

OEHI Workshop

Edge AI with Sedna

In this talk, we share our experience with Sedna and KubeEdge to deploy AI applications on heterogenous edge-cloud infrastructure. We present our multi-edge computing model which was applied to different, real-life use-cases. For demonstration, we select the object re-identification problem (ReID) as our AI application to showcase our system design, workflow, and implementation.

Edge AI with Sedna

Agenda

- Introduction
- Sedna
- Our Work with Sedna

Edge AI with Sedna

Agenda

- Introduction
- Sedna
- Our Work with Sedna

Edge AI with Sedna

About Us – Huawei MRC Edge AI Team, Sedna Contributor



Our Research

We focus on solving the challenges of managing, deploying, and monitor AI applications on an heterogenous infrastructure. To do so, we build on top of well known components such as Kubernetes, Kubeedge, and Sedna.



Edge-Cloud Synergy

We optimize the use of edge and cloud resources to support flexible deployment of AI workloads based on user requirements.



AI Applications

Supports different classes of AI applications such as object detection, classification, re-identification (ReID).



Customizable

Fully customizable by offering ML practitioners the ability to add new models to the framework and creating new ML workloads on the fly, from scratch.



Open Source

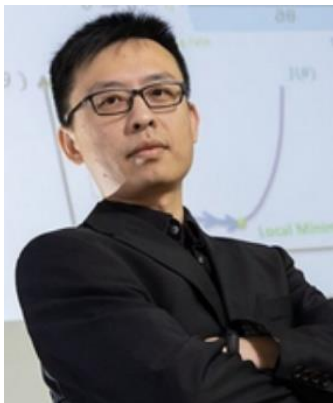
Based on fully open-source projects, our code is aswell fully available on GitHub.



Vittorio Cozzolino
PhD, Senior Software Engineer
vittorio.cozzolino@huawei.com



Zi Yang
PhD Candidate
zi.yang1@huawei.com



Haojin Yang
PhD, Scientific Advisor
yang.haojin.ext@huawei.com



Soumajit Majumder
PhD, Senior AI Researcher
soumajit.majumder@huawei.com

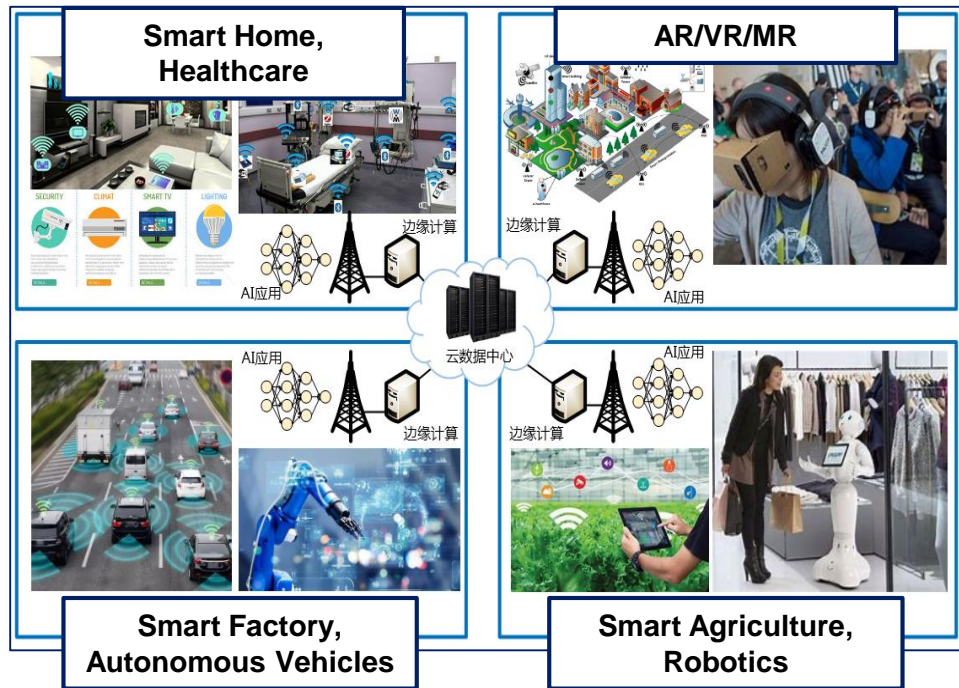
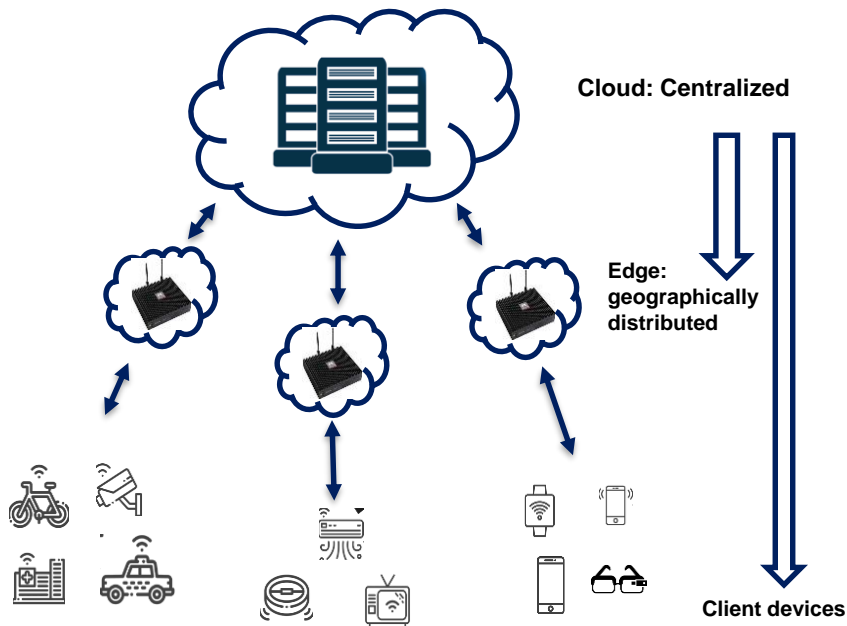


Jorge Cardoso
PhD, Chief Architect
jorge.cardoso@huawei.com

Edge AI with Sedna

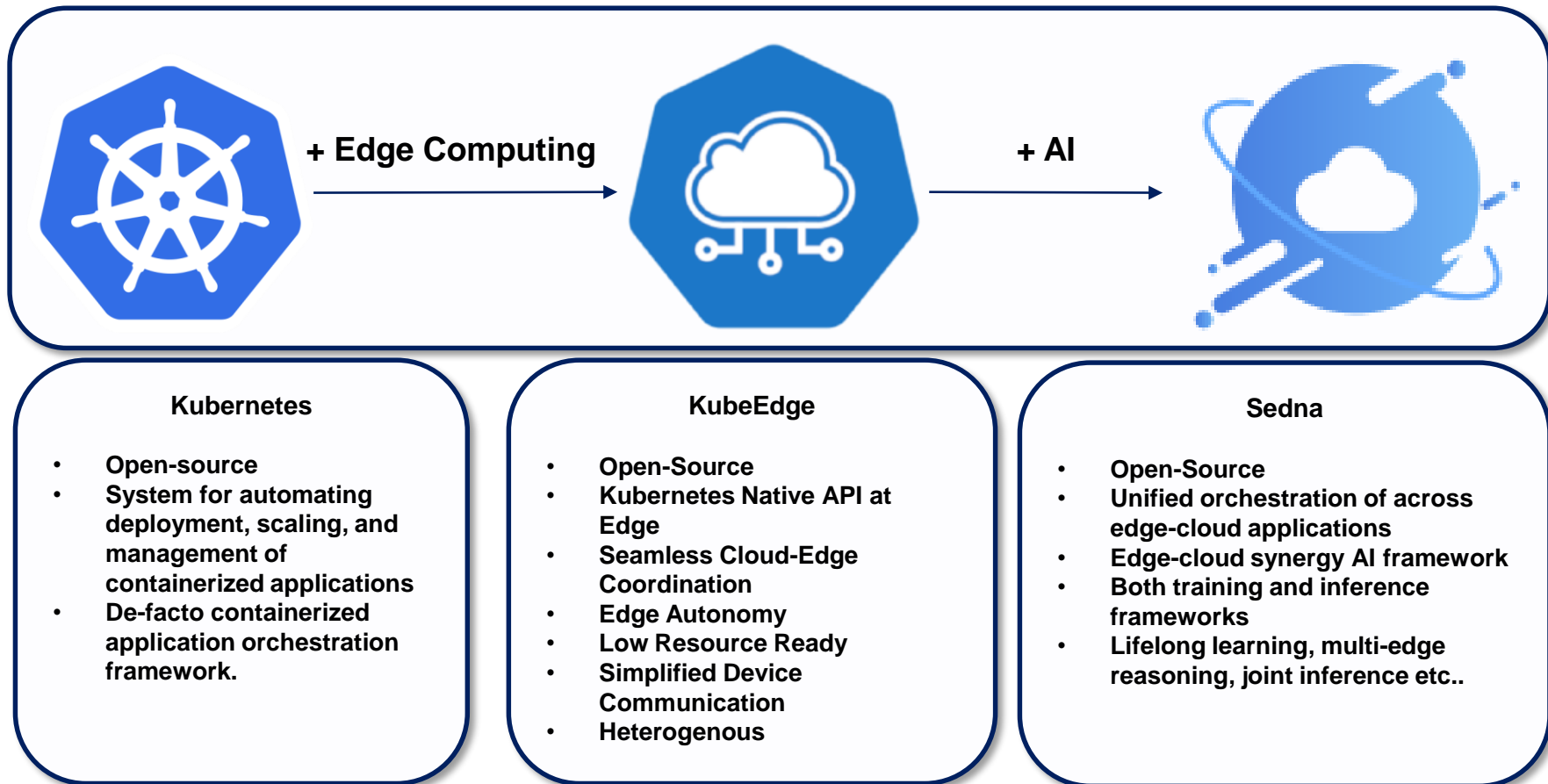
Why Edge AI?

The cloud, centralized AI computing paradigm cannot cope with the demand for real-time, accuracy, and strong interactivity of AI applications at the edge.



Edge AI with Sedna

Edge AI Building Block



Edge AI with Sedna

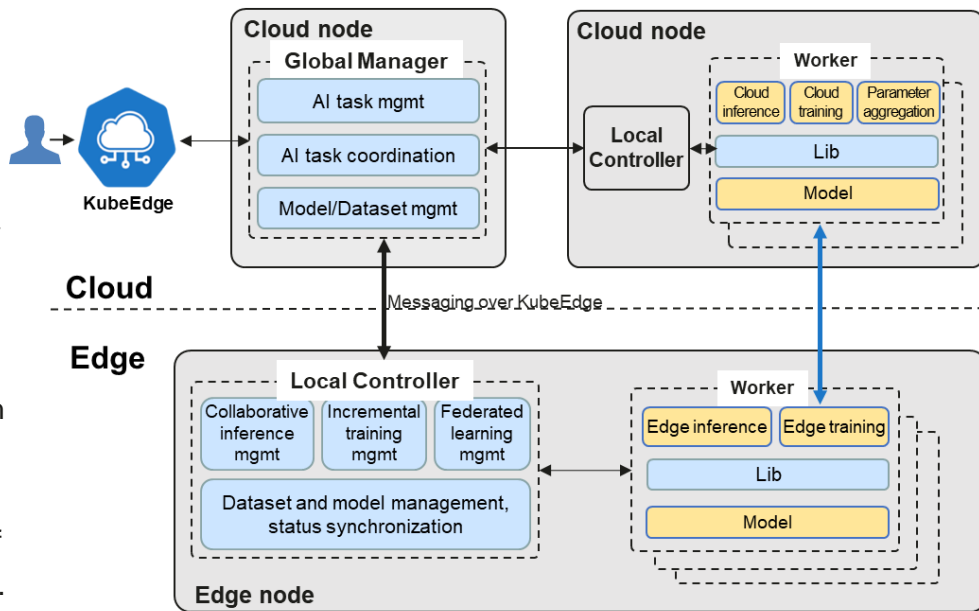
Agenda

- Introduction
- Sedna
- Our Work with Sedna

Edge AI with Sedna

Sedna – Distributed Collaborative AI

- Opensource project aimed at becoming a repository of AI models and algorithm to be easily reused and deployed in a distributed fashion.
- Based on the edge-cloud collaboration capability provided by KubeEdge (or Kubernetes).
- Supports seamless integration of existing AI applications at the edge.
- Compatible with mainstream AI frameworks such as TensorFlow, Pytorch, PaddlePaddle, and MindSpore.
- Reduces build and deployment costs by reuse of algorithms and models existing in the framework.
- Enables data privacy with local processing of sensitive data directly at the edge.



<https://github.com/kubeedge/sedna>

AI Paradigms	Compatibility
<ul style="list-style-type: none"> • Joint Inference • Incremental learning • Federated Learning • Lifelong learning • Multi-edge AI 	<ul style="list-style-type: none"> • Mainstream AI frameworks • Module algorithms • Extensible interfaces
	Framework
	<ul style="list-style-type: none"> • Dataset management • Model management •

Edge AI with Sedna

Sedna – Components

Global Manager (GM)

- Unified edge-cloud synergy AI task management
- Cross edge-cloud synergy management and collaboration
- Central configuration manager

Local Controller (LC)

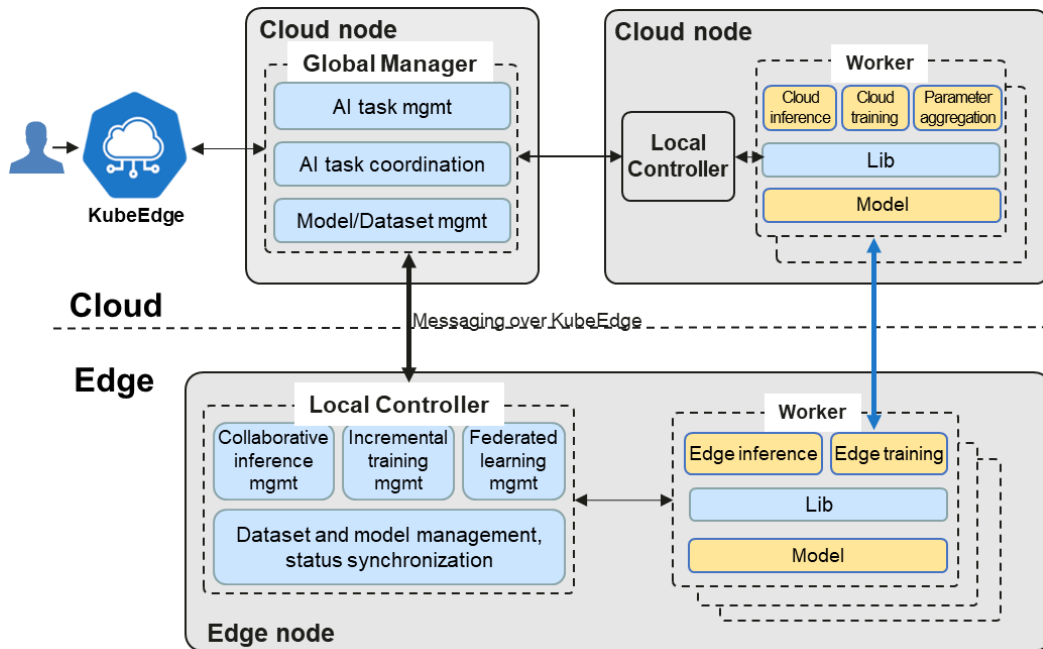
- Local process control of edge-cloud synergy AI tasks.
- Local management of models, datasets, and status synchronization.

Worker

- Do inference or training, based on existing ML framework.
- Launch on demand as Docker containers.
- Different workers for different features.
- Can run at the edge or in the cloud.

Lib

- Collection of Python classes/functions
- Exposes the Edge AI features to applications, i.e. training or inference programs



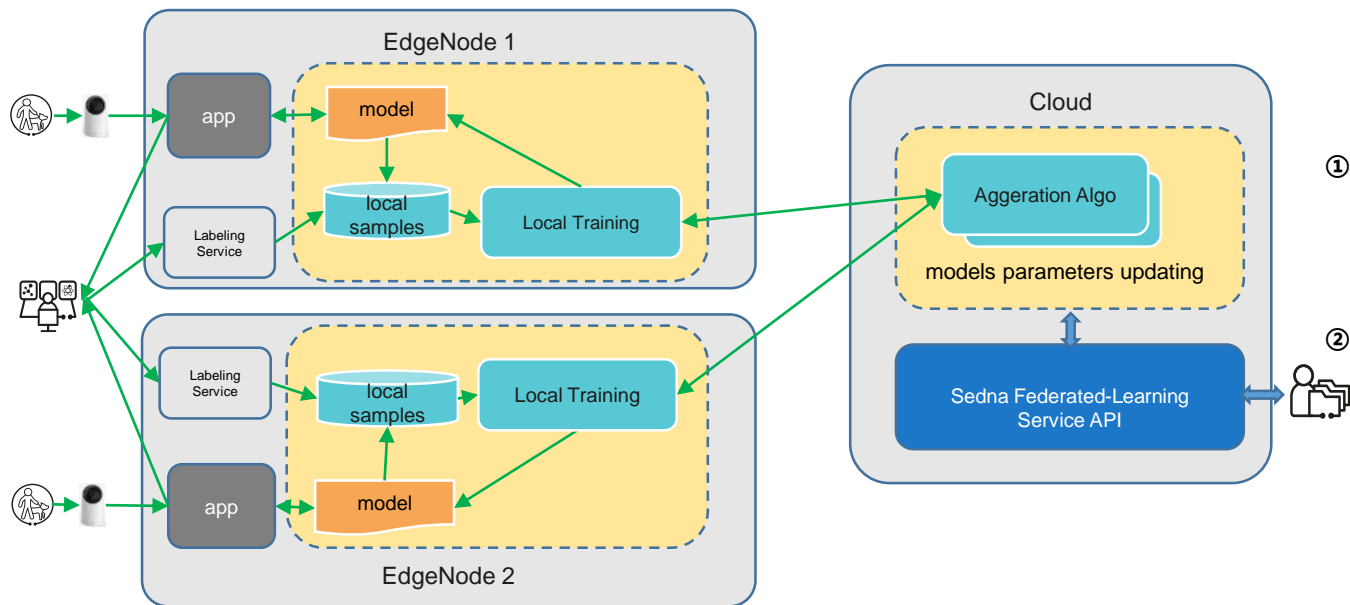
Sedna Applications: Edge-Cloud Collaborative Incremental Learning



Edge AI with Sedna

Sedna Applications: Cloud-Side Collaborative Federated Learning

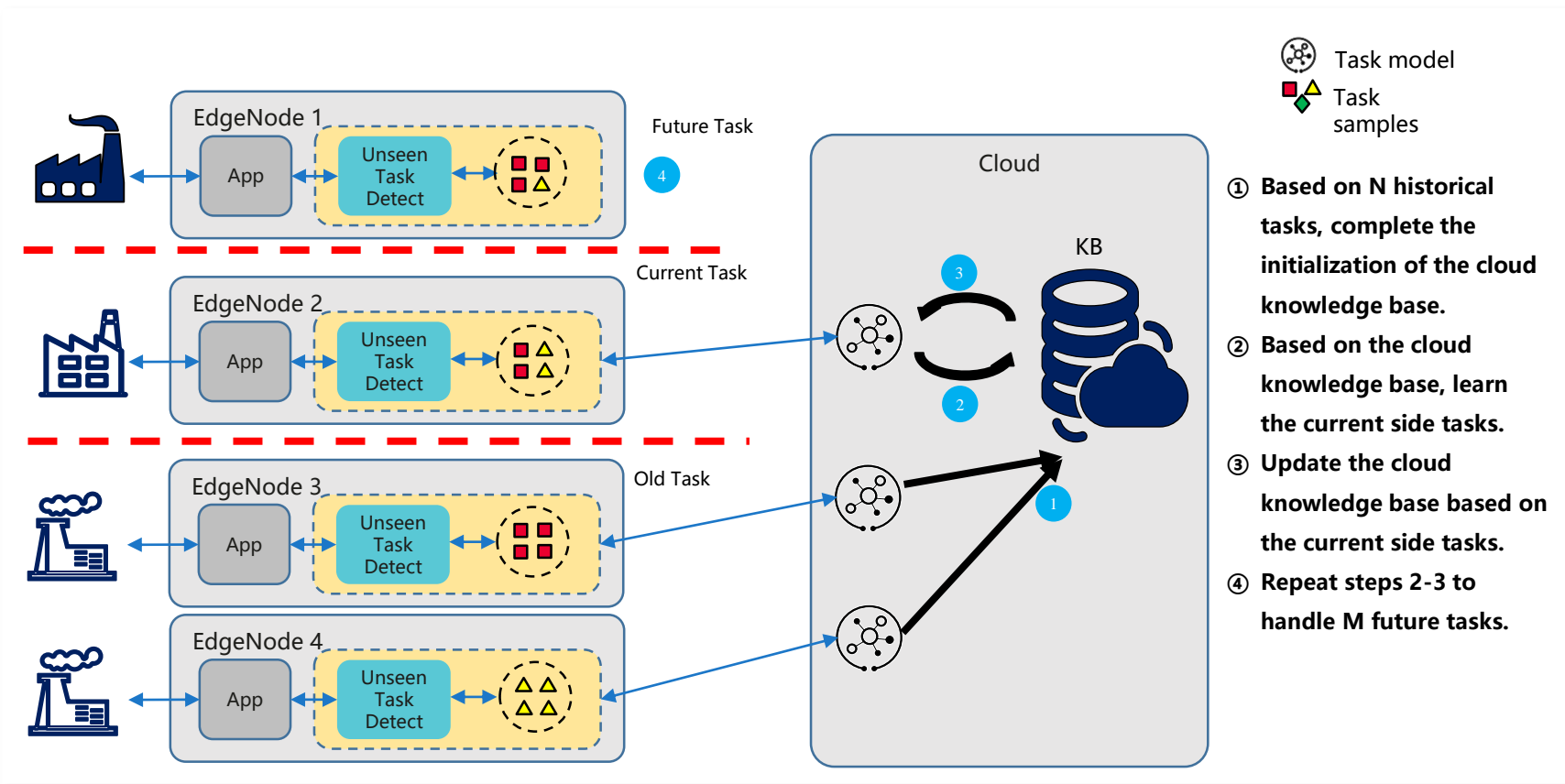
- ③ Multi-task detection, divide non-IID sample sets, and cooperate with the cloud to identify similar tasks
- ④ For local training, model parameters are uploaded to the cloud, and cross-edge migration + model aggregation algorithms are run in the cloud.



- ① Developers: Import edge AI Lib libraries and develop edge cloud collaborative federated learning programs.
- ② Launch a federated learning task and deploy the training program to the edge

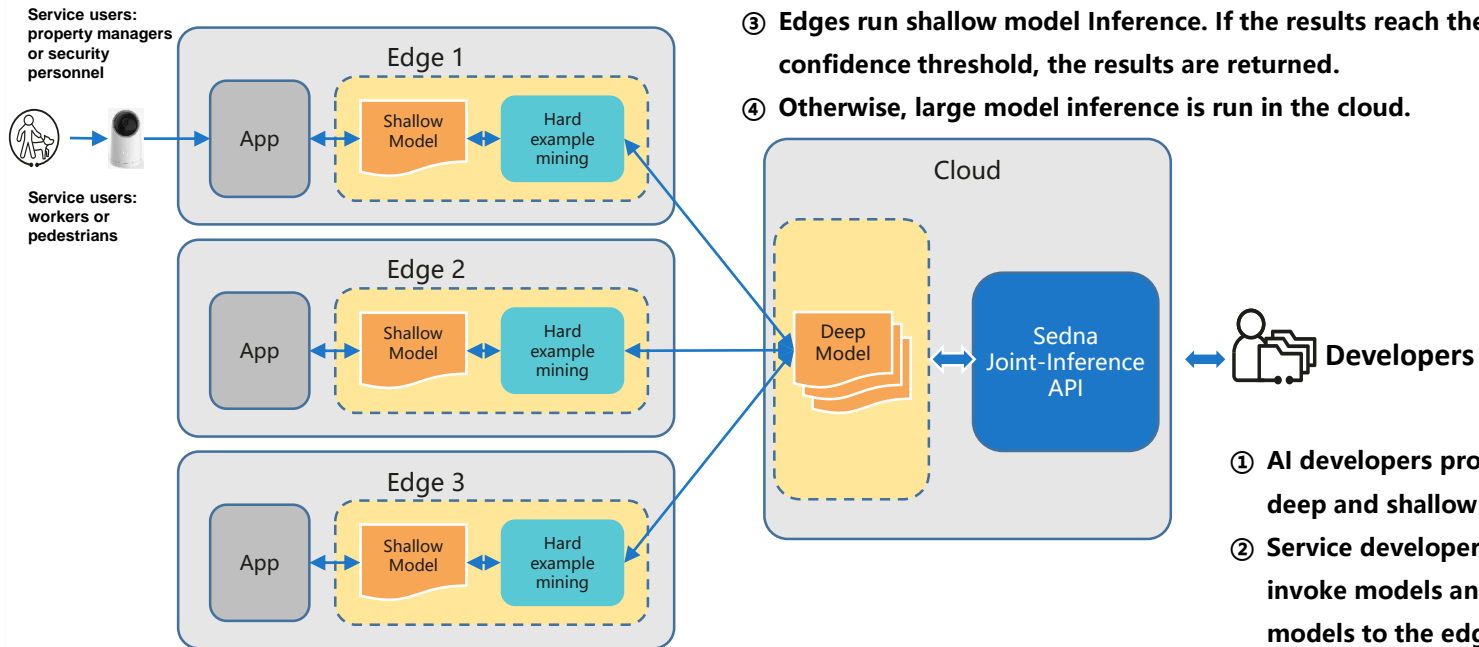
Edge AI with Sedna

Sedna Applications: Collaborative Lifelong Learning



Edge AI with Sedna

Sedna Applications: Edge-Cloud Joint Inference



Edge AI with Sedna

Agenda

- Introduction
- Sedna
- Our Work with Sedna

Edge AI with Sedna

Edge AI in Huawei MRC

Applications

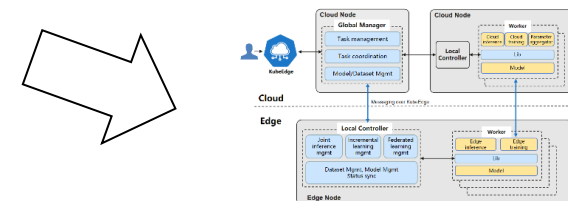
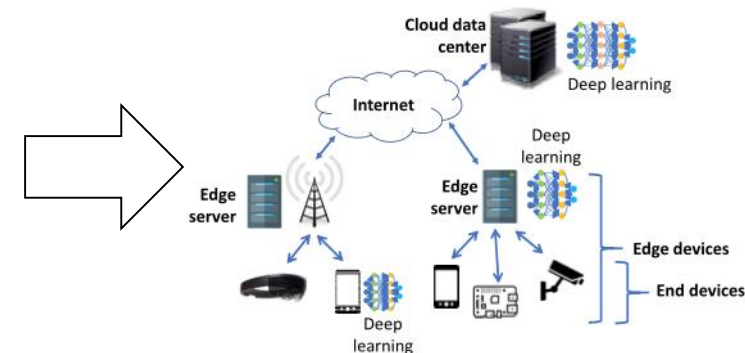
- **Object tracking, Computer vision**, Cloud Gaming, Augmented reality, and IoT use cases can be realized with edge computing architectures with machine learning (ML) techniques.
- Develop an Edge AI approach for automatic object tracking and identification in multi-camera systems.
- Comprehensively understand and analyze the use case and business problem.

System

- To support the applications, data needs to be collected, stored, processed, analyzed, and visualized across the **distributed components of edge platforms using AI**.
- Approaches are needed to support the training, deployment, and inferencing ML models in edge computing environments.
- Furthermore, ML techniques need to be optimized to provide services with low latency for end users.

Sedna Integration

- Develop a framework to systematize and drive the implementation of technology selections and to optimize end-to-end system for the computing continuum
- Translate new computing models into new functionalities of the software framework **Sedna**



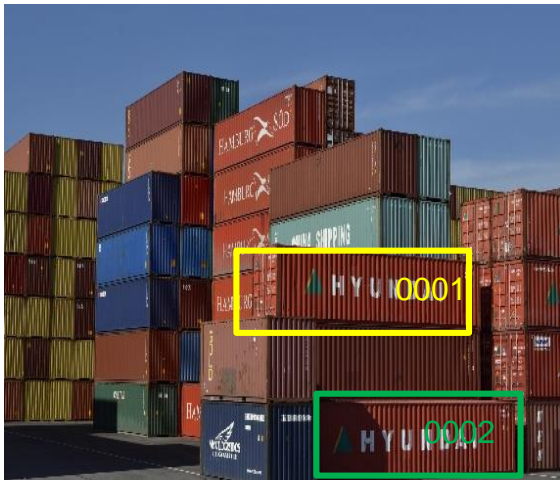
Edge AI with Sedna

Application Scenarios

Smart Port

Monitor the location of **shipment containers** in a smart port to boost logistic operations and allow quick retrieval of potentially lost goods.

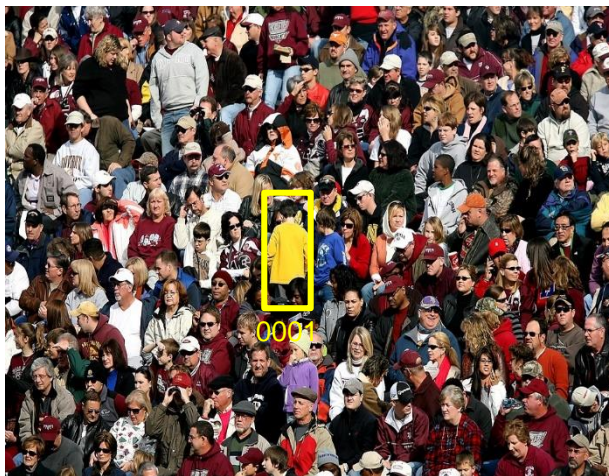
When a shipment container enters a monitored area, it's **detected**, **recognized** and **localized**.



Smart Campus

Locate **pedestrians** in a smart campus for occupancy monitoring and to offer personalized services.

When a pedestrian enters a monitored area, it's **detected**, **recognized** and **localized**.



COVID-19

Find **proximity contacts** between a potential carrier of the virus and other people to tackle the current pandemic crisis.

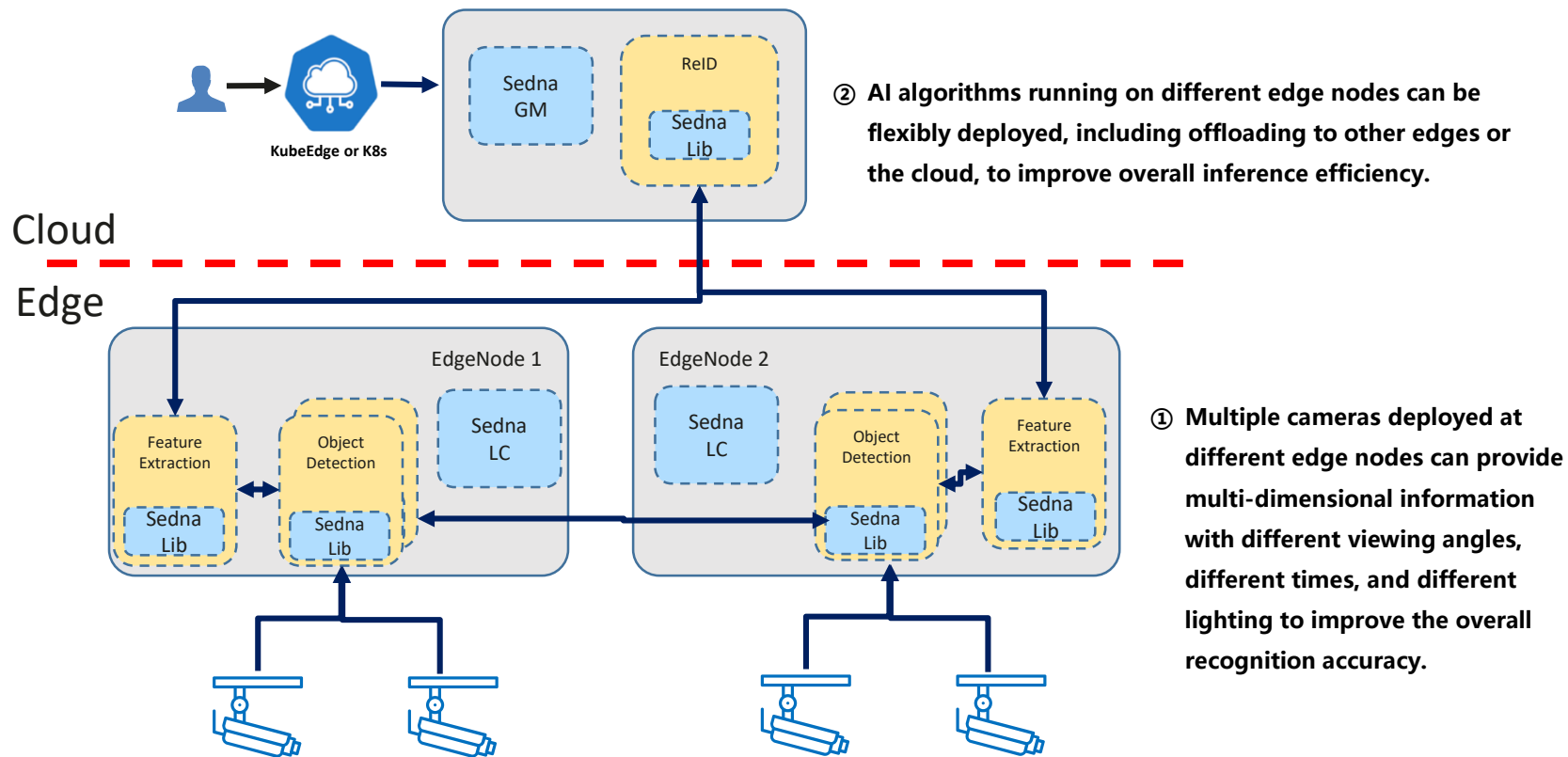
When a carrier gets in contact with someone in a monitored area, the system will identify and notify the subject of the potential risk of infection.



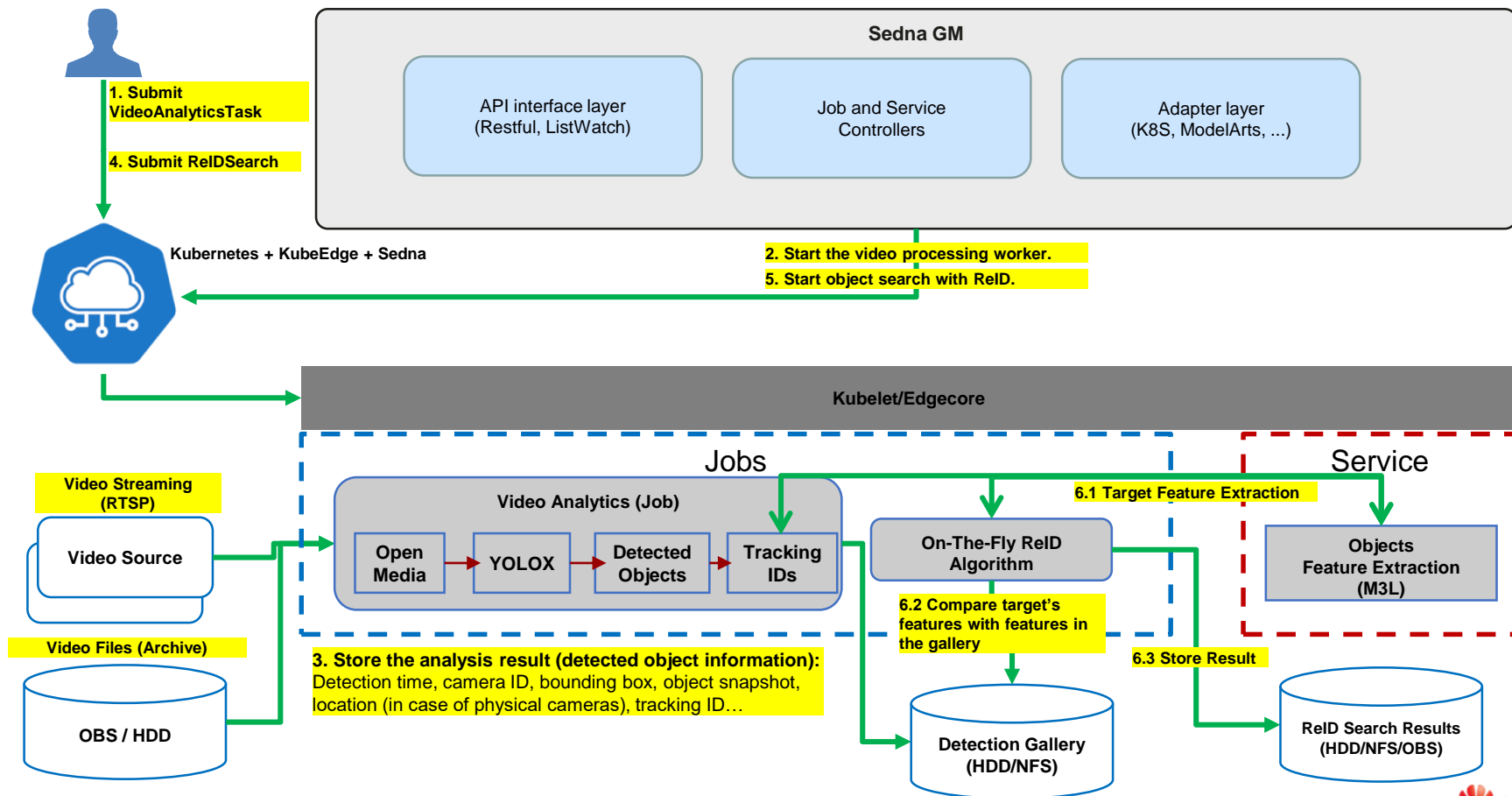
Image source: <https://pixabay.com/>

Edge AI with Sedna

Multi-Edge-Cloud Reasoning with Workload Repartition

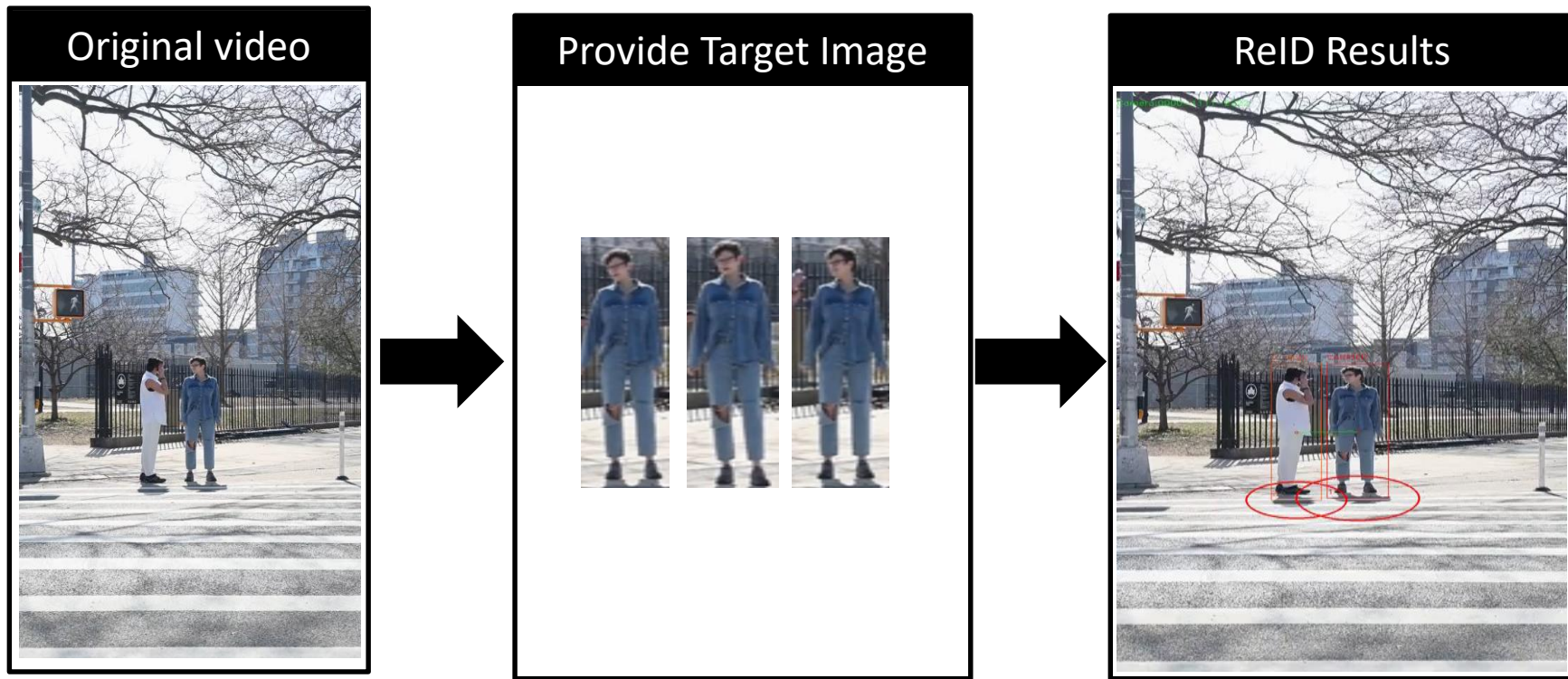


Edge AI with Sedna System Design



Edge AI with Sedna ReID in Practice

ReID can be applied to any domain, given that the feature extraction network is properly trained. In the example below, we show how it can be applied to humans due to the vast amount of data for this category.



Video from: <https://www.pexels.com/video/two-people-standing-by-the-pedestrian-crossing-while-talking-7263169/>

Edge AI with Sedna

Hands-On: Replicate ReID Application

Getting started:

- Login into your VM
- Pull the training repo @ <https://github.com/haicgu/training>
- Open the README.md under *Cloud-Edge/Sedna/*

