intelligence, and Cloud-Powered Big Data Analytics for Smarter Healthcare Solutions. Deep Science Publishing; 2025 Jun 6.
[33] Hyder Z, Siau K, Nah F. Artificial intelligence, machine learning, and autonomous technologies in mining industry. InResearch Anthology on Cross-Disciplinary Designs and Applications of Automation 2022 (pp. 478-492). IGI Global Scientific Publishing.
[34] Maguluri KK. Machine learning algorithms in personalized treatment planning. How Artificial Intelligence is Transforming Healthcare IT: Applications in Diagnostics, Treatment Planning, and Patient Monitoring. 2025 Jan 10:33.
[35] Mijwil MM, Aggarwal K, Sonia S, Al-Mistarehi AH, Alomari S, Gök M, Zein Alaabdin AM, Abdulrhman SH. Has the Future Started? The Current Growth of ArtificialIntelligence, Machine Learning, and Deep Learning. Iraqi Journal for Computer Science and Mathematics. 2022;3(1):13.
[36] Mich L. Artificial intelligence and machine learning. InHandbook of e-Tourism 2020 Sep 1 (pp. 1-21). Cham: Springer International Publishing.
[37] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27:163.
[38] Suura SR. Integrating Artificial Intelligence, Machine Learning, and Big Data with Genetic Testing and Genomic Medicine to Enable Earlier, Personalized Health Interventions. Deep Science Publishing; 2025 Apr 13.

[39] Panda SP. Relational, NoSQL, and Artificial Intelligence-Integrated Database Architectures: Foundations, Cloud Platforms, and Regulatory-Compliant Systems. Deep Science Publishing; 2025 Jun 22.
[40] Zhu S, Gilbert M, Chetty I, Siddiqui F. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: an analysis of the characteristics and intended use. International journal of medical informatics. 2022 Sep 1;165:104828.
[41] Mukhamediev RI, Popova Y, Kuchin Y, Zaitseva E, Kalimoldayev A, Symagulov A, Levashenko V, Abdoldina F, Gopejenko V, Yakunin K, Muhamedijeva E. Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges. Mathematics. 2022 Jul 22;10(15):2552.
[42] Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. NPJ digital medicine. 2020 Apr 7;3(1):53.
[43] Paramesha M, Rane N, Rane J. Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence (June 6, 2024). 2024 Jun 6.
[44] Arel I, Rose DC, Karnowski TP. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE computational intelligence magazine. 2010 Oct 18;5(4):13-8.
[45] Panda S. Observability in DevOps: Integrating AWS X-Ray, CloudWatch, and Open Telemetry. International Journal of Computer Application. 2025 Jan 1.
[46] Dunjko V, Briegel HJ. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Reports on Progress in Physics. 2018 Jun 19;81(7):074001.
[47] Cioffi R, Travaglioni M, Piscitelli G, Petrillo A, De Felice F. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. Sustainability. 2020 Jan 8;12(2):492.
[48] Tapeh AT, Naser MZ. Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. Archives of Computational Methods in Engineering. 2023 Jan;30(1):115-59.
[49] Panda SP. Artificial Intelligence Across Borders: Transforming Industries Through Intelligent Innovation. Deep Science Publishing; 2025 Jun 6.
[50] Theodosiou AA, Read RC. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. Journal of Infection. 2023 Oct 1;87(4):287-94.
[51] Kersting K. Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. Frontiers in big Data. 2018 Nov 19;1:6.
[52] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.
[53] Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nature Reviews Urology. 2019 Jul;16(7):391-403.
[54] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.

[55] Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. Journal of Oral Pathology & Medicine. 2020 Oct;49(9):849-56.
[56] Guo K, Yang Z, Yu CH, Buehler MJ. Artificial intelligence and machine learning in design of mechanical materials. Materials Horizons. 2021;8(4):1153-72.
[57] Panda SP. Securing 5G Critical Interfaces: A Zero Trust Approach for Next-Generation Network Resilience. In2025 12th International Conference on Information Technology (ICIT) 2025 May 27 (pp. 141-146). IEEE.
[58] Rane J, Chaudhari RA, Rane NL. Data Privacy and Information Security in Deep Learning Applications: Risk Assessment and Patient Safety Protocols for Big Data Analytics. Ethical Considerations and Bias Detection in Artificial Intelligence/Machine Learning Applications. 2025 Jul 10:54.
[59] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27:38.