2021 embedded VISION summit®

VIRTUAL | MAY 25-27

How Containerization Unblocks Barriers to Fast, Easy Deployment of AI-Driven Visual Recognition Applications

Rakshit Agrawal
Director of Research & Development, Camio

camio
real-time video search

Unblocking the Barriers to Fast Deployment

- From Notebooks to Chipsets
- Containers for deployment flexibility
- AI deployment pipeline
- Confidence in Iterations
- Adaptability, Speed, and Reliability
- Trade offs and Limitations

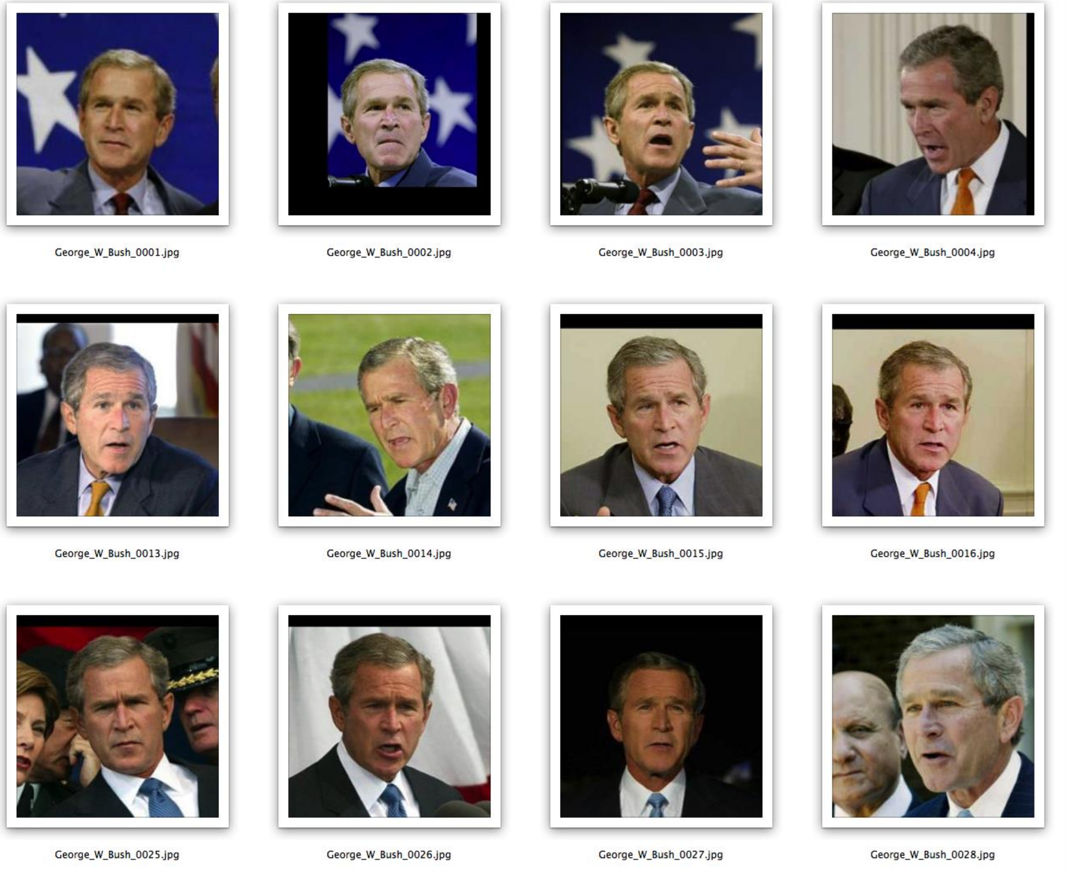
2021 embedded VISION summit VIRTUAL | MAY 25-27

Vision AI in the real world

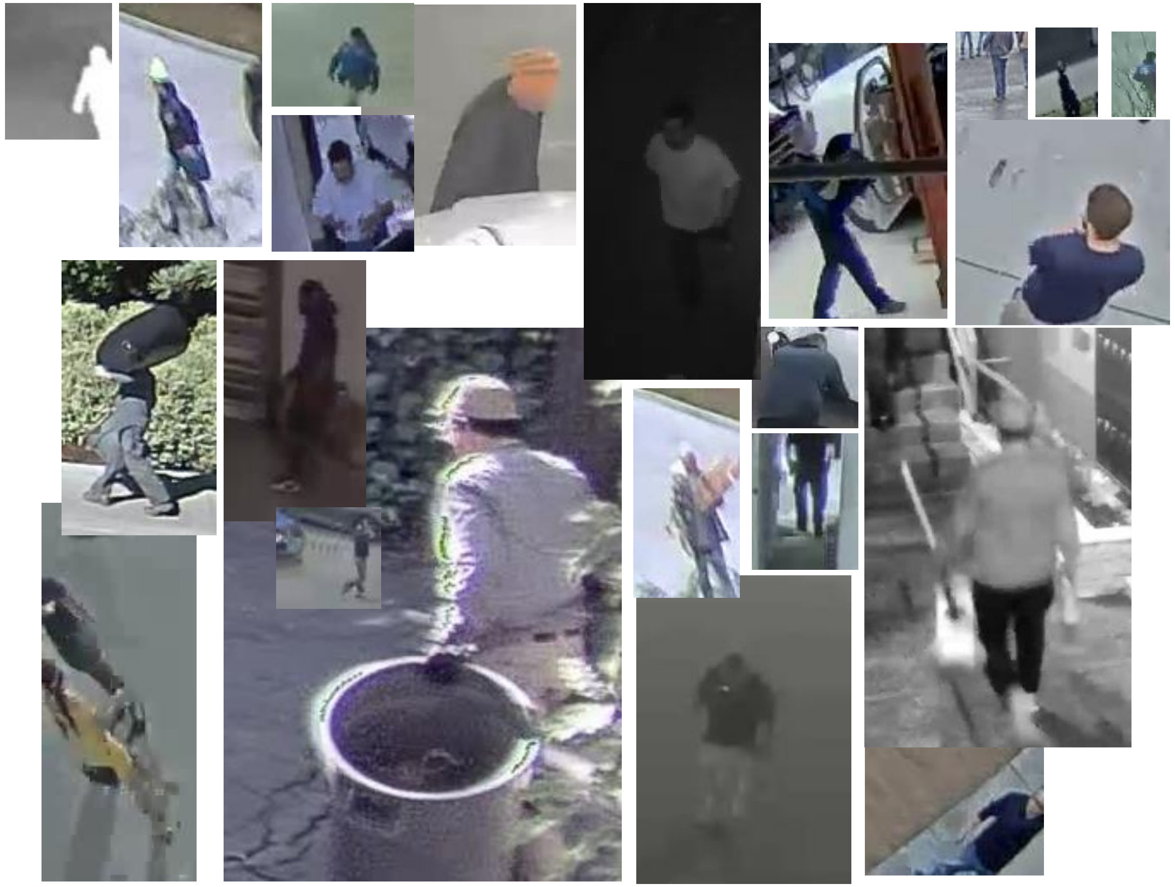
camio
real-time video search

Demo vs Reality

Demo

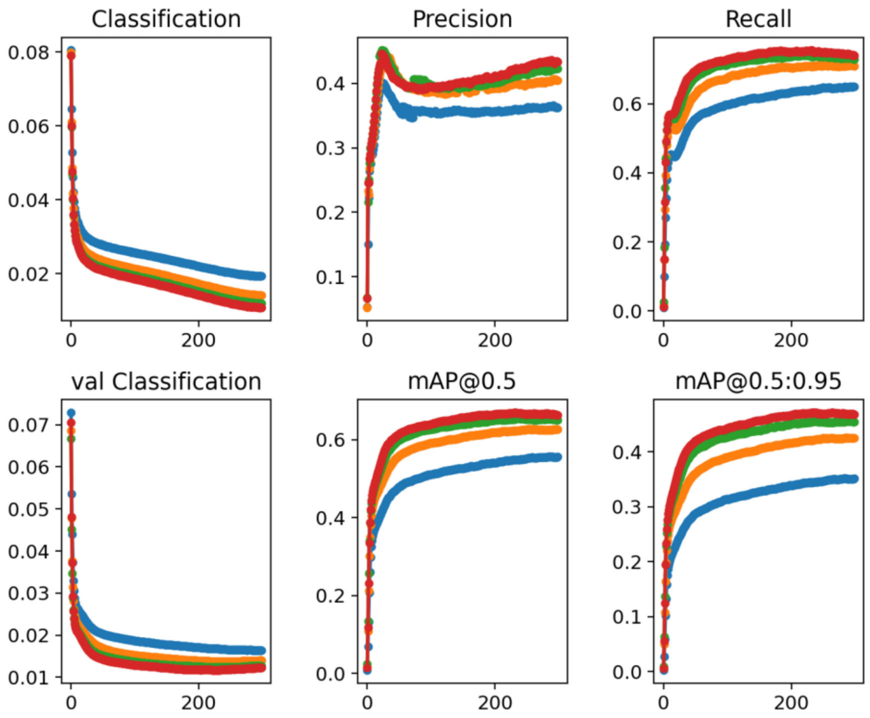


Reality

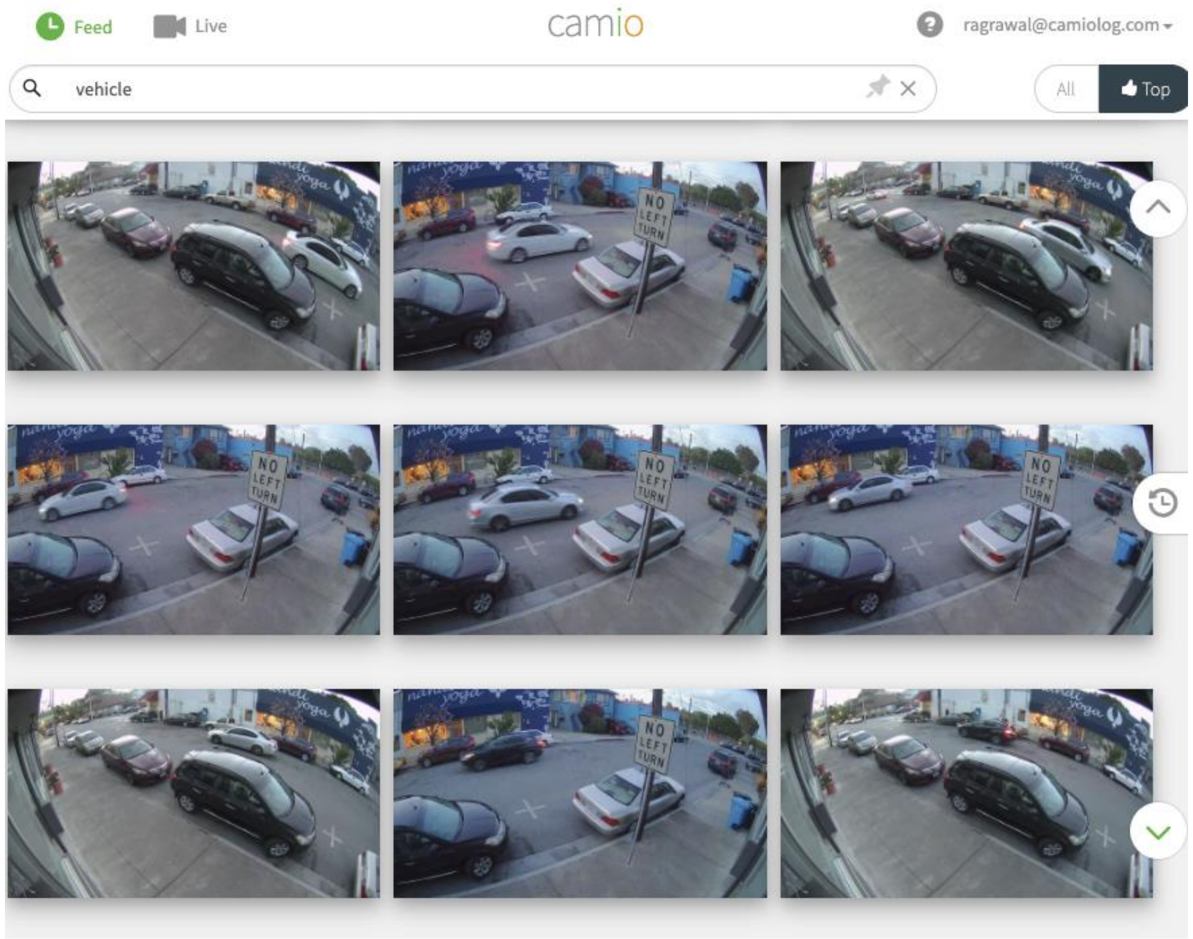


Research vs Deployment

Research dataset -> Model -> Metrics

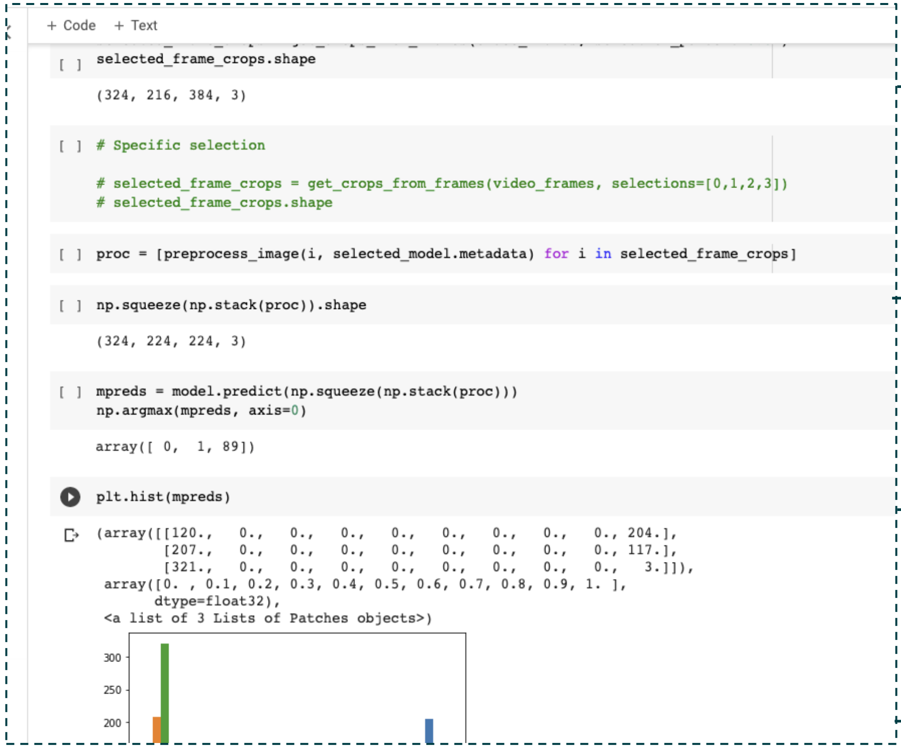


Real-Time Video Search



camio
real-time video search

From Notebooks to Chipsets



```
[ ] selected_frame_crops.shape

    (324, 216, 384, 3)

[ ] # Specific selection

    # selected_frame_crops = get_crops_from_frames(video_frames, selections=[0,1,2,3])
    # selected_frame_crops.shape

[ ] proc = [preprocess_image(i, selected_model.metadata) for i in selected_frame_crops]

[ ] np.squeeze(np.stack(proc)).shape

    (324, 224, 224, 3)

[ ] mpreds = model.predict(np.squeeze(np.stack(proc)))
    np.argmax(mpreds, axis=0)

    array([ 0,  1, 89])

    plt.hist(mpreds)

    (array([[120.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0., 204.],
            [207.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0., 117.],
            [321.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   3.]]),
     array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ],
           dtype=float32),
     <a list of 3 Lists of Patches objects>)
```

ARM based edge devices

Small form appliances

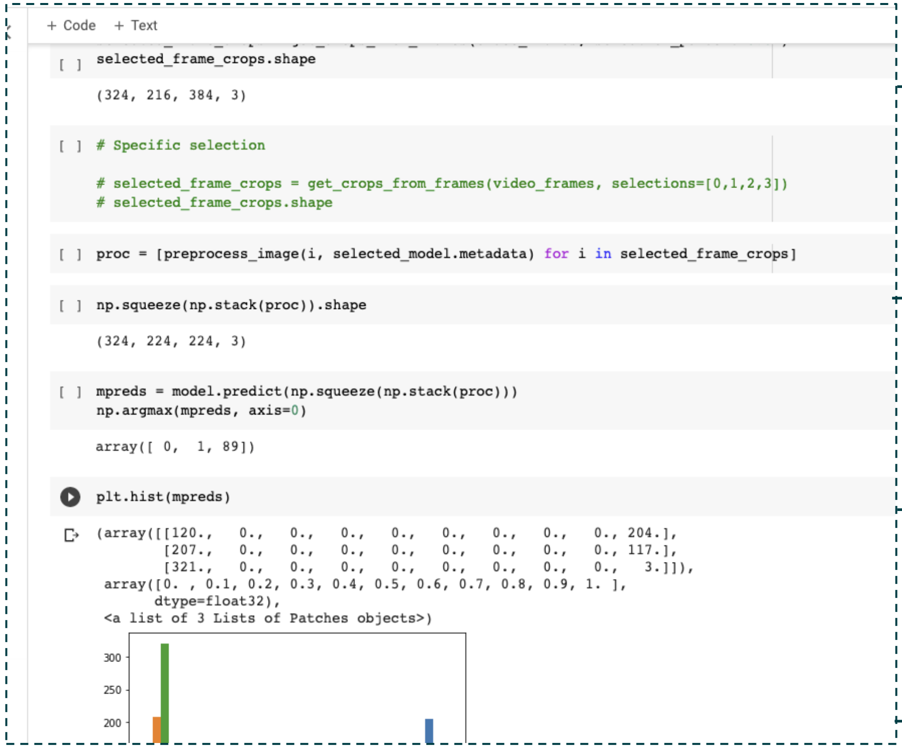
Multi-socket rack servers

Data centers

camio
real-time video search

© 2021 Camio

From Notebooks to Chipsets



```
[ ] selected_frame_crops.shape

    (324, 216, 384, 3)

[ ] # Specific selection

    # selected_frame_crops = get_crops_from_frames(video_frames, selections=[0,1,2,3])
    # selected_frame_crops.shape

[ ] proc = [preprocess_image(i, selected_model.metadata) for i in selected_frame_crops]

[ ] np.squeeze(np.stack(proc)).shape

    (324, 224, 224, 3)

[ ] mpreds = model.predict(np.squeeze(np.stack(proc)))
    np.argmax(mpreds, axis=0)

    array([ 0,  1, 89])

    plt.hist(mpreds)

    (array([[120.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0., 204.],
            [207.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0., 117.],
            [321.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   3.]]),
     array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ],
           dtype=float32),
     <a list of 3 Lists of Patches objects>)
```

Low power, low memory and compute

Consumer-grade CPU/ GPU

Xeon + Nvidia RTX

Xeon + Tesla T4 grid

camio
real-time video search

© 2021 Camio

What is Containerization?

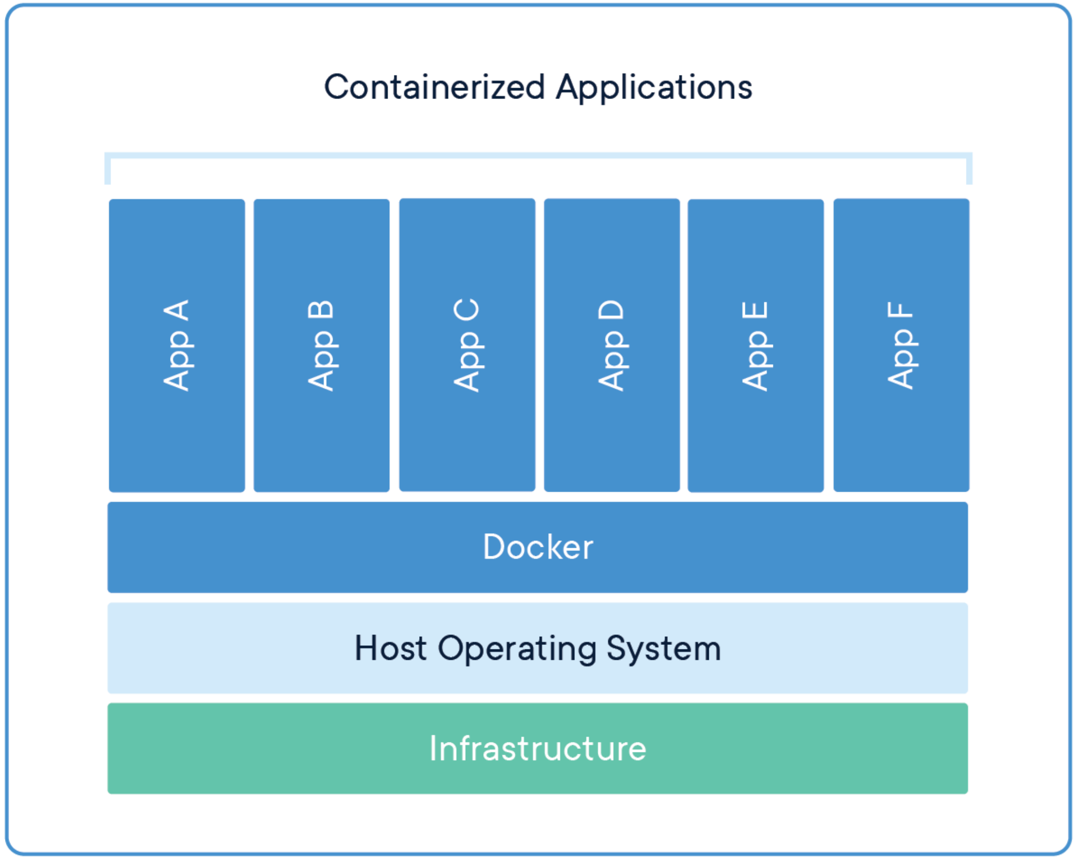
Containers for deployment flexibility

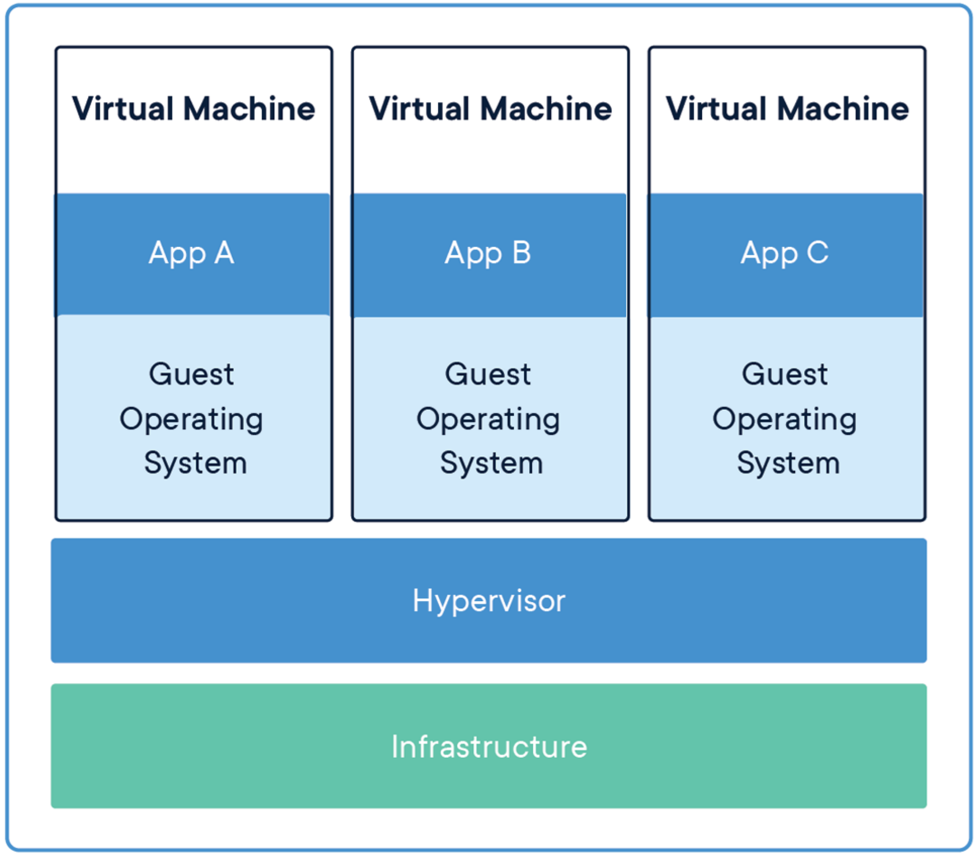
- A container is a packaged unit of code, dependencies, and environment

- Once built, the image can be shipped on any supporting runtime

- A runtime, or an orchestrator controls deployment and lifecycle of containers.

Understanding Abstractions with Containers





Source: https://www.docker.com/resources/what-container

Build, Configure, Deploy - Anywhere

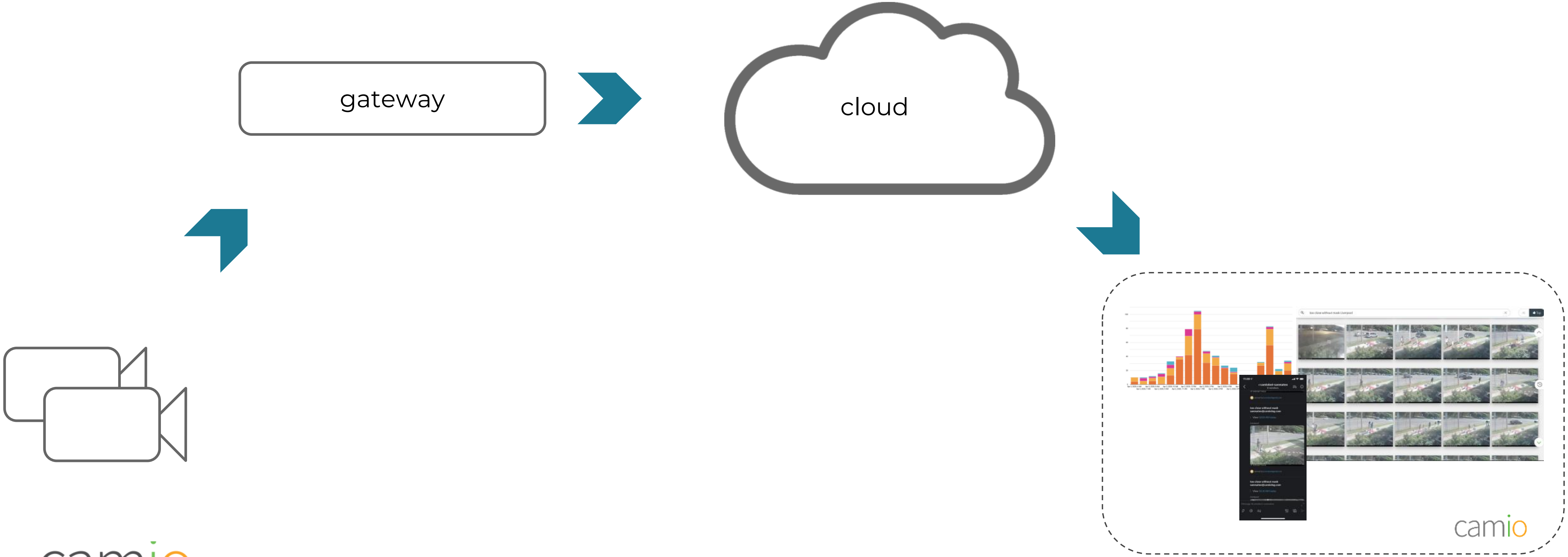
```
FORM python:3.8
RUN apt-get update && \
          apt-get install -y sudo \
          build-essential curl \
          libcurl4-openssl-dev \
          libssl-dev wget \
          python3-pip \
          git && \
          pip3 install --upgrade pip

COPY requirements.txt .
...
```

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: dep-wq-1
  labels:
    app: dep-wq-1
spec:
  template:
    metadata:
      labels:
        app:
...
```

```
kubectl config get-contexts

kubectl config use-context ...

kubectl get pods

kubectl get services

kubectl apply -f pods.yaml

kubectl apply -f services.yaml
...
```

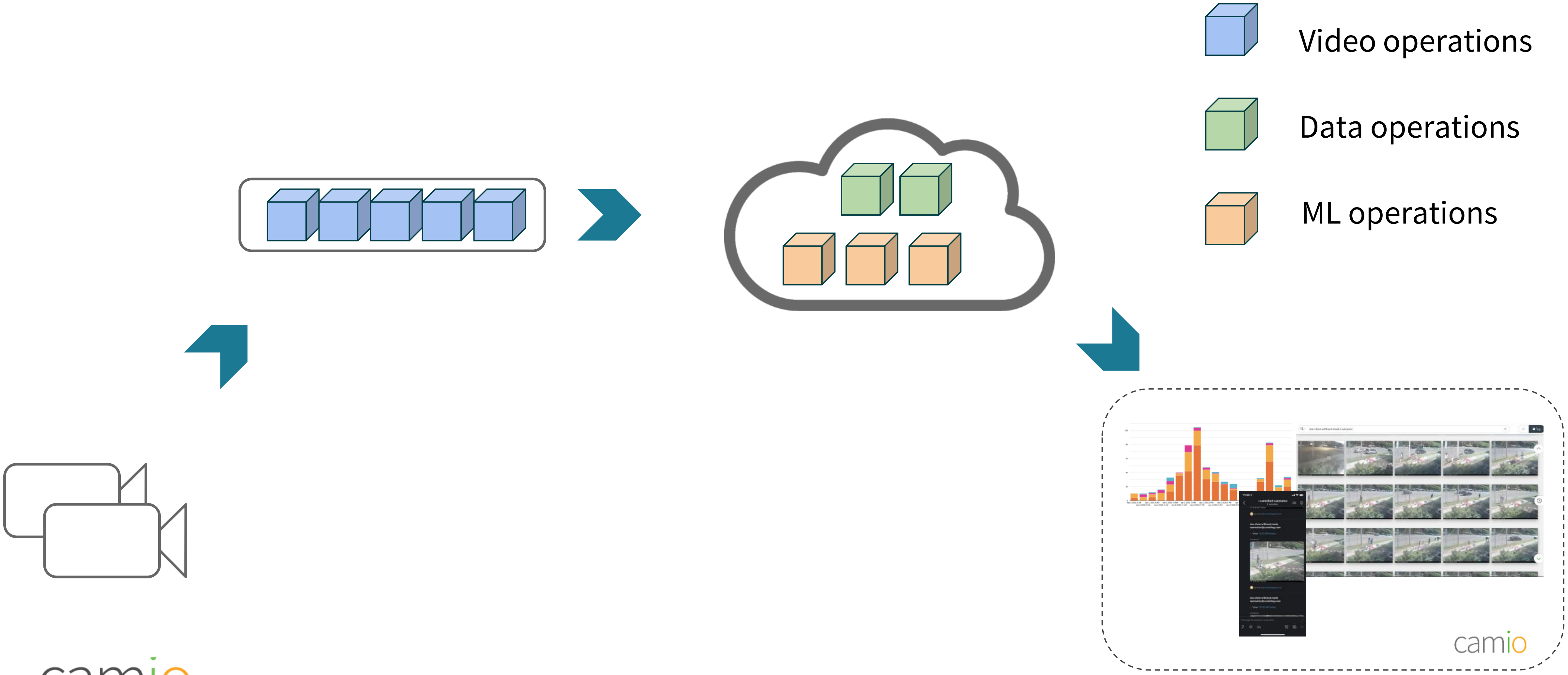
Example: AI pipeline for videos

Camio: Video AI Pipeline



camio
real-time video search

Video AI Pipeline: Traditional

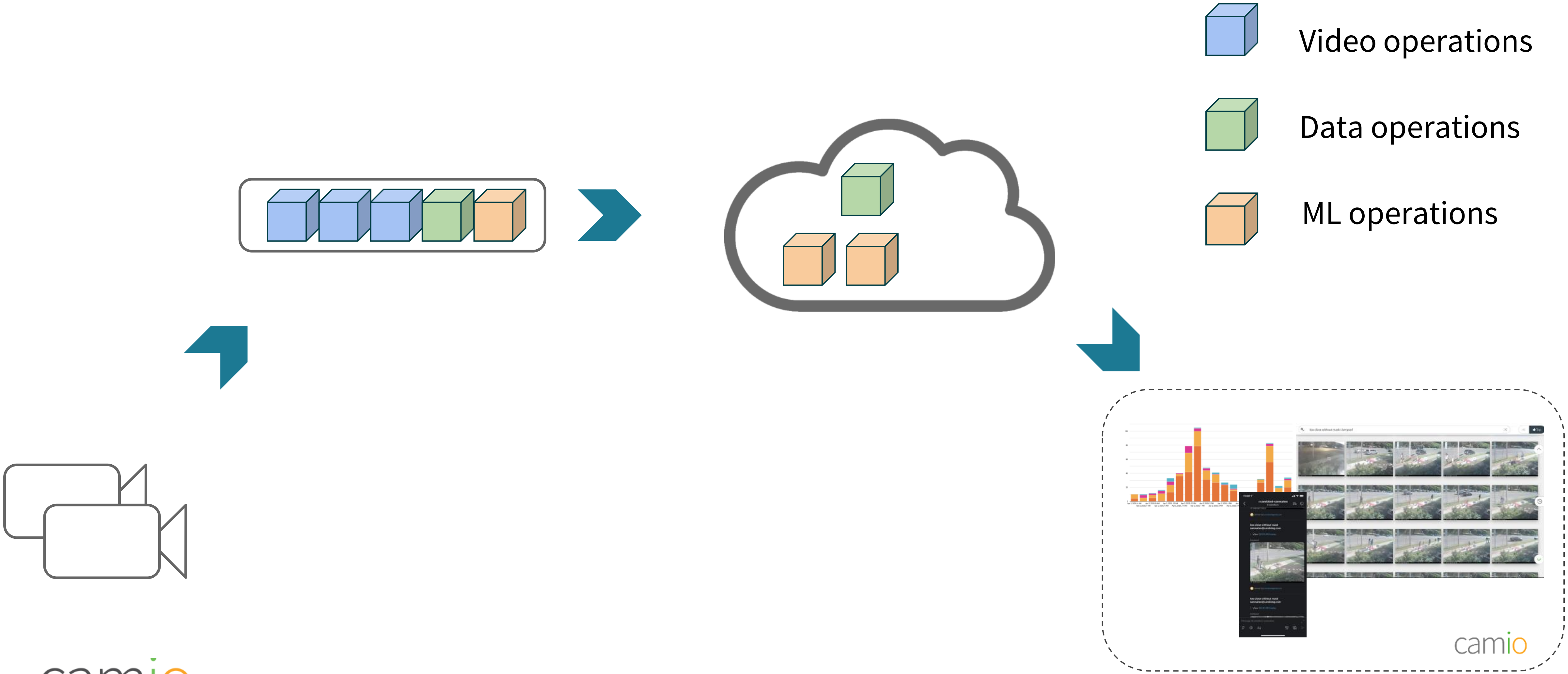


🟦 Video operations

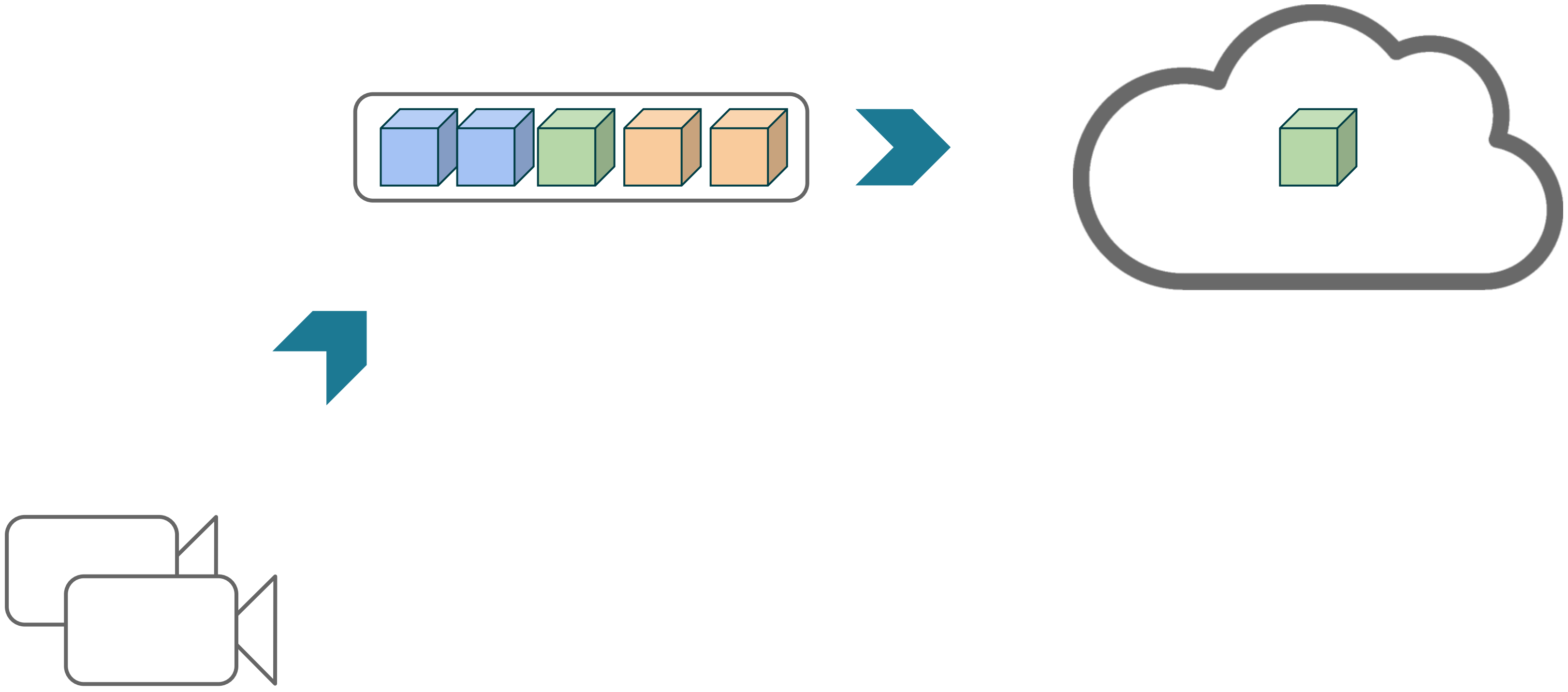
🟩 Data operations

🟧 ML operations

Video AI Pipeline: Hybrid



Video AI Pipeline: More at the Edge

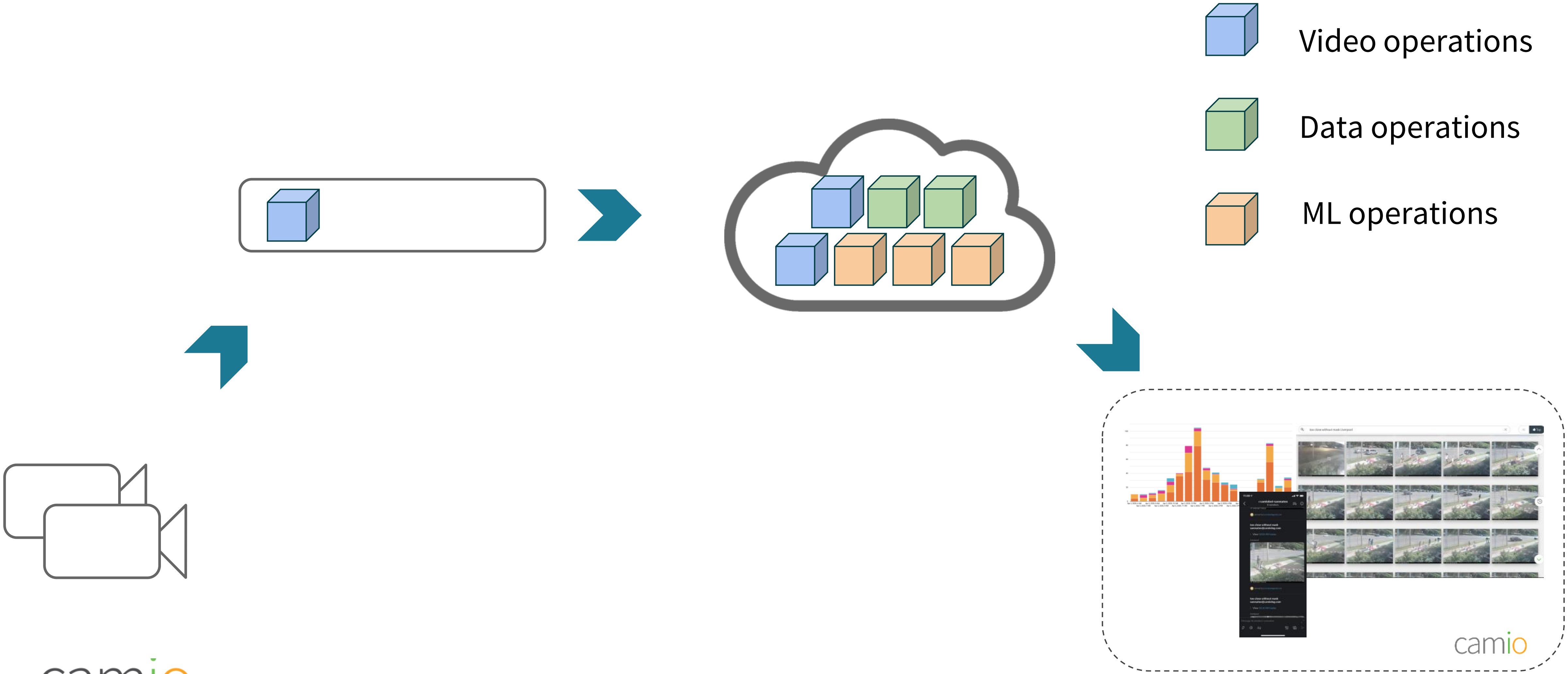


🟦 Video operations

🟩 Data operations

🟧 ML operations

Video AI Pipeline: More in the Cloud



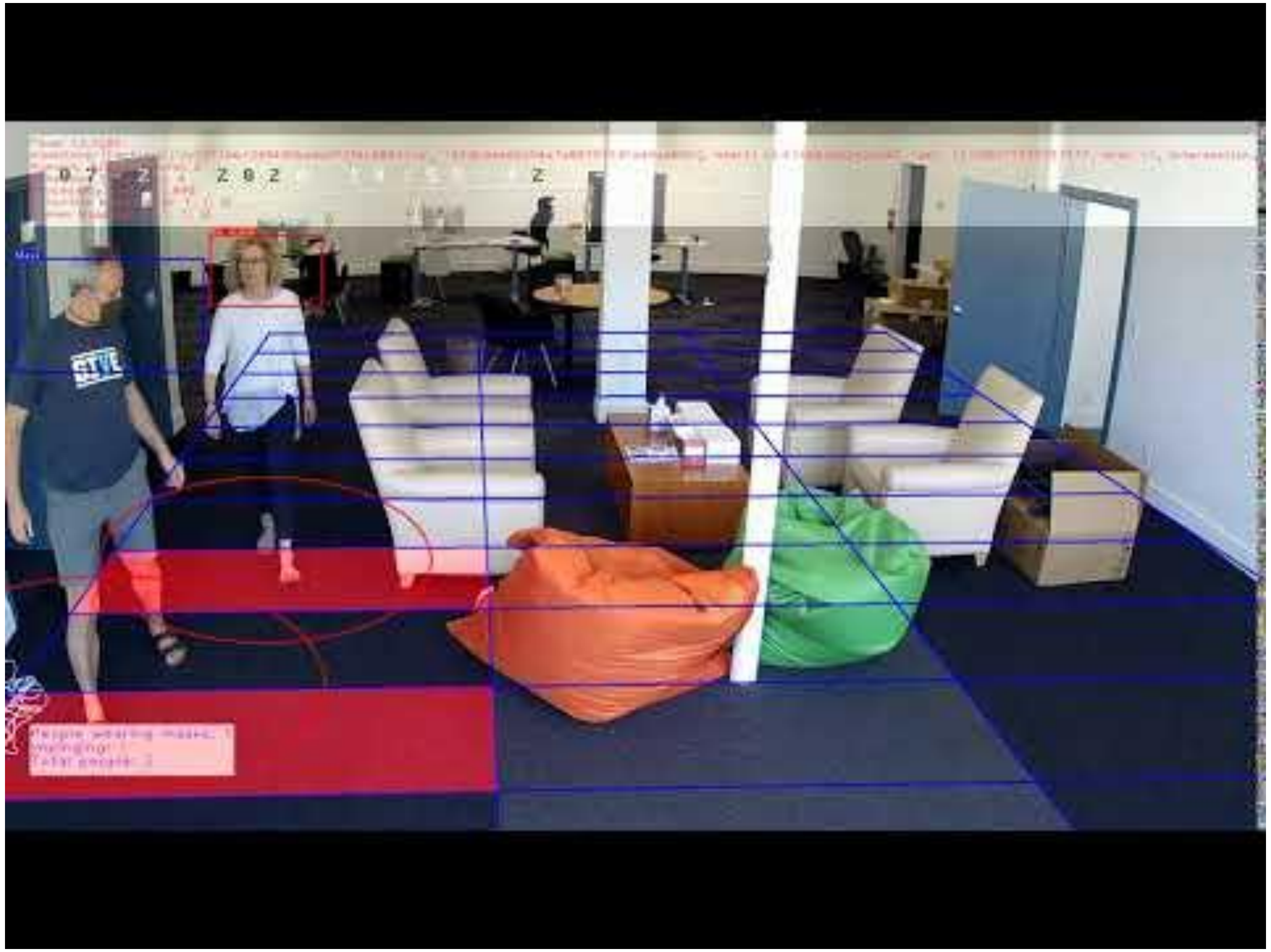
🟦 Video operations

🟩 Data operations

🟧 ML operations

camio
real-time video search

Video Vision AI in action



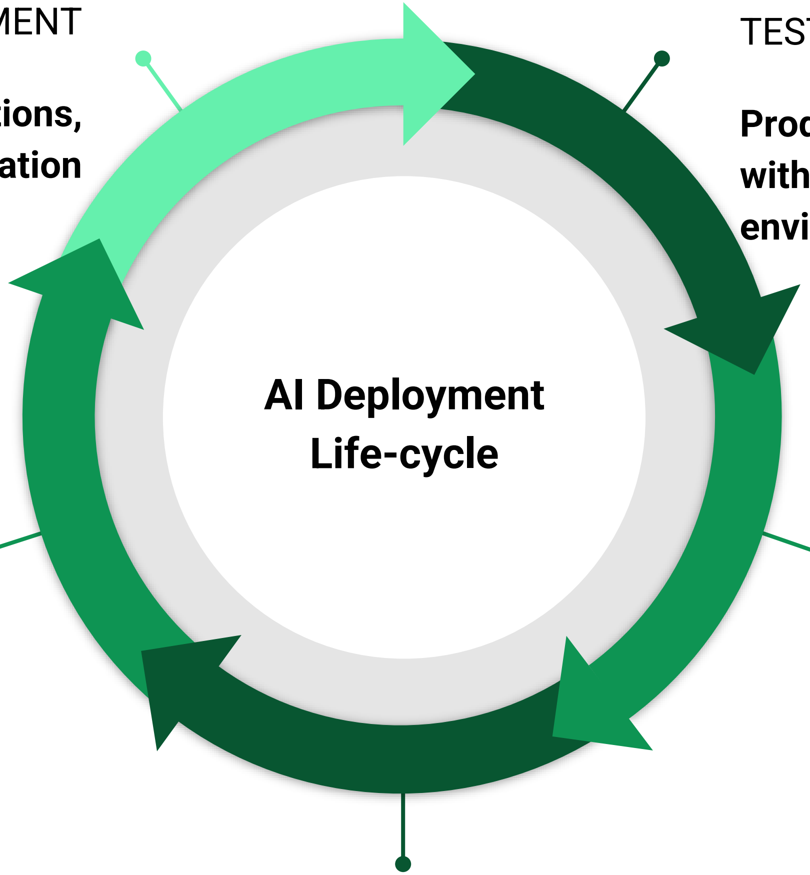
https://www.youtube.com/watch?v=OOlsxtUwMB8

2021 embedded VISION summit VIRTUAL | MAY 25-27

Lifecycle of Deployments

camio
real-time video search

Confidence in Iterations



AI Deployment Life-cycle

DEVELOPMENT

Integrations, inference, evaluation

TESTING

Product deployment with inference (Test environment)

PRODUCTION

Product deployment with inference (Production environment)

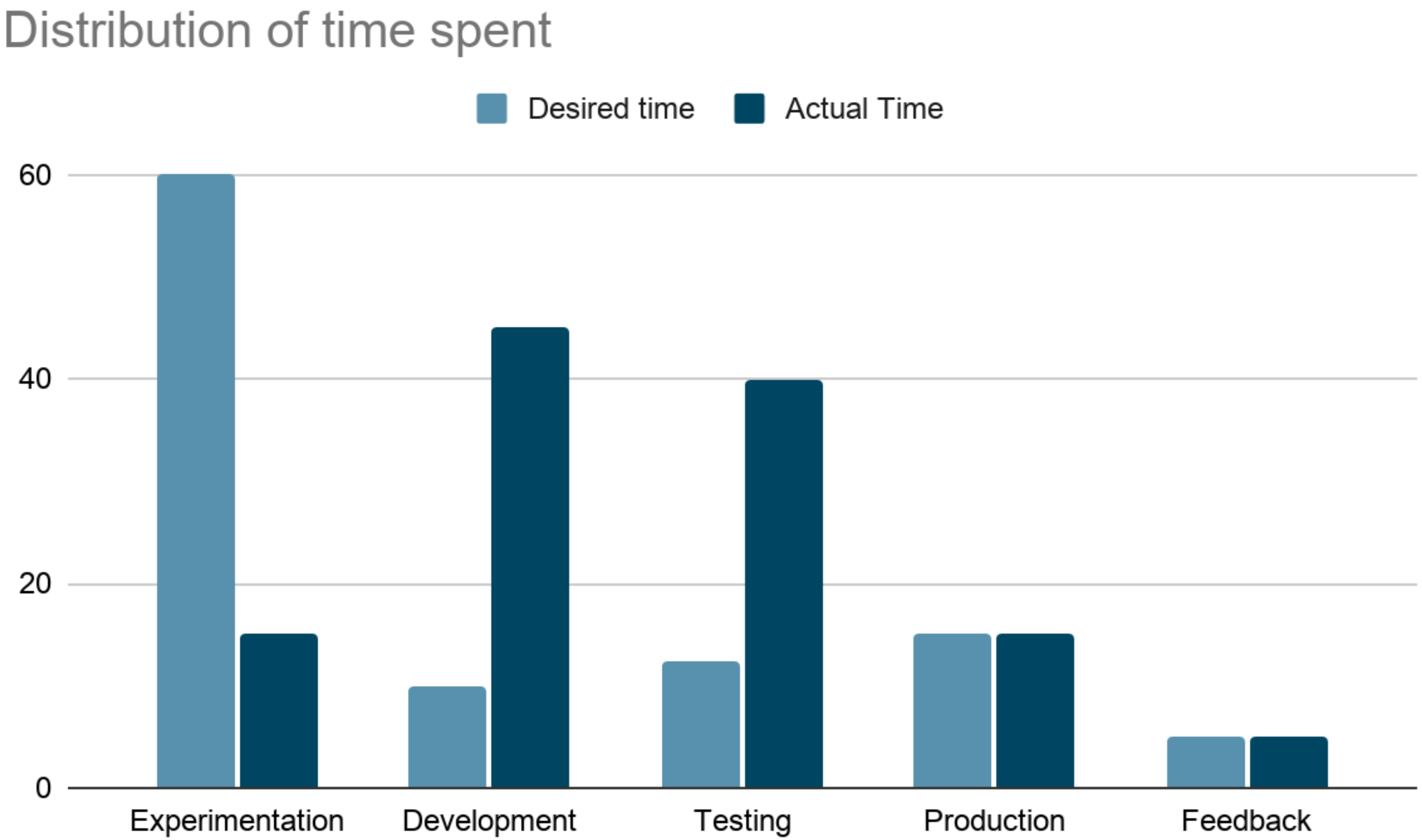
FEEDBACK

Real-world results, improvements, new data

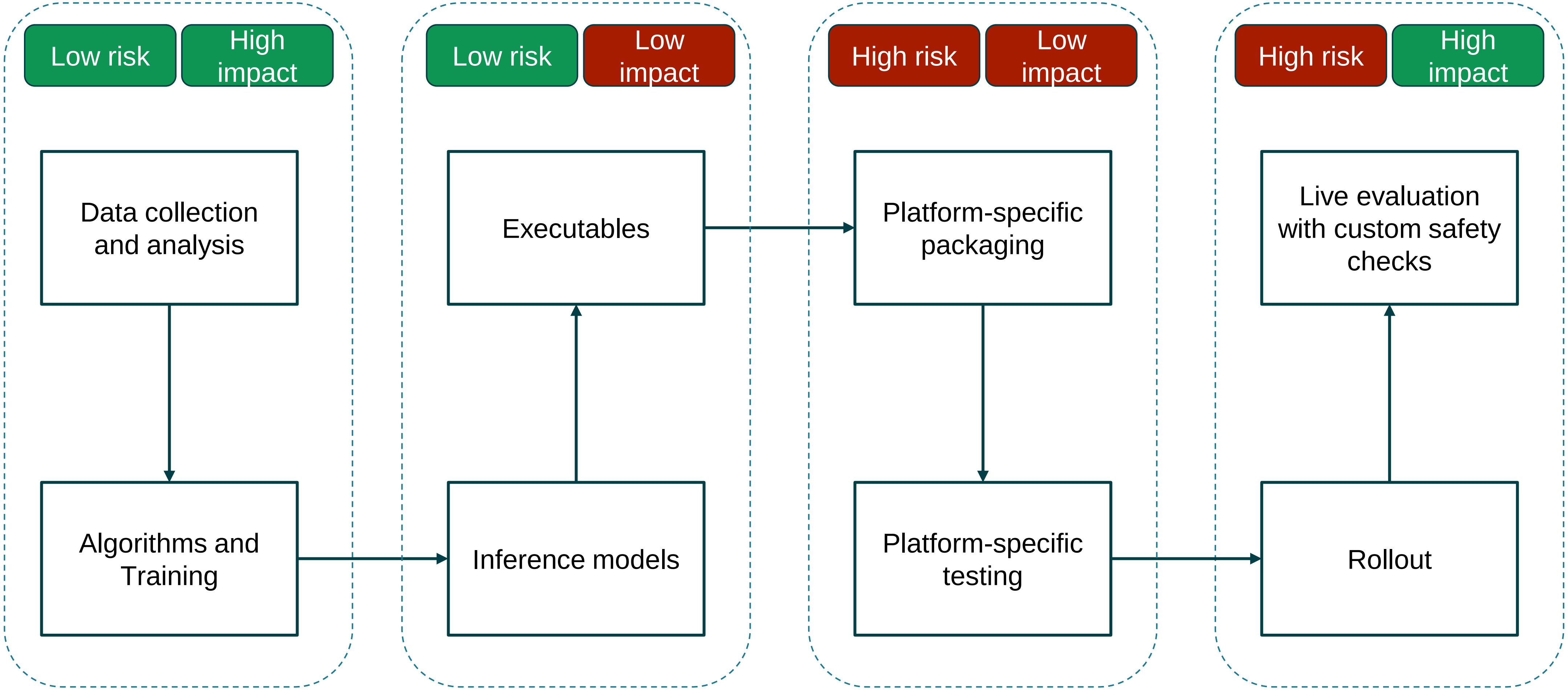
EXPERIMENTATION

Algorithms, models, data, training, analysis

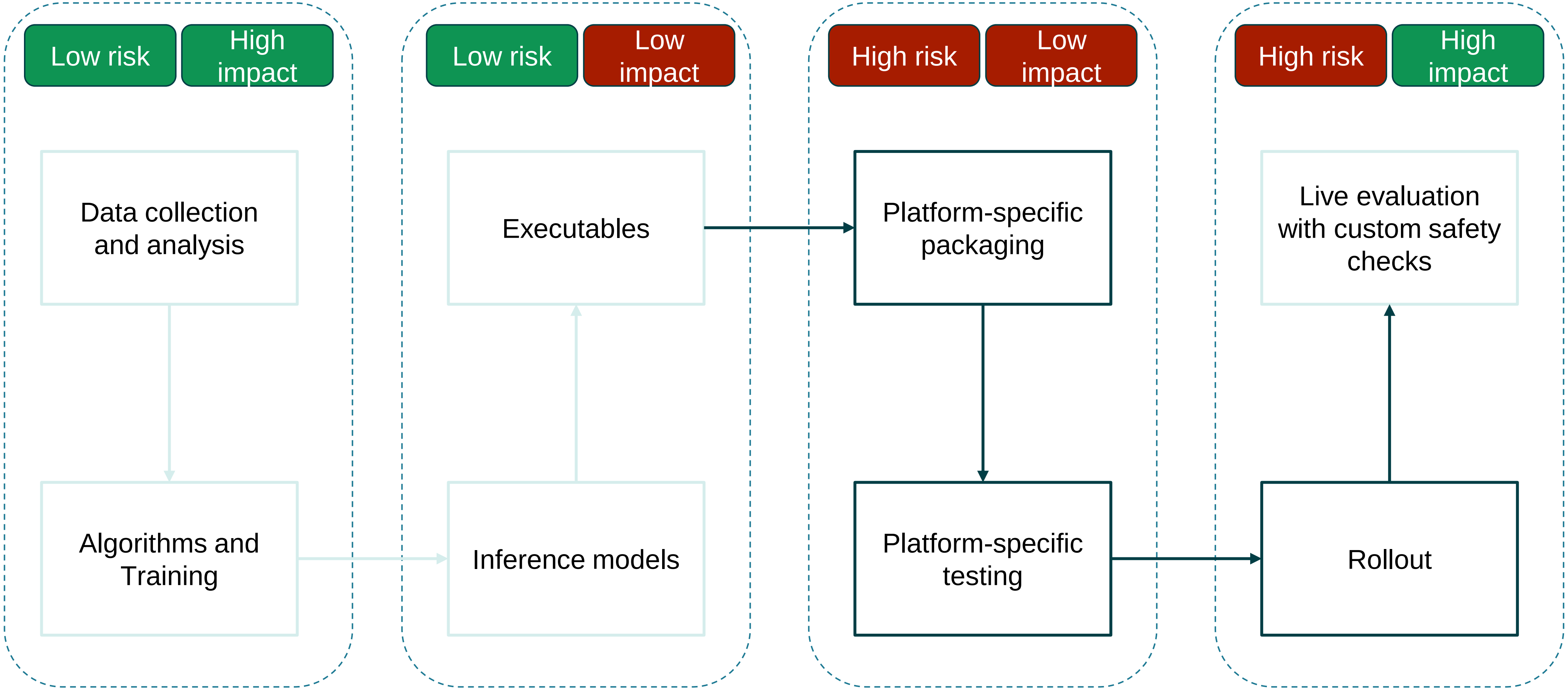
Eliminating Black Holes of Time



Deployment Process: Traditional



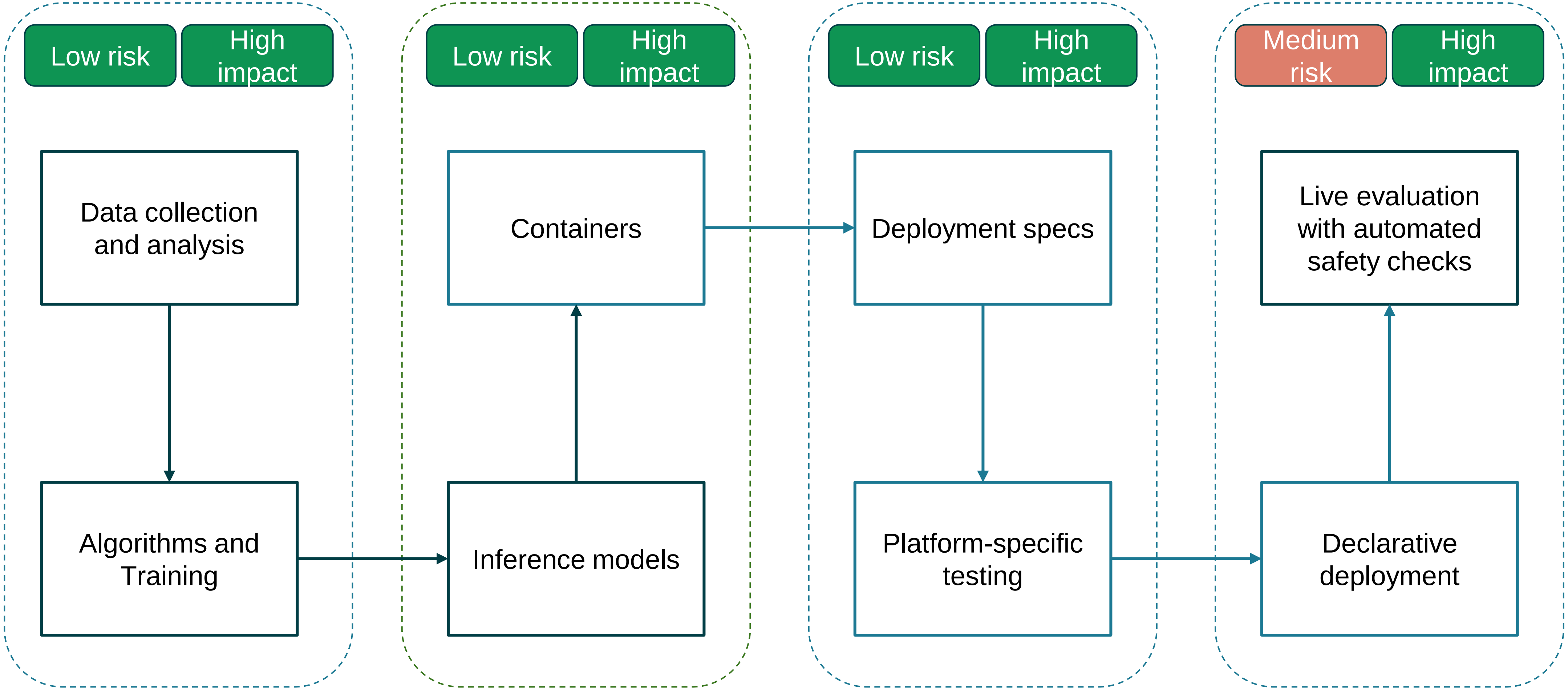
Deployment Process: Areas of Concern



camio real-time video search

© 2021 Camio

Deployment Process: Containerized



Key Takeaways

Benefits: Adaptability, Speed, and Reliability

- Adaptability
  - Containerization speeds ongoing refinements required by real-world vision AI production applications
- Speed
  - Agile pipeline operations are critical when moving from research notebooks to chipsets
- Reliability
  - With containerization, the painful process of development, packaging and deployment becomes predictable and consistent

Trade offs and Limitations

- Larger deployment payloads for low-bandwidth regions
- New paradigm for inter-process communication
  - Message passing vs RPC/ shared memory
- Excludes the low-compute edge devices (for now)

Thank you!

Learn more from

Camio

camio.com

Containerization at Camio

camio.com/technology/containers

Kubernetes

kubernetes.io/

Please contact for any questions or discussions

Rakshit Agrawal

Director of Research & Development

Camio

rakshit@camio.com

camio
real-time video search